

Doc Code - SEQ.XML
WIPO Type Code -

SLIC Placeholder Sheet for P-ELP (CMS) Content

Application Number: 18066420

Document Date: 12/15/2022

The presence of this form in the CMS record indicates that the following document type was received in electronic format on the date identified above. This content is stored in CMS.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist.

- Sequence Listing

At the time of document entry (noted above):

- USPTO employees may access CMS content via DAV or via the SLIC web page.
- External customers may access CMS content via PAIR/Patent Center Retrieval using the Supplemental Content tab.

Form Revision Date: 07 January 2022

Application Number: 18066420

Reviewed by: cSmvAsj2Eeg/p9KRnlQ7Xvx6SK8jnDazOzlUmNPrTuL7Ou65e6sKHw==

Note: This validation report concerns the sequence listing filed 12-15-2022 12:02:21 and indicates that NO errors were found by the USPTO during validation of the sequence listing. This validation report should not be construed as meaning that the sequence listing complies with all of the requirements of 37 CFR §§ 1.821-1.825, 37 CFR §§ 1.831-1.835 and 1.839, WIPO Standard ST.25, or WIPO Standard ST.26.

For any sequence listing related questions or concerns, the Applicant is invited to contact the USPTO Sequence Listing Help Desk at 571-272-2510 or SequenceHelpDesk@USPTO.GOV.

Timestamp: [year=2022; month=12; day=15; hr=16; min=27; sec=24; ms=768]

Validated By CRFValidator v 1.5.0

Application No: 18066420 Version No: 1.1

CRF Copy Request from application Version :

Input Set: /slic/CRF_WORKFLOW/18/066/420/1.1/report_18066420_1_1.xml

Output Set: /slic/CRF_WORKFLOW/18/066/420/1.1/report_18066420_1_1.xml.pdf

Started: 2022-12-15 12:06:14.000

Finished: 2022-12-15 12:06:35.013

Elapsed: 0 hr(s) 0 min(s) 21 sec(s) 13 ms

Total Warnings: 1

Total Errors: 0

No. of SeqIDs Defined: 81

Actual SeqID Count: 81

Error code	Error Description
W 261178	DataElement:PROPERTY_NAMES.ORIGINAL_FREE_TEXT_LANGUAGE_CODE;MessageKey:ORIGINAL_FREE_TEXT_LANGUAGE_CODE_MISSING;LocalizedMessage:An original free text language code has not been entered.;

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="148165.001401_SL.xml"
softwareName="WIPO Sequence" softwareVersion="2.2.0" productionDate="2022-12-15">
<ApplicationIdentification>
<IPOfficeCode>US</IPOfficeCode>
<ApplicationNumberText>18/066,420</ApplicationNumberText>
<FilingDate>2022-12-15</FilingDate>
</ApplicationIdentification>
<ApplicantFileReference>148165.001401</ApplicantFileReference>
<EarliestPriorityApplicationIdentification>
<IPOfficeCode>US</IPOfficeCode>
<ApplicationNumberText>63/289,888</ApplicationNumberText>
<FilingDate>2021-12-15</FilingDate>
</EarliestPriorityApplicationIdentification>
<ApplicantName languageCode="en">INTERIUS BIOTHERAPEUTICS, INC.</ApplicantName>
<InventionTitle languageCode="en">PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING
THE SAME, AND USES THEREOF</InventionTitle>
<SequenceTotalQuantity>81</Se
quenceTotalQuantity>
<SequenceData sequenceIDNumber="1">
<INSDSeq>
<INSDSeq_length>511</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..511</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>mol_type</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier id="q2">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>MKCLLYLAFLFIGVNCKFTIVFPHNQGNWKNVPSNYHYCPSSDLNWHNDLIGTALQVKMPKSHKAIQ
ADGWMCHASKWVTTCTDFRWYGPYIITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYT
GEWVDSQFINGKCSNYICPTVHNSTTWHSDYKVKGLCDSNLISM
DITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSL
IQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAAPKNPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISGTTTERELWD
DWAPYEDVEIGPNGVLRVTSSGYKFPLYMIGHGMLDSDLHLSKAQVFEHPHIQDAASQLPDDESILFFGDTGLSKNPIELVEGWFFSSW

```



```

KSSIASFFFIIGLIIGLFLVLRVGIHLCKIKLKHTKKRQIYTDIEMNRLGK</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="2">
<INSDSeq>
<INSDSeq_length>495</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..495</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>mol_type</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier id="q4">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</
INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>KFTIVFPHNQKGNWKNVPSNYHYCPSSSDLNWNHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPKYIITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYTGWVDSQFINGKCSNYICPTVHNSTTWHSYKVKGLCDNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPDLSYLAPKNPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMIISGTTTERELWDDWAPYEDVEIGPNGVLRVTSSGYKFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESLLFFGDTGLSKNPIELVEGWFFSSWKSSIASFFFIIGLIIGLFLVLRVGIHLCKIKLKHTKKRQIYTDIEMNRLGK</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="3">
<INSDSeq>
<INSDSeq_length>495</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..495</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>mol_type</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>

```

```

<INSDQualifier id="q6">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>KFTIVFPHNQKGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCD
FRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAEAVIVQVTPHHVLVDEYTG EWVDSQFINGKCSNY
ICPTVHNSTTWHSDYKVKGLCDSNLASMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRLPSGVWFEMADKD
LFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDSLAPKNPGTGPAFTIINGTLKYFETRY
IRVDIAAPILSRMVGMI SGTTERELWDDWAPYEDVEIGPNGVLR TSSGYKFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQDAASQL
PDDES LFFGDTGLSKNPIELVEGW FSSWKSSIASFFFIIGLIIGLFLVLRVGIHLCKIKLHTKKRQIYTDIEMNRLGK</INSDSeq
_sequence>
</INSDSeq>
</SequenceData>
<Sequ
enceData sequenceIDNumber="4">
<INSDSeq>
<INSDSeq_length>495</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..495</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>mol_type</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier id="q8">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>KFTIVFPHNQKGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCD
FRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAEAVIVQVTPHHVLVDEYTG EWVDSQFINGKCSNY
ICPTVHNSTTWHSDYKVKGLCDSNLDSMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHW
GVRLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDSLAPKNPGTGPA
FTIINGTLKYFETRYIRVDIAAPILSRMVGMI SGTTERELWDDWAPYEDVEIGPNGVLR TSSGYKFPLYMIGHGMLDSDLHLSSK
AQVFEHPHIQDAASQLPDDES LFFGDTGLSKNPIELVEGW FSSWKSSIASFFFIIGLIIGLFLVLRVGIHLCKIKLHTKKRQIYTDI
EMNRLGK</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="5">

```

```

<INSDSeq>
<INSDSeq_length>495</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..495</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>mol_type</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier id="q10">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</INSDQualifier_value>
</INSDQualifie
r>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>KFTIVFPHNQKGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAEAVIVQVTPHHVLVDEYTG EWVDSQFINGKCSNYICPTVHNSTTWHSDYKVKGLCDNLESMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDSL SYLAPKNPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMI SGTTERELWDDWAPYEDVEIGPNGVLR TSSGYKFPLYMIGHGMLSD LHLSSKAQVFEHPHIQDAASQLPDES LFFGDTGLSKNPIELVEGW FSSWKSSIASFFFIIGLIIGLFLVLRVGIHL C I K L K H T K K R Q I Y T D I E M N R L G K </INSDSeq_sequence>
</INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="6">
<INSDSeq>
<INSDSeq_length>495</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..495</INSDFeature_location>
<INSDFeature_qual>
<INSDQualif
ier>
<INSDQualifier_name>mol_type</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier id="q12">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</INSDQualifier_value>

```

```

</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>KFTIVFPANQKGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPQSHKAIQADGWMCHASKWVTTCD
FRWYGPKYIITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYTG EWVDSQFINGKCSNY
ICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRLPSGVWFEMADKD
LFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPKNPGTGPAFTIINGTLKYFETRY
IRVDIAAPILSRMVGMI SGTTERELWDDWAPYEDVEIGPNGVLR TSSGYKFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQDAASQL
PDDES LFFGDTGLSKNPIELVEGW FSSWKSSIASFFFIIGLIIGLFLVLRVGIHLCKIKLHTKKRQIYTDIEMNRLGK</INSDSeq
_sequence>
</INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="7">
<INSDSe
q>
<INSDSeq_length>495</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..495</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>mol_type</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier id="q14">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>KFTIVFPHNAKGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCD
FRWYGPKYIITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYTG EWVDSQFINGKCSNY
ICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRLPSGVWFEMADKD
LFAAARFPECPEGSSISAPSQTSV
DVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPKNPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMI SGTTER
ELWDDWAPYEDVEIGPNGVLR TSSGYKFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDES LFFGDTGLSKNPIELVEGW
FSSWKSSIASFFFIIGLIIGLFLVLRVGIHLCKIKLHTKKRQIYTDIEMNRLGK</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="8">
<INSDSeq>
<INSDSeq_length>495</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>

```

```

<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..495</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>mol_type</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier id="q16">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>KFTIVFPHNRKGNWKNVPSNYHYCPSSSDLNWNHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAEAVIVQVTPHHVLVDEYTG EWVDSQFINGKCSNYICPTVHNSTTWHSDYKVKGLCDNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISFVDLSYLAPKNPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISSGTTTERELWDDWAPYEDVEIGPNGVLRVTSSGYKFPLYMIGHGMLSDSLHLSSKAQVFEHPHIQDAASQLPDDESLFFGDTGLSKNPIELVEGWFFSSWKSSIASFFFIIGLIIGLFLVLRVGIHLCKIKLHKKRQIYTDIEMNRLGK</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="9">
<INSDSeq>
<INSDSeq_length>495</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..495</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>mol_type
e</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier id="q18">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>

```

```

</INSDSeq_feature-table>
<INSDSeq_sequence>KFTIVFPHNKKGNWKNVPSNYHYCPSSSDLNWNHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYTG EWVDSQFINGKCSNYICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPKNPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMI SGTTERELWDDWAPYEDVEIGPNGVLRITSSGYKFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDES LFFGDTGLSKNPIELVEGW FSSWKSSIASFFFIIGLIIGLFLVLRVGIHLCKIKLHTKKRQIYTDIEMNRLGK</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="10">
<INSDSeq>
<INSDSeq_length>517</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..517</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>mol_type</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier id="q20">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>MLSYLIFALVVSPILGKIEIVFPQHTTGDWKRVPHEYNYCPTSADKN SHGTQTGIPVELTMPKGLTTHQVDGFMCHSALWMTTCDFRWYGPKYITHSIHNEPTDYQCLEAIKAYKDGVSFNPGFPPQSCGYGTVTDAAEHIVTVTPHSVKVDEYTG EWIDPHFIGGRCKGQICETVHNSTKWFTSSDGESVCSQLFTLVGGTFFSDSEEITSMGLPETGIRSNYFPYVSTEGICKMPFCRKG YK LKNDLWFQITDPDLDKTVRDLPHIKDCDLSSSIVTPGEHATDISLISDVERILDYALCQNTWSKIEAGEPITPVDLSYLGPNPGAGPVFTIINGSLHYFMSKYL RVELESVPVPRMEGKVAGTRIVRQLWDQWFPFGEVEIGPNGV LKTKQGYKFPLHIIGTGEVDNDIKMERIVKHWEPHIEAAQTFLKKDDTEEVLYYGD TGVSKNPVELVEGW FSGWRSSIMGVLA VII GFVILIFLIRLIGVLSS LFRQKRRPIYKSDVEMAHFR</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="11">
<INSDSeq>
<INSDSeq_length>501</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>

```

```

<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..501</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>mol_type</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier id="q22">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</
INSDSeq_feature-table>
<INSDSeq_sequence>KIEIVFPQHTTGDWKRVPHEYNYCPTSADKNSHGTQTGIPVELTMPKGLTTHQVDGFMCHSALWMTTCD
FRWYGPKYIITHSIHNEETDYQCLEAIKAYKDGVSFNPGFPPQSCGYGTVTDAEAHIVTVTPHSVKVDEYTG EWIDPHFIGGRCKGQ
ICETVHNSTKWFTSSDGESVCSQLFTLVGGTFFSDSEEITSMGLPETGIRSNYFPYVSTEGICKMPFCRKPGYKLNLDLWFQITDPD
LDKTVRDLPHIKDCDLSSSIVTPGEHATDISLISDVERILDYALCQNTWSKIEAGEPITPVDLSYLGPKNPGAGPVFTIINGSLHYF
MSKYLRVELESFVIPRMEGKVAGTRIVRQLWDQWFPFGEVEIGPNGVLKTKQGYKFPLHIIGTGEVDNDIKMERIVKHWEHPHIEAA
QTFLKKDDTEEVLYYGD TGVSKNPV ELVEGWFGWRSSIMGVLAVIIGFVILIFLIRLIGVLSSLF RQKRRPIYKSDVEMAHFR</I
NSDSeq_sequence>
</INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="12">
<INSDSeq>
<INSDSeq_length>512</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..512</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>mol_type</INSDQu
alifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier id="q24">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>MLRLFLFCFLALGAH SKFTIVFP HHQKGNWKNVPSTYHYCPSSSDQNWHNDLTGVSLHV KIPKSHKAIQ
ADGWMCHA AKWVTTCDFRWYGPKYIITHSIHMSPTLEQCKTSIEQTKQGVWINPGFPPQSCGYATVTD AEVVVVQATPHHVLVDEYT

```

```

GEWIDSQVLGGKCSKEVCQTVHNSTVWHADYKITGLCESNLASVDITFFSEDGQKTSLGKPNTGFRSNHFAYESGEKACRMQYCTQW
GIRLPSGVWFELVDKDLFQAAKLPECPRGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAKLPVSPVDLSYLAPKNPGSGP
AFTIINGTLKYFETRYIRVDISNPIIPHMVGTMSTGTTTERELWNDWYPYEDVEIGPNGVLKTPTGFKFPLYMIGHGMLDSDLHKSSQ
AQVFEHHPHAKDAASQLPDDETLFFGDTGLSKNPVELVEGWFSWKSTLASFFLIIGLVALIFIIRIIVAIRYKYKGRKTQKIYNDV
EMSR LGNK</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="13">
<INSDSeq>
<INSDSeq_length>496</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..496</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>mol_type</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier id="q26">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_feature-table>
<INSDSeq_sequence>KFTIVFPHHQKGNWKNVPSTYHYCPSSSDQNWHDLTGVSLHVKIPKSHKAIQADGWMCHAAKWVTTCD
FRWYGPKYITHSIHMSPTLEQCKTSIEQTKQGVWINPGFPPQSCGYATVTD AEVVVVQATPHHVLVDEYTG EWIDSQVLGGKCSKE
VCQTVHNSTVWHADYKITGLCESNLASVDITFFSEDGQKTSLGKPNTGFRSNHFAYESGEKACRMQYCTQWGIRLPSGVWFELVDKDL
FQAAKLPECPRGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAKLPVS
PVDLSYLAPKNPGSGPAFTIINGTLKYFETRYIRVDISNPIIPHMVGTMSTGTTTERELWNDWYPYEDVEIGPNGVLKTPTGFKFPLY
MIGHGMLDSDLHKSSQAQVFEHHPHAKDAASQLPDDETLFFGDTGLSKNPVELVEGWFSWKSTLASFFLIIGLVALIFIIRIIVAI
RYKYKGRKTQKIYNDVEMSR LGNK</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="14">
<INSDSeq>
<INSDSeq_length>523</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..523</INSDFeature_location>
<INSDFeature_qual>

```



```

<INSDQualifier>
<INSDQualifier_name>mol_type</INSDQualifier_name>
<INSDQualifier_value>protein</INSDQualifier_value>
</INSDQualifier>
<INSDQualifier id="q28">
<INSDQualifier_name>organism</INSDQualifier_name>
<INSDQualifier_value>synthetic construct</INSDQualifier_value>
</INSDQualifier>
</INSDFeature_qual>
</INSDFeature>
</INSDSeq_featu
re-table>
<INSDSeq_sequence>MKMKMVIAGLILCIGILPAIGKITISFPQSLKGDWRPVPKGYNYCPTSADKNLHGDLIDIGLRLRAPKS
FKGISADGWMCHAARWITTCDFRWYGPKYITHSIHSFRPSNDQCKEAIRLTNEGNWINPGFPPQSCGYASVTDSESVVVTVTKHQVL
VDEYSGSWIDSQFPGGSCCTSPICDTVHNSTLWHADHTLDSICDQEFVAMDAVLFTESGKFEEFGKPNSGIRSNYFPYESLKDVCQMD
FCKRKGFKLPSGVWFEIEDAEKSHKAQVELKIKRCPHGAVISAPNQNAADINLIMDVERILDYSLCQATWSKIQNKEALTPIDISYL
GPKNPGPGPAFTIINGTLHYFNTRYIRVDIAGPVTKEITGFVSGTSTSRVLWDQWFPYGENSIGPNLLKTASGYKYPLFMVGTGVL
DADIHKLGEATVIEHPHAKEAQKVDDSEVIFFGDTGVSKNPVEVVEGWFSGWRSSLMSIFGIILLIVCLVLIVRILIALKYCCVRH
KKRTIYKEDLEMGRIARRA</INSDSeq_sequence>
</INSDSeq>
</SequenceData>
<SequenceData sequenceIDNumber="15">
<INSDSeq>
<INSDSeq_length>502</INSDSeq_length>
<INSDSeq_moltype>AA</INSDSeq_moltype>
<INSDSeq_division>PAT</INSDSeq_division>
<INSDSeq_feature-table>
<INSDFeature>
<INSDFeature_key>source</INSDFeature_key>
<INSDFeature_location>1..502</INSDFeature_location>
<INSDFeature_qual>
<INSDQualifier>
<INSDQualifier_name>mol_typ

```

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of : Ronnie M. Russell et al.
Application No. : Filed herewith
Filed : Filed herewith
Group Art Unit : Not yet assigned
Confirmation No. : Filed herewith
Examiner : Not yet assigned

Title: PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME,
AND USES THEREOF

**REQUEST FOR PRIORITIZED EXAMINATION UNDER 37 C.F.R. §1.102(E) AND
PRELIMINARY AMENDMENT**

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Commissioner:

**REQUEST FOR PRIORITIZED EXAMINATION UNDER 37 C.F.R. § 1.102(e)(1) AND
PRELIMINARY AMENDMENT**

Prior to examination on the merits and calculation of fees, Applicant respectfully requests that the application be amended as follows. After the Preliminary Amendment is entered, and consistent with the accompanying certification, *the present Application does not contain more than four independent claims, more than thirty total claims, or any multiple dependent claims*. Applicant also submits herewith USPTO form PTO/AIA/424, which is the Certification and Request for Prioritized Examination under 37 C.F.R. § 1.102(e)(a).

Amendments to the Claims begin on page 2; and
Remarks begin on page 5.

AMENDMENTS TO THE CLAIMS

This listing of the claims will replace all prior versions, or listings of the claims in the present application.

1 - 70. (Cancelled)

71. (New) A polypeptide comprising an amino acid sequence that is at least 95% identical to the amino acid sequence of SEQ ID NO: 2 and comprises a mutation that corresponds to a mutation at position 182 of SEQ ID NO: 2.

72. (New) The polypeptide of claim 1, wherein the polypeptide comprises an amino acid sequence that is at least 96% identical to the amino acid sequence of SEQ ID NO: 2 and comprises a mutation that corresponds to a mutation at position 182 of SEQ ID NO: 2.

73. (New) The polypeptide of claim 1, wherein the polypeptide comprises an amino acid sequence that is at least 97% identical to the amino acid sequence of SEQ ID NO: 2 and comprises a mutation that corresponds to a mutation at position 182 of SEQ ID NO: 2.

74. (New) The polypeptide of claim 1, wherein the polypeptide comprises an amino acid sequence that is at least 98% identical to the amino acid sequence of SEQ ID NO: 2 and comprises a mutation that corresponds to a mutation at position 182 of SEQ ID NO: 2.

75. (New) The polypeptide of claim 1, wherein the polypeptide comprises an amino acid sequence that is at least 99% identical to the amino acid sequence of SEQ ID NO: 2 and comprises a mutation that corresponds to a mutation at position 182 of SEQ ID NO: 2.

76. (New) The polypeptide of claim 1, wherein the mutation at position is a I182E mutation as compared to SEQ ID NO: 2.

77. (New) The polypeptide of claim 1, wherein the mutation at position 182 comprises a I182D mutation as compared to SEQ ID NO: 2.

78. (New) The polypeptide of claim 1, wherein the polypeptide further comprises a mutation that corresponds to a mutation at position 214, a mutation at position 352, or a mutation at both position 214 and position 352 of SEQ ID NO: 2.

79. (New) The polypeptide of claim 1, wherein the polypeptide further comprises a T214N mutation, a T352A mutation, or a combination thereof, as compared to SEQ ID NO: 2.

80. (New) The polypeptide of claim 1, wherein the polypeptide further comprises a T214N mutation and a T352A mutation as compared to SEQ ID NO: 2.

81. (New) The polypeptide of claim 1, wherein the mutation at position 182 comprises a I182E or I182D mutation as compared to SEQ ID NO: 2 and wherein the polypeptide further comprises a T214N mutation, a T352A mutation, or a combination thereof, as compared to SEQ ID NO: 2.

82. (New) The polypeptide of claim 1, wherein the mutation at position 182 comprises a I182E mutation as compared to SEQ ID NO: 2 and wherein the polypeptide further comprises a T214N mutation, a T352A mutation, or a combination thereof, as compared to SEQ ID NO: 2.

83. (New) The polypeptide of claim 1, wherein the mutation at position 182 comprises a I182E or I182D mutation as compared to SEQ ID NO: 2 and wherein the polypeptide further comprises a T214N mutation and a T352A mutation as compared to SEQ ID NO: 2.

84. (New) The polypeptide of claim 1, wherein the mutation at position 182 comprises a I182E mutation as compared to SEQ ID NO: 2 and wherein the polypeptide further comprises a T214N mutation and a T352A mutation as compared to SEQ ID NO: 2.

85. (New) The polypeptide of claim 1, wherein the polypeptide comprises the amino acid sequence of SEQ ID NO: 23.
86. (New) A nucleic acid molecule encoding the polypeptide of claim 1.
87. (New) A vector comprising the nucleic acid molecule of claim 86.
88. (New) A cell comprising the nucleic acid molecule of claim 86.
89. (New) A viral particle comprising the polypeptide of claim 1.
90. (New) The viral particle of claim 89, wherein the viral particle is a lentivirus comprising the polypeptide.
91. (New) The viral particle of claim 89, wherein the viral particle further comprises a heterologous nucleic acid molecule encoding a heterologous molecule of interest.
92. (New) The viral particle of claim 91, wherein the heterologous molecule of interest is an siRNA, an shRNA, a non-coding RNA, a polypeptide, a viral payload, a viral genome, or a combination thereof.
93. (New) The viral particle of claim 91, wherein the heterologous molecule of interest is a chimeric antigen receptor ("CAR").

REMARKS

Claims 1-70 were pending. Claims 1-70 have been canceled herein without prejudice and solely for the purpose of claim fee calculation. New claims 71-93 have been added. Support for the new claims may be found in the as filed specification and claims. Accordingly, no new matter has been added. Applicant has not, by way of any amendment or cancelation of any claim, abandoned any subject matter that may have been removed from the claims, and Applicant reserves the right to reintroduce any such subject matter in the future. Upon entry of the present amendment, claims 71-93 will be pending.

It is respectfully requested that this Preliminary Amendment be entered before examination of this application. Applicant respectfully submits that all amendments are made without prejudice or disclaimer solely in order to expedite prosecution. Applicant further expressly reserves the right to file additional claims in further applications, whether such claims have been previously canceled, withdrawn, or are newly presented.

The present application does not contain more than four independent claims, more than thirty total claims, or any multiple dependent claims. Therefore, Applicant respectfully requests that the Request for Prioritized Examination under 37 C.F.R. § 1.102(e) be granted.

AUTHORIZATION

In view of the foregoing, Applicant respectfully submits that the claims are in condition for allowance. A notice of allowance is, therefore, earnestly solicited. If the Examiner believes for any reason that personal communication will expedite prosecution of this application, the Examiner is invited to telephone the undersigned directly.

The Commissioner is hereby authorized to charge any additional fees which may be required for this submission, or credit any overpayment, to Deposit Account No. 50-0436.

Respectfully submitted,

/Christopher R. Bodle /

Christopher R. Bodle, Ph.D.

Registration No. 80,876

TROUTMAN PEPPER HAMILTON SANDERS LLP

Suite 300

501 Grant Street

Pittsburgh, PA 15219-4429

Telephone: (412) 454-5061

Date: December 15, 2022

**PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME,
AND USES THEREOF**

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application Ser. No. 63/289,888 filed 15 December 2021, and U.S. Provisional Application Ser. No. 63/289,977 filed 15 December 2021, and U.S. Provisional Application Ser. No. 63/266,044 filed 27 December 2021, and U.S. Provisional Application Ser. No. 63/267,039 filed 21 January 2022, each of which is hereby incorporated by reference in their entirety.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

[0002] The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. Said XML copy, created on 28 November 2022, is named “148165.001401_SL.xml” and is 86.3 kilobytes in size.

BACKGROUND

[0003] Vesicular stomatitis virus (VSV) is an enveloped, negative-strand RNA virus that belongs to the Vesiculovirus genus of the Rhabdovirus family. It is an arbovirus which can infect insects, cattle, horses and pigs. VSV genome encodes five structural proteins among which include a single transmembrane glycoprotein (G). The glycoprotein is a classic type I membrane glycoprotein with an amino-terminal signal peptide, an ectodomain of about 450 amino acids, a single alpha helical transmembrane segment and a small intraviral carboxy-terminal domain. The signal peptide is cleaved in the lumen of the endoplasmic reticulum and the native glycoprotein consists in the ectodomain, the transmembrane domain and the intraviral domain.

[0004] G plays a critical role during the initial steps of virus infection (Molecular and Cellular Aspects of Rhabdovirus Entry. Viruses 4, 117-139.), which is hereby incorporated by reference in its entirety. First, it is responsible for virus attachment to specific receptors. After binding, virions enter the cell by a clathrin-mediated endocytic pathway. In the acidic environment of the endocytic vesicle, G triggers the fusion between the viral and endosomal membranes, which releases the genome in the cytosol for the subsequent steps of infection. Fusion is catalyzed by a low-pH-induced large structural transition from a pre-toward a post-fusion conformation which

are both trimeric (Roche, S., Bressanelli, S., Rey, F. A., and Gaudin, Y. (2006). Crystal structure of the low-pH form of the vesicular stomatitis virus glycoprotein G. *Science* 313, 187-191. Roche, S., Rey, F. A., Gaudin, Y., and Bressanelli, S. (2007). Structure of the prefusion form of the vesicular stomatitis virus glycoprotein g. *Science* 315, 843-848), each of which is hereby incorporated by reference in its entirety).

[0005] The polypeptide chain of G ectodomain folds into three distinct domains which are the fusion domain (FD), the pleckstrin homology domain (PHD), and the trimerization domain (TrD). During the structural transition, the FD, the PHD and the TrD retain their tertiary structure. Nevertheless, they undergo large rearrangements in their relative orientation due to secondary changes in hinge segments (S1 to S5) which refold during the low-pH induced conformational change (Roche et al., 2006; Roche et al., 2007).

[0006] It has been shown that low-density lipoprotein receptor (LDL-R) and other members of this receptor family serve as VSV receptors (Finkelshtein, D., Werman, A., Novick, D., Barak, S., and Rubinstein, M. (2013). LDL receptor and its family members serve as the cellular receptors for vesicular stomatitis virus. *Proceedings of the National Academy of Sciences of the United States of America* 110, 7306-7311, which is hereby incorporated by reference in its entirety). VSV-G can be used for pseudotyping other viruses and VSV-G-pseudotyped lentiviruses (VSV-G-LVs) exhibit the same broad tropism as VSV. However, this broad tropism can inhibit the selective targeting of specific cell types. Therefore, there is a need, for modified (mutated or mutant) VSV-G proteins that can be used to pseudotype viruses that abrogate its binding to the LDL receptor. The present embodiments, fulfill these needs as well as others.

[0007] The human CD7 molecule is a cell surface glycoprotein with a molecular weight of approximately 40 kDa belonging to the immunoglobulin superfamily. The CD7 molecule is mainly expressed on the surface of most thymocytes, more than 85% of the surface of peripheral blood T lymphocytes and the surface of natural killer cells. The embodiments disclosed herein provide for polypeptide and antibodies against CD7, compositions comprising the same, and uses thereof.

[0008] CD8 (cluster of differentiation 8) is a transmembrane glycoprotein which is a specific marker for a subclass of T-cells (which includes cytotoxic T-cells). Without wishing to be bound to a particular theory, CD8 assembles as either a heterodimer of the CD8 alpha and CD8 beta subunits or a CD8 alpha homodimer. The assembled dimeric CD8 complex acts as a co-receptor together with the T cell receptor (TCR) to recognize antigen presentation by MHC class

I cells. CD8 plays a role in the development of T-cells and activation of mature T-cells. Changes in T-cell localization can reflect the progression of an immune response and can occur over time. The embodiments disclosed herein provide for polypeptide and antibodies against CD8, compositions comprising the same, and uses thereof.

BRIEF SUMMARY

[0009] In some embodiments, a VSV-G polypeptide is provided. In some embodiments, the VSV-G polypeptide comprises a mutation at position 182 of SEQ ID NO: 2. In some embodiments, the polypeptide comprises an I182E or I182D mutation as compared to SEQ ID NO: 2. In some embodiments, the VSV-G polypeptide comprises the sequence of SEQ ID NO: 4. In some embodiments, the VSV-G polypeptide comprises the sequence of SEQ ID NO: 5.

[0010] In some embodiments, the VSV-G polypeptide further comprises a mutation in the VSV-G protein that corresponds to a position of 8, 10, 47, 209, and/or 354 as compared to SEQ ID NO: 2. In some embodiments, the VSV-G polypeptide comprises a H8A and/or a K47Q substitution. In some embodiments, the VSV-G polypeptide comprises a substitution at position 10 as compared to SEQ ID NO: 2. In some embodiments, the substitution at position 10 is Q10A, Q10R, or Q10K.

[0011] In some embodiments, a VSV-G polypeptide is provided. In some embodiments, the VSV-G polypeptide comprises a substitution at position I182 and at least one of T214, and T352 of SEQ ID NO: 2. In some embodiments, the VSV-G polypeptide comprises a substitution at position I182, T214, and T352 of SEQ ID NO: 2. In some embodiments, the VSV-G polypeptide comprises the sequence of SEQ ID NO: 22. In some embodiments, the VSV-G polypeptide comprises the sequence of SEQ ID NO: 23. In some embodiments, the VSV-G polypeptide comprises the sequence of SEQ ID NO: 24. In some embodiments, the VSV-G polypeptide comprises the sequence of SEQ ID NO: 25.

[0012] In some embodiments, a nucleic acid molecule is provided. In some embodiments, the nucleic acid molecule encodes for the VSV-G polypeptide as provided for herein.

[0013] In some embodiments, a vector is provided comprising a nucleic acid molecule as provided for herein.

[0014] In some embodiments, a plasmid is provided comprising a nucleic acid molecule as provided for herein.

[0015] In some embodiments, a viral particle is provided. In some embodiments, the viral particle comprises a VSV-G polypeptide as provided for herein. In some embodiments, the viral particle further comprises a targeting moiety. In some embodiments, the viral particle further comprises a nucleic acid molecule encoding for a heterologous molecule of interest.

[0016] In some embodiments, a method of delivering a heterologous molecule of interest to a target cell is provided. In some embodiments, the method comprises contacting the cell with a viral vector comprising a VSV-G protein as provided for herein, a targeting moiety that binds to the target cell, and a nucleic acid molecule encoding a heterologous molecule of interest. In some embodiments, the heterologous molecule of interest is a chimeric antigen receptor.

[0017] In some embodiments, a method of delivering a heterologous molecule of interest to a target cell in a subject is provided. In some embodiments, the method comprises contacting the cell with a viral vector comprising a VSV-G protein as provided for herein, a targeting moiety that binds to the target cell, and a nucleic acid molecule encoding a heterologous molecule of interest. In some embodiments, the heterologous molecule of interest is a chimeric antigen receptor.

[0018] In some embodiments, a method of treating cancer in a subject is provided. In some embodiments, the method comprises contacting the cell with a viral vector comprising a VSV-G protein as provided for herein, a targeting moiety that binds to the target cell, and a nucleic acid molecule encoding a heterologous molecule of interest. In some embodiments, the heterologous molecule of interest is a chimeric antigen receptor.

[0019] In some embodiments, the targeting moiety of the viral particle binds to CD7. In some embodiments, the targeting moiety comprises a polypeptide comprising a heavy chain CDR1 having the amino acid sequence of SEQ ID NO: 35, a heavy chain CDR2 having the amino acid sequence of SEQ ID NO: 36, a heavy chain CDR3 having the amino acid sequence of SEQ ID NO: 37, a light chain CDR1 having the amino acid sequence of SEQ ID NO: 38, a light chain CDR2 having the amino acid sequence of SEQ ID NO: 39, a light chain CDR3 having the amino acid sequence of SEQ ID NO: 40. In some embodiments, the targeting moiety comprises a polypeptide comprising a heavy chain variable region having the amino acid sequence of SEQ ID NO: 47 and a light chain variable region having the amino acid sequence of SEQ ID NO: 48.

[0020] In some embodiments, the targeting moiety that binds to CD7 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 51, at

least 95% sequence identity to SEQ ID NO: 51, at least 99% sequence identity to SEQ ID NO: 51, or a sequence as set forth in SEQ ID NO: 51. In some embodiments, the targeting moiety that binds to CD7 comprises a polypeptide comprising the sequence of SEQ ID NO: 51.

[0021] In some embodiments, the targeting moiety that binds to CD7 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 52, at least 95% sequence identity to SEQ ID NO: 52, at least 99% sequence identity to SEQ ID NO: 52, or a sequence as set forth in SEQ ID NO: 52. In some embodiments, the targeting moiety that binds to CD7 comprises a polypeptide comprising the sequence of SEQ ID NO: 52.

[0022] In some embodiments, the targeting moiety of the viral particle binds to CD8. In some embodiments, the targeting moiety comprises a polypeptide comprising a heavy chain CDR1 having the amino acid sequence of SEQ ID NO: 55, a heavy chain CDR2 having the amino acid sequence of SEQ ID NO: 56, a heavy chain CDR3 having the amino acid sequence of SEQ ID NO: 57, a light chain CDR1 having the amino acid sequence of SEQ ID NO: 58, a light chain CDR2 having the amino acid sequence of SEQ ID NO: 59, a light chain CDR3 having the amino acid sequence of SEQ ID NO: 60. In some embodiments, the targeting moiety comprises a polypeptide comprising a heavy chain variable region having the amino acid sequence of SEQ ID NO: 67 and a light chain variable region having the amino acid sequence of SEQ ID NO: 68.

[0023] In some embodiments, the targeting moiety that binds to CD8 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 69, at least 95% sequence identity to SEQ ID NO: 69, at least 99% sequence identity to SEQ ID NO: 69, or a sequence as set forth in SEQ ID NO: 69. In some embodiments, the targeting moiety that binds to CD7 comprises a polypeptide comprising the sequence of SEQ ID NO: 69.

[0024] In some embodiments, the targeting moiety that binds to CD7 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 70, at least 95% sequence identity to SEQ ID NO: 70, at least 99% sequence identity to SEQ ID NO: 70, or a sequence as set forth in SEQ ID NO: 70. In some embodiments, the targeting moiety that binds to CD7 comprises a polypeptide comprising the sequence of SEQ ID NO: 70.

[0025] In some embodiments, a method of delivering a heterologous molecule of interest to a target cell is provided. In some embodiments, the method comprises contacting the cell with a viral vector comprising a VSV-G protein as provided for herein, a targeting moiety as provided for herein that binds to the target cell, and a nucleic acid molecule encoding a heterologous

molecule of interest. In some embodiments the targeting moiety binds to CD7. In some embodiments, the targeting moiety binds to CB8. In some embodiments, the heterologous molecule of interest is a chimeric antigen receptor.

[0026] In some embodiments, a method of delivering a heterologous molecule of interest to a target cell in a subject is provided. In some embodiments, the method comprises contacting the cell with a viral vector comprising a VSV-G protein as provided for herein, a targeting moiety as provided for herein that binds to the target cell, and a nucleic acid molecule encoding a heterologous molecule of interest. In some embodiments the targeting moiety binds to CD7. In some embodiments, the targeting moiety binds to CB8. In some embodiments, the heterologous molecule of interest is a chimeric antigen receptor.

[0027] In some embodiments, a method of treating cancer in a subject is provided. In some embodiments, the method comprises contacting the cell with a viral vector comprising a VSV-G protein as provided for herein, a targeting moiety as provided for herein that binds to the target cell, and a nucleic acid molecule encoding a heterologous molecule of interest. In some embodiments the targeting moiety binds to CD7. In some embodiments, the targeting moiety binds to CB8. In some embodiments, the heterologous molecule of interest is a chimeric antigen receptor.

BRIEF DESCRIPTION OF FIGURES

[0028] FIG. 1 illustrates crystal structures of VSV-G bound to LDL-R. FIG. 1A illustrates the crystal structure of VSV-G bound to CR3 of the LDL-R. FIG. 1B illustrates the crystal structure of VSV-G bound to CR2 of the LDL-R.

[0029] FIG. 2 illustrates the effect of adding negatively charged amino acids to the VSV-G:LDL-R binding interface on native tropism and fusogenicity. FIG. 2A illustrates the titration of VSV-G constructs on SupT1 cells. FIG. 2B illustrates functional titer of each construct calculated from the titration in FIG. 2A.

[0030] FIG. 3 illustrates an alignment of the ectodomains of different VSV-G proteins from different strains.

[0031] FIG. 4 illustrates the effect of various VSV-G mutations on the serum stability of viral constructs in combination with a CD7 binder.

[0032] FIG. 5 illustrates the effect of various VSV-G mutations on the serum stability of viral constructs in combination with a CD7 binder.

[0033] FIG. 6A-L show flow cytometry data of human PBMCs transduced with exemplary vectors comprising CD7 binders as disclosed herein.

[0034] FIG. 6M shows flow cytometry data of human PBMCs transduced with exemplary vectors comprising CD7 binders as disclosed herein.

[0035] FIG. 7A-L shows flow cytometry data of non-human primate PBMCs transduced with exemplary vectors comprising CD7 binders as disclosed herein.

[0036] FIG. 7M shows flow cytometry data of non-human primate PBMCs transduced with exemplary vectors comprising CD7 binders as disclosed herein.

[0037] FIG. 8A shows flow cytometry data of human PBMCs transduced with exemplary vectors comprising CD8 binders as disclosed herein.

[0038] FIG. 8B shows flow cytometry data of non-human primate PBMCs transduced with exemplary vectors comprising CD8 binders as disclosed herein.

[0039] FIG. 9 illustrates GFP or CAR transduction of SupT1 cells utilizing VSV-G pseudotyped lentiviral particles harboring CD7 or CD8 binders as provided for herein. FIG. 9A illustrates GFP transduction. FIG. 9B illustrates CAR transduction.

[0040] FIG. 10A illustrates the ability of VSV-G* pseudotyped lentiviral particles harboring a CD7 binder to transduce SupT1 cells as well as human and non-human primate PBMCs. FIG. 10B illustrates the transduction of cells in the absence of the CD7 binder. FIG. 10C illustrates the transduction of human and non-human primate PBMCs in terms of MOI calculated from SupT1 titration. VSV-G* denotes VSV-G polypeptide comprising I182E, T214N, T352A, which corresponds to a mature polypeptide of SEQ ID NO: 23 or SEQ ID NO: 25 (with I196E, T230N and T368A mutations (with leader sequence and adjusted numbering)).

[0041] FIG. 11 illustrates the off target transduction of VSV-G* pseudotyped lentiviral particles harboring a CD7 binder in a panel of B-cell cell lines as compared to control SupT1 cells. FIG. 11A illustrates the data for GFP transduction. FIG. 11B illustrates the data for CAR20-T2A-GFP transduction. VSV-G* denotes a VSV-G polypeptide comprising I182E, T214N, and T352A mutations as provided for herein.

[0042] FIG. 12 illustrates the ability of VSV-G* pseudotyped lentiviral particles utilizing CD7 binders and a CD20-CAR transgene as provided for herein to kill Duadi lymphoma cells (FIG. 12A) or Raji lymphoma cells (FIG. 12B). CAR20 constructs utilized antigen binding domains comprising rituximab, or CD20AB1 SEQ ID NOs 75 or 76 as provided for herein.

[0043] FIG. 13 illustrates the ability of VSV-G* pseudotyped lentiviral particles utilizing CD7 binders and a CD20-CAR transgene as provided for herein to deplete B-cells *in vivo* in huCD34 NSG mice. B-cells were detected via assessment for CD20 (FIG. 13A) or CD19 (FIG. 13B).

[0044] FIG. 14 illustrates the ability of VSV-G* pseudotyped lentiviral particles utilizing CD7 binders and a CD20-CAR transgene as provided for herein to prevent tumor formation *in vivo*. FIG. 14A depicts the experimental design. FIG. 14B illustrates that mice receiving i.v. lentiviral particles as provided for herein prior to Raji tumor infusion had significantly lower tumor burden than mice receiving control (GFP) vector or untreated mice. Tumor burden measured via IVIS imaging.

[0045] FIG. 15 illustrates the ability of VSV-G* pseudotyped lentiviral particles utilizing CD7 binders and a CD20-CAR transgene as provided for herein to eliminate established Raji tumors *in vivo*. FIG. 15A depicts the experimental design. FIG. 15B illustrates that mice receiving i.v. lentiviral particles as provided for herein after 6 days after Raji tumor infusion had their tumor burdens diminished below the limit of detection. Tumor burden measured via IVIS imaging.

DETAILED DESCRIPTION

[0046] Provided for herein are mutant VSV-G proteins that can be used, for example, to pseudotype a virus, such as a lentivirus. In some embodiments, the pseudotyped viral-like particles are pseudotyped using viral glycoproteins of a vesicular stomatitis New Jersey virus strain, a vesicular stomatitis Indiana virus strain, a vesicular stomatitis Alagoas virus strain, a vesicular stomatitis Maraba virus strain, or a vesicular stomatitis Carajas virus strain.

[0047] The pseudotyped viruses comprising the mutant VSV-G proteins can be used in conjunction with a targeting moiety to facilitate the fusion of the pseudotyped virus with a specific cell or tissue based on the expression of the target on the cell or the tissue.

[0048] Unless defined otherwise, all technical and scientific terms have the same meaning as is commonly understood by one of ordinary skill in the art to which the embodiments disclosed belongs.

[0049] As used herein, the terms “a” or “an” means that “at least one” or “one or more” unless the context clearly indicates otherwise.

[0050] As used herein, the term “about” means that the numerical value is approximate

and small variations would not significantly affect the practice of the disclosed embodiments. Where a numerical limitation is used, unless indicated otherwise by the context, “about” means the numerical value can vary by $\pm 10\%$ and remain within the scope of the disclosed embodiments. Additionally, where a phrase recites “about x to y,” the term “about” modifies both x and y and can be used interchangeably with the phrase “about x to about y” unless context dictates differently.

[0051] “Activation,” as used herein in reference to a T cell, refers to the state of a T cell that has been sufficiently stimulated to induce detectable cellular proliferation. Activation can also be associated with induced cytokine production, and detectable effector functions. The term “activated T cells” refers to, among other things, T cells that are undergoing cell division.

[0052] As used herein, to “alleviate” a disease means reducing the severity of one or more symptoms of the disease.

[0053] The term “antigen” as used herein is defined as a molecule that provokes an immune response. This immune response may involve either antibody production, or the activation of specific immunologically-competent cells, or both. The skilled artisan will understand that any macromolecule, including virtually all proteins or peptides, can serve as an antigen. The term “antigen” can also refer to a molecule that an antibody or antibody-like molecule can bind to or is recognized by the antibody or antibody-like molecule.

[0054] The term “antibody molecule,” “antibody” or antigen binding domain, as that term is used herein, refers to a polypeptide, e.g., an immunoglobulin chain or fragment thereof, comprising at least one functional immunoglobulin variable domain sequence. An antibody molecule encompasses antibodies (e.g., full-length antibodies) and antibody fragments. In some embodiments, an antibody molecule comprises an antigen binding or functional fragment of a full-length antibody, or a full-length immunoglobulin chain. For example, a full-length antibody is an immunoglobulin (Ig) molecule (e.g., an IgG antibody) that is naturally occurring or formed by normal immunoglobulin gene fragment recombinatorial processes. In embodiments, an antibody molecule refers to an immunologically active, antigen binding portion of an immunoglobulin molecule, such as an antibody fragment. An antibody fragment, e.g., functional fragment, comprises a portion of an antibody, e.g., Fab, Fab', F(ab')₂, F(ab)₂, variable fragment (Fv), domain antibody (dAb), or single chain variable fragment (scFv). A functional antibody fragment binds to the same antigen as that recognized by the intact (e.g., full-length) antibody. The terms “antibody

fragment” or “functional fragment” also include isolated fragments consisting of the variable regions, such as the “Fv” fragments consisting of the variable regions of the heavy and light chains or recombinant single chain polypeptide molecules in which light and heavy variable regions are connected by a peptide linker (“scFv proteins”). In some embodiments, an antibody fragment does not include portions of antibodies without antigen binding activity, such as Fc fragments or single amino acid residues. Exemplary antibody molecules include full-length antibodies and antibody fragments, e.g., dAb (domain antibody), single chain, Fab, Fab’, and F(ab’)2 fragments, and single chain variable fragments (scFvs).

[0055] The term “antibody molecule” also encompasses whole or antigen binding fragments of domain, or single domain, antibodies, which can also be referred to as “sdAb” or “VHH.” Domain antibodies comprise either VH or VL that can act as stand-alone, antibody fragments. Additionally, domain antibodies include heavy-chain-only antibodies (HCAbs). Domain antibodies also include a CH2 domain of an IgG as the base scaffold into which CDR loops are grafted. It can also be generally defined as a polypeptide or protein comprising an amino acid sequence that is comprised of four framework regions interrupted by three complementarity determining regions. This is represented as FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4. sdAbs can be produced in camelids such as llamas, but can also be synthetically generated using techniques that are well known in the art. The numbering of the amino acid residues of a sdAb or polypeptide is according to the general numbering for VH domains given by Kabat et al. (“Sequence of proteins of immunological interest,” US Public Health Services, NIH Bethesda, MD, Publication No. 91, which is hereby incorporated by reference). According to this numbering, FR1 of a sdAb comprises the amino acid residues at positions 1-30, CDR1 of a sdAb comprises the amino acid residues at positions 31-36, FR2 of a sdAb comprises the amino acids at positions 36-49, CDR2 of a sdAb comprises the amino acid residues at positions 50-65, FR3 of a sdAb comprises the amino acid residues at positions 66-94, CDR3 of a sdAb comprises the amino acid residues at positions 95-102, and FR4 of a sdAb comprises the amino acid residues at positions 103-113. Domain antibodies are also described in WO2004041862 and WO2016065323, each of which is hereby incorporated by reference. The domain antibodies can be a targeting moiety as described herein.

[0056] As used herein, unless otherwise indicated, “antibody fragment” or “antigen binding fragment” refers to antigen binding fragments of antibodies, *i.e.* antibody fragments that

retain the ability to bind specifically to the antigen bound by the full-length antibody, *e.g.* fragments that retain one or more CDR regions. Examples of antibody binding fragments include, but are not limited to, Fab, Fab', F(ab')₂, and Fv fragments; diabodies; linear antibodies; single-chain antibody molecules, *e.g.*, sc-Fv; nanobodies (single domain antibody) and multispecific antibodies formed from antibody fragments.

[0057] A “Fab fragment” is comprised of one light chain and the C_H1 and variable regions of one heavy chain. The heavy chain of a Fab molecule cannot form a disulfide bond with another heavy chain molecule.

[0058] An “Fc” region contains two heavy chain fragments comprising the C_H2 and C_H3 domains of an antibody. The two heavy chain fragments are held together by two or more disulfide bonds and by hydrophobic interactions of the C_H3 domains.

[0059] A “Fab' fragment” contains one light chain and a portion or fragment of one heavy chain that contains the V_H domain and the C_H1 domain and also the region between the C_H1 and C_H2 domains, such that an interchain disulfide bond can be formed between the two heavy chains of two Fab' fragments to form a F(ab')₂ molecule.

[0060] A “F(ab')₂ fragment” contains two light chains and two heavy chains containing a portion of the constant region between the C_H1 and C_H2 domains, such that an interchain disulfide bond is formed between the two heavy chains. A F(ab')₂ fragment thus is composed of two Fab' fragments that are held together by a disulfide bond between the two heavy chains.

[0061] The “Fv region” comprises the variable regions from both the heavy and light chains, but lacks the constant regions.

[0062] The term “single-chain Fv” or “scFv” antibody refers to antibody fragments comprising the V_H and V_L domains of an antibody, wherein these domains are present in a single polypeptide chain. Generally, the Fv polypeptide further comprises a polypeptide linker between the V_H and V_L domains which enables the scFv to form the desired structure for antigen binding. For a review of scFv, see Pluckthun (1994) THE PHARMACOLOGY OF MONOCLONAL ANTIBODIES, vol. 113, Rosenberg and Moore eds. Springer-Verlag, New York, pp. 269-315. See also, International Patent Application Publication No. WO 88/01649 and U.S. Pat. Nos. 4,946, 778 and 5,260,203.

[0063] Antibody molecules can be monospecific (*e.g.*, monovalent or bivalent), bispecific (*e.g.*, bivalent, trivalent, tetravalent, pentavalent, or hexavalent), trispecific (*e.g.*, trivalent,

tetravalent, pentavalent, or hexavalent), or with higher orders of specificity (e.g., tetraspecific) and/or higher orders of valency beyond hexavalency. An antibody molecule can comprise a functional fragment of a light chain variable region and a functional fragment of a heavy chain variable region, or heavy and light chains may be fused together into a single polypeptide.

[0064] Furthermore, antigens can be derived from recombinant or genomic DNA. A skilled artisan will understand that any DNA, which comprises a nucleotide sequences or a partial nucleotide sequence encoding a protein that elicits an immune response therefore encodes an “antigen” as that term is used herein. Furthermore, one skilled in the art will understand that an antigen need not be encoded solely by a full length nucleotide sequence of a gene. Moreover, a skilled artisan will understand that an antigen need not be encoded by a “gene” at all. It is readily apparent that an antigen can be generated synthesized or can be derived from a biological sample. Such a biological sample can include, but is not limited to a tissue sample, a tumor sample, a cell or a biological fluid.

[0065] In certain embodiments, monoclonal antibodies herein also include camelized single domain antibodies. *See, e.g., Muyldermans et al. (2001) Trends Biochem. Sci. 26:230; Reichmann et al. (1999) J. Immunol. Methods 231:25; WO 94/04678; WO 94/25591; U.S. Pat. No. 6,005,079.* In one embodiment, the present invention provides single domain antibodies comprising two V_H domains with modifications such that single domain antibodies are formed.

[0066] As used herein, the term “diabodies” refers to small antibody fragments with two antigen-binding sites, which fragments comprise a heavy chain variable domain (V_H) connected to a light chain variable domain (V_L) in the same polypeptide chain (V_H-V_L or V_L-V_H). By using a linker that is too short to allow pairing between the two domains on the same chain, the domains are forced to pair with the complementary domains of another chain and create two antigen-binding sites. Diabodies are described more fully in, *e.g., EP 404,097; WO 93/11161; and Holliger et al. (1993) Proc. Natl. Acad. Sci. USA 90: 6444-6448.* For a review of engineered antibody variants generally see Holliger and Hudson (2005) *Nat. Biotechnol.* 23:1126-1136.

[0067] “Isolated antibody” refers to the purification status of a binding compound and in such context means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally, the term “isolated” is not intended to refer to a complete absence of such material or to an absence of water, buffers, or salts, unless they are present in amounts that substantially interfere

with experimental or therapeutic use of the binding compound as described herein.

[0068] The term “monoclonal antibody”, as used herein, refers to population of substantially homogeneous antibodies, *i.e.*, the antibody molecules comprising the population are identical in amino acid sequence except for possible naturally occurring mutations and/or post-translational modifications that may be present in minor amounts. In contrast, conventional (polyclonal) antibody preparations typically include a multitude of different antibodies having different amino acid sequences in their variable domains, particularly their CDRs, that are often specific for different epitopes. The modifier “monoclonal” indicates the character of the antibody as being obtained from a substantially homogeneous population of antibodies, and is not to be construed as requiring production of the antibody by any particular method. For example, the monoclonal antibodies to be used in accordance with the present invention may be made by the hybridoma method first described by Kohler *et al.* (1975) *Nature* 256: 495, or may be made by recombinant DNA methods (*see, e.g.*, U.S. Pat. No. 4,816,567). The “monoclonal antibodies” may also be isolated from phage antibody libraries using the techniques described in Clackson *et al.* (1991) *Nature* 352: 624-628 and Marks *et al.* (1991) *J. Mol. Biol.* 222: 581-597, for example. *See also* Presta (2005) *J. Allergy Clin. Immunol.* 116:731.

[0069] As used herein, the term “humanized antibody” refers to forms of antibodies that contain sequences from both human and non-human (*e.g.*, murine, rat) antibodies. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the hypervariable loops correspond to those of a non-human immunoglobulin, and all or substantially all of the framework (FR) regions are those of a human immunoglobulin sequence. The humanized antibody may optionally comprise at least a portion of a human immunoglobulin constant region (Fc).

[0070] The term “fully human antibody” refers to an antibody that comprises human immunoglobulin protein sequences only. A fully human antibody may contain murine carbohydrate chains if produced in a mouse, in a mouse cell, or in a hybridoma derived from a mouse cell. Similarly, “mouse antibody” refers to an antibody that comprises mouse immunoglobulin sequences only. Alternatively, a fully human antibody may contain rat carbohydrate chains if produced in a rat, in a rat cell, or in a hybridoma derived from a rat cell. Similarly, “rat antibody” refers to an antibody that comprises rat immunoglobulin sequences only.

[0071] In some embodiments, the basic antibody structural unit comprises a tetramer.

Each tetramer includes two identical pairs of polypeptide chains, each pair having one “light” (about 25 kDa) and one “heavy” chain (about 50-70 kDa). The amino-terminal portion of each chain includes a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The carboxy-terminal portion of the heavy chain may define a constant region primarily responsible for effector function. Typically, human light chains are classified as kappa and lambda light chains. Furthermore, human heavy chains are typically classified as mu, delta, gamma, alpha, or epsilon, and define the antibody's isotype as IgM, IgD, IgG, IgA, and IgE, respectively. Within light and heavy chains, the variable and constant regions are joined by a “J” region of about 12 or more amino acids, with the heavy chain also including a “D” region of about 10 more amino acids. See generally, *Fundamental Immunology* Ch. 7 (Paul, W., ed., 2nd ed. Raven Press, N.Y. (1989)).

[0072] The variable regions of each light/heavy chain pair form the antibody binding site. Thus, in general, an intact antibody has two binding sites. However, in bifunctional or bispecific antibodies, the two binding sites are, in general, not the same.

[0073] Typically, the variable domains of both the heavy and light chains comprise three hypervariable regions, also called complementarity determining regions (CDRs), located within relatively conserved framework regions (FR). The CDRs are usually aligned by the framework regions, enabling binding to a specific epitope. In general, from N-terminal to C-terminal, both light and heavy chains variable domains comprise FR1, CDR1, FR2, CDR2, FR3, CDR3 and FR4. The assignment of amino acids to each domain is, generally, in accordance with the definitions of *Sequences of Proteins of Immunological Interest*, Kabat, *et al.*; National Institutes of Health, Bethesda, Md. ; 5th ed.; NIH Publ. No. 91-3242 (1991); Kabat (1978) *Adv. Prot. Chem.* 32:1-75; Kabat, *et al.*, (1977) *J. Biol. Chem.* 252:6609-6616; Chothia, *et al.*, (1987) *J Mol. Biol.* 196:901-917 or Chothia, *et al.*, (1989) *Nature* 342:878-883.

[0074] As used herein, the term “hypervariable region” refers to the amino acid residues of an antibody that are responsible for antigen-binding. The hypervariable region comprises amino acid residues from a “complementarity determining region” or “CDR” (*i.e.* residues 24-34 (CDRL1), 50-56 (CDRL2) and 89-97 (CDRL3) in the light chain variable domain and residues 31-35 (CDRH1), 50-65 (CDRH2) and 95-102 (CDRH3) in the heavy chain variable domain; Kabat *et al.* (1991) *Sequences of Proteins of Immunological Interest*, 5th Ed. Public Health Service, National Institutes of Health, Bethesda, Md.) and/or those residues from a “hypervariable loop”

(i.e. residues 26-32 (CDRL1), 50-52 (CDRL2) and 91-96 (CDRL3) in the light chain variable domain and 26-32 (CDRH1), 53-55 (CDRH2) and 96-101 (CDRH3) in the heavy chain variable domain; Chothia and Lesk (1987) *J. Mol. Biol.* 196: 901-917). The CDRs can also be referenced according to the IMGT system for the identification of CDRs, which is described in Lefranc MP. Unique database numbering system for immunogenetic analysis. *Immunol Today* (1997) 18:509. As used herein, the term “framework” or “FR” residues refers to those variable domain residues other than the hypervariable region residues defined herein as CDR residues. CDRs provide the majority of contact residues for the binding of the antibody to the antigen or epitope. CDRs of interest can be derived from donor antibody variable heavy and light chain sequences, and include analogs of the naturally occurring CDRs, which analogs also share or retain the same antigen binding specificity and/or neutralizing ability as the donor antibody from which they were derived.

[0075] As used herein, “specific binding” or “immunospecific binding” or “binds immunospecifically” refer to antibody binding to a predetermined antigen at a much higher affinity than for another antigen(s). In some embodiments, the antibody binds the predetermined antigen with a dissociation constant (K_D) of 10^{-7} M or less, and such K_D is at least two-fold less than its K_D for binding to a non-specific antigen (e.g., BSA, casein, or another non-specific polypeptide).

[0076] Methods for determining mAb specificity and affinity by competitive inhibition can be found in Harlow, et al., *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1988), Colligan et al., eds., *Current Protocols in Immunology*, Greene Publishing Assoc. and Wiley Interscience, N.Y., (1992, 1993), and Muller, *Meth. Enzymol.* 92:589 601 (1983), which references are entirely incorporated herein by reference.

[0077]

[0078] As used herein, the term “individual” or “subject,” or “patient” used interchangeably, means any animal, including mammals, such as mice, rats, other rodents, rabbits, dogs, cats, swine, cattle, sheep, horses, or primates, such as humans. In some embodiments, the subject is a human. A subject that is “in need thereof” refers to a subject that has been identified as requiring treatment for the condition that is to be treated and is treated with the specific intent of treating such condition. The conditions can be, for example, any of the conditions described herein.

[0079] As used herein, the terms “comprising” (and any form of comprising, such as “comprise”, “comprises”, and “comprised”), “having” (and any form of having, such as “have”

and “has”), “including” (and any form of including, such as “includes” and “include”), or “containing” (and any form of containing, such as “contains” and “contain”), are inclusive or open-ended and do not exclude additional, unrecited elements or method steps. Any step or composition that uses the transitional phrase of “comprise” or “comprising” can also be said to describe the same with the transitional phrase of “consisting of” or “consists.”

[0080] As used herein, the term “contacting” means bringing together of two elements in an in vitro system or an in vivo system. For example, “contacting” virus or vector described herein with an individual or patient or cell includes the administration of the virus to an individual or patient, such as a human, as well as, for example, introducing a compound into a sample containing a cellular or purified preparation containing the cell.

[0081] As used herein, the term “fused” or “linked” when used in reference to a protein having different domains or heterologous sequences means that the protein domains are part of the same peptide chain that are connected to one another with either peptide bonds or other covalent bonding. The domains or section can be linked or fused directly to one another or another domain or peptide sequence can be between the two domains or sequences and such sequences would still be considered to be fused or linked to one another. In some embodiments, the various domains or proteins provided for herein are linked or fused directly to one another or a linker sequences, such as the glycine/serine sequences described herein link the two domains together.

[0082] A “disease” is a state of health of an animal wherein the animal cannot maintain homeostasis, and wherein if the disease is not ameliorated then the animal’s health continues to deteriorate. In contrast, a “disorder” in an animal is a state of health in which the animal is able to maintain homeostasis, but in which the animal’s state of health is less favorable than it would be in the absence of the disorder. Left untreated, a disorder does not necessarily cause a further decrease in the animal’s state of health.

[0083] “Effective amount” or “therapeutically effective amount” are used interchangeably herein, and refer to an amount of a compound, formulation, material, or composition, as described herein effective to achieve a particular biological result or provides a therapeutic or prophylactic benefit. Such results may include, but are not limited to an amount that when administered to a mammal, causes a detectable level of immune cell activation compared to the immune cell activation detected in the absence of the composition. The immune response can be readily assessed by a plethora of art-recognized methods. The skilled artisan would understand that the

amount of the composition administered herein varies and can be readily determined based on a number of factors such as the disease or condition being treated, the age and health and physical condition of the mammal being treated, the severity of the disease, the particular compound being administered, and the like.

[0084] “Encoding” refers to the inherent property of specific sequences of nucleotides in a polynucleotide, such as a gene, a cDNA, or an mRNA, to serve as templates for synthesis of other polymers and macromolecules in biological processes having either a defined sequence of nucleotides (i.e., rRNA, tRNA and mRNA) or a defined sequence of amino acids and the biological properties resulting therefrom. Thus, a gene encodes a protein if transcription and translation of mRNA corresponding to that gene produces the protein in a cell or other biological system. Both the coding strand, the nucleotide sequence of which is identical to the mRNA sequence and is usually provided in sequence listings, and the non-coding strand, used as the template for transcription of a gene or cDNA, can be referred to as encoding the protein or other product of that gene or cDNA.

[0085] “Expression vector” refers to a vector comprising a recombinant polynucleotide comprising expression control sequences operatively linked to a nucleotide sequence to be expressed. An expression vector comprises sufficient cis-acting elements for expression; other elements for expression can be supplied by the host cell or in an in vitro expression system. Expression vectors include all those known in the art, such as cosmids, plasmids (e.g., naked or contained in liposomes) and viruses (e.g., Sendai viruses, lentiviruses, retroviruses, adenoviruses, and adeno-associated viruses) that incorporate the recombinant polynucleotide.

[0086]

[0087] As used herein, the phrase “ex vivo” in reference to a cell being transduced, transfected or transformed ex vivo, refers to a cell being transduced, transfected or transformed outside of the subject, that is with the cells being removed from the subject before such cells are transduced, transfected or transformed.

[0088] “Identity” as used herein refers to the subunit sequence identity between two polymeric molecules such as between two nucleic acid or amino acid molecules, such as, between two polynucleotide or polypeptide molecules. When two amino acid sequences have the same residues at the same positions; e.g., if a position in each of two polypeptide molecules is occupied by an Arginine, then they are identical at that position. The identity or extent to which two amino

acid or two nucleic acid sequences have the same residues at the same positions in an alignment is often expressed as a percentage. The identity between two amino acid or two nucleic acid sequences is a direct function of the number of matching or identical positions; e.g., if half of the positions in two sequences are identical, the two sequences are 50% identical; if 90% of the positions (e.g., 9 of 10), are matched or identical, the two amino acids sequences are 90% identical.

[0089] By "substantially identical" is meant a polypeptide or nucleic acid molecule exhibiting at least 50% identity to a reference amino acid sequence (for example, any one of the amino acid sequences described herein) or nucleic acid sequence (for example, any one of the nucleic acid sequences described herein). In some embodiments, such a sequence is at least 60%, 80% or 85%, or 90%, 95% or even 99% identical at the amino acid level or nucleic acid to the sequence used for comparison. Other percentages of identity in reference to specific sequences are described herein.

[0090] Sequence identity can be measured/determined using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. In an exemplary approach to determining the degree of identity, a BLAST program may be used, with a probability score between e3 and e100 indicating a closely related sequence. In some embodiments, sequence identity is determined by using BLAST with the default settings.

[0091] To the extent embodiments provided for herein, includes composition comprising various proteins, these proteins may, in some instances, comprise amino acid sequences that have sequence identity to the amino acid sequences disclosed herein. Therefore, in certain embodiments, depending on the particular sequence, the degree of sequence identity is preferably greater than 50% (e.g. 60%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more) to the SEQ ID NOs disclosed herein. In addition to these percentages, other percentages of identity are provided for herein. Identity between polypeptides can be determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford

Molecular), using an affine gap search with parameters gap open penalty – 12 and gap extension penalty = 1.

[0092] These proteins may, compared to the disclosed proteins, include one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, etc.) conservative amino acid replacements i.e. replacements of one amino acid with another which has a related side chain. Genetically-encoded amino acids are generally divided into four families: (1) acidic i.e. aspartate, glutamate; (2) basic i.e. lysine, arginine, histidine; (3) non polar i.e. alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar i.e. glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In general, Substitution of single amino acids within these families does not have a major effect on the biological activity. The proteins may have one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, etc.) single amino acid deletions relative to the disclosed protein sequences. The proteins may also include one or more (e.g. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, etc.) insertions (e.g. each of 1, 2, 3, 4 or 5 amino acids) relative to the disclosed protein sequences.

[0093] As used herein, the phrase “in vivo” in reference to a cell being transduced, transfected or transformed in vivo, refers to a cell being transduced, transfected or transformed in the subject without the cells being removed from the subject before such cells are transduced, transfected or transformed.

[0094] “Isolated” means altered or removed from the natural state. For example, a nucleic acid or a peptide naturally present in a living animal is not “isolated,” but the same nucleic acid or peptide partially or completely separated from the coexisting materials of its natural state is “isolated.” An isolated nucleic acid or protein can exist in substantially purified form, or can exist in a non-native environment such as, for example, a host cell.

[0095] A “lentivirus” as used herein refers to a genus of the Retroviridae family that is able to infect non-dividing cells. Non-limiting examples of lentiviruses are HIV, SIV, and FIV. Vectors or viral-like particles derived from lentiviruses can be used to transduce cells and deliver genes or other molecules and have them expressed in a cell either in vitro (ex-vivo) or in vivo.

[0096] By the term “modified” as used herein, is meant a changed state or structure of a molecule or cell as provided herein. Molecules may be modified in many ways, including chemically, structurally, and functionally, such as mutations, substitutions, insertions, or deletions (e.g. internal deletions truncations). Cells may be modified through the introduction of nucleic

acids or the expression of heterologous proteins.

[0097] By the term “modulating,” as used herein, is meant mediating an increase or decrease in the level of a response in a subject compared with the level of a response in the subject in the absence of a treatment or compound, and/or compared with the level of a response in an otherwise identical but untreated subject. The term encompasses perturbing and/or affecting a native signal or response thereby mediating a beneficial therapeutic response in a subject, such as, a human.

[0098] As used herein, the following abbreviations for the commonly occurring nucleic acid bases are used: “A” refers to adenosine, “C” refers to cytosine, “G” refers to guanosine, “T” refers to thymidine, and “U” refers to uridine.

[0099] Unless otherwise specified, a “nucleotide sequence encoding an amino acid sequence” includes all nucleotide sequences that are degenerate versions of each other and that encode the same amino acid sequence. The phrase nucleotide sequence that encodes a protein or an RNA may also include introns to the extent that the nucleotide sequence encoding the protein may in some version contain an intron(s).

[0100] The “Nipah virus” (NiV) is member of the family Paramyxoviridae, genus Henipavirus. Nipah virus is an enveloped virus with negative-stranded polarity and a non-segmented RNA genome consisting of helical nucleocapsids. Two strains of Nipah virus include, but are not limited to, the Malaysian (MY) and the Bangladesh (BD) strains.

[0101] The term “oligonucleotide” typically refers to short polynucleotides. It will be understood that when a nucleotide sequence is represented by a DNA sequence (i.e., A, T, C, G), this also provides the corresponding RNA sequence (i.e., A, U, C, G) in which “U” replaces “T.”

[0102] “Parenteral” administration of a composition includes, e.g., subcutaneous (s.c.), intravenous (i.v.), intramuscular (i.m.), or intrasternal injection, or infusion techniques.

[0103] The term “polynucleotide” as used herein is defined as a chain of nucleotides. Furthermore, nucleic acids are polymers of nucleotides. Thus, the terms “nucleic acids” and “polynucleotides” as used herein are interchangeable. As used herein polynucleotides include, but are not limited to, all nucleic acid sequences which are obtained by any methods available in the art, including, without limitation, recombinant methods, i.e., the cloning of nucleic acid sequences from a recombinant library or a cell genome, using cloning technology and PCR, and the like, and by synthetic means.

[0104] As used herein, the terms “peptide,” “polypeptide,” and “protein” are used interchangeably, and refer to a compound comprised of a plurality of amino acid residues covalently linked by peptide bonds. As used herein, the term refers to both short chains, which also commonly are referred to in the art as peptides, oligopeptides and oligomers, for example, and to longer chains, which generally are referred to in the art as proteins, of which there are many types. “Polypeptides” include, for example, biologically active fragments, substantially homologous polypeptides, oligopeptides, homodimers, heterodimers, variants of polypeptides, modified polypeptides, derivatives, analogs, fusion proteins, among others. The polypeptides include natural peptides, recombinant peptides, synthetic peptides, or a combination thereof.

[0105] The term “pseudotyped” or “pseudotyped viral particle”, as used herein, refers to a viral particle bearing glycoproteins derived from other viruses having envelopes or a viral vector encoding envelope glycoproteins from a virus that is different from the parental virus. The host range of the vector particles can thus be expanded or altered depending on the type of cell surface receptor used by the glycoprotein. For example, a virus can be pseudotyped with a VSV-G mutant protein as provided for herein or other virus glycoproteins, such as those provided for herein.

[0106] By the term “specifically binds,” as used herein with respect to an antibody, is meant an antibody which recognizes a specific antigen, but does not substantially recognize or bind other molecules in a sample. For example, an antibody that specifically binds to an antigen from one species may also bind to that antigen from one or more species. But, such cross-species reactivity does not itself alter the classification of an antibody as specific. In another example, an antibody that specifically binds to an antigen may also bind to different allelic forms of the antigen. However, such cross reactivity does not itself alter the classification of an antibody as specific. In some instances, the terms “specific binding” or “specifically binding,” can be used in reference to the interaction of an antibody, a protein, or a peptide with a second chemical species, to mean that the interaction is dependent upon the presence of a particular structure (e.g., an antigenic determinant or epitope) on the chemical species; for example, an antibody recognizes and binds to a specific protein structure rather than to proteins generally. If an antibody is specific for epitope “A”, the presence of a molecule containing epitope A (or free, unlabeled A), in a reaction containing labeled “A” and the antibody, will reduce the amount of labeled A bound to the antibody. In some embodiments, the targeting moieties described herein that can be used to target the viral particles comprising the mutant VSV-G protein can specifically bind to their target.

[0107] The term “subject” includes living organisms, including those in which an immune response can be elicited (e.g., mammals). A “subject” or “patient,” as used therein, may be a human or non-human mammal. Non-human mammals include, for example, livestock and pets, such as ovine, bovine, porcine, canine, non-human primates, feline and murine mammals. In some embodiments, the subject is human.

[0108] The term “therapeutic” as used herein means a treatment and/or prophylaxis. A therapeutic effect is obtained by suppression, remission, or eradication of a disease state.

[0109] The term “transfected” or “transformed” or “transduced” as used herein refers to a process by which exogenous nucleic acid is transferred or introduced into a cell. A “transfected” or “transformed” or “transduced” cell is one which has been transfected, transformed or transduced with exogenous nucleic acid. The cell includes the primary subject cell and its progeny. In some embodiments, the transfection, transformation, or transduction is performed or occurs in vivo.

[0110] To “treat” a disease as the term is used herein, means to reduce the frequency or severity of at least one sign or symptom of a disease or disorder experienced by a subject. As used herein, “inhibit” or “treat” or “treatment” also includes a postponement of development of the symptoms associated with a disorder and/or a reduction in the severity of the symptoms of such disorder. The terms further include ameliorating existing uncontrolled or unwanted symptoms, preventing additional symptoms, and ameliorating or preventing the underlying causes of such symptoms. Thus, the terms denote that a beneficial result has been conferred on a vertebrate subject with a disorder, disease or symptom, or with the potential to develop such a disorder, disease or symptom.

[0111] A “vector” is a composition of matter which comprises an isolated nucleic acid encoding a protein or a peptide. Numerous vectors are known in the art including, but not limited to, linear polynucleotides, plasmids, DNA, and RNA. Examples of viral vectors include, but are not limited to, Sendai viral vectors, adenoviral vectors, adeno-associated virus vectors, retroviral vectors, lentiviral vectors, and the like.

[0112] A “carrier” or “delivery vehicle” includes viral particles, viruses, polylysine compounds, and liposomes, which facilitate transfer of nucleic acid into cells. A carrier or delivery vehicle can also be used to deliver a protein or peptide to a cell.

[0113] Ranges: throughout this disclosure, various aspects of the embodiments can be presented in a range format. It should be understood that the description in range format is merely

for convenience and brevity and should not be construed as an inflexible limitation. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 2.7, 3, 4, 5, 5.3, and 6. This applies regardless of the breadth of the range. Unless otherwise explicitly stated to the contrary, a range that is disclosed also includes the endpoints of the range.

[0114] Without being bound to any particular theory, the embodiments provided for herein have been found to show that a mutant VSV-G protein comprising a mutation at position 182 can be used to pseudotype a virus and transduce a cell when the virus comprises a targeting moiety. This mutation inhibits or reduces the VSV-G affinity to its natural co-receptor, the LDL-R. The mutant VSV-G proteins as provided can be used, in some embodiments, to transduce a target cell and deliver a heterologous molecule to the targeted cells.

[0115] In some embodiments, a VSV-G protein is provided that comprises a mutation at position 198 as compared to SEQ ID NO: 1 or at position 182 as compared to SEQ ID NO: 2. SEQ ID NO: 1 is the full length protein and SEQ ID NO: 2 is the ectodomain of the VSV-G protein. The 16-mer signal peptide of MKCLLYLAFLFIGVNC (SEQ ID NO: 26) as shown at the N-terminus of SEQ ID NO: 1 is cleaved leaving a protein of SEQ ID NO: 2. Thus, although a mutation may be referred to in the context of SEQ ID NO: 2, it should be understood to also be made in the context of SEQ ID NO: 1, which contains the leader sequence, and thus would be a position number that is 16 more than the position recited for SEQ ID NO: 2. In some embodiments, the mutation inhibits or decreases the binding of the VSV-G protein to the LDL receptor (LDL-R). In some embodiments, the mutation is a I182D mutation as compared to SEQ ID NO: 2. In some embodiments, the mutation is a I182E mutation as compared to SEQ ID NO: 2.

[0116] In some embodiments, a VSV-G protein is provided that comprises a mutation at position 198 as compared to SEQ ID NO: 10 or at position 182 as compared to SEQ ID NO: 11. SEQ ID NO: 10 is the full length protein and SEQ ID NO: 11 is the ectodomain of the VSV-G protein. The 16-mer signal peptide of MLSYLIFALVVSPILG (SEQ ID NO: 27) as shown at the N-terminus of SEQ ID NO: 10 is cleaved leaving a protein of SEQ ID NO: 11. Thus, although a mutation may be referred to in the context of SEQ ID NO: 11, it should be understood to also be

made in the context of SEQ ID NO: 10, which contains the leader sequence, and thus would be a position number that is 16 more than the position recited for SEQ ID NO: 11. In some embodiments, the mutation inhibits or decreases the binding of the VSV-G protein to the LDL receptor (LDL-R). In some embodiments, the mutation is a T182D mutation as compared to SEQ ID NO: 11. In some embodiments, the mutation is a T182E mutation as compared to SEQ ID NO: 11.

[0117] In some embodiments, a VSV-G protein is provided that comprises a mutation at position 198 as compared to SEQ ID NO: 12 or at position 182 as compared to SEQ ID NO: 13. SEQ ID NO: 12 is the full length protein and SEQ ID NO: 13 is the ectodomain of the VSV-G protein. The 16-mer signal peptide of MLRLFLFCFLALGAHS (SEQ ID NO: 28) as shown at the N-terminus of SEQ ID NO: 12 is cleaved leaving a protein of SEQ ID NO: 13. Thus, although a mutation may be referred to in the context of SEQ ID NO: 13, it should be understood to also be made in the context of SEQ ID NO: 12, which contains the leader sequence, and thus would be a position number that is 16 more than the position recited for SEQ ID NO: 13. In some embodiments, the mutation inhibits or decreases the binding of the VSV-G protein to the LDL receptor (LDL-R). In some embodiments, the mutation is a A182D mutation as compared to SEQ ID NO: 13. In some embodiments, the mutation is a A182E mutation as compared to SEQ ID NO: 13.

[0118] In some embodiments, a VSV-G protein is provided that comprises a mutation at position 203 as compared to SEQ ID NO: 14 or at position 182 as compared to SEQ ID NO: 15. SEQ ID NO: 14 is the full length protein and SEQ ID NO: 15 is the ectodomain of the VSV-G protein. The 21-mer signal peptide of MKMKMVIAGLILCIGILPAIG (SEQ ID NO: 29) as shown at the N-terminus of SEQ ID NO: 14 is cleaved leaving a protein of SEQ ID NO: 15. Thus, although a mutation may be referred to in the context of SEQ ID NO: 15, it should be understood to also be made in the context of SEQ ID NO: 14, which contains the leader sequence, and thus would be a position number that is 21 more than the position recited for SEQ ID NO: 15. In some embodiments, the mutation inhibits or decreases the binding of the VSV-G protein to the LDL receptor (LDL-R). In some embodiments, the mutation is a V182D mutation as compared to SEQ ID NO: 15. In some embodiments, the mutation is a V182E mutation as compared to SEQ ID NO: 15.

[0119] In some embodiments, a VSV-G protein is provided that comprises a mutation at

position 199 as compared to SEQ ID NO: 16 or at position 182 as compared to SEQ ID NO: 17. SEQ ID NO: 16 is the full length protein and SEQ ID NO: 17 is the ectodomain of the VSV-G protein. The 17-mer signal peptide of MTPAFILCMLLAGSSWA (SEQ ID NO: 30) as shown at the N-terminus of SEQ ID NO: 16 is cleaved leaving a protein of SEQ ID NO: 17. Thus, although a mutation may be referred to in the context of SEQ ID NO: 17, it should be understood to also be made in the context of SEQ ID NO: 16, which contains the leader sequence, and thus would be a position number that is 17 more than the position recited for SEQ ID NO: 17. In some embodiments, the mutation inhibits or decreases the binding of the VSV-G protein to the LDL receptor (LDL-R). In some embodiments, the mutation is a V182D mutation as compared to SEQ ID NO: 17. In some embodiments, the mutation is a V182E mutation as compared to SEQ ID NO: 17.

[0120] In some embodiments, a VSV-G protein is provided that comprises a mutation at position 199 as compared to SEQ ID NO: 18 or at position 182 as compared to SEQ ID NO: 19. SEQ ID NO: 18 is the full length protein and SEQ ID NO: 19 is the ectodomain of the VSV-G protein. The 17-mer signal peptide of MNFLLLTFFIVLPLCSHA (SEQ ID NO: 31) as shown at the N-terminus of SEQ ID NO: 18 is cleaved leaving a protein of SEQ ID NO: 19. Thus, although a mutation may be referred to in the context of SEQ ID NO: 19, it should be understood to also be made in the context of SEQ ID NO: 18, which contains the leader sequence, and thus would be a position number that is 17 more than the position recited for SEQ ID NO: 19. In some embodiments, the mutation inhibits or decreases the binding of the VSV-G protein to the LDL receptor (LDL-R). In some embodiments, the mutation is a V182D mutation as compared to SEQ ID NO: 19. In some embodiments, the mutation is a V182E mutation as compared to SEQ ID NO: 19.

[0121] In some embodiments, a VSV-G protein is provided that comprises a mutation at position 199 as compared to SEQ ID NO: 20 or at position 182 as compared to SEQ ID NO: 21. SEQ ID NO: 20 is the full length protein and SEQ ID NO: 21 is the ectodomain of the VSV-G protein. The 17-mer signal peptide of MLVLYLLLSLLALGAQC (SEQ ID NO: 32) as shown at the N-terminus of SEQ ID NO: 20 is cleaved leaving a protein of SEQ ID NO: 21. Thus, although a mutation may be referred to in the context of SEQ ID NO: 21, it should be understood to also be made in the context of SEQ ID NO: 20, which contains the leader sequence, and thus would be a position number that is 17 more than the position recited for SEQ ID NO: 21. In some

embodiments, the mutation inhibits or decreases the binding of the VSV-G protein to the LDL receptor (LDL-R). In some embodiments, the mutation is a I182D mutation as compared to SEQ ID NO: 21. In some embodiments, the mutation is a I182E mutation as compared to SEQ ID NO: 21.

[0122] As used herein, when a polypeptide is said to have a mutation as compared to a reference sequence, such comparison is based on an alignment such as using BlastP or ClustalW or ClustalOmega alignment software using default parameters. For example, position 182 can be found in SEQ ID NO: 2 and also as compared to the other strains as illustrated in FIG. 3. FIG. 3 illustrates a clustal alignment of the wild-type sequences of the ectodomains of the various strains of the VSV-G protein. The residue that is bolded and underlined are the residues that align to position 182 of SEQ ID NO: 2 of the various strains. SEQ ID NO: 2 refers to ectodomain of the VSV-G protein of the Indiana strain. SEQ ID NO: 11 refers to ectodomain of the VSV-G protein of the New Jersey strain. SEQ ID NO: 13 refers to ectodomain of the VSV-G protein of the Marraba strain. SEQ ID NO: 15 refers to ectodomain of the VSV-G protein of the Carajas strain. SEQ ID NO: 17 refers to ectodomain of the VSV-G protein of the Alagoa strain. SEQ ID NO: 19 refers to ectodomain of the VSV-G protein of the Cocal strain. SEQ ID NO: 21 refers to ectodomain of the VSV-G protein of the Morreton strain. Accordingly, the residue that aligns to residues 182 as compared to SEQ ID NO: 2 can also be mutated as provided for herein.

[0123] In some embodiments, the mutation at position 182 as compared to SEQ ID NO: 2 is not an alanine. In some embodiments, the mutation at position 182 as compared to SEQ ID NO: 2 is not a valine.

[0124] In some embodiments, the mutation at position 182 as compared to SEQ ID NO: 2 is I182S, I182H, I182T, I182Q, or I182N. In some embodiments, the mutation at position 182 as compared to SEQ ID NO: 11 is T182S, T182H, T182Q, or T182N. In some embodiments, the mutation at position 182 as compared to SEQ ID NO: 13 is A182S, A182H, A182T, A182Q, or A182N. In some embodiments, the mutation at position 182 as compared to SEQ ID NO: 15 is V182S, V182H, V182T, V182Q, or V182N. In some embodiments, the mutation at position 182 as compared to SEQ ID NO: 17 is V182S, V182H, V182T, V182Q, or V182N. In some embodiments, the mutation at position 182 as compared to SEQ ID NO: 19 is V182S, V182H, V182T, V182Q, or V182N. In some embodiments, the mutation at position 182 as compared to SEQ ID NO: 21 is I182S, I182H, I182T, I182Q, or I182N. In some embodiments, the mutation at

position 182 is not a hydrophobic residue. In some embodiments, the mutation at position 182 is a charged residue. In some embodiments, the mutation at position 182 is a negatively charged residue.

[0125] Although, the mutations may be described in reference to SEQ ID NO: 1 or SEQ ID NO: 2, which is the VSV-G protein from the Indiana strain, the mutation can also be used in other strains of the VSV-G protein. For example, the mutation can be made in the New Jersey Strain of VSV-G, the Marraba strain of VSV-G, the Carajas strain of VSV-G, the Alagoa strain of VSV-G, the Cocal strain of VSV-G, or the Morreton strain of VSV-G. In some embodiments, the sequences of each are as provided herein. Examples of these can be found, for example in U.S. Patent Application Publication No. 20200216502, which is hereby incorporated by reference. For example, the wild-type full length or ectodomain of the New Jersey Strain of VSV-G are SEQ ID NO: 10 and SEQ ID NO: 11, respectively, the wild-type full length or ectodomain of Marraba strain of VSV-G are SEQ ID NO: 12 and SEQ ID NO: 13, respectively, the wild-type full length or ectodomain of Carajas strain of VSV-G are SEQ ID NO: 14 and SEQ ID NO: 15, respectively, the wild-type full length or ectodomain of Alagoa strain of VSV-G are SEQ ID NO: 16 and SEQ ID NO: 17, respectively, the wild-type full length or ectodomain of Cocal strain of VSV-G are SEQ ID NO: 18 and SEQ ID NO: 19, respectively, or the wild-type full length or ectodomain of Morreton strain of VSV-G are SEQ ID NO: 20 and SEQ ID NO: 21, respectively.

[0126] A VSV-G protein comprising a mutation at position 182 as compared to SEQ ID NO: 2 can also comprise other mutations, such as those described in U.S. Patent Application Publication No. 2020/0216502, which is hereby incorporated by reference in its entirety. For example, the VSV-G protein can comprise a mutation at a position that corresponds to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2.

[0127] In some embodiments, the substitution at position 8 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except Y. In some embodiments, the substitution at position 8 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except H. In some embodiments, the substitution at position 8 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except Q. In some embodiments, the substitution at position 8 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except Y or H or Q. In some embodiments, the substitution at position

8 is selected from alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, glycine, isoleucine, valine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, or tryptophan. In some embodiments, the substitution at position 8 is selected from 8A, 8I, 8V, 8L, and the like. In some embodiments, the substitution at position 8 is selected from H8A, H8I, H8V, H8L, and the like.

[0128] In some embodiments, the substitution at position 209 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except H. In some embodiments, the substitution at position 209 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except Y. In some embodiments, the substitution at position 209 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except H or Y. In some embodiments, the substitution at position 209 is selected from alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, isoleucine, valine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, or tryptophan.

[0129] In some embodiments, the substitution at position 47 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except K. In some embodiments, the substitution at position 47 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except R. In some embodiments, the substitution at position 47 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except K or R. In some embodiments, the substitution at position 47 is selected from alanine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, histidine, isoleucine, valine, leucine, methionine, phenylalanine, proline, serine, threonine, tryptophan, or tyrosine. In some embodiments, the substitution at position 47 is selected from A, G, F, N, or Q. In some embodiments, the substitution at position 47 is selected from Q or N.

[0130] In some embodiments, the substitution at position 354 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except K. In some embodiments, the substitution at position 354 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except R. In some embodiments, the substitution at position 354 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except K or R. In some embodiments,

the substitution at position 354 is selected from alanine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, histidine, isoleucine, valine, leucine, methionine, phenylalanine, proline, serine, threonine, tryptophan, or tyrosine. In some embodiments, the substitution at position 354 is selected from A, G, F, N, or Q.

[0131] In some embodiments, the substitution is at position 47 or at position 354, or at both positions 47 and 354 are, independently, substituted by A, G, F, N or Q. In some embodiments, the substitution at position 47 or at position 354, or at both positions 47 and 354 is, independently, A or Q or N.

[0132] In some embodiments, the VSV-G protein can comprise a mutation at a position that corresponds to positions of 8, 47, 209 and/or 354. In some embodiments, the VSV-G protein can comprise a mutation at a position that corresponds to position 8 as provided for herein, at position 47 as provided for herein, at position 209 as provided for herein, and/or position 354 as provided for herein. In some embodiments, the VSV-G protein can comprise a mutation at a position that corresponds to position 8 as provided for herein and at position 47 as provided for herein. In some embodiments, the VSV-G protein can comprise a mutation at a position that corresponds to position 8 as provided for herein and at position 209 as provided for herein. In some embodiments, the VSV-G protein can comprise a mutation at a position that corresponds to position 8 as provided for herein and at position 354 as provided for herein. In some embodiments, the VSV-G protein can comprise a mutation at a position that corresponds to position 47 as provided for herein and at position 209 as provided for herein. In some embodiments, the VSV-G protein can comprise a mutation at a position that corresponds to position 47 as provided for herein and at position 354 as provided for herein. In some embodiments, the VSV-G protein can comprise a mutation at a position that corresponds to position 209 as provided for herein and at position 354 as provided for herein. In some embodiments, the VSV-G protein can comprise a mutation at a position that corresponds to position 8 as provided for herein, at position 47 as provided for herein, and position 209 as provided for herein. In some embodiments, the VSV-G protein can comprise a mutation at a position that corresponds to position 8 as provided for herein, at position 47 as provided for herein, and position 354 as provided for herein. In some embodiments, the VSV-G protein can comprise a mutation at a position that corresponds to position 8 as provided for herein, at position 209 as provided for herein, and position 354 as provided for herein. In some embodiments, the VSV-G protein can comprise a mutation at a

position that corresponds to position 47 as provided for herein, at position 209 as provided for herein, and position 354 as provided for herein. In some embodiments, the VSV-G protein can comprise a mutation at a position that corresponds to position 8 as provided for herein, at position 47 as provided for herein, at position 209 as provided for herein, and at position 354 as provided for herein.

[0133] In some embodiments, the substitution at position 8 is an alanine, i.e., H8A.

[0134] In some embodiments, the substitution at position 47 is Q or N or A, i.e., K47Q or K47N or K47A.

[0135] In some embodiments, the protein comprises a substitution at position 8 and/or a substitution at position 47. In some embodiments, the protein comprises a substitution at position 8 and a substitution at position 47. In some embodiments, the protein comprises a substitution at position 8 or a substitution at position 47. In some embodiments, the substitution at position 8 and/or a substitution at position 47 comprises a H8A and/or K47Q mutation.

[0136] In some embodiments, the protein comprises a mutation (substitution) at position 10. In some embodiments, the substitution/mutation is Q10A, Q10R, or Q10K.

[0137] In some non-limiting embodiments, the substitution at position 209 is A, i.e., Y209A.

[0138] In some non-limiting embodiments, the substitution at position 354 is A or Q, i.e., R354A or R354Q.

[0139] In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 11 can also comprise other mutations, such as those described in U.S. Patent Application Publication No. 2020/0216502. For example, the VSV-G protein can comprise a mutation at a position analogous to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2. In some embodiments, the mutation in SEQ ID NO: 11 at a position analogous to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2 are as provided for herein.

[0140] In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 13 can also comprise other mutations, such as those described in U.S. Patent Application Publication No. 2020/0216502. For example, the VSV-G protein can comprise a mutation at a position analogous to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2. In some embodiments, the mutation in SEQ ID NO: 13 at a position analogous to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2 are as provided for herein.

[0141] In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 15 can also comprise other mutations, such as those described in U.S. Patent Application Publication No. 2020/0216502. For example, the VSV-G protein can comprise a mutation at a position analogous to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2. In some embodiments, the mutation in SEQ ID NO: 15 at a position analogous to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2 are as provided for herein.

[0142] In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 17 can also comprise other mutations, such as those described in U.S. Patent Application Publication No. 2020/0216502. For example, the VSV-G protein can comprise a mutation at a position analogous to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2. In some embodiments, the mutation in SEQ ID NO: 17 at a position analogous to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2 are as provided for herein.

[0143] In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 19 can also comprise other mutations, such as those described in U.S. Patent Application Publication No. 2020/0216502. For example, the VSV-G protein can comprise a mutation at a position analogous to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2. In some embodiments, the mutation in SEQ ID NO: 19 at a position analogous to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2 are as provided for herein.

[0144] In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 21 can also comprise other mutations, such as those described in U.S. Patent Application Publication No. 2020/0216502. For example, the VSV-G protein can comprise a mutation at a position analogous to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2. In some embodiments, the mutation in SEQ ID NO: 21 at a position analogous to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2 are as provided for herein.

[0145] In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 2 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 2 (or SEQ ID NO: 1 if using the full length protein). In some embodiments, the protein is at least, or about, 80% identical to SEQ ID NO: 2. In some embodiments, the protein is at least, or about, 85% identical to SEQ ID NO: 2. In some embodiments, the protein is at least, or about,

90% identical to SEQ ID NO: 2. In some embodiments, the protein is at least, or about, 95% identical to SEQ ID NO: 2. In some embodiments, the mutation at position 182 is I182D or I182E mutation. In some embodiments, the mutation is I182D. In some embodiments, the mutation is I182E mutation. In some embodiments, the mutation is I182S, I182H, I182T, I182Q, or I182N mutation.

[0146] In some embodiments, a VSV-G is provided wherein the protein comprises the amino acid sequence of SEQ ID NO: 2 with a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 2.

[0147] In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 11 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 11 (or SEQ ID NO: 10 if using the full length protein). In some embodiments, the polypeptide comprises a T182D or T182E mutation. In some embodiments, the VSV-G protein comprises a T182S, T182H, T182Q, or T182N mutation.

[0148] In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 13 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 13 (or SEQ ID NO: 12 if using the full length protein). In some embodiments, the polypeptide comprises a A182D or A182E mutation. In some embodiments, the VSV-G protein comprises a A182S, A182H, A182T, A182Q, or A182N mutation.

[0149] In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 15 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 15 (or SEQ ID NO: 14 if using the full length protein). In some embodiments, the polypeptide comprises a V182D or V182E mutation. In some embodiments, the VSV-G protein comprises a V182S, V182H, V182T, V182Q, or V182N mutation.

[0150] In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 17 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 17 (or SEQ ID NO: 16 if using the full length protein). In some embodiments, the polypeptide comprises a V182D or V182E mutation. In some embodiments, the VSV-G protein comprises a V182S, V182H, V182T, V182Q, or V182N mutation.

[0151] In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 19 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 19 (or SEQ ID NO: 18 if using the full length protein). In some embodiments, the polypeptide comprises a V182D or V182E mutation. In some embodiments, the VSV-G protein comprises a V182S, V182H, V182T, V182Q, or V182N mutation.

[0152] In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 21 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 21 (or SEQ ID NO: 20 if using the full length protein). In some embodiments, the polypeptide comprises a I182D or I182E mutation. In some embodiments, the VSV-G protein comprises a I182S, I182H, I182T, I182Q, or I182N mutation.

Viral Particles

[0153] The mutant VSV-G proteins can be used, for example, to pseudotype a virus, such as, but not limited to a lentivirus. Accordingly, in some embodiments, a viral particle comprising a mutant VSV-G protein as provided herein are provided. In some embodiments, the viral particle comprises a VSV-G protein comprising a mutation at position 198 as compared to SEQ ID NO: 1. In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 2 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 2 (or SEQ ID NO: 1 if using the full length protein). In some embodiments, the polypeptide comprises

a I182D or I182E mutation as compared to SEQ ID NO: 2. In some embodiments, the VSV-G protein comprises a I182S, I182H, I182T, I182Q, or I182N mutation.

[0154] In some embodiments, the viral particle comprises a VSV-G protein comprising a mutation at position 198 as compared to SEQ ID NO: 10. In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 11 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 11 (or SEQ ID NO: 10 if using the full length protein). In some embodiments, the polypeptide comprises a T182D or T182E mutation as compared to SEQ ID NO: 11. In some embodiments, the VSV-G protein comprises a T182S, T182H, T182Q, or T182N mutation.

[0155] In some embodiments, the viral particle comprises a VSV-G protein comprising a mutation at position 198 as compared to SEQ ID NO: 12. In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 13 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 13 (or SEQ ID NO: 12 if using the full length protein). In some embodiments, the polypeptide comprises a A182D or A182E mutation as compared to SEQ ID NO: 13. In some embodiments, the VSV-G protein comprises a A182S, A182H, A182T, A182Q, or A182N mutation.

[0156] In some embodiments, the viral particle comprises a VSV-G protein comprising a mutation at position 203 as compared to SEQ ID NO: 14. In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 15 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 15 (or SEQ ID NO: 14 if using the full length protein). In some embodiments, the polypeptide comprises a V182D or V182E mutation as compared to SEQ ID NO: 15. In some embodiments, the VSV-G protein comprises a V182S, V182H, V182T, V182Q, or V182N mutation.

[0157] In some embodiments, the viral particle comprises a VSV-G protein comprising a mutation at position 199 as compared to SEQ ID NO: 16. In some embodiments, a protein

comprising a mutation at position 182 as compared to SEQ ID NO: 17 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 17 (or SEQ ID NO: 16 if using the full length protein). In some embodiments, the polypeptide comprises a V182D or V182E mutation as compared to SEQ ID NO: 17. In some embodiments, the VSV-G protein comprises a V182S, V182H, V182T, V182Q, or V182N mutation.

[0158] In some embodiments, the viral particle comprises a VSV-G protein comprising a mutation at position 199 as compared to SEQ ID NO: 18. In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 19 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 19 (or SEQ ID NO: 18 if using the full length protein). In some embodiments, the polypeptide comprises a V182D or V182E mutation as compared to SEQ ID NO: 19. In some embodiments, the VSV-G protein comprises a V182S, V182H, V182T, V182Q, or V182N mutation.

[0159] In some embodiments, the viral particle comprises a VSV-G protein comprising a mutation at position 199 as compared to SEQ ID NO: 20. In some embodiments, a protein comprising a mutation at position 182 as compared to SEQ ID NO: 21 comprises a mutation at position 182 and at least, or about, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical as compared to SEQ ID NO: 21 (or SEQ ID NO: 20 if using the full length protein). In some embodiments, the polypeptide comprises a I182D or I182E mutation as compared to SEQ ID NO: 21. In some embodiments, the VSV-G protein comprises a I182S, I182H, I182T, I182Q, or I182N mutation.

[0160] In some embodiments, the VSV-G protein further comprises a mutation at position that corresponds to positions 214 and/or 352 of SEQ ID NO: 2. In some embodiments, the residue that corresponds to position 214 of SEQ ID NO: 2 is T214. In some embodiments, the residue that corresponds to position 352 of SEQ ID NO: 2 is T352. In some embodiments, the VSV-G protein comprises mutation that corresponds to T214N mutation as compared to SEQ ID NO: 2. In some embodiments, the VSV-G protein comprises mutation that corresponds to T352A mutation as

compared to SEQ ID NO: 2. In some embodiments, the VSV-G protein comprises a T214N and T352A mutations as compared to SEQ ID NO: 2. These mutations can be combined with any other mutations as provided for herein. In some embodiments, the T214N and/or T352A mutations are combined with the I182E or I182D mutations. In some embodiments, a VSV-G protein comprises an amino acid sequence of SEQ ID NO: 22 and SEQ ID NO: 23, which combines the I182D or I182E, respectively, with the T214N and T352A mutations. The sequences are also illustrated below with the leader sequences, which are removed during protein processing.

VSV-G Protein_ I196D, T230N and T368A mutations (with leader sequence and adjusted numbering)

MKCLLYLAFLFIGVNCKFTIVFPHNQKGNWKNVPSNYHYCPSSSDLNWHNDLIG
TALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQ
CKESIEQTKQGTWLNPGFPPQSCGYATVTDAEAVIVQVTPHHVLVDEYTGWVVD
SQFINGKCSNYICPTVHNSTTWHSDYKVKGLCDSNLDSMDITFFSEDGELSSLGK
EGTGFRSNYFAYENGGKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPE
CPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPKN
PGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISGTTAERELWDDWAPYED
VEIGPNGVLRRTSSGYKFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDD
ESLFFGDTGLSKNPIELVEGWFFSSWKSSIASFFFIIGLIIGLFLVLRVGIHLCLIKLHT
KKRQIYTDIEMNRLGK (SEQ ID NO: 24)

VSV-G Protein_I182D, T214N and T352A mutations (without leader sequence)

KFTIVFPHNQKGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQ
ADGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLN
PGFPPQSCGYATVTDAEAVIVQVTPHHVLVDEYTGWVVD
SQFINGKCSNYICPTV
HNSTTWHSDYKVKGLCDSNLDSMDITFFSEDGELSSLGKEGTGFRSNYFAYENG
GKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVD
VSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPKNPGTGPAFTIINGTLKY
FETRYIRVDIAAPILSRMVGMISGTTAERELWDDWAPYEDVEIGPNGVLRRTSSGY
KFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESLFFGDTGLSKNPI

ELVEGWFWSSWKSSIASFFFIIGLIIGLFLVLRVGIHLCKIKLKHTKKRQIYTDIEMNRL
GK (SEQ ID NO: 22)

VSV-G Protein with I196E, T230N and T368A mutations (with leader sequence and adjusted numbering)

MKCLLYLAFLFIGVNCKFTIVFPHNQKGNWKNVPSNYHYCPSSSDLNWHNDLIG
TALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQ
CKESIEQTKQGTWLNPGFPPQSCGYATVTDAEAVIVQVTPHHVLVDEYTGWVVD
SQFINGKCSNYICPTVHNSTTWHSYKVKGLCDNLESMDITFFSEDGELSSLGK
EGTGFRSNFYAYENGGKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPE
CPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPKN
PGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISGTTAERELWDDWAPYED
VEIGPNGVLRRTSSGYKFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDD
ESLFFGDTGLSKNPIELVEGWFWSSWKSSIASFFFIIGLIIGLFLVLRVGIHLCKIKLKHT
KKRQIYTDIEMNRLGK (SEQ ID NO: 25)

VSV-G Protein with I182E, T214N and T352A mutations (without leader sequences)

KFTIVFPHNQKGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQ
ADGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLN
PGFPPQSCGYATVTDAEAVIVQVTPHHVLVDEYTGWVVDVDSQFINGKCSNYICPTV
HNSTTWHSYKVKGLCDNLESMDITFFSEDGELSSLGKEGTGFRSNFYAYENG
GKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVD
VSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPKNPGTGPAFTIINGTLKY
FETRYIRVDIAAPILSRMVGMISGTTAERELWDDWAPYEDVEIGPNGVLRRTSSGY
KFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESLELFFGDTGLSKNPI
ELVEGWFWSSWKSSIASFFFIIGLIIGLFLVLRVGIHLCKIKLKHTKKRQIYTDIEMNRL
GK (SEQ ID NO: 23)

The other strains of the VSV-G protein as described herein can also comprises the mutations that correspond to T214N and/or T352A in SEQ ID NO: 2 and as illustrated in SEQ ID NO: 22 and SEQ ID NO: 23.

[0161] In some embodiments, the composition comprises a mutation as described in Hwang et al., Gene Ther 2013 Aug;20(8):807-15. (Epub 2013 Jan 31), which is hereby incorporated by reference in its entirety. For example, the mutations can be at positions 230, 368, 66, and/or 162 that corresponds to SEQ ID NO: 1. The positions will be 16 positions less as compared to SEQ ID NO: 2, when the leader sequence is removed. In some embodiments, the mutations at those positions are, for example, T230N, T368A, K66T, S162T, or any combination thereof. In some embodiments, the VSV-G protein comprises a T230N and a T368A mutation. In some embodiments, the VSV-G polypeptide comprises a K66T, S162T, T230N, and a T368A. These positions are those that correspond to the positions in the full length protein (SEQ ID NO: 1). In some embodiments, the VSV-G protein comprises T230N mutation, a T368A mutation, a K66T mutation, a S162T mutation, or any combination thereof. In some embodiments, the VSV-G protein further comprises one or more mutations in addition to the mutation that corresponds to position 182 of SEQ ID NO: 2, such as those described in U.S. Patent Application Publication No. 20200216502, which is hereby incorporated by reference in its entirety. For example, the VSV-G protein can further comprise a mutation at a position that corresponds to positions of 8, 47, 209 and/or 354 of SEQ ID NO: 2.

[0162] In some embodiments, the substitution at position 8 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except Y. In some embodiments, the substitution at position 209 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except H. In some embodiments, the substitution at position 47 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except K or R. In some embodiments, the substitution at position 354 is by any amino acid different from the amino acid indicated at that position in the sequence SEQ ID NO: 2, except K or R. In some embodiments, the substitution is at position 47 or at position 354, or at both positions 47 and 354 are substituted by A, G, F or Q. In some embodiments, the substitution is A or Q. In some embodiments, the substitution at position 8 is an alanine, i.e., H8A. In some embodiments, the substitution at position 47 is Q or N, i.e., K47Q or K47N. In some embodiments, the protein comprises a mutation (substitution) at position 10. In some embodiments, the substitution/mutation is Q10A, Q10R, or Q10K.

[0163] In some embodiments, the viral particle comprises a targeting moiety. The targeting moiety can be used to target the viral particle comprising the mutant VSV-G protein to a cell that

expresses the target to which the targeting moiety binds to. In some embodiments, the targeting moiety is an antibody, a scFv antibody, an antigen binding domain, an ankyrin repeat (e.g., DARPIN), a VHH domain antibody, a nanobody, single domain antibody, a FN3 domain, or any combination thereof. The targeting moiety can be attached to the viral surface through an IgG Fc stalk. In some embodiments, the stalk comprises a transmembrane domain. In some embodiments, the transmembrane domain comprises a CD8 transmembrane domain. In some embodiments, the transmembrane domain comprises a CD28 transmembrane domain. In some embodiments, the targeting moiety is attached (fused or linked) an envelope glycoprotein G or H of a virus of the Paramyxoviridae family, such as a morbillivirus, such as Measles virus, or a henipavirus, such as Nipah virus, Cedar virus, or Hendra virus. In some embodiments, the targeting moiety can be attached (fused or linked) to a glycoprotein of a virus of the Rhabdoviridae family, such as a vesicular stomatitis New Jersey virus, a vesicular stomatitis Indiana virus, a vesicular stomatitis Alagoas virus, a vesicular stomatitis Maraba virus, a vesicular stomatitis Carajas virus, Parainfluenza virus, Spodoptera frugiperda rhabdovirus isolate Sf G, Drosophila obscura sigmavirus 10A, Wuhan insect virus 7, Perch virus, or Spring viremia of carp virus. In some embodiments, the VSV protein is the mutated proteins, such as those provided for herein. In some embodiments, the targeting moiety is attached to a glycoprotein of a virus of the Filoviridae family, such as Ebola virus or a glycoprotein of a virus of the Arenaviridae family, such as Machupo virus.

[0164] In some embodiments, the targeting moiety is a scFv. In some embodiments, the targeting moiety is a single domain antibody. In some embodiments, the targeting moiety is a VHH.

[0165] In some embodiments, the targeting moiety binds to CD7, CD8, cKit (CD117), CD4, CD3, CD5, CD6, CD2, TCR alpha, TCR beta, TCR gamma, TCR delta, CD10, CD34, CD110, CD33, CD14, CD68, CCR7, CD62L, CD25, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, or CXCR3, A glycosylated CD43 epitope expressed on acute leukemia or lymphoma but not on hematopoietic progenitors; A glycosylated CD43 epitope expressed on non-hematopoietic cancers; A kinase anchor protein 4 (AKAP-4); Adrenoceptor beta 3 (ADRB3); AFP; Anaplastic lymphoma kinase (ALK); Androgen receptor; Angiopoietin-binding cell surface receptor 2 (Tie 2); Auto antibody to desmoglein 1 (Dsg1); Auto antibody to desmoglein 3 (Dsg3); B7H3 (CD276); Biotin; Bone marrow stromal cell antigen 2 (BST2); BST1/CD157; Cancer/testis antigen 1 (NY-ESO-1); Cancer/testis antigen 2 (LAGE-la); Carbonic anhydrase IX (CA1X); Carcinoembryonic

antigen (CEA); CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS or Brother of the Regulator of Imprinted Sites); CCR4; CD5; CD19; CD20; CD22; CD24; CD30; CD32 (FCGR2A); CD33; CD34; CD38; CD44v6; CD72; CD79a; CD79b; CD97; CD99; CD123; CD171; CD179a; CD179b-IGLL1; CD200R; CD276/B7H3; CD300 molecule-like family member f (CD300LF); CDH1-CD324; CDH6; CDH17; CDH19; Chromosome X open reading frame 61 (CXORF61); Claudin 6 (CLDN6); Claudin18.2 (CLDN18A2 or CLDN18A.2); CMV pp65; C-MYC epitope Tag; Cripto; CS1 (also referred to as CD2 subset 1 or CRACC or SLAMF7 or CD319 or 19A24); CSF2RA (GM-CSFR-alpha); C-type lectin domain family 12 member A (CLEC12A); C-type lectin-like molecule-1 (CLL-1 or CLECL1); Cyclin B1; Cytochrome P450 IB 1 (CYP1B 1); DLL3; EBV-EBNA3c; EGF-bke module- containing mucin-like hormone receptor-like 2 (EMR2); Elongation factor 2 mutated (ELF2M); Ephrin B2; Ephrin type-A receptor 2 (EphA2); Epidermal growth factor receptor (EGFR); Epidermal growth factor receptor variant III (EGFRviii); Epithelial cell adhesion molecule (EPCAM); ERG; ETS translocation-variant gene 6 located on chromosome 12p (ETV6-AML); Fc fragment of IgA receptor (FCAR or CD89); Fc receptor-like 5 (FCRL5); Fibroblast activation protein alpha (FAP); FITC; Fms Like Tyrosine Kinase 3 (FLT3); Folate receptor alpha (FRa or FR1); Folate receptor beta (FRb); Follicle stimulating hormone receptor (FSHR); Fos-related antigen 1; Fucosyl-GM1; G protein coupled receptor class C group 5 member D (GPRC5D); G protein-coupled receptor 20 (GPR20); GAD; Ganglioside G2 (GD2) ; Ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlc(1-1)Cer); Ganglioside GM3 (aNeu5Ac(2-3)bDCIalp(1- 4)bDGlc(1-1)Cer); GD3; GFRalpha4; Glycoprotein 100 (gp100); Glypican-3 (GPC3); Gonadotropin Hormone receptor (CGHR or GR); GpA33; GpNMB; GPRC5D; Guanylyl cyclase C (GCC); Heat shock protein 70-2 mutated (mut hsp70-2); Hepatitis A virus cellular receptor 1 (HAVCR1); Hexasaccharide portion of globoH glycosphingolipid (GloboH); High molecular weight-melanoma associated antigen (HMWMAA); HIV1 envelope glycoprotein; HLA; HLA-DOA; HLA-A; HLA-A2; HLA-B; HLA-C; HLA-DM; HLA-DOB; HLA-DP; HLA-DQ; HLA-DR; HLA-G; HTLV1-Tax; Human papilloma virus E6 (HPV E6); Human papilloma virus E7 (HPV E7); Human Telomerase reverse transcriptase (hTERT); IgE; IL13Ra2; IL1 lRa; Immunoglobulin lambda-like polypeptide 1 (IGLL1); Influenza A hemagglutinin (HA); Insulin-like growth factor 1 receptor (IGF-I receptor); Interleukin 11 receptor alpha (IL-11Ra); Interleukin- 13 receptor subunit alpha-2 (IL-13Ra2 or CD213A2); Intestinal carboxyl esterase; KIT (CD117); KSHV K8.1; KSHV-gH; LAMP1 ; Legumain; Leukocyte

immunoglobulin-like receptor subfamily A member 2 (LILRA2); Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1); Leutenizing hormone receptor (LHR); Lewis(Y) antigen; Lews Ag; Livl; Locus K 9 (LY6K); Low conductance chloride channel; Lymphocyte antigen 6 complex; Lymphocyte antigen 75 (LY75); Lymphocyte-specific protein tyrosine kinase (LCK); Mammary gland differentiation antigen (NY-BR-1); Melanoma antigen recognized by T cells 1 (MelanA or MARTI); Melanoma- associated antigen 1 (MAGE-A1); Melanoma cancer testis antigen-1 (MAD-CT-1); Melanoma cancer testis antigen-2 (MAD-CT-2); Melanoma inhibitor of apoptosis (ML-IAP); Mesothelin; MPL; Mucin 1 cell surface associated (MUC1); N-Acetyl glucosaminyl-transferase V (NA17); Nectin-4; Neural cell adhesion molecule (NCAM); NKG2D; NYBR1; O-acetyl-GD2 ganglioside (OAcGD2); Olfactory receptor 51E2 (OR51E2); Oncogene fusion protein consisting of breakpoint cluster region (BCR) and Abelson murine leukemia viral oncogene homolog 1 (Abl) (bcr-abl); P53 mutant; Paired box protein Pax-3 (PAX3); Paired box protein Pax-5 (PAX5); Pannexin 3 (PANX3); PDL1; P-glycoprotein; Placenta-specific 1 (PLAC1); Platelet-derived growth factor receptor beta (PDGFR-beta); Polysialic acid; Proacrosin binding protein sp32 (OY-TES1); Prostase; Prostate carcinoma tumor antigen-1 (PCT A-1 or Galectin 8); Prostate stem cell antigen (PSCA); Prostate-specific membrane antigen (PSMA); Prostatic acid phosphatase (PAP); Prostein; Protease Serine 21 (Testisin or PRSS21); Proteasome (Prosome Macropain) Subunit Beta Type 9 (LMP2); PTK7; Ras G12V; Ras Homolog Family Member C (RhoC); Rat sarcoma (Ras) mutant; Receptor for Advanced Glycation Endproducts (RAGE-1); Receptor tyrosine kinase-like orphan receptor 1 (ROR1); Receptor tyrosine-protein kinase ERBB2 or Her-22/neu; Renal ubiquitous 1 (RU1); Renal ubiquitous 2 (RU2); Sarcoma translocation breakpoints; Serine 2 (TMPRSS2) ETS fusion gene; Sialyl Lewis adhesion molecule (sLe); SLAMF4; SLAMF6; Slea (CA19.9 or Sialyl Lewis Antigen); Sperm protein 17 (SPA17); Squamous Cell Carcinoma Antigen Recognized By T Cells 3 (SART3); Stage-specific embryonic antigen-4 (SSEA-4); STEAP1; Survivin; Synovial sarcoma X breakpoint 2 (SSX2); TCR Gamma Alternate Reading Frame Protein (TARP); TCR-beta1 chain; TCR-beta2 chain; TCR-delta chain; TCR-gamma chain; TCRgamma-delta; Telomerase; TGFbetaR2; The antigen recognized by TNT antibody; Thyroid stimulating hormone receptor (TSHR); Timl-/HVCR1; Tissue Factor 1 (TF1); Tn ag; Tn antigen ((Tn Ag) or (GalNAc-Ser/Thr)); TNF receptor family member B cell maturation (BCMA); Transglutaminase 5 (TGS5); Transmembrane protease; TROP2; Tumor endothelial marker 1 (TEM1/CD248); Tumor

endothelial marker 7-related (TEM7R); Tumor protein p53 (p53); Tumor-associated glycoprotein 72 (TAG72); Tyrosinase; Tyrosinase-related protein 2 (TRP-2); Uroplakin 2 (UPK2); Vascular endothelial growth factor receptor 2 (VEGFR2); V-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN); Wilms tumor protein (WT1); or X Antigen Family Member 1A (XAGE1). In some embodiments, the targeting moiety binds to CD7. In some embodiments, the targeting moiety binds to CD8.

[0166] In some embodiments, the targeting moiety binds to a target that is present on a cell, such as an immune cell. In some embodiments, the cell is an immune cell, such as, but not limited to, T cell, B cell; NK cell, dendritic cell, neutrophils, macrophages, a cancer cell; or, for example, CD3+ T cell; CD4+ T cell; CD7+ T cell, CD8+ T cell; CD19+ B cell; CD19+ cancer cell; CD20+ B cell; CD20+ cancer cell; CD30+ lung epithelial cell; CD34+ haematopoietic stem cell; CD105+ endothelial cell; CD105+ haematopoietic stem cell; CD117+ haematopoietic stem cell; CD133+ cancer cell; EpCAM+ cancer cell; GluA2+ neuron; GluA4+ neuron; Haematopoietic stem cell; Hepatocyte; Her2/Neu+ cancer cell; NKG2D+ natural killer cell; SLC1A3+ astrocyte; SLC7A10+ adipocyte. In some embodiments, the cell is a T cell. In some embodiments, the cell is a B cell. In some embodiments, the cell is a CD7+ T cell and/or CD8+ T cell.

[0167] In some embodiments, the targeting moiety (a polypeptide) can bind to CD7.

[0168] In some embodiments, the polypeptide binds to CD7. In some embodiments, the polypeptide that binds to CD7 is an antibody which binds to non-human primate CD7. In some embodiments, the polypeptide that binds to CD7 is an antibody which binds to human CD7. The sequence of human CD7 (UniProtKB P09564) is as follows (SEQ ID NO: 33):

MAGPPRLLLLPLLLALARGLPALAAQEVQQSPHCTTPVVGASVNITCSTSGGLR
GIYLRQLGPQPQDIYYEDGVVPTTDRRFRGRIDFSGSQDNLTITMHLQLSDTGT
YTCQAITEVNVYGSGLTLVLVTEEQSQGWHRCS DAPPRASALPAPPTGSALPDPQT
ASALPDPPAASALPAALAVISFLLGLGLGVACVLARTQIKKLCSWRDKNSAACV
VYEDMSHSRCNTLSSPNQYQ (SEQ ID NO: 33).

[0169] In some embodiments, the CD7 that the polypeptide binds to is expressed on the surface of a cell. In some embodiments, the cell is an immune cell. In some embodiments, the immune cell is a CD7+ T cell, CD4+ T cell, CD8+ T cell, NK cell, alpha-beta T cell, gamma-delta T cell, lymphoid progenitor cell, hematopoietic stem cell, myeloid cell, monocyte, macrophage,

central memory T cell, effector memory T cell, stem-cell like memory T cells, naïve T cell, activated T cell, regulatory T cell (TReg), terminally differentiated effector memory T cell (TEMRA), resident memory T cell (TRM) or a T-cellCD8+CCR7+.

[0170] In some embodiments, the antibody comprises a Fc region. The Fc region can be linked to the heavy or light chain of the antibody. In some embodiments, the Fc region is an IgG Fc. In some embodiments, the IgG is selected from IgG1, IgG2, IgG3, or IgG4. In some embodiments, the IgG fc is IgG1 Fc. In some embodiments, the antibody comprises an Fc constant region of SEQ ID NO: 34 as set forth below.

```
ASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPP
AVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHT
CPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYV
DGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIE
KTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQP
ENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSL
SLSPGK (SEQ ID NO: 34)
```

[0171] In some embodiments, the IgG fc is IgG2 Fc. In some embodiments, the antibody comprises an Fc constant region of SEQ ID NO: 71 as set forth below.

```
STKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAV
LQSSGLYSLSSVVTVPSSSLGTQTYTCNVDHKPSNTKVDKTVERKCCVECPPCA
PPVAGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVQFNWYVDGVEVHN
AKTKPREEQFNSTFRVSVLTVVHQDWLNGKEYKCKVSNKGLPAPIEKTISKTK
GQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKT
PPMLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK
(SEQ ID NO: 71)
```

[0172] In some embodiments, the IgG fc is IgG4 Fc. In some embodiments, the antibody comprises an Fc constant region of SEQ ID NO: 72 as set forth below.

```
STKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAV
LQSSGLYSLSSVVTVPSSSLGTQTYTCNVDHKPSNTKVDKRVESKYGPPCPSCPA
PEFLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSQEDPEVQFNWYVDGVEVH
NAKTKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKGLPSSIEKTISKA
KGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKT
```

TPPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALHNHYTQKSLSLGLGK
(SEQ ID NO: 72)

[0173] In some embodiments, polypeptides (e.g. CD7-binding polypeptide) are provided herein. In some embodiments, antibodies (e.g. an anti-CD7 antibody) are provided herein. In some embodiments, the antibody is a recombinant antibody that binds to CD7. In some embodiments, the CD7 protein is a human CD7 protein. In some embodiments, the CD7 protein is a non-human CD7 protein (e.g., mouse, rat, pig, dog, non-human primate). As used herein, the term “recombinant antibody” refers to an antibody that is not naturally occurring. In some embodiments, the term “recombinant antibody” refers to an antibody that is not isolated from a human subject.

[0174] In some embodiments, an antibody, or antigen binding fragment thereof is provided, wherein the antibody or antibody fragment comprises a peptide selected from the following table, which illustrate the CDRs based on Chothia numbering.

Chothia CDRs						
Ab ID No	HCDR1	HCDR2	HCDR3	LCDR1	LCDR2	HCDR3
CD7AB1	GYPFTSY (SEQ ID NO: 35)	DPNSGD (SEQ ID NO: 36)	SPYYSDNSMDY (SEQ ID NO: 37)	RASQSIGT SIH (SEQ ID NO: 38)	YASESIS (SEQ ID NO: 39)	QQNSWPTT (SEQ ID NO: 40)

[0175] In some embodiments, an antibody, or antigen binding fragment thereof is provided, wherein the antibody or antibody fragment comprises a peptide selected from the following table, which illustrate the CDRs based on Kabat numbering.

Kabat CDRs						
Ab ID No	HCDR1	HCDR2	HCDR3	LCDR1	LCDR2	HCDR3
CD7AB1	SYWIH (SEQ ID NO: 41)	RIDPNSGDKYNEKFKN (SEQ ID NO: 42)	SPYYSDNSMDY (SEQ ID NO: 37)	RASQSIGT SIH (SEQ ID NO: 38)	YASESIS (SEQ ID NO: 39)	QQNSWPTT (SEQ ID NO: 40)

[0176] In some embodiments, an antibody, or antigen binding fragment thereof is provided, wherein the antibody or antibody fragment comprises a peptide selected from the following table, which illustrate the CDRs based on IMGT numbering.

IMGT CDRs						
Ab ID No	HCDR1	HCDR2	HCDR3	LCDR1	LCDR2	HCDR3

CD7ABI	GYPFTS YW (SEQ ID NO: 43)	IDPNSGDT (SEQ ID NO: 44)	ARSPYYSNDN SMDY (SEQ ID NO: 45)	QSIGTS (SEQ ID NO: 46)	YA	QQSNSWPTT (SEQ ID NO: 40)
--------	---------------------------------------	--------------------------------	---------------------------------------	------------------------------	----	---------------------------------

[0177] In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy or light chain CDR as provided in the tables above. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy or light chain CDR as provided in the tables above and binds to non-human primate CD7. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy or light chain CDR as provided in the tables above and binds to human CD7. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a light chain CDR having a sequence selected from SEQ ID NO: 38-40. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a light chain CDR having a sequence of SEQ ID NO: 38. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a light chain CDR having a sequence of SEQ ID NO: 39. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a light chain CDR having a sequence of SEQ ID NO: 40. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain CDR having a sequence selected from SEQ ID NO: 35-37. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain CDR having a sequence of SEQ ID NO: 35. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain CDR having a sequence of SEQ ID NO: 36. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain CDR having a sequence of SEQ ID NO: 37. The CDRs referenced in the embodiments throughout the present specification can be interchanged with the CDRs that are characterized by different formats, such as Kabat and IMGT.

[0178] In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a light chain variable region having a LCDR1, a LCDR2, and a LCDR3, wherein the LCDR1 has a sequence of SEQ ID NO: 38, the LCDR2 has a sequence of SEQ ID NO: 39, and the LCDR3 has a sequence of SEQ ID NO: 40.

[0179] In some embodiments, a polypeptide, an antibody, or antibody binding fragment

thereof, comprises a light chain variable region having a LCDR1, a LCDR2, and a LCDR3, wherein the LCDR1 has a sequence of SEQ ID NO: 46, the LCDR2 has a sequence of YA, and the LCDR3 has a sequence of SEQ ID NO: 40.

[0180] In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain variable region having a HCDR1, a HCDR2, and a HCDR3, wherein the HCDR1 has a sequence of SEQ ID NO: 35, the HCDR2 has a sequence of SEQ ID NO: 36, and the HCDR3 has a sequence of SEQ ID NO: 37.

[0181] In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain variable region having a HCDR1, a HCDR2, and a HCDR3, wherein the HCDR1 has a sequence of SEQ ID NO: 41, the HCDR2 has a sequence of SEQ ID NO: 42, and the HCDR3 has a sequence of SEQ ID NO: 37.

[0182] In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain variable region having a HCDR1, a HCDR2, and a HCDR3, wherein the HCDR1 has a sequence of SEQ ID NO: 43, the HCDR2 has a sequence of SEQ ID NO: 44, and the HCDR3 has a sequence of SEQ ID NO: 45.

[0183] In some embodiments, a polypeptide, an antibody or antibody binding fragment thereof, comprises: (i) a light chain having any one of the foregoing recited combinations of LCDR1, LCDR2, and LCDR3 sequences; and (ii) a heavy chain having any one of the foregoing recited combinations of HCDR1, HCDR2, and HCDR3 sequences.

[0184] The different CDR motifs can be combined in any combination including those not depicted in the table above. For example, the following embodiments are provided as non-limiting examples of such combinations.

[0185] In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises: (i) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 38; the light chain CDR2 has the amino acid sequence of SEQ ID NO: 39; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 40; and (ii) a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 35; the heavy chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 36; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 37; or variants of any of the foregoing.

[0186] In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises: (i) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 38; the light chain CDR2 has the amino acid sequence of SEQ ID NO: 39; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 40; and (ii) a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 41; the heavy chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 42; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 37; or variants of any of the foregoing.

[0187] In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises: (i) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 46; the light chain CDR2 has the amino acid sequence of YA; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 40; and (ii) a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 43; the heavy chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 44; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 45; or variants of any of the foregoing.

[0188] In some embodiments, the light chain variable region CDR1 is replaced with any of the other light chain CDR1 sequences. In some embodiments, the light chain variable region CDR2 is replaced with any of the other light chain CDR2 sequences. In some embodiments, the light chain variable region CDR3 is replaced with any of the other light chain CDR3 sequences. In some embodiments, the heavy chain variable region CDR1 is replaced with any of the other heavy chain CDR1 sequences. In some embodiments, the heavy chain variable region CDR2 is replaced with any of the other heavy chain CDR2 sequences. In some embodiments, the heavy chain variable region CDR3 is replaced with any of the other heavy chain CDR3 sequences.

[0189] In some embodiments, the polypeptide comprises a heavy chain variable region peptide having one of the following sequences, or a variant thereof:

SEQ ID NO:	AB ID NO.	Sequence
47	CD7AB1	QVQLQQPGAELVKPGASVKLSCKASGYPFTSYWIHWVKQRPGRGLEWLGRIDPNSGDTKYNEKFKNKATLTVDKSSTTAYMQLSSLTSEDSAVYYCARSPYYSDNSMDYWGQGTSTVTVSS

[0190] In some embodiments, the polypeptide comprises a light chain variable region peptide having one of the following sequences, or a variant thereof:

SEQ ID NO:	AB ID NO.	Sequence
48	CD7AB1	DILLTQSPAILSVSPGERVSFSCRASQSIGTSHHWYQORTNDSPR LLIKYASESISGIPSRFSGSGSGTDFTLSINSVESEDIADYYCQQ SNSWPTTFGGGKLEIKR

[0191] In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises a V_H peptide of SEQ ID NO: 47. In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises a V_L peptide of SEQ ID NO: 48. In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises a V_H peptide and a V_L peptide, wherein the wherein the V_H peptide comprises a sequence of SEQ ID NO: 47, or a variant thereof; and the V_L peptide comprises a sequence of SEQ ID NO: 48, or a variant thereof. In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises a V_H peptide and a V_L peptide, wherein the wherein the V_H peptide comprises a sequence of SEQ ID NO: 47, or a variant thereof; and the V_L peptide comprises a sequence of SEQ ID NO: 48, or a variant thereof, and the polypeptide, the antibody, or antigen binding fragment thereof, binds to non-human primate CD7. In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises a V_H peptide and a V_L peptide, wherein the wherein the V_H peptide comprises a sequence of SEQ ID NO: 47, or a variant thereof; and the V_L peptide comprises a sequence of SEQ ID NO: 48, or a variant thereof, and the polypeptide, the antibody, or antigen binding fragment thereof, binds to human CD7. In some embodiments, the V_H peptide comprises a sequence of SEQ ID NO: 47; and the V_L peptide comprises a sequence of SEQ ID NO: 48.

[0192] The V_H and the V_L sequences can be in any format, including, but not limited to an scFv format where the V_H and V_L regions are linked with a peptide linker. Examples of peptide linkers that can be used to link various peptides provided for herein include, but are not limited to: (GGGS)_n (SEQ ID NO: 49), wherein each n is independently 1-5. In some embodiments, n is 1. In some embodiments, n is 2. In some embodiments, n is 3. In some embodiments, n is 4. In some embodiments, n is 5. In some embodiments, the variable regions are not linked with a peptide linker. In some embodiments, the polypeptide comprises SEQ ID NO: 47 and SEQ ID NO: 48.

[0193] In some embodiments, the VH and VL polypeptides are linked to a Fc region. In some embodiments, the Fc region is as provided for herein. In some embodiments, the Fc region comprises an amino acid sequence of SEQ ID NO: 34, SEQ ID NO: 71, or SEQ ID NO: 72 as provided for herein. As provided for herein, the heavy chain can be linked to a Fc region. Non-limiting mutations in the Fc region are provided for herein. In some embodiments, the Fc region further comprises a transmembrane domain. Examples of transmembrane domains include, but are not limited to, a CD8 or CD28 (“CD8/CD28”) transmembrane domain. In some embodiments, the Fc region further comprises a CD8 transmembrane domain. In some embodiments, the Fc region further comprises a CD28 transmembrane domain. In some embodiments, the Fc region comprising a transmembrane domain further comprises an Env incorporation motif. In some embodiments, the Fc region comprising a CD8/CD28 transmembrane domain further comprises an Env incorporation motif. In some embodiments, the VH and VL polypeptides provided herein are linked to an Fc region comprising a transmembrane domain. In some embodiments, the VH and VL polypeptides provided herein are linked to an Fc region comprising a CD8/CD28 transmembrane domain. In some embodiments, the VH and VL polypeptides provided herein are linked to an Fc region comprising a CD8/CD28 transmembrane domain and an Env incorporation motif. In some embodiments, the VH and VL polypeptides provided herein linked to an Fc region comprising a CD8/CD28 transmembrane domain are anchored to the plasma membrane on the surface of a cell. In some embodiments, the cell is an immune cell, such as those provided herein. In some embodiments, the VH having a sequence as set forth in SEQ ID NO: 47 and VL having a sequence as set forth in SEQ ID NO: 48 are linked to an Fc region comprising a transmembrane domain. In some embodiments, the VH having a sequence as set forth in SEQ ID NO: 47 and VL having a sequence as set forth in SEQ ID NO: 48 are linked to an Fc region comprising a CD8/CD28 transmembrane domain. In some embodiments, the VH having a sequence as set forth in SEQ ID NO: 47 and VL having a sequence as set forth in SEQ ID NO: 48 are linked to an Fc region comprising a CD8/CD28 transmembrane domain and an Env incorporation motif. In some embodiments, the VH having a sequence as set forth in SEQ ID NO: 47 and VL having a sequence as set forth in SEQ ID NO: 48 linked to an Fc region comprising a CD8/CD28 transmembrane domain are anchored to the plasma membrane on the surface of a cell. In some embodiments, the cell is an immune cell, such as those provided herein.

[0194] In some embodiments, the V_H peptide comprising a sequence that is at least 85%,

86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48 are linked to an Fc region comprising a transmembrane domain. In some embodiments, the V_H peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48 are linked to an Fc region comprising a CD8/CD28 transmembrane domain. In some embodiments, the V_H peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48 are linked to an Fc region comprising a CD8/CD28 transmembrane domain and an Env incorporation motif. In some embodiments, the V_H peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48 linked to an Fc region comprising a CD8/CD28 transmembrane domain are anchored to the plasma membrane on the surface of a cell. In some embodiments, the cell is an immune cell, such as those provided herein.

[0195] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48.

[0196] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48;

provided that the V_H peptide and a V_L peptide comprises a light chain CDR having a sequence of SEQ ID NO: 38-40; and/or a heavy chain CDR having a sequence of SEQ ID NO: 35-37. In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48; provided that the V_H peptide and a V_L peptide comprise a light chain CDR1 having a sequence of SEQ ID NO: 38; a light chain CDR2 having a sequence of SEQ ID NO: 39; a light chain CDR3 having a sequence of SEQ ID NO: 40; and/or a heavy chain CDR1 having a sequence of SEQ ID NO: 35; a heavy chain CDR2 having a sequence of SEQ ID NO: 36; and a heavy chain CDR3 having a sequence of SEQ ID NO: 37. In some embodiments, the CDRs in the V_H or V_L chain are as set forth in the combinations provided for herein.

[0197] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48; provided that the V_L peptide comprises a LCDR1 having a sequence of SEQ ID NO: 38; a LCDR2 having a sequence of SEQ ID NO: 39; and a LCDR3 having a sequence of SEQ ID NO: 40; and the V_H peptide comprises a HCDR1 having a sequence of SEQ ID NO: 35; a HCDR2 having a sequence of SEQ ID NO: 36; and a HCDR3 having a sequence of SEQ ID NO: 37.

[0198] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48; provided that the V_L peptide comprises a LCDR1 having a sequence of SEQ ID NO: 38, wherein the LCDR1 comprises at most 1 conservative amino acid substitution, a LCDR2 having a sequence of SEQ ID NO: 39, wherein the LCDR2 comprises at most 1 conservative amino acid substitution, and a LCDR3 having a sequence of SEQ ID NO: 40, wherein the LCDR3 comprises at most 1

conservative amino acid substitution; and the V_H peptide comprises a HCDR1 having a sequence of SEQ ID NO: 35, wherein the HCDR1 comprises at most 1 conservative amino acid substitution, a HCDR2 having a sequence of SEQ ID NO: 36, wherein the HCDR2 comprises at most 1 conservative amino acid substitution, and a HCDR3 having a sequence of SEQ ID NO: 37, wherein the HCDR3 comprises at most 1 conservative amino acid substitution.

[0199] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48; provided that the V_H peptide and a V_L peptide comprise a light chain CDR1 having a sequence of SEQ ID NO: 38; a light chain CDR2 having a sequence of SEQ ID NO: 39; a light chain CDR1 having a sequence of SEQ ID NO: 40; and/or a heavy chain CDR1 having a sequence of SEQ ID NO: 41; a heavy chain CDR2 having a sequence of SEQ ID NO: 42; and a heavy chain CDR3 having a sequence of SEQ ID NO: 37. In some embodiments, the CDRs in the V_H or V_L chain are as set forth in the combinations provided for herein.

[0200] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48; provided that the V_L peptide comprises a LCDR1 having a sequence of SEQ ID NO: 38; a LCDR2 having a sequence of SEQ ID NO: 39; and a LCDR3 having a sequence of SEQ ID NO: 40; and the V_H peptide comprises a HCDR1 having a sequence of SEQ ID NO: 41; a HCDR2 having a sequence of SEQ ID NO: 42; and a HCDR3 having a sequence of SEQ ID NO: 37.

[0201] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48; provided that the V_L peptide comprises a LCDR1 having a sequence of SEQ ID NO: 38, wherein

the LCDR1 comprises at most 1 conservative amino acid substitution, a LCDR2 having a sequence of SEQ ID NO: 39, wherein the LCDR2 comprises at most 1 conservative amino acid substitution, and a LCDR3 having a sequence of SEQ ID NO: 40, wherein the LCDR3 comprises at most 1 conservative amino acid substitution; and the V_H peptide comprises a HCDR1 having a sequence of SEQ ID NO: 41, wherein the HCDR1 comprises at most 1 conservative amino acid substitution, a HCDR2 having a sequence of SEQ ID NO: 42, wherein the HCDR2 comprises at most 1 conservative amino acid substitution, and a HCDR3 having a sequence of SEQ ID NO: 37, wherein the HCDR3 comprises at most 1 conservative amino acid substitution.

[0202] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48; provided that the V_H peptide and a V_L peptide comprise a light chain CDR1 having a sequence of SEQ ID NO: 46; a light chain CDR2 having a sequence of YA; a light chain CDR1 having a sequence of SEQ ID NO: 40; and/or a heavy chain CDR1 having a sequence of SEQ ID NO: 43; a heavy chain CDR2 having a sequence of SEQ ID NO: 44; and a heavy chain CDR3 having a sequence of SEQ ID NO: 45. In some embodiments, the CDRs in the V_H or V_L chain are as set forth in the combinations provided for herein.

[0203] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48; provided that the V_L peptide comprises a LCDR1 having a sequence of SEQ ID NO: 46; a LCDR2 having a sequence of YA; and a LCDR3 having a sequence of SEQ ID NO: 40; and the V_H peptide comprises a HCDR1 having a sequence of SEQ ID NO: 43; a HCDR2 having a sequence of SEQ ID NO: 44; and a HCDR3 having a sequence of SEQ ID NO: 45.

[0204] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 47;

and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 48; provided that the V_L peptide comprises a LCDR1 having a sequence of SEQ ID NO: 46, wherein the LCDR1 comprises at most 1 conservative amino acid substitution, a LCDR2 having a sequence of YA, wherein the LCDR2 comprises at most 1 conservative amino acid substitution, and a LCDR3 having a sequence of SEQ ID NO: 40, wherein the LCDR3 comprises at most 1 conservative amino acid substitution; and the V_H peptide comprises a HCDR1 having a sequence of SEQ ID NO: 43, wherein the HCDR1 comprises at most 1 conservative amino acid substitution, a HCDR2 having a sequence of SEQ ID NO: 44, wherein the HCDR2 comprises at most 1 conservative amino acid substitution, and a HCDR3 having a sequence of SEQ ID NO: 45, wherein the HCDR3 comprises at most 1 conservative amino acid substitution.

[0205] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence of SEQ ID NO: 47 and the V_L peptide comprises a sequence of SEQ ID NO: 48.

[0206] In some embodiments, a polypeptide as provided herein binds to non-human primate CD7. In some embodiments, a polypeptide as provided herein binds to human CD7.

[0207] As provided for herein, the different polypeptides (V_H or V_L) described herein can be linked with a peptide linker or not linked with a peptide linker and instead for a continuous sequence. In some embodiments, the peptide linker comprises a sequence of (GGGGS)_n (SEQ ID NO: 49), wherein each n is independently 1-5. In some embodiments, n is 1. In some embodiments, n is 2. In some embodiments, n is 3. In some embodiments, n is 4. In some embodiments, n is 5. The linked peptide format can be represented by a formula of V_H-Z-V_L or V_L-Z-V_H, wherein Z is the peptide linker. In some embodiments, Z is (GGGGS)_n (SEQ ID NO: 49), wherein each n is independently 1-5.

[0208] In some embodiments, a polypeptide comprising the linked peptide represented by a formula of V_L-Z-V_H comprises a heavy chain variable region as set forth in SEQ ID NO: 47 linked via a linker sequence GGGGSGGGGSGGGGSGGGGS (SEQ ID NO: 50) to a light chain variable region as set forth in SEQ ID NO: 48. In some embodiments, a polypeptide comprising a V_L linked via a peptide linker to a V_H has the sequence as set forth below,

DILLTQSPAILSVSPGERVSFSCRASQSIGTSIHWYQQRNTDSPRLLIKYASESISGIP
 SRFSGSGSGTDFTLSINSVESEDIADYYCQQSNSWPTTFGGGKLEIKRGGGGSGG
 GGSGGGSGGGGSQVQLQQPGAELVKPGASVKLSCKASGYPFTSYWIHWVKQR
 PGRGLEWLGRIDPNSGDTKYNEKFKNKATLTVDKSSTTAYMQLSSLTS
 EDSAVYYCARSPYYSDNSMDYWGQTSVTVSS (SEQ ID NO: 51).

[0209] In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 51. In some embodiments, a polypeptide comprises a sequence that is at least 90% identical to a sequence of SEQ ID NO: 51. In some embodiments, a polypeptide comprises a sequence that is at least 95% identical to a sequence of SEQ ID NO: 51. In some embodiments, a polypeptide comprises a sequence that is at least 99% identical to a sequence of SEQ ID NO: 51. In some embodiments, a polypeptide comprises a sequence as set forth in SEQ ID NO: 51. In some embodiments, the polypeptide as set forth in SEQ ID NO: 51 is an antibody, or an antigen binding fragment thereof. In some embodiments, the antibody is an anti-CD7 antibody.

[0210] In some embodiments, a polypeptide comprising the linked peptide represented by a formula of V_H -Z- V_L comprises a light chain variable region as set forth in SEQ ID NO: 48 linked via a linker sequence GGGGSGGGGSGGGGSGGGGS (SEQ ID NO: 50) to a heavy chain variable region as set forth in SEQ ID NO: 49. In some embodiments, a polypeptide comprising a V_H linked via a peptide linker to a V_L has the sequence as set forth below,

QVQLQQPGAELVKPGASVKLSCKASGYPFTSYWIHWVKQRPGRGLEWLGRIDP
 NSGDTKYNEKFKNKATLTVDKSSTTAYMQLSSLTSEDSAVYYCARSPYYSDNS
 MDYWGQTSVTVSSGGGGSGGGGSGGGGSGGGGSDILLTQSPAILSVSPGERVS
 FSCRASQSIGTSIHWYQQRNTDSPRLLIKYASESISGIPSRFSGSGSGTDFTLSINSV
 ESEDIADYYCQQSNSWPTTFGGGKLEIKR (SEQ ID NO: 52)

[0211] In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 52. In some embodiments, a polypeptide comprises a sequence that is at

least 90% identical to a sequence of SEQ ID NO: 52. In some embodiments, a polypeptide comprises a sequence that is at least 95% identical to a sequence of SEQ ID NO: 52. In some embodiments, a polypeptide comprises a sequence that is at least 99% identical to a sequence of SEQ ID NO: 52. In some embodiments, a polypeptide comprises a sequence as set forth in SEQ ID NO: 52. In some embodiments, the polypeptide as set forth in SEQ ID NO: 52 is an antibody, or an antigen binding fragment thereof. In some embodiments, the antibody is an anti-CD7 antibody. In some embodiments, the anti-CD7 antibody binds to non-human primate CD7. In some embodiments, the anti-CD7 antibody binds to human CD7.

[0212] In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 51 and comprises an Fc region, such as those provided herein. In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 51 and comprises an Fc region, such as those provided herein, and a transmembrane domain, such as those provided herein. In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 51 and comprises an Fc region, such as those provided herein, a transmembrane domain, such as those provided herein, and an Env incorporation motif.

[0213] In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 52 and comprises an Fc region, such as those provided herein. In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 52 and comprises an Fc region, such as those provided herein, and a transmembrane domain, such as those provided herein. In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 52 and comprises an Fc region, such as those provided

herein, and a CD8/CD28 transmembrane domain. In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 52 and comprises an Fc region, such as those provided herein, a transmembrane domain, such as those provided herein, and an Env incorporation motif.

[0214] As provided for herein, the polypeptide, antibodies, or antigen binding fragments thereof can be variants of the sequences.

[0215] The sequences of the polypeptides or antibodies can be modified to yield human IgG antibodies. The conversion of the sequences provided herein can be modified to yield other types of antibodies. The CDRs can also be linked to other antibodies, proteins, or molecules to create antibody fragments that bind CD7.

[0216] In some embodiments, a polypeptide or an antibody as provided herein is a targeting moiety on the surface of an engineered viral particle. In some embodiments, the targeting moiety allows for binding to a target cell. In some embodiments, the targeting moiety is a CD7 binding moiety, such as a polypeptide or an antibody as provided herein. In some embodiments, the targeting moiety comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 51. In some embodiments, the targeting moiety comprises a sequence that is at least 90% identical to a sequence of SEQ ID NO: 51. In some embodiments, the targeting moiety comprises a sequence that is at least 95% identical to a sequence of SEQ ID NO: 51. In some embodiments, the targeting moiety comprises a sequence that is at least 99% identical to a sequence of SEQ ID NO: 51. In some embodiments, the targeting moiety comprises a sequence as set forth in SEQ ID NO: 51. In some embodiments, the targeting moiety as set forth in SEQ ID NO: 51 is an antibody, or an antigen binding fragment thereof. In some embodiments, the targeting moiety is an anti-CD7 antibody.

[0217] In some embodiments, a polypeptide or an antibody as provided for herein is a targeting moiety on the surface of an engineered viral particle. In some embodiments, the engineered viral particle is a pseudotyped viral-like particle. In some embodiments, the targeting moiety allows for binding to a target cell. In some embodiments, the targeting moiety is a CD7 binding moiety, such as a polypeptide or an antibody as provided herein. In some embodiments, the targeting moiety comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%,

92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 52. In some embodiments, the targeting moiety comprises a sequence that is at least 90% identical to a sequence of SEQ ID NO: 52. In some embodiments, the targeting moiety comprises a sequence that is at least 95% identical to a sequence of SEQ ID NO: 52. In some embodiments, the targeting moiety comprises a sequence that is at least 99% identical to a sequence of SEQ ID NO: 52. In some embodiments, the targeting moiety comprises a sequence as set forth in SEQ ID NO: 52. In some embodiments, the targeting moiety as set forth in SEQ ID NO: 52 is an antibody, or an antigen binding fragment thereof. In some embodiments, the targeting moiety is an anti-CD7 antibody. In some embodiments, the anti-CD7 antibody binds to non-human primate CD7. In some embodiments, the anti-CD7 antibody binds to human CD7.

[0218] In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 51 and comprises an Fc region, such as those provided herein. In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 51 and comprises an Fc region, such as those provided herein, and a transmembrane domain, such as those provided herein. In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 51 and comprises an Fc region, such as those provided herein, and a CD8/CD28 transmembrane domain. In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 51 and comprises an Fc region, such as those provided herein, a transmembrane domain, such as those provided herein, and an Env incorporation motif.

[0219] In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 52 and comprises an Fc region, such as those provided herein. In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 52 and comprises an Fc region, such as those provided herein, and a transmembrane domain, such as those provided herein. In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 52 and comprises an Fc region, such as those provided herein, and a CD8/CD28 transmembrane domain. In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 52 and comprises an Fc region, such as those provided herein, a transmembrane domain, such as those provided herein, and an Env incorporation motif.

[0220] In some embodiments, the targeting moiety (a polypeptide) can bind to CD8.

[0221] In some embodiments, the polypeptide binds to CD8. In some embodiments, the polypeptide binds to CD8-alpha. In some embodiments, the polypeptide binds to CD8-beta. In some embodiments, the polypeptide binds to CD8 heterodimer. In some embodiments, the CD8 heterodimer comprises CD8-alpha and CD8-beta subunits. In some embodiments, the polypeptide binds to CD8-alpha homodimer. In some embodiments, the polypeptide that binds to CD8 is an antibody which binds to non-human primate CD8. In some embodiments, the antibody that binds to non-human primate CD8 is an antibody which binds to non-human primate CD8-alpha. In some embodiments, the antibody that binds to non-human primate CD8 is an antibody which binds to non-human primate CD8-beta. In some embodiments, the antibody that binds to non-human primate CD8 is an antibody which binds to non-human primate CD8-alpha homodimer. In some embodiments, the antibody that binds to non-human primate CD8 is an antibody which binds to non-human primate CD8 heterodimer. In some embodiments, the polypeptide that binds to CD8 is an antibody which binds to human CD8. In some embodiments, the antibody that binds to human CD8 is an antibody which binds to human CD8-alpha. In some embodiments, the antibody that binds to human CD8 is an antibody which binds to human CD8-beta. In some embodiments, the antibody that binds to human CD8 is an antibody which binds to human CD8-alpha homodimer. In some embodiments, the antibody that binds to human CD8 is an antibody which binds to human CD8 heterodimer. The sequence of human CD8-alpha (UniProtKB Q8TAW8) is as follows (SEQ ID NO: 53):

MALPVTALLLPLALLLHAARPSQFRVSPLDRTWNLGETVELKCQVLLSNPTSGCS
WLFQPRGAAASPTFLLYLSQNKPKAAEGLDTQRFSGKRLGDTFVLTLSDFRREN
EGCYFCSALSNSIMYFSHFVPVFLPAKPTTTPAPRPPTPAPTIASQPLSLRPEACRP
AAGGAVHTRGLDFACDIYIWAPLAGTCGVLLLSLVITLYCNHRNRRRVCKCPRP
VVKSGDKPSLSARYV (SEQ ID NO: 53).

[0222] The sequence of human CD8-beta (UniProtKB Q8TD28) is as follows (SEQ ID NO: 54):

MRPRLWLLLAQLTVLHGNSVLQQTPAYIKVQTNKMVMLSCEAKISLSNMRIY
WLRQRQAPSSDSHHEFLALWDSAKGTIHGEEVEQEKIAVFRDASRFILNLTSVKP
EDSGIYFCMIVGSPELTFGKGTQLSVVDLPTTAQPTKKSTLKKRVCRLRPETQK

GPLCSPITLGLLVAGVLVLLVSLGVAIHLCCRRRRARLRFMKQLYK (SEQ ID NO: 54).

[0223] In some embodiments, the CD8 that the polypeptide binds to is expressed on the surface of a cell. In some embodiments, the cell is an immune cell. In some embodiments, the immune cell is a CD7+ T cell, CD4+ T cell, CD8+ T cell, NK cell, alpha-beta T cell, gamma-delta T cell, lymphoid progenitor cell, hematopoietic stem cell, myeloid cell, monocyte, macrophage, central memory T cell, effector memory T cell, stem-cell like memory T cells, naïve T cell, activated T cell, regulatory T cell (TReg), terminally differentiated effector memory T cell (TEMRA), resident memory T cell (TRM) or a T-cellCD8+CCR7+. In some embodiments, the cell is a CD8+ T cell. In some embodiments, the cell is a CD8+ cell.

[0224] In some embodiments, the antibody comprises a Fc region. The Fc region can be linked to the heavy or light chain of the antibody. In some embodiments, the Fc region is an IgG Fc. In some embodiments, the IgG is selected from IgG1, IgG2, IgG3, or IgG4. In some embodiments, the IgG Fc is IgG1 Fc. In some embodiments, the antibody comprises an Fc constant region as set forth below.

ASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPA
AVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKHT
CPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVDVSHEDPEVKFNWYV
DGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIE
KTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQP
ENNYKTTTPVLDSGDSFFLYSKLTVDKSRWQQGNVVFSCSVMHEALHNHYTQKSL
SLSPGK (SEQ ID NO: 34)

[0225] In some embodiments, the IgG fc is IgG2 Fc. In some embodiments, the antibody comprises an Fc constant region of SEQ ID NO: 71 as set forth below.

STKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAV
LQSSGLYSLSSVVTVPSSSLGTQTYTCNVDHKPSNTKVDKTVERKCCVECPPCA
PPVAGPSVFLFPPKPKDTLMISRTPEVTCVVDVSHEDPEVQFNWYVDGVEVHN
AKTKPREEQFNSTFRVSVLTVVHQDWLNGKEYKCKVSNKGLPAPIEKTISKTK
GQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTT
PPMLDSGDSFFLYSKLTVDKSRWQQGNVVFSCSVMHEALHNHYTQKSLSLSPGK
(SEQ ID NO: 71)

[0226] In some embodiments, the IgG fc is IgG4 Fc. In some embodiments, the antibody comprises an Fc constant region of SEQ ID NO: 72 as set forth below.

STKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAV
LQSSGLYSLSSVVTVPSSSLGKTYTCNVDHKPSNTKVDKRVESKYGPPCPSCPA
PEFLGGPSVFLFPPKPKDTLMISRTPEVTCVVDVSQEDPEVQFNWYVDGVEVH
NAKTKPREEQFNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKGLPSSIEKTISKA
KGQPREPQVYTLPPSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKT
TPPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALHNHYTQKSLSLGLK
(SEQ ID NO: 72)

[0227] In some embodiments, polypeptides (e.g. CD8-binding polypeptide) are provided herein. In some embodiments, antibodies (e.g. an anti-CD8 antibody) are provided herein. In some embodiments, the antibody is a recombinant antibody that binds to CD8. In some embodiments, the CD8 protein is a human CD8 protein. In some embodiments, the CD8 protein is a non-human CD8 protein (e.g., mouse, rat, pig, dog, non-human primate). As used herein, the term “recombinant antibody” refers to an antibody that is not naturally occurring. In some embodiments, the term “recombinant antibody” refers to an antibody that is not isolated from a human subject.

[0228] In some embodiments, an antibody, or antigen binding fragment thereof is provided, wherein the antibody or antibody fragment comprises a peptide selected from the following table, which illustrate the CDRs based on Chothia numbering.

Chothia CDRs						
Ab ID No	HCDR1	HCDR2	HCDR3	LCDR1	LCDR2	HCDR3
CD8AB1	RYTFTD Y (SEQ ID NO: 55)	YPYNGG (SEQ ID NO: 56)	DHRYNEGVSF DY (SEQ ID NO: 57)	RASESVDG FGNSFMN (SEQ ID NO: 58)	LASNLES (SEQ ID NO: 59)	QQNNEDPYT (SEQ ID NO: 60)

[0229] In some embodiments, an antibody, or antigen binding fragment thereof is provided, wherein the antibody or antibody fragment comprises a peptide selected from the following table, which illustrate the CDRs based on Kabat numbering.

Kabat CDRs						
Ab ID No	HCDR1	HCDR2	HCDR3	LCDR1	LCDR2	HCDR3
CD8AB1	DYNLH (SEQ ID	FIYPYNGGT GYNQKFKN	DHRYNEGVSF DY (SEQ ID	RASESVDG FGNSFMN	LASNLES (SEQ ID NO: 59)	QQNNEDPYT (SEQ ID

	NO: 61)	(SEQ ID NO: 62)	NO: 57)	(SEQ ID NO: 58)		NO: 60)
--	---------	--------------------	---------	--------------------	--	---------

[0230] In some embodiments, an antibody, or antigen binding fragment thereof is provided, wherein the antibody or antibody fragment comprises a peptide selected from the following table, which illustrate the CDRs based on IMGT numbering.

IMGT CDRs						
Ab ID No	HCDR1	HCDR2	HCDR3	LCDR1	LCDR2	HCDR3
CD8AB1	RYTFTD YN (SEQ ID NO: 63)	IYPYNGGT (SEQ ID NO: 64)	ARDHRYNEGV SFDY (SEQ ID NO: 65)	ESVDGFGN SF (SEQ ID NO: 66)	LA	QQNNEDPYT (SEQ ID NO: 60)

[0231] In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy or light chain CDR as provided in the tables above. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy or light chain CDR as provided in the tables above and binds to non-human primate CD8. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy or light chain CDR as provided in the tables above and binds to human CD8. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a light chain CDR having a sequence selected from SEQ ID NO: 58-60. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a light chain CDR having a sequence of SEQ ID NO: 58. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a light chain CDR having a sequence of SEQ ID NO: 59. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a light chain CDR having a sequence of SEQ ID NO: 60. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain CDR having a sequence selected from SEQ ID NO: 55-57. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain CDR having a sequence of SEQ ID NO: 55. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain CDR having a sequence of SEQ ID NO: 56. In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain CDR having a sequence of SEQ ID NO: 57. The CDRs referenced in the embodiments throughout the present specification can be interchanged with the CDRs that are characterized by

different formats, such as Kabat and IMGT.

[0232] In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a light chain variable region having a LCDR1, a LCDR2, and a LCDR3, wherein the LCDR1 has a sequence of SEQ ID NO: 58, the LCDR2 has a sequence of SEQ ID NO: 59, and the LCDR3 has a sequence of SEQ ID NO: 60.

[0233] In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a light chain variable region having a LCDR1, a LCDR2, and a LCDR3, wherein the LCDR1 has a sequence of SEQ ID NO: 66, the LCDR2 has a sequence of LA, and the LCDR3 has a sequence of SEQ ID NO: 60.

[0234] In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain variable region having a HCDR1, a HCDR2, and a HCDR3, wherein the HCDR1 has a sequence of SEQ ID NO: 55, the HCDR2 has a sequence of SEQ ID NO: 56, and the HCDR3 has a sequence of SEQ ID NO: 57.

[0235] In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain variable region having a HCDR1, a HCDR2, and a HCDR3, wherein the HCDR1 has a sequence of SEQ ID NO: 61, the HCDR2 has a sequence of SEQ ID NO: 62, and the HCDR3 has a sequence of SEQ ID NO: 57.

[0236] In some embodiments, a polypeptide, an antibody, or antibody binding fragment thereof, comprises a heavy chain variable region having a HCDR1, a HCDR2, and a HCDR3, wherein the HCDR1 has a sequence of SEQ ID NO: 63, the HCDR2 has a sequence of SEQ ID NO: 64, and the HCDR3 has a sequence of SEQ ID NO: 65.

[0237] In some embodiments, a polypeptide, an antibody or antibody binding fragment thereof, comprises: (i) a light chain having any one of the foregoing recited combinations of LCDR1, LCDR2, and LCDR3 sequences; and (ii) a heavy chain having any one of the foregoing recited combinations of HCDR1, HCDR2, and HCDR3 sequences.

[0238] The different CDR motifs can be combined in any combination including those not depicted in the table above. For example, the following embodiments are provided as non-limiting examples of such combinations.

[0239] In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises: (i) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID

NO: 58; the light chain CDR2 has the amino acid sequence of SEQ ID NO: 59; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 60; and (ii) a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 55; the heavy chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 56; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 57; or variants of any of the foregoing.

[0240] In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises: (i) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 58; the light chain CDR2 has the amino acid sequence of SEQ ID NO: 59; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 60; and (ii) a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 61; the heavy chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 62; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 57; or variants of any of the foregoing.

[0241] In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises: (i) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 66; the light chain CDR2 has the amino acid sequence of LA; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 60; and (ii) a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 63; the heavy chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 64; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 65; or variants of any of the foregoing.

[0242] In some embodiments, the light chain variable region CDR1 is replaced with any of the other light chain CDR1 sequences. In some embodiments, the light chain variable region CDR2 is replaced with any of the other light chain CDR2 sequences. In some embodiments, the light chain variable region CDR3 is replaced with any of the other light chain CDR3 sequences. In some embodiments, the heavy chain variable region CDR1 is replaced with any of the other heavy chain CDR1 sequences. In some embodiments, the heavy chain variable region CDR2 is replaced with any of the other heavy chain CDR2 sequences. In some embodiments, the heavy

chain variable region CDR3 is replaced with any of the other heavy chain CDR3 sequences.

[0243] In some embodiments, the polypeptide comprises a heavy chain variable region peptide having one of the following sequences, or a variant thereof:

SEQ ID NO:	AB ID NO.	Sequence
67	CD8AB1	EVQLQQSGPELVKPGASVKISCKASRYTFTDYNLHWVKSHEKSL EWIGFLIYPYNGGTGYNQKFKNKAKLTVDYSSSTAYMELRSLTSVD AAVYYCARDHRYNEGVSFDYWGQGTTTLTVSS

[0244] In some embodiments, the polypeptide comprises a light chain variable region peptide having one of the following sequences, or a variant thereof:

SEQ ID NO:	AB ID NO.	Sequence
68	CD8AB1	NIVLTQSPASLAVSLGQRATISCRASESVDFGNSFMNWWYQQKPG QSPKLLIYLASNLESGVPARFSGSGSRDFTLTIDPVEADDAATY YCQQNNEDPYTFGGGTKLEIKR

[0245] In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises a V_H peptide of SEQ ID NO: 67. In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises a V_L peptide of SEQ ID NO: 68. In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises a V_H peptide and a V_L peptide, wherein the wherein the V_H peptide comprises a sequence of SEQ ID NO: 67, or a variant thereof; and the V_L peptide comprises a sequence of SEQ ID NO: 68, or a variant thereof. In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises a V_H peptide and a V_L peptide, wherein the wherein the V_H peptide comprises a sequence of SEQ ID NO: 67, or a variant thereof; and the V_L peptide comprises a sequence of SEQ ID NO: 68, or a variant thereof, and the polypeptide, the antibody, or antigen binding fragment thereof, binds to non-human primate CD8. In some embodiments, a polypeptide, an antibody, or antigen binding fragment thereof, comprises a V_H peptide and a V_L peptide, wherein the wherein the V_H peptide comprises a sequence of SEQ ID NO: 67, or a variant thereof; and the V_L peptide comprises a sequence of SEQ ID NO: 68, or a variant thereof, and the polypeptide, the antibody, or antigen binding fragment thereof, binds to human CD8. In some embodiments, the V_H peptide comprises a sequence of SEQ ID NO: 67; and the V_L peptide comprises a sequence of SEQ ID NO: 68.

[0246] The V_H and the V_L sequences can be in any format, including, but not limited to

an scFv format where the VH and VL regions are linked with a peptide linker. Examples of peptide linkers that can be used to link various peptides provided for herein include, but are not limited to: (GGGGS)_n (SEQ ID NO: 49), wherein each n is independently 1-5. In some embodiments, n is 1. In some embodiments, n is 2. In some embodiments, n is 3. In some embodiments, n is 4. In some embodiments, n is 5. In some embodiments, the variable regions are not linked with a peptide linker. In some embodiments, the polypeptide comprises SEQ ID NO: 67 and SEQ ID NO: 68.

[0247] In some embodiments, the VH and VL polypeptides are linked to a Fc region. In some embodiments, the Fc region is as provided for herein. As provided for herein, the heavy chain can be linked to a Fc region. Non-limiting mutations in the Fc region are provided for herein. In some embodiments, the Fc region further comprises a transmembrane domain. Examples of transmembrane domains include, but are not limited to, a CD8/CD28 transmembrane domain. In some embodiments, the Fc region further comprises a CD8 or CD28 (“CD8/CD28”) transmembrane domain. In some embodiments, the Fc region further comprises a CD8 transmembrane domain. In some embodiments, the Fc region further comprises a CD28 transmembrane domain. In some embodiments, the Fc region comprising a transmembrane domain further comprises an Env incorporation motif. In some embodiments, the Fc region comprising a CD8/CD28 transmembrane domain further comprises an Env incorporation motif. In some embodiments, the VH and VL polypeptides provided herein are linked to an Fc region comprising a transmembrane domain. In some embodiments, the VH and VL polypeptides provided herein are linked to an Fc region comprising a CD8/CD28 transmembrane domain. In some embodiments, the VH and VL polypeptides provided herein are linked to an Fc region comprising a CD8/CD28 transmembrane domain and an Env incorporation motif. In some embodiments, the VH and VL polypeptides provided herein linked to an Fc region comprising a CD8/CD28 transmembrane domain are anchored to the plasma membrane on the surface of a cell. In some embodiments, the cell is an immune cell, such as those provided herein. In some embodiments, the VH having a sequence as set forth in SEQ ID NO: 67 and VL having a sequence as set forth in SEQ ID NO: 68 are linked to an Fc region comprising a transmembrane domain. In some embodiments, the VH having a sequence as set forth in SEQ ID NO: 67 and VL having a sequence as set forth in SEQ ID NO: 68 are linked to an Fc region comprising a CD8/CD28 transmembrane domain. In some embodiments, the VH having a sequence as set forth in SEQ ID NO: 67 and VL having a sequence as set forth in SEQ ID NO: 68 are linked to an Fc region

comprising a CD8/CD28 transmembrane domain and an Env incorporation motif. In some embodiments, the V_H having a sequence as set forth in SEQ ID NO: 67 and V_L having a sequence as set forth in SEQ ID NO: 68 linked to an Fc region comprising a CD8/CD28 transmembrane domain are anchored to the plasma membrane on the surface of a cell. In some embodiments, the cell is an immune cell, such as those provided herein.

[0248] In some embodiments, the V_H peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68 are linked to an Fc region comprising a transmembrane domain. In some embodiments, the V_H peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68 are linked to an Fc region comprising a CD8/CD28 transmembrane domain. In some embodiments, the V_H peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68 are linked to an Fc region comprising a CD8/CD28 transmembrane domain and an Env incorporation motif. In some embodiments, the V_H peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprising a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68 linked to an Fc region comprising a CD8/CD28 transmembrane domain are anchored to the plasma membrane on the surface of a cell. In some embodiments, the cell is an immune cell, such as those provided herein.

[0249] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%,

92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68.

[0250] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68; provided that the V_H peptide and a V_L peptide comprises a light chain CDR having a sequence of SEQ ID NO: 58-60; and/or a heavy chain CDR having a sequence of SEQ ID NO: 55-57. In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68; provided that the V_H peptide and a V_L peptide comprise a light chain CDR1 having a sequence of SEQ ID NO: 58; a light chain CDR2 having a sequence of SEQ ID NO: 59; a light chain CDR3 having a sequence of SEQ ID NO: 60; and/or a heavy chain CDR1 having a sequence of SEQ ID NO: 55; a heavy chain CDR2 having a sequence of SEQ ID NO: 56; and a heavy chain CDR3 having a sequence of SEQ ID NO: 57. In some embodiments, the CDRs in the V_H or V_L chain are as set forth in the combinations provided for herein.

[0251] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68; provided that the V_L peptide comprises a LCDR1 having a sequence of SEQ ID NO: 58; a LCDR2 having a sequence of SEQ ID NO: 59; and a LCDR3 having a sequence of SEQ ID NO: 60; and the V_H peptide comprises a HCDR1 having a sequence of SEQ ID NO: 55; a HCDR2 having a sequence of SEQ ID NO: 56; and a HCDR3 having a sequence of SEQ ID NO: 57.

[0252] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67;

and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68; provided that the V_L peptide comprises a LCDR1 having a sequence of SEQ ID NO: 58, wherein the LCDR1 comprises at most 1 conservative amino acid substitution, a LCDR2 having a sequence of SEQ ID NO: 59, wherein the LCDR2 comprises at most 1 conservative amino acid substitution, and a LCDR3 having a sequence of SEQ ID NO: 60, wherein the LCDR3 comprises at most 1 conservative amino acid substitution; and the V_H peptide comprises a HCDR1 having a sequence of SEQ ID NO: 55, wherein the HCDR1 comprises at most 1 conservative amino acid substitution, a HCDR2 having a sequence of SEQ ID NO: 56, wherein the HCDR2 comprises at most 1 conservative amino acid substitution, and a HCDR3 having a sequence of SEQ ID NO: 57, wherein the HCDR3 comprises at most 1 conservative amino acid substitution.

[0253] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68; provided that the V_H peptide and a V_L peptide comprise a light chain CDR1 having a sequence of SEQ ID NO: 58; a light chain CDR2 having a sequence of SEQ ID NO: 59; a light chain CDR3 having a sequence of SEQ ID NO: 60; and/or a heavy chain CDR1 having a sequence of SEQ ID NO: 61; a heavy chain CDR2 having a sequence of SEQ ID NO: 62; and a heavy chain CDR3 having a sequence of SEQ ID NO: 57. In some embodiments, the CDRs in the V_H or V_L chain are as set forth in the combinations provided for herein.

[0254] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68; provided that the V_L peptide comprises a LCDR1 having a sequence of SEQ ID NO: 58; a LCDR2 having a sequence of SEQ ID NO: 59; and a LCDR3 having a sequence of SEQ ID NO: 60; and the V_H peptide comprises a HCDR1 having a sequence of SEQ ID NO: 61; a HCDR2 having a sequence of SEQ ID NO: 62; and a HCDR3 having a sequence of SEQ ID NO: 57.

[0255] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68; provided that the V_L peptide comprises a LCDR1 having a sequence of SEQ ID NO: 58, wherein the LCDR1 comprises at most 1 conservative amino acid substitution, a LCDR2 having a sequence of SEQ ID NO: 59, wherein the LCDR2 comprises at most 1 conservative amino acid substitution, and a LCDR3 having a sequence of SEQ ID NO: 60, wherein the LCDR3 comprises at most 1 conservative amino acid substitution; and the V_H peptide comprises a HCDR1 having a sequence of SEQ ID NO: 61, wherein the HCDR1 comprises at most 1 conservative amino acid substitution, a HCDR2 having a sequence of SEQ ID NO: 62, wherein the HCDR2 comprises at most 1 conservative amino acid substitution, and a HCDR3 having a sequence of SEQ ID NO: 57, wherein the HCDR3 comprises at most 1 conservative amino acid substitution.

[0256] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68; provided that the V_H peptide and a V_L peptide comprise a light chain CDR1 having a sequence of SEQ ID NO: 66; a light chain CDR2 having a sequence of LA; a light chain CDR3 having a sequence of SEQ ID NO: 60; and/or a heavy chain CDR1 having a sequence of SEQ ID NO: 63; a heavy chain CDR2 having a sequence of SEQ ID NO: 64; and a heavy chain CDR3 having a sequence of SEQ ID NO: 65. In some embodiments, the CDRs in the V_H or V_L chain are as set forth in the combinations provided for herein.

[0257] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68; provided that the V_L peptide comprises a LCDR1 having a sequence of SEQ ID NO: 66; a LCDR2

having a sequence of LA; and a LCDR3 having a sequence of SEQ ID NO: 60; and the V_H peptide comprises a HCDR1 having a sequence of SEQ ID NO: 63; a HCDR2 having a sequence of SEQ ID NO: 64; and a HCDR3 having a sequence of SEQ ID NO: 65.

[0258] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 67; and the V_L peptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 68; provided that the V_L peptide comprises a LCDR1 having a sequence of SEQ ID NO: 66, wherein the LCDR1 comprises at most 1 conservative amino acid substitution, a LCDR2 having a sequence of LA, wherein the LCDR2 comprises at most 1 conservative amino acid substitution, and a LCDR3 having a sequence of SEQ ID NO: 60, wherein the LCDR3 comprises at most 1 conservative amino acid substitution; and the V_H peptide comprises a HCDR1 having a sequence of SEQ ID NO: 63, wherein the HCDR1 comprises at most 1 conservative amino acid substitution, a HCDR2 having a sequence of SEQ ID NO: 64, wherein the HCDR2 comprises at most 1 conservative amino acid substitution, and a HCDR3 having a sequence of SEQ ID NO: 65, wherein the HCDR3 comprises at most 1 conservative amino acid substitution.

[0259] In some embodiments, a polypeptide comprises a V_H peptide and a V_L peptide, wherein the V_H peptide comprises a sequence of SEQ ID NO: 67 and the V_L peptide comprises a sequence of SEQ ID NO: 68.

[0260] In some embodiments, a polypeptide as provided herein binds to non-human primate CD8. In some embodiments, a polypeptide as provided herein binds to human CD8.

[0261] As provided for herein, the different polypeptides (V_H or V_L) described herein can be linked with a peptide linker or not linked with a peptide linker and instead for a continuous sequence. In some embodiments, the peptide linker comprises a sequence of (GGGGS)_n (SEQ ID NO: 49), wherein each n is independently 1-5. In some embodiments, n is 1. In some embodiments, n is 2. In some embodiments, n is 3. In some embodiments, n is 4. In some embodiments, n is 5. The linked peptide format can be represented by a formula of V_H-Z-V_L or V_L-Z-V_H, wherein Z is the peptide linker. In some embodiments, Z is (GGGGS)_n (SEQ ID NO: 49), wherein each n is independently 1-5.

[0262] In some embodiments, a polypeptide comprising the linked peptide represented by

a formula of V_L -Z- V_H comprises a heavy chain variable region as set forth in SEQ ID NO: 67 linked via a linker sequence GGGGSGGGGSGGGGSGGGGS (SEQ ID NO: 50) to a light chain variable region as set forth in SEQ ID NO: 68. In some embodiments, a polypeptide comprising a V_L linked via a peptide linker to a V_H has the sequence as set forth below,

NIVLTQSPASLAVSLGQRATISCRASESVDGFGNSFMNWFYQQKPGQSPKLLIYLA
SNLESGVPARFSGSGSRDFTLTIDPVEADDAATYYCQQNNEDPYTFGGGGTKLEI
KRGGGGSGGGGSGGGGSGGGGSEVQLQQSGPELVKPGASVKISCKASRYTFTDY
NLHWVKLSHEKSLEWIGFIYPYNGGTGYNQKFKNKAKLTVDYSSSTAYMELRSL
TSVDAAVYYCARDHRYNEGVSFDYWGGQTTLTVSS (SEQ ID NO: 69).

[0263] In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 69. In some embodiments, a polypeptide comprises a sequence that is at least 90% identical to a sequence of SEQ ID NO: 69. In some embodiments, a polypeptide comprises a sequence that is at least 95% identical to a sequence of SEQ ID NO: 69. In some embodiments, a polypeptide comprises a sequence that is at least 99% identical to a sequence of SEQ ID NO: 69. In some embodiments, a polypeptide comprises a sequence as set forth in SEQ ID NO: 69. In some embodiments, the polypeptide as set forth in SEQ ID NO: 69 is an antibody, or an antigen binding fragment thereof. In some embodiments, the antibody is an anti-CD8 antibody.

[0264] In some embodiments, a polypeptide comprising the linked peptide represented by a formula of V_H -Z- V_L comprises a light chain variable region as set forth in SEQ ID NO: 68 linked via a linker sequence GGGGSGGGGSGGGGSGGGGS (SEQ ID NO: 50) to a heavy chain variable region as set forth in SEQ ID NO: 67. In some embodiments, a polypeptide comprising a V_H linked via a peptide linker to a V_L has the sequence as set forth below,

EVQLQQSGPELVKPGASVKISCKASRYTFTDYNLHWVKLSHEKSLEWIGFIYPYN
GGTGYNQKFKNKAKLTVDYSSSTAYMELRSLTSVDAAVYYCARDHRYNEGVSF
DYWGQTTLTVSSGGGGSGGGGSGGGGSGGGGSNIVLTQSPASLAVSLGQRATI

SCRASESVDGFGNSFMNWDYQQKPGQSPKLLIYLASNLESGVPARFSGSGSRDFT
LTIDPVEADDAATYYCQQNEDPYTFGGGTKLEIKR (SEQ ID NO: 70).

[0265] In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 70. In some embodiments, a polypeptide comprises a sequence that is at least 90% identical to a sequence of SEQ ID NO: 70. In some embodiments, a polypeptide comprises a sequence that is at least 95% identical to a sequence of SEQ ID NO: 70. In some embodiments, a polypeptide comprises a sequence that is at least 99% identical to a sequence of SEQ ID NO: 70. In some embodiments, a polypeptide comprises a sequence as set forth in SEQ ID NO: 70. In some embodiments, the polypeptide as set forth in SEQ ID NO: 70 is an antibody, or an antigen binding fragment thereof. In some embodiments, the antibody is an anti-CD8 antibody. In some embodiments, the anti-CD8 antibody binds to non-human primate CD8. In some embodiments, the anti-CD8 antibody binds to human CD8.

[0266] In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 69 and comprises an Fc region, such as those provided herein. In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 69 and comprises an Fc region, such as those provided herein, and a transmembrane domain, such as those provided herein. In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 69 and comprises an Fc region, such as those provided herein, and a CD8/CD28 transmembrane domain. In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 69 and comprises an Fc region, such as those provided herein, a transmembrane domain, such as those provided herein, and an Env incorporation motif.

[0267] In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 70 and comprises an Fc region, such as those provided herein. In some

embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 70 and comprises an Fc region, such as those provided herein, and a transmembrane domain, such as those provided herein. In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 70 and comprises an Fc region, such as those provided herein, and a CD8/CD28 transmembrane domain. In some embodiments, a polypeptide comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 70 and comprises an Fc region, such as those provided herein, a transmembrane domain, such as those provided herein, and an Env incorporation motif.

[0268] As provided for herein, the polypeptide, antibodies, or antigen binding fragments thereof can be variants of the sequences.

[0269] The sequences of the polypeptides or antibodies can be modified to yield human IgG antibodies. The conversion of the sequences provided herein can be modified to yield other types of antibodies. The CDRs can also be linked to other antibodies, proteins, or molecules to create antibody fragments that bind CD8.

[0270] In some embodiments, a polypeptide or an antibody as provided herein is a targeting moiety on the surface of an engineered viral particle. In some embodiments, the targeting moiety allows for binding to a target cell. In some embodiments, the targeting moiety is a CD8 binding moiety, such as a polypeptide or an antibody as provided herein. In some embodiments, the targeting moiety comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 69. In some embodiments, the targeting moiety comprises a sequence that is at least 90% identical to a sequence of SEQ ID NO: 69. In some embodiments, the targeting moiety comprises a sequence that is at least 95% identical to a sequence of SEQ ID NO: 69. In some embodiments, the targeting moiety comprises a sequence that is at least 99% identical to a sequence of SEQ ID NO: 69. In some embodiments, the targeting moiety comprises a sequence as set forth in SEQ ID NO: 69. In some embodiments, the targeting moiety as set forth in SEQ ID NO: 69 is an antibody, or an antigen binding fragment thereof. In some embodiments, the targeting moiety is an anti-CD8 antibody.

[0271] In some embodiments, a polypeptide or an antibody as provided for herein is a targeting moiety on the surface of an engineered viral particle. In some embodiments, the engineered viral particle is a pseudotyped viral-like particle. In some embodiments, the targeting moiety allows for binding to a target cell. In some embodiments, the targeting moiety is a CD8 binding moiety, such as a polypeptide or an antibody as provided herein. In some embodiments, the targeting moiety comprises a sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to a sequence of SEQ ID NO: 70. In some embodiments, the targeting moiety comprises a sequence that is at least 90% identical to a sequence of SEQ ID NO: 70. In some embodiments, the targeting moiety comprises a sequence that is at least 95% identical to a sequence of SEQ ID NO: 70. In some embodiments, the targeting moiety comprises a sequence that is at least 99% identical to a sequence of SEQ ID NO: 70. In some embodiments, the targeting moiety comprises a sequence as set forth in SEQ ID NO: 70. In some embodiments, the targeting moiety as set forth in SEQ ID NO: 70 is an antibody, or an antigen binding fragment thereof. In some embodiments, the targeting moiety is an anti-CD8 antibody. In some embodiments, the anti-CD8 antibody binds to non-human primate CD8. In some embodiments, the anti-CD8 antibody binds to human CD8.

[0272] In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 69 and comprises an Fc region, such as those provided herein. In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 69 and comprises an Fc region, such as those provided herein, and a transmembrane domain, such as those provided herein. In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 69 and comprises an Fc region, such as those provided herein, and a CD8/CD28 transmembrane domain. In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 69 and comprises an Fc region, such as those provided herein, a transmembrane domain, such as those provided herein, and an Env incorporation motif.

[0273] In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 70 and comprises an Fc region, such as those provided herein. In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 70 and comprises an Fc region, such as those provided herein, and a transmembrane domain, such as those provided herein. In some embodiments, a polypeptide comprises a sequence having a

sequence as set forth in SEQ ID NO: 70 and comprises an Fc region, such as those provided herein, and a CD8/CD28 transmembrane domain. In some embodiments, a polypeptide comprises a sequence having a sequence as set forth in SEQ ID NO: 70 and comprises an Fc region, such as those provided herein, a transmembrane domain, such as those provided herein, and an Env incorporation motif.

[0274] Although, the binders that bind to CD7 and CD8 are illustrated, in part, with a mutant VSV-G protein, other fusogenic proteins can also be used in place of the VSV-G protein. For example, in some embodiments, the pseudotyped viral-like particles can be pseudotyped using viral glycoproteins other than VSV-G.

[0275] In some embodiments, viral particle can be pseudotyped using a viral glycoprotein from, for example, VSV. In some embodiments, viral particle can be pseudotyped using a viral glycoprotein from, for example, the Paramyxoviridae family. In some embodiments, the pseudotyped viral-like particles are pseudotyped using viral glycoproteins of morbillivirus, such as Measles virus. In some embodiments, the pseudotyped viral-like particles are pseudotyped using viral glycoproteins of the Measles virus. In some embodiments, the pseudotyped viral-like particles are pseudotyped using viral glycoproteins of henipavirus, such as Nipah virus, Cedar virus, or Hendra virus. In some embodiments, the pseudotyped viral-like particles are pseudotyped using viral glycoproteins of the Nipah virus. In some embodiments, a polypeptide or an antibody as provided herein is linked via a linker to an envelope glycoprotein G or H of a virus of the Paramyxoviridae family. In some embodiments, the virus of the Paramyxoviridae family is a morbillivirus, such as Measles virus. In some embodiments, the virus of the Paramyxoviridae family is a henipavirus, such as Nipah virus, Cedar virus, or Hendra virus.

[0276] In some embodiments, the pseudotyped viral-like particles comprising the targeting moieties provided for herein are pseudotyped with viral glycoproteins of viruses of the Rhabdoviridae family. In some embodiments, the pseudotyped viral-like particles are pseudotyped using viral glycoproteins of a vesicular stomatitis New Jersey virus, a vesicular stomatitis Indiana virus, a vesicular stomatitis Alagoas virus, a vesicular stomatitis Maraba virus, or a vesicular stomatitis Carajas virus. In some embodiments, a polypeptide or an antibody as provided herein is linked via a linker to a glycoprotein of a virus of the Rhabdoviridae family. In some embodiments, the virus of the Rhabdoviridae family is the vesicular stomatitis New Jersey virus, the vesicular stomatitis Indiana virus, the vesicular stomatitis Alagoas virus, the vesicular stomatitis Maraba

virus, or the vesicular stomatitis Carajas virus.

[0277] In some embodiments, the pseudotyped viral-like particles comprising the targeting moieties provided for herein are pseudotyped with viral glycoproteins of the Parainfluezna virus, Spodoptera frugiperda virus, Drosophila obscura sigmavirus, Wuhan insect virus 7, Spring viremia of carp virus, or Perch virus. In some embodiments, a polypeptide or an antibody as provided herein is linked via a linker to a glycoprotein of a virus of the Rhabdoviridae family. In some embodiments, the virus of the Rhabdoviridae family is a Parainfluezna virus. In some embodiments, the virus of the Rhabdoviridae family is a Spodoptera frugiperda virus. In some embodiments, the virus of the Rhabdoviridae family is a Drosophila obscura sigmavirus. In some embodiments, the virus of the Rhabdoviridae family is a Wuhan insect virus 7. In some embodiments, the virus of the Rhabdoviridae family is a Spring viremia of carp virus. In some embodiments, the virus of the Rhabdoviridae family is a Perch virus. In some embodiments, a polypeptide or an antibody as provided herein is linked via a linker to a Parainfluezna virus glycoprotein. In some embodiments, a polypeptide or an antibody as provided herein is linked via a linker to a Spodoptera frugiperda rhabdovirus isolate Sf G virus glycoprotein.

[0278] In some embodiments, the pseudotyped viral-like particles comprising the targeting moieties provided for herein are pseudotyped with viral glycoproteins of viruses of the Arenaviridae family. In some embodiments, the pseudotyped viral-like particles are pseudotyped using viral glycoproteins of the Machupo virus. In some embodiments, a polypeptide or an antibody as provided herein is linked via a linker to a glycoprotein of a virus of the Arenaviridae family. In some embodiments, the virus of the Arenaviridae family is a Machupo virus.

[0279] In some embodiments, the viral particle comprising a mutant VSV-G protein (or other viral glycoprotein) and/or targeting moiety as provided for herein comprises a nucleic acid molecule encoding for a heterologous molecule of interest or “cargo.” For example, heterologous molecule of interest is meant to refer to any product that may be encoded by a nucleic acid molecule. As non-limiting examples, “cargo” or “heterologous molecule of interest” may refer to an siRNA, an shRNA, a peptide, a polypeptide, a protein, a viral payload, a viral genome, or a combination thereof. In some embodiments, the polypeptide is a chimeric antigen receptor (“CAR”). In some embodiments, the heterologous molecule of interest is an siRNA, an shRNA, a non-coding RNA (e.g. a guide RNA for a CRISPR system), a peptide, a polypeptide, a protein, a viral payload, a viral genome, a chimeric antigen receptor (“CAR”), or a combination thereof.

In some embodiments, the heterologous molecule of interest is a CAR.

[0280] A “chimeric antigen receptor” or “CAR” as used herein refers to an antigen-binding domain that is fused, directly, or indirectly (e.g. via a hinge or transmembrane domain to an intracellular signaling domain capable of activating or stimulating an immune cell. Most commonly, the CAR's extracellular binding domain is composed of a single chain variable fragment (scFv) derived from fusing the variable heavy and light regions of a murine or humanized monoclonal antibody. Alternatively, scFvs may be used that are derived from Fab's (instead of from an antibody, e.g., obtained from Fab libraries). In various embodiments, this scFv is fused to a transmembrane domain and then to an intracellular signaling domain. However, the antigen binding domain can be any molecule that can bind to the target on the cell. For example, the antigen binding domain of a CAR can be an antibody, a scFv antibody, an antigen binding domain, an ankyrin repeat (e.g. DARPIN), a VHH domain antibody, a nanobody, single domain antibody, a FN3 domain, or any combination thereof. In some embodiments, a CAR includes those that solely provide CD3 ζ signals upon antigen binding. In some embodiments, the CAR includes those that provide both costimulation (e.g. CD28 or CD137) and activation (CD3 ζ). In some embodiments, the CARs include those that provide multiple costimulation (e.g. CD28 and CD137) and activation (CD3 ζ). In various embodiments, the CAR is selected to have high affinity or avidity for the antigen. In some embodiments, the antigen-binding domain binds to CD20. In some embodiments, the antigen-binding domain comprises a CD20 antibody, or fragment thereof. In some embodiments, antibody fragments are as provided for herein, such as but not limited to a scFv antibody, an antigen binding domain, an ankyrin repeat (e.g. DARPIN), a VHH domain antibody, a nanobody, single domain antibody, a FN3 domain, or any combination thereof.

[0281] In some embodiments, the antigen-binding domain of the CAR comprises a V_H domain, a V_L domain, or a V_H and a V_L domain. In some embodiments, the V_H domain comprises an amino acid sequence having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 73, or any value or range in-between.

EVQLVESGGGLVQPGRSLRLSCAASGFTFNDYAMHWVRQAPGKGLEWVSTISW
NSGSIGYADSVKGRFTISRDNAKKSLYLQMNSLRAEDTALYYCAKDIQYGNYYY
GMDVWGQGTTVTVSS (SEQ ID NO: 73)

In some embodiments, the V_H domain comprises an amino acid sequence having at least 90% identity to SEQ ID NO: 73. In some embodiments, the V_H domain comprises an amino acid sequence having at least 95% identity to SEQ ID NO: 73. In some embodiments, the V_H domain comprises an amino acid sequence having at least 98% identity to SEQ ID NO: 73. In some embodiments, the V_H domain comprises an amino acid sequence having at least 99% identity to SEQ ID NO: 73. In some embodiments, the V_H domain comprises an amino acid sequence having the sequence of SEQ ID NO: 73.

[0282] In some embodiments, the V_L domain comprises an amino acid sequence having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 74, or any value or range in-between.

EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRA
TGIPARFSGSGSGTDFTLTISSLEPEDFAVYYCQQRSNWPITFGQGTRLEIK (SEQ
ID NO: 74)

In some embodiments, the V_L domain comprises an amino acid sequence having at least 90% identity to SEQ ID NO: 74. In some embodiments, the V_L domain comprises an amino acid sequence having at least 95% identity to SEQ ID NO: 74. In some embodiments, the V_L domain comprises an amino acid sequence having at least 98% identity to SEQ ID NO: 74. In some embodiments, the V_L domain comprises an amino acid sequence having at least 99% identity to SEQ ID NO: 74. In some embodiments, the V_L domain comprises an amino acid sequence having the sequence of SEQ ID NO: 74.

[0283] In some embodiments, the antigen-binding domain of the CAR comprises a V_H domain and a V_L domain. In some embodiments, the V_H and V_L domain are not linked by a linker peptide. In some embodiments, the V_H and V_L domain are linked by a linker peptide, such as those as provided for herein, including but not limited to: (GGGGS)_n (SEQ ID NO: 49), wherein each n is independently 1-5. In some embodiment n is 1. In some embodiment n is 2. In some embodiment n is 3. In some embodiment n is 4. In some embodiment n is 5.

[0284] In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 73, and comprises a V_L having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%,

80

V_H domain having at least 99% identity to SEQ ID NO: 73, and comprises a V_L having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 74. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having an amino acid sequence of SEQ ID NO: 73, and comprises a V_L having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 74. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 90% identity to SEQ ID NO: 73, and comprises a V_L having at least 90% identity to SEQ ID NO: 74. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 95% identity to SEQ ID NO: 73, and comprises a V_L having at least 90% identity to SEQ ID NO: 74. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 90% identity to SEQ ID NO: 73, and comprises a V_L having at least 95% identity to SEQ ID NO: 74. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 95% identity to SEQ ID NO: 73, and comprises a V_L having at least 95% identity to SEQ ID NO: 74. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 98% identity to SEQ ID NO: 73, and comprises a V_L having at least 98% identity to SEQ ID NO: 74. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 99% identity to SEQ ID NO: 73, and comprises a V_L having at least 99% identity to SEQ ID NO: 74. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having an amino acid sequence of SEQ ID NO: 73, and comprises a V_L having an amino acid sequence of SEQ ID NO: 74.

[0285] In some embodiments, the antigen-binding domain of the CAR comprises a formula of V_H-Z-V_L, wherein V_H is a heavy chain variable region comprising the amino acid sequence of SEQ ID NO: 73, Z is a linker comprising the amino acid sequence GGGGSGGGGSGGGGS (SEQ ID NO: 77), and V_L is a light chain variable region comprising the amino acid sequence of SEQ ID NO: 74. In some embodiments, the antigen-binding domain of the CAR comprising a formula of V_H-Z-V_L has an amino acid sequence as set forth below:

EVQLVESGGGLVQPGRSLRLSCAASGFTFNDYAMHWVRQAPGKGLEWVSTISW
 NSGSIGYADSVKGRFTISRDNAAKSLYLQMNSLRAEDTALYYCAKDIQYGNYYY
 GMDVWGQGTTTVTVSSGGGGSGGGSGGGGSEIVLTQSPATLSLSPGERATLSCR
 ASQSVSSYLAWYQQKPGQAPRLLIYDASNRAITGIPARFSGSGSGTDFTLTISLEP
 EDFAVYYCQQRSNWPITFGQGTRLEIK (SEQ ID NO: 75)

[0286] In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a sequence of SEQ ID NO: 75. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 90% identity to a sequence of SEQ ID NO: 75. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 95% identity to a sequence of SEQ ID NO: 75. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 98% identity to a sequence of SEQ ID NO: 75. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 99% identity to a sequence of SEQ ID NO: 75. In some embodiments, the antigen-binding domain of the CAR comprises the amino acid sequence of SEQ ID NO: 75.

[0287] In some embodiments, the antigen-binding domain of the CAR comprises a formula of V_L -Z- V_H , wherein V_L is a light chain variable region comprising the amino acid sequence of SEQ ID NO: 74, Z is a linker comprising the amino acid sequence GGGGSGGGSGGGGS (SEQ ID NO: 77), and V_H is a light chain variable region comprising the amino acid sequence of SEQ ID NO: 73. In some embodiments, the antigen-binding domain of the CAR comprising a formula of V_L -Z- V_H has an amino acid sequence as set forth below:

EIVLTQSPATLSLSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRA
 TGIPARFSGSGSGTDFTLTISLEPEDFAVYYCQQRSNWPITFGQGTRLEIKGGGGG
 SGGGSGGGGSEVQLVESGGGLVQPGRSLRLSCAASGFTFNDYAMHWVRQAPG
 KGLEWVSTISWNSGSIGYADSVKGRFTISRDNAAKSLYLQMNSLRAEDTALYYC
 AKDIQYGNYYYGMDVWGQGTTTVTVSS (SEQ ID NO: 76)

[0288] In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,

97%, 98%, 99%, or 100% identity to a sequence of SEQ ID NO: 76. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 90% identity to a sequence of SEQ ID NO: 76. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 95% identity to a sequence of SEQ ID NO: 76. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 98% identity to a sequence of SEQ ID NO: 76. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 99% identity to a sequence of SEQ ID NO: 76. In some embodiments, the antigen-binding domain of the CAR comprises the amino acid sequence of SEQ ID NO: 76.

[0289] In some embodiments, the antigen-binding domain of the CAR comprises a V_H domain, a V_L domain, or a V_H and a V_L domain. In some embodiments, the V_H domain comprises an amino acid sequence having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 78, or any value or range in-between.

DIVLTQSPAILSASPGEKVTMTCRASSSVNYMDWYQKKPGSSPKPWIYATSNLAS
GVPARFSGSGSGTSYSLTISRVEAEDAATYYCQQWSFNPPTFGGGKLEIKGSTS
(SEQ ID NO: 78)

In some embodiments, the V_H domain comprises an amino acid sequence having at least 90% identity to SEQ ID NO: 78. In some embodiments, the V_H domain comprises an amino acid sequence having at least 95% identity to SEQ ID NO: 78. In some embodiments, the V_H domain comprises an amino acid sequence having at least 98% identity to SEQ ID NO: 78. In some embodiments, the V_H domain comprises an amino acid sequence having at least 99% identity to SEQ ID NO: 78. In some embodiments, the V_H domain comprises an amino acid sequence having the sequence of SEQ ID NO: 78.

[0290] In some embodiments, the V_L domain comprises an amino acid sequence having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 79, or any value or range in-between.

EVQLQQSGAELVKPGASVKMSCKASGYTFTSYNMHWVKQTPGQGLEWIGAIYP
GNGDTSYNQKFKGKATLTADKSSSTAYMQLSSLTSEDSADYYCARSNYYGSSY
WFFDVWGAGTTVTVSS (SEQ ID NO: 79)

In some embodiments, the V_L domain comprises an amino acid sequence having at least 90% identity to SEQ ID NO: 79. In some embodiments, the V_L domain comprises an amino acid sequence having at least 95% identity to SEQ ID NO: 79. In some embodiments, the V_L domain comprises an amino acid sequence having at least 98% identity to SEQ ID NO: 79. In some embodiments, the V_L domain comprises an amino acid sequence having at least 99% identity to SEQ ID NO: 79. In some embodiments, the V_L domain comprises an amino acid sequence having the sequence of SEQ ID NO: 79.

[0291] In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 78, and comprises a V_L having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 78, and comprises a V_L having at least 90% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 78, and comprises a V_L having at least 95% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 78, and comprises a V_L having at least 98% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 78, and comprises a V_L having at least 99% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 78, and comprises a V_L having the sequence of SEQ ID NO: 79. In some

embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 90% identity to SEQ ID NO: 78, and comprises a V_L having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 95% identity to SEQ ID NO: 78, and comprises a V_L having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 98% identity to SEQ ID NO: 78, and comprises a V_L having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 99% identity to SEQ ID NO: 78, and comprises a V_L having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having an amino acid sequence of SEQ ID NO: 78, and comprises a V_L having at least 75%, 80%, 85%, 85%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 90% identity to SEQ ID NO: 78, and comprises a V_L having at least 90% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 95% identity to SEQ ID NO: 78, and comprises a V_L having at least 90% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 90% identity to SEQ ID NO: 78, and comprises a V_L having at least 95% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 95% identity to SEQ ID NO: 78, and comprises a V_L having at least 95% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 98% identity to SEQ ID NO: 78, and comprises a V_L having at least 98% identity to SEQ ID NO: 79. In some

embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having at least 99% identity to SEQ ID NO: 78, and comprises a V_L having at least 99% identity to SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a V_H domain and a V_L domain comprises a V_H domain having an amino acid sequence of SEQ ID NO: 78, and comprises a V_L having an amino acid sequence of SEQ ID NO: 79.

[0292] In some embodiments, the antigen-binding domain of the CAR comprises a formula of V_H-Z-V_L, wherein V_H is a heavy chain variable region comprising the amino acid sequence of SEQ ID NO: 78, Z is a linker comprising the amino acid sequence GGGGSGGGGSGGGGS (SEQ ID NO: 77), and V_L is a light chain variable region comprising the amino acid sequence of SEQ ID NO: 79. In some embodiments, the antigen-binding domain of the CAR comprising a formula of V_H-Z-V_L has an amino acid sequence as set forth below:

DIVLTQSPAILSASPGEKVTMTCRASSSVNYMDWYQKKPGSSPKPWIYATSNLAS
GVPARFSGSGSGTSYSLTISRVEAEDAATYYCQQWSFNPPTFGGGKLEIKGSTS
GGGGSGGGGSGGGGSSEVQLQQSGAELVKPGASVKMSCKASGYTFTSYNMHW
VKQTPGQGLEWIGAIYPGNGDTSYNQKFKGKATLTADKSSSTAYMQLSSLTSED
SADYYCARSNYYGSSYWFFDVWGAGTTTVTVSS (SEQ ID NO: 80)

[0293] In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a sequence of SEQ ID NO: 80. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 90% identity to a sequence of SEQ ID NO: 80. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 95% identity to a sequence of SEQ ID NO: 80. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 98% identity to a sequence of SEQ ID NO: 80. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 99% identity to a sequence of SEQ ID NO: 80. In some embodiments, the antigen-binding domain of the CAR comprises the amino acid sequence of SEQ ID NO: 80.

[0294] In some embodiments, the antigen-binding domain of the CAR comprises a formula of V_L-Z-V_H, wherein V_L is a light chain variable region comprising the amino acid sequence of

SEQ ID NO: 79, Z is a linker comprising the amino acid sequence GGGGSGGGGSGGGGS (SEQ ID NO: 77), and V_H is a light chain variable region comprising the amino acid sequence of SEQ ID NO: 78. In some embodiments, the antigen-binding domain of the CAR comprising a formula of V_L-Z-V_H has an amino acid sequence as set forth below:

SEVQLQQSGAELVKPGASVKMSCASGYTFTSYNMHWVKQTPGQGLEWIGAIY
 PNGGDTSYNQKFKGKATLTADKSSSTAYMQLSSLTSEDSADYYCARSNYYGSSY
 WFFDVWGAGTTVTVSSGGGGSGGGGSGGGGSDIVLTQSPAILSASPGEKVTMTC
 RASSSVNYMDWYQKKPGSSPKPWIYATSNLASGVPARFSGSGSGTSYSLTISRVE
 AEDAATYYCQQWSFNPPTFGGGTKLEIKGSTS (SEQ ID NO: 81)

[0295] In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to a sequence of SEQ ID NO: 81. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 90% identity to a sequence of SEQ ID NO: 81. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 95% identity to a sequence of SEQ ID NO: 81. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 98% identity to a sequence of SEQ ID NO: 81. In some embodiments, the antigen-binding domain of the CAR comprises an amino acid sequence having at least 99% identity to a sequence of SEQ ID NO: 81. In some embodiments, the antigen-binding domain of the CAR comprises the amino acid sequence of SEQ ID NO: 81.

[0296] In some embodiments, the antigen-binding domain of the CAR comprises rituximab, ocrelizumab, obinutuzumab, ofatumumab, ibritumomab tiuxetan, tositumomab, or ublituximab. In some embodiment, the antigen-binding domain comprises rituximab. In some embodiment, the antigen-binding domain comprises ofatumumab. In some embodiments, the CAR comprises the 4-1BB domain as well. These are merely illustrative in nature and are not limiting to the present embodiments and any chimeric antigen receptor can be delivered in conjunction with the viral particles and vectors provided for herein. These are non-limiting examples of CARs and any CAR construct could be encoded for by the nucleic acid molecule.

[0297] In some embodiments, the pseudotyped viral particle further comprises a heterologous nucleic acid molecule encoding a cargo of interest. The nucleic acid molecule may

be useful for modulating the expression of a target gene. In some embodiments, the cargo can be used to modulate the activity of a cell or express a protein that is trafficked to the surface of the target cell. Therefore, in some embodiments, the nucleic acid may comprise an siRNA or an shRNA. The nucleic acid may also encode for a cargo of interest. Therefore, in some embodiments, the cargo of interest may comprise a polypeptide or portion thereof, a protein or portion thereof, a chimeric antigen receptor or portion thereof, or a tumor antigen or a portion thereof. In some embodiments, the cargo of interest is an antibody that is produced by the virus, which can then be secreted by the cell that is infected with the virus. The term “protein” can refer to any polypeptide that carries a native function in a cellular environment. Therefore, in some embodiments, the protein encoded by the nucleic acid cargo of interest may comprise an enzyme, a nuclear receptor, a transporter, a ribosomal protein, a membrane bound protein, a cytoplasmic protein, a G-protein coupled receptor, a voltage gated ion channel, a secretory protein, a mitochondria protein, a cytokine, a chimeric antigen receptor, a tumor antigen, or a portion or chimeric species thereof.

[0298] Without being bound to any particular theory, the viral particle comprising the mutant VSV-G protein or other viral glycoprotein as provided for herein that comprises a targeting moiety can be used to express the heterologous molecule of interest in the target cell. Thus, for example, the CAR can be expressed in a T cell that is targeted by a viral particle pseudotyped with a VSV-G protein as provided for herein. Where the T cell is the intended target the viral particle can comprise a targeting moiety that binds to a target on the surface of a T cell, such as, but not limited to CD2, CD3, CD4, CD5, CD7 or CD8. In some embodiments, the target is CD2. In some embodiments, the target is CD3. In some embodiments, the target is CD4. In some embodiments, the target is CD5. In some embodiments, the target is CD6. In some embodiments, the target is CD7. In some embodiments, the target is CD8.

[0299] Also provided herein are viral particles comprising a targeting moiety that binds to CD7 or CD8, or a pharmaceutical composition comprising the same. In some the viral particle comprises a targeting moiety that binds to CD7. . In some the viral particle comprises a targeting moiety that binds to CD7.

[0300] In some embodiments, the viral particle that comprises a targeting moiety that binds to CD7 or CD8 is a pseudotyped viral particle. The viral particle can be pseudotyped with a viral glycoprotein, such as those, but not limited to, described herein. In some embodiments, the viral

particle is pseudotyped with a VSV-G protein or mutant thereof, including, but not limited to, those provided for herein. The targeting moiety can be any antibody that or binder that binds to CD7 or CD8. Non-limiting examples are provided for herein.

[0301] In addition to being pseudotyped, the viral particles comprising a targeting moiety that binds to CD7 or CD8 can comprise a heterologous nucleic acid molecule encoding a molecule of interest. This can be, for example, a siRNA, an shRNA, a non-coding RNA (e.g. a guide RNA for a CRISPR system), a peptide, a polypeptide, a protein, a viral payload, a viral genome, or a combination thereof. In some embodiments, the heterologous molecule of interest is a chimeric antigen receptor (“CAR”).

[0302] In some embodiments, the pseudotyped viral particle is a recombinant lentivirus. In some embodiments, the recombinant pseudotyped viral particle is replication competent. In some embodiments, the recombinant pseudotyped viral particle is replication incompetent.

[0303] In some embodiments, a pharmaceutical composition is provided comprising any composition, particle, or component provided for herein, such as an envelope pseudotyped viral particles or vectors as provided for herein, i.e. a particle that comprises a VSV-G mutant protein provided for herein or other viral glycoprotein, or particles that comprise a targeting moiety and/or a heterologous nucleic acid molecule as provided for herein.

[0304] In some embodiments, methods of delivering a cargo of interest to a cell are provided. In some embodiments, the methods comprise contacting the cell with the pseudotyped viral-like particles or viral vectors as provided for herein, or a pharmaceutical composition comprising the same.

[0305] In some embodiments, methods of delivering a cargo of interest to a cell in a subject are provided. In some embodiments, the methods comprise administering to the subject the pseudotyped viral-like particles or viral vectors as provided for herein, or a pharmaceutical composition comprising the same. In some embodiments, the cargo is a chimeric antigen receptor or as otherwise provided for herein.

[0306] In some embodiments, methods for of delivering a chimeric antigen receptor to a T-cell in a subject are provided. In some embodiments, the methods comprising administering to the subject the pseudotyped viral-like particles or viral vectors as provided for herein, or a pharmaceutical composition comprising the same, wherein the pseudotyped viral-like particle or

viral vector comprises a heterologous nucleic acid molecule encoding the chimeric antigen receptor.

[0307] Also provided herein are nucleic acid molecules encoding a mutant VSV-G protein as provided for herein.

[0308] Also provided herein are nucleic acid molecules encoding a targeting moiety, such as those provided for herein.

[0309] Methods of making the viral like particles or vectors are also provided. In some embodiments, the methods comprise making a viral like particles or vectors comprising a mutant VSV-G protein or other viral glycoprotein. In some embodiments, the methods comprise transfecting or transducing a packaging cell line with the nucleic acid molecules encoding a mutant VSV-G protein or another viral glycoprotein as provided for herein under conditions sufficient to produce the pseudotyped viral-like particles or viral vectors. In some embodiments, the methods comprise transfecting or transducing a packaging cell line with the plurality of nucleic acid molecules provided for herein under conditions sufficient to produce the pseudotyped viral-like particles or viral vectors and optionally with a targeting moiety, such as one that binds to a target antigen, such as those provided for herein. In some embodiments, the targeting moiety binds to CD7 or CD8. In some embodiments, the targeting moiety comprises a sequence that binds to CD7 or CD8, such as those provided for herein. In some embodiments, methods further comprise isolating the pseudotyped viral-like particle or viral vector. In some embodiments, the nucleic acid molecules also comprise a nucleic acid molecule encoding a targeting moiety and/or a cargo that is to be delivered by the viral vector that is produced.

[0310] Accordingly, in some embodiments, cells are provided comprising heterologous nucleic acid molecules that encode the components to produce the virus. Packaging cell lines are known in the art and can be modified with the molecules of interest to produce the viral particle of interest.

[0311] For example, a cell or population of cells is provided that comprises one or more heterologous nucleic acid molecules encoding for a targeting moiety and/or a viral glycoprotein. In some embodiments, the cell comprises heterologous nucleic acid molecules encoding for a targeting moiety and a viral glycoprotein. In some embodiments, the heterologous nucleic acid molecules comprise a nucleic acid sequence encoding for a targeting moiety that binds to CD7 or CD8. In some embodiments, the targeting moiety that is encoded for comprises an amino acid

sequence as provided for herein. In some embodiments, the viral glycoprotein encoded for is any viral glycoprotein that can function as a fusogenic protein to facilitate entry of the viral particle into a cell that expresses CD7 or CD8. In some embodiments, the viral glycoprotein is wild-type or a mutant VSV-G protein, such as those provided for herein. Other viral glycoproteins that can be encoded for are also described herein and can be used.

[0312] The viral particles provided for herein can be produced or made by, for example, culturing the cell under conditions sufficient to make the viral particle. In some embodiments, the cell is what is called a packaging cell line that provides components of the virus to produce the viral particle. These can be structural or non-structural viral components or proteins.

[0313] Also provided for herein are methods of treating cancer in a subject. In some embodiments, the methods comprise administering to the subject the pseudotyped viral-like particles or viral vectors as provided for herein, or a pharmaceutical composition comprising the same, wherein the pseudotyped viral-like particle or viral vector comprises a heterologous nucleic acid molecule encoding the chimeric antigen receptor.

[0314] Also provided herein are methods of treating a disease in a subject in need thereof.

[0315] In some embodiments, the methods provided include, but are not limited to, methods of treating a disease in a subject in need thereof, comprising administering to the subject the viral particle(s) provided herein to treat the disease.

[0316] In certain embodiments, the disease is a cancer. In addition, the compositions provided for herein can be used in methods for the treatment of any condition related to a cancer, such as a cell-mediated immune response against a tumor cell(s), where it is desirable to treat or alleviate the disease. The types of cancers to be treated include, but are not limited to, carcinoma, blastoma, sarcoma, certain leukemia or lymphoid malignancies, benign and malignant tumors, malignancies e.g., sarcomas, carcinomas, and melanomas. Other exemplary cancers include, but are not limited to, breast cancer, prostate cancer, ovarian cancer, cervical cancer, skin cancer, pancreatic cancer, colorectal cancer, renal cancer, liver cancer, brain cancer, lymphoma, leukemia, lung cancer, thyroid cancer, and the like. In some embodiments, the cancer is the cancer is carcinoma, blastoma, sarcoma, leukemia, lymphoid malignancies, benign tumors, malignant tumors, sarcoma, carcinoma, melanoma, breast cancer, prostate cancer, ovarian cancer, cervical cancer, skin cancer, pancreatic cancer, colorectal cancer, renal cancer, liver cancer, brain cancer, lymphoma, leukemia, lung cancer, thyroid cancer, B cell related cancer or T cell related cancer.

The cancers may be non-solid tumors (such as hematological tumors) or solid tumors. Adult tumors/cancers and pediatric tumors/cancers are also included. In one embodiment, the cancer is a hematological tumor. In one embodiment, the cancer is a carcinoma. In one embodiment, the cancer is a sarcoma. In one embodiment, the cancer is a leukemia. In one embodiment the cancer is a solid tumor.

[0317] Solid tumors are abnormal masses of tissue that usually do not contain cysts or liquid areas. Solid tumors can be benign or malignant. Different types of solid tumors are named for the type of cells that form them (such as sarcomas, carcinomas, and lymphomas). Examples of solid tumors, such as sarcomas and carcinomas, include fibrosarcoma, myxosarcoma, liposarcoma, chondrosarcoma, osteosarcoma, synovioma, mesothelioma, Ewing's tumor, leiomyosarcoma, rhabdomyosarcoma, colon carcinoma, lymphoid malignancy, pancreatic cancer, breast cancer, lung cancers, ovarian cancer, prostate cancer, hepatocellular carcinoma, squamous cell carcinoma, basal cell carcinoma, adenocarcinoma, sweat gland carcinoma, medullary thyroid carcinoma, papillary thyroid carcinoma, pheochromocytomas sebaceous gland carcinoma, papillary carcinoma, papillary adenocarcinomas, medullary carcinoma, bronchogenic carcinoma, renal cell carcinoma, hepatoma, bile duct carcinoma, choriocarcinoma, Wilms' tumor, cervical cancer, testicular tumor, seminoma, bladder carcinoma, melanoma, CNS tumors (such as a glioma (such as brainstem glioma and mixed gliomas), glioblastoma (also known as glioblastoma multiforme) astrocytoma, CNS lymphoma, germinoma, medulloblastoma, Schwannoma craniopharyngioma, ependymoma, pinealoma, hemangioblastoma, acoustic neuroma, oligodendroglioma, meningioma, neuroblastoma, retinoblastoma and brain metastases).

[0318] Carcinomas that can be amenable to therapy by the methods disclosed herein include, but are not limited to, esophageal carcinoma, hepatocellular carcinoma, basal cell carcinoma (a form of skin cancer), squamous cell carcinoma (various tissues), bladder carcinoma, including transitional cell carcinoma (a malignant neoplasm of the bladder), bronchogenic carcinoma, colon carcinoma, colorectal carcinoma, gastric carcinoma, lung carcinoma, including small cell carcinoma and non-small cell carcinoma of the lung, adrenocortical carcinoma, thyroid carcinoma, pancreatic carcinoma, breast carcinoma, ovarian carcinoma, prostate carcinoma, adenocarcinoma, sweat gland carcinoma, sebaceous gland carcinoma, papillary carcinoma, papillary adenocarcinoma, cystadenocarcinoma, medullary carcinoma, renal cell carcinoma, ductal carcinoma in situ or bile duct carcinoma, choriocarcinoma, seminoma, embryonal

carcinoma, Wilm's tumor, cervical carcinoma, uterine carcinoma, testicular carcinoma, osteogenic carcinoma, epithelial carcinoma, and nasopharyngeal carcinoma.

[0319] In certain exemplary embodiments, the compositions provided herein can be used in methods to treat a myeloma, or a condition related to myeloma. Examples of myeloma or conditions related thereto include, without limitation, light chain myeloma, non-secretory myeloma, monoclonal gamopathy of undertermined significance (MGUS), plasmacytoma (e.g., solitary, multiple solitary, extramedullary plasmacytoma), amyloidosis, and multiple myeloma. In some embodiments, methods of treating multiple myeloma are provided. In some embodiments, the multiple myeloma is refractory myeloma. In some embodiments, the multiple myeloma is relapsed myeloma.

[0320] In certain exemplary embodiments, the in vivo modified immune cells produced using the compositions provided herein are used to treat a melanoma, or a condition related to melanoma. Examples of melanoma or conditions related thereto include, without limitation, superficial spreading melanoma, nodular melanoma, lentigo maligna melanoma, acral lentiginous melanoma, amelanotic melanoma, or melanoma of the skin (e.g., cutaneous, eye, vulva, vagina, rectum melanoma). In some embodiments, the melanoma is cutaneous melanoma. In some embodiments, the melanoma is refractory melanoma. In some embodiments, the melanoma is relapsed melanoma.

[0321] In some embodiments, the compositions provided herein are used to treat a sarcoma, or a condition related to sarcoma. Examples of sarcoma or conditions related thereto include, without limitation, angiosarcoma, chondrosarcoma, chordoma, endotheliosarcoma, Ewing's sarcoma, fibrosarcoma, gastrointestinal stromal tumor, leiomyosarcoma, liposarcoma, lymphangiosarcoma, lymphangioendotheliosarcoma, mesothelioma, malignant peripheral nerve sheath tumor, myxosarcoma, osteogenic sarcoma, osteosarcoma, pleomorphic sarcoma, rhabdomyosarcoma, synovioma, synovial sarcoma, and other soft tissue sarcomas. In some embodiments, the sarcoma is synovial sarcoma. In some embodiments, the sarcoma is liposarcoma such as myxoid/round cell liposarcoma, differentiated/dedifferentiated liposarcoma, or pleomorphic liposarcoma. In some embodiments, the sarcoma is myxoid/round cell liposarcoma. In some embodiments, the sarcoma is refractory sarcoma. In some embodiments, the sarcoma is relapsed sarcoma.

[0322] In some embodiments, the subject has been treated with a therapeutic agent

targeting the disease or condition, e.g. the tumor, prior to administration of the composition. In some aspects, the subject is refractory or non-responsive to the other therapeutic agent. In some embodiments, the subject has persistent or relapsed disease, e.g., following treatment with another therapeutic intervention, including chemotherapy, radiation, and/or hematopoietic stem cell transplantation (HSCT), e.g., allogenic HSCT. In some embodiments, the administration effectively treats the subject despite the subject having become resistant to another therapy.

[0323] In some embodiments, the subject is responsive to the other therapeutic agent, and treatment with the therapeutic agent reduces disease burden. In some aspects, the subject is initially responsive to the therapeutic agent, but exhibits a relapse of the disease or condition over time. In some embodiments, the subject has not relapsed. In some such embodiments, the subject is determined to be at risk for relapse, such as at a high risk of relapse, and thus the composition is administered prophylactically, e.g., to reduce the likelihood of or prevent relapse. In some aspects, the subject has not received prior treatment with another therapeutic agent.

[0324] The administration of the compositions may be carried out in any convenient manner known to those of skill in the art. For example, the compositions may be administered to a subject by aerosol inhalation, injection, ingestion, transfusion, implantation or transplantation. The compositions described herein may be administered to a patient transarterially, subcutaneously, intradermally, intratumorally, intranodally, intramedullary, intramuscularly, by intravenous (i.v.) injection, intraperitoneally, intranasally, intracranially, or intraosseously. In other instances, the compositions are injected directly into a site of a local disease site in the subject, a lymph node, an organ, a tumor, and the like.

[0325] For the prevention or treatment of disease, the appropriate dosage may depend on the type of disease to be treated, the severity and course of the disease, whether the composition is administered for preventive or therapeutic purposes, previous therapy, the subject's clinical history and response to the treatment, and the discretion of the attending physician. The composition is, in some embodiments, suitably administered to the subject at one time or over a series of treatments.

[0326] In some embodiments, the composition is administered as part of a combination treatment, such as simultaneously with or sequentially with, in any order, another therapeutic intervention, such as an antibody or produced cell or receptor or agent, such as a cytotoxic or therapeutic agent. The composition(s), in some embodiments, is co-administered with one or more

additional therapeutic agents or in connection with another therapeutic intervention, either simultaneously or sequentially in any order. In some contexts, the composition is co-administered with another therapy sufficiently close in time such that the composition enhances the effect of one or more additional therapeutic agents, or vice versa. In some embodiments, the composition is administered prior to the one or more additional therapeutic agents. In some embodiments, the composition is administered after the one or more additional therapeutic agents. In some embodiments, the one or more additional agents includes a cytokine, such as IL-2, for example, to enhance persistence. In some embodiments, the methods comprise administration of a chemotherapeutic agent. In some embodiments, the methods do not comprise the administration of a chemotherapeutic agent.

[0327] In certain embodiments, the compositions may be administered to a subject in combination with an immune checkpoint antibody (e.g., an anti-PD1, anti-CTLA-4, or anti-PDL1 antibody). For example, viral vectors may be administered in combination with an antibody or antibody fragment targeting, for example, PD-1 (programmed death 1 protein). Examples of anti-PD-1 antibodies include, but are not limited to, pembrolizumab (KEYTRUDA®, formerly lambrolizumab, also known as MK-3475), and nivolumab (BMS-936558, MDX-1106, ONO-4538, OPDIVO®) or an antigen-binding fragment thereof. In certain embodiments, the compositions may be administered in combination with an anti-PD-L1 antibody or antigen-binding fragment thereof. Examples of anti-PD-L1 antibodies include, but are not limited to, BMS-936559, MPDL3280A (TECENTRIQ®, Atezolizumab), and MEDI4736 (Durvalumab, Imfinzi). In certain embodiments, the composition may be administered in combination with an anti-CTLA-4 antibody or antigen-binding fragment thereof. An example of an anti-CTLA-4 antibody includes, but is not limited to, Ipilimumab (trade name Yervoy). Other types of immune checkpoint modulators may also be used including, but not limited to, small molecules, siRNA, miRNA, and CRISPR systems. Immune checkpoint modulators may be administered before, after, or concurrently with the viral vector. In certain embodiments, combination treatment comprising an immune checkpoint modulator may increase the therapeutic efficacy of a therapy comprising a composition as provided herein. The other therapeutic can be administered simultaneously, before, or after the compositions provided herein are administered to the subject.

[0328] In certain embodiments, the subject is provided a secondary treatment. Secondary treatments include but are not limited to chemotherapy, radiation, surgery, and medications. In

some embodiments, the subject is not provided a secondary treatment.

[0329] In some embodiments, the methods are performed without a lymphodepletion step, such as the administration of cyclophosphamide and/or fludarabine.

[0330] In some embodiments, the subject can be administered a conditioning therapy after the administration of the compositions to kill certain immune cells that are not transduced with the CAR encoded by the compositions. This can be done by including a selection marker that is encoded by the nucleic acid cargo of interest. In some embodiments, the conditioning therapy comprises administering an effective amount of cyclophosphamide to the subject. In some embodiments, the conditioning therapy comprises administering an effective amount of fludarabine to the subject. In some embodiments, the conditioning therapy comprises administering an effective amount of a combination of cyclophosphamide and fludarabine to the subject.

[0331] In some embodiments, a specific dosage regimen of the present disclosure includes a lymphodepletion step after the administration of the composition. In an exemplary embodiment, the lymphodepletion step includes administration of cyclophosphamide and/or fludarabine.

[0332] In some embodiments, the lymphodepletion step includes administration of cyclophosphamide at a dose of between about 200 mg/m²/day and about 2000 mg/m²/day (e.g., 200 mg/m²/day, 300 mg/m²/day, or 500 mg/m²/day). In an exemplary embodiment, the dose of cyclophosphamide is about 300 mg/m²/day. In some embodiments, the lymphodepletion step includes administration of fludarabine at a dose of between about 20 mg/m²/day and about 900 mg/m²/day (e.g., 20 mg/m²/day, 25 mg/m²/day, 30 mg/m²/day, or 60 mg/m²/day). In an exemplary embodiment, the dose of fludarabine is about 30 mg/m²/day.

[0333] In some embodiment, the lymphodepletion step includes administration of cyclophosphamide at a dose of between about 200 mg/m²/day and about 2000 mg/m²/day (e.g., 200 mg/m²/day, 300 mg/m²/day, or 500 mg/m²/day), and fludarabine at a dose of between about 20 mg/m²/day and about 900 mg/m²/day (e.g., 20 mg/m²/day, 25 mg/m²/day, 30 mg/m²/day, or 60 mg/m²/day). In an exemplary embodiment, the lymphodepletion step includes administration of cyclophosphamide at a dose of about 300 mg/m²/day, and fludarabine at a dose of about 30 mg/m²/day.

[0334] In an exemplary embodiment, the dosing of cyclophosphamide is 300 mg/m²/day over three days, and the dosing of fludarabine is 30 mg/m²/day over three days.

[0335] It is known in the art that one of the adverse effects of the use of CAR T cells can be the onset of immune activation, known as cytokine release syndrome (CRS). CRS is immune activation resulting in elevated inflammatory cytokines. CRS is a known on-target toxicity, development of which likely correlates with efficacy. Clinical and laboratory measures range from mild CRS (constitutional symptoms and/or grade-2 organ toxicity) to severe CRS (sCRS; grade ≥ 3 organ toxicity, aggressive clinical intervention, and/or potentially life threatening). Clinical features include: high fever, malaise, fatigue, myalgia, nausea, anorexia, tachycardia/hypotension, capillary leak, cardiac dysfunction, renal impairment, hepatic failure, and disseminated intravascular coagulation. Dramatic elevations of cytokines including interferon-gamma, granulocyte macrophage colony-stimulating factor, IL-10, and IL-6 have been shown following CAR T-cell infusion. One CRS signature is elevation of cytokines including IL-6 (severe elevation), IFN-gamma, TNF-alpha (moderate), and IL-2 (mild). Elevations in clinically available markers of inflammation including ferritin and C-reactive protein (CRP) have also been observed to correlate with the CRS syndrome. The presence of CRS generally correlates with expansion and progressive immune activation of adoptively transferred cells. It has been demonstrated that the degree of CRS severity is dictated by disease burden at the time of infusion as patients with high tumor burden experience a more sCRS.

[0336] Accordingly, in some embodiments, the methods comprise, following the diagnosis of CRS, appropriate CRS management strategies to mitigate the physiological symptoms of uncontrolled inflammation without dampening the antitumor efficacy of the in vivo generated cells (e.g., CAR T cells). CRS management strategies are known in the art. For example, systemic corticosteroids may be administered to rapidly reverse symptoms of sCRS (e.g., grade 3 CRS) without compromising initial antitumor response.

[0337] In some embodiments, an anti-IL-6R antibody may be administered. An example of an anti-IL-6R antibody is the Food and Drug Administration-approved monoclonal antibody tocilizumab, also known as atilizumab (marketed as Actemra, or RoActemra). Tocilizumab is a humanized monoclonal antibody against the interleukin-6 receptor (IL-6R). Administration of tocilizumab has demonstrated near-immediate reversal of CRS.

[0338] CRS is generally managed based on the severity of the observed syndrome and interventions are tailored as such. CRS management decisions may be based upon clinical signs and symptoms and response to interventions, not solely on laboratory values alone.

[0339] Mild to moderate cases generally are treated with symptom management with fluid therapy, non-steroidal anti-inflammatory drug (NSAID) and antihistamines as needed for adequate symptom relief. More severe cases include patients with any degree of hemodynamic instability; with any hemodynamic instability, the administration of tocilizumab is recommended. The first-line management of CRS may be tocilizumab, in some embodiments, at the labeled dose of 8 mg/kg IV over 60 minutes (not to exceed 800 mg/dose); tocilizumab can be repeated Q8 hours. If suboptimal response to the first dose of tocilizumab, additional doses of tocilizumab may be considered. Tocilizumab can be administered alone or in combination with corticosteroid therapy. Patients with continued or progressive CRS symptoms, inadequate clinical improvement in 12-18 hours or poor response to tocilizumab, may be treated with high-dose corticosteroid therapy, generally hydrocortisone 100 mg IV or methylprednisolone 1-2 mg/kg. In patients with more severe hemodynamic instability or more severe respiratory symptoms, patients may be administered high-dose corticosteroid therapy early in the course of the CRS. CRS management guidance may be based on published standards (Lee et al. (2019) *Biol Blood Marrow Transplant*, doi.org/10.1016/j.bbmt.2018.12.758; Neelapu et al. (2018) *Nat Rev Clin Oncology*, 15:47; Teachey et al. (2016) *Cancer Discov*, 6(6):664-679).

[0340] Features consistent with Macrophage Activation Syndrome (MAS) or Hemophagocytic lymphohistiocytosis (HLH) have been observed in patients treated with CAR-T therapy (Henter, 2007), coincident with clinical manifestations of the CRS. MAS appears to be a reaction to immune activation that occurs from the CRS, and should therefore be considered a manifestation of CRS. MAS is similar to HLH (also a reaction to immune stimulation). The clinical syndrome of MAS is characterized by high grade non-remitting fever, cytopenias affecting at least two of three lineages, and hepatosplenomegaly. It is associated with high serum ferritin, soluble interleukin-2 receptor, and triglycerides, and a decrease of circulating natural killer (NK) activity.

[0341] In some embodiments, methods of treating cancer in a subject in need thereof are provided, the methods comprising administering to the subject any of the compositions, such as the viral particle(s), provided herein.

[0342] The compositions disclosed herein can comprise a pharmaceutical composition, and for example include a pharmaceutically acceptable carrier, and/or a pharmaceutical formulation.

[0343] The term “pharmaceutical formulation” refers to a preparation which is in such

form as to permit the biological activity of an active ingredient contained therein to be effective, and which contains no additional components which are unacceptably toxic to a subject to which the formulation would be administered. A “pharmaceutically acceptable carrier” refers to an ingredient in a pharmaceutical formulation, other than an active ingredient, which is nontoxic to a subject. A pharmaceutically acceptable carrier includes, but is not limited to, a buffer, excipient, stabilizer, or preservative. In some aspects, the choice of carrier is determined in part by the particular cell and/or by the method of administration. Accordingly, there are a variety of suitable formulations. For example, the pharmaceutical composition can contain preservatives. Suitable preservatives may include, for example, methylparaben, propylparaben, sodium benzoate, and benzalkonium chloride. In some aspects, a mixture of two or more preservatives is used. The preservative or mixtures thereof are typically present in an amount of about 0.0001% to about 2% by weight of the total composition. Carriers are described, e.g., by Remington's Pharmaceutical Sciences 16th edition, Osol, A. Ed. (1980). Pharmaceutically acceptable carriers are generally nontoxic to recipients at the dosages and concentrations employed, and include, but are not limited to: buffers such as phosphate, citrate, and other organic acids; antioxidants including ascorbic acid and methionine; preservatives (such as octadecyldimethylbenzyl ammonium chloride; hexamethonium chloride; benzalkonium chloride; benzethonium chloride; phenol, butyl or benzyl alcohol; alkyl parabens such as methyl or propyl paraben; catechol; resorcinol; cyclohexanol; 3-pentanol; and m-cresol); low molecular weight (less than about 10 residues) polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids such as glycine, glutamine, asparagine, histidine, arginine, or lysine; monosaccharides, disaccharides, and other carbohydrates including glucose, mannose, or dextrans; chelating agents such as EDTA; sugars such as sucrose, mannitol, trehalose or sorbitol; salt-forming counter-ions such as sodium; metal complexes (e.g. Zn-protein complexes); and/or non-ionic surfactants such as polyethylene glycol (PEG).

[0344] Buffering agents in some aspects are included in the compositions. Suitable buffering agents include, for example, citric acid, sodium citrate, phosphoric acid, potassium phosphate, and various other acids and salts. In some aspects, a mixture of two or more buffering agents is used. The buffering agent or mixtures thereof are typically present in an amount of about 0.001% to about 4% by weight of the total composition. Methods for preparing administrable pharmaceutical compositions are known. Exemplary methods are described in more detail in, for

example, Remington: The Science and Practice of Pharmacy, Lippincott Williams & Wilkins; 21st ed. (May 1, 2005).

[0345] The formulations can include aqueous solutions. The formulation or composition may also contain more than one active ingredient useful for the particular indication, disease, or condition being treated with the composition, preferably those with activities complementary to the composition, where the respective activities do not adversely affect one another. Such active ingredients are suitably present in combination in amounts that are effective for the purpose intended. Thus, in some embodiments, the pharmaceutical composition further includes other pharmaceutically active agents or drugs, such as chemotherapeutic agents, e.g., asparaginase, busulfan, carboplatin, cisplatin, daunorubicin, doxorubicin, fluorouracil, gemcitabine, hydroxyurea, methotrexate, paclitaxel, rituximab, vinblastine, and/or vincristine. The pharmaceutical composition in some embodiments contains the composition in amounts effective to treat or prevent the disease or condition, such as a therapeutically effective or prophylactically effective amount. Therapeutic or prophylactic efficacy in some embodiments is monitored by periodic assessment of treated subjects. The desired dosage can be delivered by a single bolus administration of the composition, by multiple bolus administrations of the composition, or by continuous infusion administration of the composition. In some embodiments, the pharmaceutical composition does not include a chemotherapeutic.

[0346] Formulations include those for oral, intravenous, intraperitoneal, subcutaneous, pulmonary, transdermal, intramuscular, intranasal, buccal, sublingual, or suppository administration. In some embodiments, the composition is administered parenterally. The term “parenteral,” as used herein, includes intravenous, intramuscular, subcutaneous, rectal, vaginal, and intraperitoneal administration. In some embodiments, the composition is administered to the subject using peripheral systemic delivery by intravenous, intraperitoneal, or subcutaneous injection. Compositions in some embodiments are provided as sterile liquid preparations, e.g., isotonic aqueous solutions, suspensions, emulsions, dispersions, or viscous compositions, which may in some aspects be buffered to a selected pH. Liquid preparations are normally easier to prepare than gels, other viscous compositions, and solid compositions. Additionally, liquid compositions are somewhat more convenient to administer, especially by injection. Viscous compositions, on the other hand, can be formulated within the appropriate viscosity range to provide longer contact periods with specific tissues. Liquid or viscous compositions can comprise

carriers, which can be a solvent or dispersing medium containing, for example, water, saline, phosphate buffered saline, polyol (for example, glycerol, propylene glycol, liquid polyethylene glycol) and suitable mixtures thereof.

[0347] Sterile injectable solutions can be prepared by incorporating the composition in a solvent, such as in admixture with a suitable carrier, diluent, or excipient such as sterile water, physiological saline, glucose, dextrose, or the like. The compositions can contain auxiliary substances such as wetting, dispersing, or emulsifying agents (e.g., methylcellulose), pH buffering agents, gelling or viscosity enhancing additives, preservatives, flavoring agents, and/or colors, depending upon the route of administration and the preparation desired. Standard texts may in some aspects be consulted to prepare suitable preparations.

[0348] Various additives which enhance the stability and sterility of the compositions, including antimicrobial preservatives, antioxidants, chelating agents, and buffers, can be added. Prevention of the action of microorganisms can be ensured by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, and sorbic acid. Prolonged absorption of the injectable pharmaceutical form can be brought about by the use of agents delaying absorption, for example, aluminum monostearate and gelatin.

[0349] The formulations to be used for in vivo administration are generally sterile. Sterility may be readily accomplished, e.g., by filtration through sterile filtration membranes.

[0350] The embodiments provided for herein can be used for many purposes, since the a pseudotyped virus capable of fusing with a target cell can be used to deliver a gene or other heterologous sequence of interest.

ENUMERATED EMBODIMENTS

[0351] In some embodiments, the following embodiments are provided:

1. A VSV-G polypeptide comprising a mutation that corresponds to a mutation at position 182 of SEQ ID NO: 2.
2. The VSV-G polypeptide of embodiment 1, wherein the protein comprises an amino acid sequence of SEQ ID NO: 2 with a mutation at position 182 and has at least 70% identity to SEQ ID NO: 2.

3. The VSV-G polypeptide of embodiments 1 or 2, wherein the polypeptide comprises a I182E or I182D mutation as compared to SEQ ID NO: 2.
4. The VSV-G polypeptide of any one of embodiments 1-3, wherein the polypeptide comprises an amino acid sequence of SEQ ID NO: 1 with a mutation at position 198 and at least 70% identity to SEQ ID NO: 2.
5. The VSV-G polypeptide of any one of embodiments 1-4, wherein the polypeptide comprises a mutation that corresponds to I182D or I182E as compared to a sequence of SEQ ID NO: 2.
6. The VSV-G polypeptide of any one of embodiments 1-5, wherein the polypeptide comprises an amino acid sequence at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical to an amino acid sequence of SEQ ID NO: 4.
7. The VSV-G polypeptide of any one of embodiments 1-5, wherein the polypeptide comprises a sequence of SEQ ID NO: 4.
8. The VSV-polypeptide of any one of embodiments 1-5, wherein the polypeptide comprises an amino acid sequence at least 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical to an amino acid sequence of SEQ ID NO: 5.
9. The VSV-G polypeptide of any one of embodiments 1-5, wherein the polypeptide comprises a sequence of SEQ ID NO: 5.
10. The VSV-G polypeptide of any one of embodiments 1-9, further comprising a mutation in the VSV-G protein that corresponds to a mutation as described in US 2020/0216502, which is hereby incorporated by reference in its entirety.

11. The VSV-G polypeptide of any one of embodiments 1-10, further comprising a mutation in the VSV-G protein that corresponds to a position of 8, 10, 47, 209 and/or 354 as compared to SEQ ID NO: 2.

12. The VSV-G polypeptide of any one of embodiments 1-11, further comprising a mutation that corresponds to position 8 in SEQ ID NO: 2, wherein the mutation is any amino acid different from the amino acid indicated at that position in SEQ ID NO: 2, except Y.

13. The VSV-G polypeptide of any one of embodiments 1-12, further comprising a mutation that corresponds to position 209 in SEQ ID NO: 2, wherein the mutation is any amino acid different from the amino acid indicated at that position in SEQ ID NO: 2, except H.

14. The VSV-G polypeptide of any one of embodiments 1-13, further comprising a mutation that corresponds to position 47 in SEQ ID NO: 2, wherein the mutation is any amino acid different from the amino acid indicated at that position in SEQ ID NO: 2, except K or R.

15. The VSV-G polypeptide of any one of embodiments 1-14, further comprising a mutation that corresponds to position 354 in SEQ ID NO: 2, wherein the mutation is any amino acid different from the amino acid indicated at that position in SEQ ID NO: 2, except K or R.

16. The VSV-G polypeptide of any one of embodiments 1-15, further comprising a mutation that corresponds to position 10 in SEQ ID NO: 2, wherein the mutation is any amino acid different from the amino acid indicated at that position in SEQ ID NO: 2, except Q or N.

17. The VSV-G polypeptide of any one of embodiments 1-16, wherein the protein comprises a substitution at position 47 or at position 354, or at both positions 47 and 354, wherein each position is, independently, substituted by A, G, F, Q, or N.

18. The VSV-G polypeptide of any one of embodiments 1-17, wherein the protein comprises a substitution at position 8, wherein the substitution is H8A, H8I, H8V, H8L, and the like.

19. The VSV-G polypeptide of any one of embodiments 1-18, wherein the protein comprises a substitution at position 47, wherein the substitution is K47Q or K47N.
20. The VSV-G polypeptide of any one of embodiments 1-19, wherein the protein comprises a substitution H8A and/or K47Q mutation.
21. The VSV-G polypeptide of any one of embodiments 1-20, wherein the protein comprises a substitution at position 10, such as Q10A, Q10R, or Q10K substitution.
22. The VSV-G polypeptide of any one of embodiments 1-21, further comprising a mutation that corresponds to a mutation at positions 214 and/or 352 of SEQ ID NO: 2.
23. The VSV-G polypeptide of embodiment 22, wherein the polypeptide comprises a T214N and/or T352A mutation.
24. A VSV-G polypeptide comprising a substitution at positions I182 and at least one of T214, and T352 of SEQ ID NO: 2.
25. The VSV-G polypeptide of embodiment 24, wherein the polypeptide comprises substitutions at positions I182, T214, and T352 of SEQ ID NO: 2.
26. The VSV-G polypeptide of embodiments 24 or 25, wherein the substitution at position 182 is I182D or I182E, the substitution at position 214 is T214N, and the substitution at position 352 is T352A.
27. The VSV-G polypeptide of any one of embodiments 24-26, wherein the polypeptide comprises a sequence of SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24 or SEQ ID NO: 25.
28. The VSV-G polypeptide of any one of embodiments 24-26, wherein the polypeptide comprises a sequence of SEQ ID NO: 22.

29. The VSV-G polypeptide of any one of embodiments 24-26, wherein the polypeptide comprises a sequence of SEQ ID NO: 23.
30. The VSV-G polypeptide of any one of embodiments 24-26, wherein the polypeptide comprises a sequence of SEQ ID NO: 24.
31. The VSV-G polypeptide of any one of embodiments 24-26, wherein the polypeptide comprises a sequence of SEQ ID NO: 25.
32. A nucleic acid molecule encoding the VSV-G polypeptide of any one of embodiments 1-31.
33. A vector comprising the nucleic acid molecule of embodiment 32.
34. A plasmid comprising the nucleic acid molecule of embodiment 32.
35. A viral particle comprising the polypeptide of any one of embodiments 1-31.
36. The viral particle of embodiment 35, further comprising a targeting moiety.
37. The viral particle of embodiments 35 or 36, wherein the viral particle is a pseudotyped lentivirus.
38. The viral particle of any one of embodiments 35-37, wherein the viral particle further comprises a nucleic acid molecule encoding a heterologous molecule of interest.
39. The viral particle of embodiment 38, wherein the heterologous molecule of interest is an siRNA, an shRNA, a non-coding RNA (e.g. a guide RNA for a CRISPR system), a peptide, a polypeptide, a protein, a viral payload, a viral genome, or a combination thereof.

40. The viral particle of embodiments 38 or 39, wherein the heterologous molecule of interest is a chimeric antigen receptor (“CAR”).

41. The viral particle of any one of embodiments 35-40, wherein the targeting moiety binds to an immune cell, such as a T cell, B cell; NK cell, dendritic cell, neutrophils, macrophages, a cancer cell; or, for example, CD3+ T cell; CD4+ T cell; CD7+ T cell, CD8+ T cell; CD19+ B cell; CD19+ cancer cell; CD20+ B cell; “CD20+ cancer cell, CD30+ lung epithelial cell; CD34+ haematopoietic stem cell; CD105+ endothelial cell; CD105+ haematopoietic stem cell; CD117+ haematopoietic stem cell; CD133+ cancer cell; EpCAM+ cancer cell; GluA2+ neuron; GluA4+ neuron; Haematopoietic stem cell; Hepatocyte; Her2/Neu+ cancer cell; NKG2D+ natural killer cell; SLC1A3+ astrocyte; SLC7A10+ adipocyte.

42. The viral particle of any one of embodiments 35-41, wherein the targeting moiety binds to CD7, CD8, cKit (CD117), CD4, CD3, CD5, CD6, CD2, TCR alpha, TCR beta, TCR gamma, TCR delta, CD10, CD34, CD110, CD33, CD14, CD68, CCR7, CD62L, CD25, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, or CXCR3, A glycosylated CD43 epitope expressed on acute leukemia or lymphoma but not on hematopoietic progenitors; A glycosylated CD43 epitope expressed on non-hematopoietic cancers; A kinase anchor protein 4 (AKAP-4); Adrenoceptor beta 3 (ADRB3); AFP; Anaplastic lymphoma kinase (ALK); Androgen receptor; Angiopoietin-binding cell surface receptor 2 (Tie 2); Auto antibody to desmoglein 1 (Dsg1); Auto antibody to desmoglein 3 (Dsg3); B7H3 (CD276); Biotin; Bone marrow stromal cell antigen 2 (BST2); BST1/CD157; Cancer/testis antigen 1 (NY-ESO-1); Cancer/testis antigen 2 (LAGE-la); Carbonic anhydrase IX (CA1X); Carcinoembryonic antigen (CEA); CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS or Brother of the Regulator of Imprinted Sites); CCR4; CD5; CD19; CD20; CD22; CD24; CD30; CD32 (FCGR2A); CD33; CD34; CD38; CD44v6; CD72; CD79a; CD79b; CD97; CD99; CD123; CD171; CD179a; CD179b-IGL1; CD200R; CD276/B7H3; CD300 molecule-like family member f (CD300LF); CDH1-CD324; CDH6; CDH17; CDH19; Chromosome X open reading frame 61 (CXORF61); Claudin 6 (CLDN6); Claudin18.2 (CLD18A2 or CLDN18A.2); CMV pp65; C-MYC epitope Tag; Cripto; CS1 (also referred to as CD2 subset 1 or CRACC or SLAMF7 or CD319 or 19A24); CSF2RA (GM-CSFR-alpha); C-type lectin domain family 12 member A (CLEC12A); C-type lectin-like molecule-1 (CLL-1 or CLECL1); Cyclin B1; Cytochrome P450 IB 1 (CYP1B 1);

DLL3; EBV-EBNA3c; EGF-bke module- containing mucin-like hormone receptor-like 2 (EMR2); Elongation factor 2 mutated (ELF2M); Ephrin B2; Ephrin type-A receptor 2 (EphA2); Epidermal growth factor receptor (EGFR); Epidermal growth factor receptor variant III (EGFRviii); Epithelial cell adhesion molecule (EPCAM); ERG; ETS translocation-variant gene 6 located on chromosome 12p (ETV6-AML); Fc fragment of IgA receptor (FCAR or CD89); Fc receptor-like 5 (FCRL5); Fibroblast activation protein alpha (FAP); FITC; Fms Like Tyrosine Kinase 3 (FLT3); Folate receptor alpha (FRa or FR1); Folate receptor beta (FRb); Follicle stimulating hormone receptor (FSHR); Fos-related antigen 1; Fucosyl-GM1; G protein coupled receptor class C group 5 member D (GPRC5D); G protein-coupled receptor 20 (GPR20); GAD; Ganglioside G2 (GD2) ; Ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlc(1-1)Cer); Ganglioside GM3 (aNeu5Ac(2-3)bDClalp(1- 4)bDGlc(1-1)Cer); GD3; GFRalpha4; Glycoprotein 100 (gp100); Glypican-3 (GPC3); Gonadotropin Hormone receptor (CGHR or GR); GpA33; GpNMB; GPRC5D; Guanylyl cyclase C (GCC); Heat shock protein 70-2 mutated (mut hsp70-2); Hepatitis A virus cellular receptor 1 (HAVCR1); Hexasaccharide portion of globoH glycosphingolipid (GloboH); High molecular weight-melanoma associated antigen (HMWMAA); HIV1 envelope glycoprotein; HLA; HLA-DOA; HLA-A; HLA-A2; HLA-B; HLA-C; HLA-DM; HLA-DOB; HLA-DP; HLA-DQ; HLA-DR; HLA-G; HTLV1-Tax; Human papilloma virus E6 (HPV E6); Human papilloma virus E7 (HPV E7); Human Telomerase reverse transcriptase (hTERT); IgE; IL13Ra2; IL1 lRa; Immunoglobulin lambda-like polypeptide 1 (IGLL1); Influenza A hemagglutinin (HA); Insulin-like growth factor 1 receptor (IGF-I receptor); Interleukin 11 receptor alpha (IL-11Ra); Interleukin- 13 receptor subunit alpha-2 (IL-13Ra2 or CD213A2); Intestinal carboxyl esterase; KIT (CD117); KSHV K8.1; KSHV-gH; LAMP1 ; Legumain; Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2); Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1); Leutenizing hormone receptor (LHR); Lewis(Y) antigen; Lews Ag; Liv1; Locus K 9 (LY6K); Low conductance chloride channel; Lymphocyte antigen 6 complex; Lymphocyte antigen 75 (LY75); Lymphocyte-specific protein tyrosine kinase (LCK); Mammary gland differentiation antigen (NY-BR-1); Melanoma antigen recognized by T cells 1 (MelanA or MART1); Melanoma- associated antigen 1 (MAGE-A1); Melanoma cancer testis antigen-1 (MAD-CT-1); Melanoma cancer testis antigen-2 (MAD-CT-2); Melanoma inhibitor of apoptosis (ML-IAP); Mesothelin; MPL; Mucin 1 cell surface associated (MUC1); N-Acetyl glucosaminyl-transferase V (NA17); Nectin-4; Neural cell adhesion molecule

(NCAM); NKG2D; NYBR1; O-acetyl-GD2 ganglioside (OAcGD2); Olfactory receptor 51E2 (OR51E2); Oncogene fusion protein consisting of breakpoint cluster region (BCR) and Abelson murine leukemia viral oncogene homolog 1 (Abl) (bcr-abl); P53 mutant; Paired box protein Pax-3 (PAX3); Paired box protein Pax-5 (PAX5); Pannexin 3 (PANX3); PDL1; P-glycoprotein; Placenta-specific 1 (PLAC1); Platelet-derived growth factor receptor beta (PDGFR-beta); Polysialic acid; Proacrosin binding protein sp32 (OY-TES1); Prostase; Prostate carcinoma tumor antigen-1 (PCT A-1 or Galectin 8); Prostate stem cell antigen (PSCA); Prostate-specific membrane antigen (PSMA); Prostatic acid phosphatase (PAP); Prostein; Protease Serine 21 (Testisin or PRSS21); Proteasome (Prosome Macropain) Subunit Beta Type 9 (LMP2); PTK7; Ras G12V; Ras Homolog Family Member C (RhoC); Rat sarcoma (Ras) mutant; Receptor for Advanced Glycation Endproducts (RAGE-1); Receptor tyrosine kinase-like orphan receptor 1 (ROR1); Receptor tyrosine-protein kinase ERBB2 or Her-22/neu; Renal ubiquitous 1 (RU1); Renal ubiquitous 2 (RU2); Sarcoma translocation breakpoints; Serine 2 (TMPRSS2) ETS fusion gene; Sialyl Lewis adhesion molecule (sLe); SLAMF4; SLAMF6; Slea (CA19.9 or Sialyl Lewis Antigen); Sperm protein 17 (SPA17); Squamous Cell Carcinoma Antigen Recognized By T Cells 3 (SART3); Stage-specific embryonic antigen-4 (SSEA-4); STEAP1; Survivin; Synovial sarcoma X breakpoint 2 (SSX2); TCR Gamma Alternate Reading Frame Protein (TARP); TCR-beta1 chain; TCR-beta2 chain; TCR-delta chain; TCR-gamma chain; TCRgamma-delta; Telomerase; TGFbetaR2; The antigen recognized by TNT antibody; Thyroid stimulating hormone receptor (TSHR); Timl-/HVCR1; Tissue Factor 1 (TF1); Tn ag; Tn antigen ((Tn Ag) or (GalNAc-Ser/Thr)); TNF receptor family member B cell maturation (BCMA); Transglutaminase 5 (TGS5); Transmembrane protease; TROP2; Tumor endothelial marker 1 (TEM1/CD248); Tumor endothelial marker 7-related (TEM7R); Tumor protein p53 (p53); Tumor-associated glycoprotein 72 (TAG72); Tyrosinase; Tyrosinase-related protein 2 (TRP-2); Uroplakin 2 (UPK2); Vascular endothelial growth factor receptor 2 (VEGFR2); V-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN); Wilms tumor protein (WT1); or X Antigen Family Member 1A (XAGE1).

43. The viral particle of any one of embodiments 35-42, wherein the targeting moiety is an antibody, a scFv antibody, an antigen binding domain, an ankyrin repeat (e.g., DARPIN), a VHH domain antibody, a nanobody, single domain antibody, a FN3 domain, or any combination thereof.

44. The viral particle of any one of embodiments 35-43, wherein the targeting moiety is attached to the viral surface through

an IgG Fc stalk;

an envelope glycoprotein G or H of a virus of the Paramyxoviridae family, such as a morbillivirus, such as Measles virus, or a henipavirus, such as Nipah virus, Cedar virus, or Hendra virus;

a glycoprotein of a virus of the Rhabdoviridae family, such as a vesicular stomatitis New Jersey virus, a vesicular stomatitis Indiana virus, a vesicular stomatitis Alagoas virus, a vesicular stomatitis Maraba virus, a vesicular stomatitis Carajas virus, Parainfluenza virus, Spodoptera frugiperda rhabdovirus isolate Sf G, Drosophila obscura sigmavirus 10A, Wuhan insect virus 7, Perch virus, or Spring viremia of carp virus;

a glycoprotein of a virus of the Filoviridae family, such as Ebola virus; or

a glycoprotein of a virus of the Arenaviridae family, such as Machupo virus.

45. The viral particle of embodiment 44, wherein the stalk comprises a transmembrane domain, such as, but not limited to a CD8 or CD28 transmembrane domain.

46. A method of delivering a heterologous molecule to a target cell, the method comprising contacting the cell with a viral vector comprising:

a) the VSV-G protein of any one of embodiments 1-31;

b) a targeting moiety that binds to the target cell; and

c) a nucleic acid molecule encoding the heterologous molecule.

47. The method of embodiment 46, wherein the target cell is an immune cell, such as a T cell, B cell; NK cell, dendritic cell, neutrophils, macrophages, a cancer cell; or, for example, CD3+ T cell; CD4+ T cell; CD7+ T cell, CD8+ T cell; CD19+ B cell; CD19+ cancer cell; CD20+ B cell; CD30+ lung epithelial cell; CD34+ haematopoietic stem cell; CD105+ endothelial cell; CD105+ haematopoietic stem cell; CD117+ haematopoietic stem cell; CD133+ cancer cell; EpCAM+ cancer cell; GluA2+ neuron; GluA4+ neuron; Haematopoietic stem cell; Hepatocyte; Her2/Neu+

cancer cell; NKG2D+ natural killer cell; SLC1A3+ astrocyte; SLC7A10+ adipocyte.

48. The method of embodiments 46 or 47, wherein the targeting moiety binds to CD7, CD8, cKit (CD117), CD4, CD3, CD5, CD6, CD2, TCR alpha, TCR beta, TCR gamma, TCR delta, CD10, CD34, CD110, CD33, CD14, CD68, CCR7, CD62L, CD25, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, or CXCR3, A glycosylated CD43 epitope expressed on acute leukemia or lymphoma but not on hematopoietic progenitors; A glycosylated CD43 epitope expressed on non-hematopoietic cancers; A kinase anchor protein 4 (AKAP-4); Adrenoceptor beta 3 (ADRB3); AFP; Anaplastic lymphoma kinase (ALK); Androgen receptor; Angiopoietin-binding cell surface receptor 2 (Tie 2); Auto antibody to desmoglein 1 (Dsg1); Auto antibody to desmoglein 3 (Dsg3); B7H3 (CD276); Biotin; Bone marrow stromal cell antigen 2 (BST2); BST1/CD157; Cancer/testis antigen 1 (NY-ESO-1); Cancer/testis antigen 2 (LAGE-la); Carbonic anhydrase IX (CA1X); Carcinoembryonic antigen (CEA); CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS or Brother of the Regulator of Imprinted Sites); CCR4; CD5; CD19; CD20; CD22; CD24; CD30; CD32 (FCGR2A); CD33; CD34; CD38; CD44v6; CD72; CD79a; CD79b; CD97; CD99; CD123; CD171; CD179a; CD179b-IGL1; CD200R; CD276/B7H3; CD300 molecule-like family member f (CD300LF); CDH1-CD324; CDH6; CDH17; CDH19; Chromosome X open reading frame 61 (CXORF61); Claudin 6 (CLDN6); Claudin18.2 (CLDN18A2 or CLDN18A.2); CMV pp65; C-MYC epitope Tag; Cripto; CS1 (also referred to as CD2 subset 1 or CRACC or SLAMF7 or CD319 or 19A24); CSF2RA (GM-CSFR-alpha); C-type lectin domain family 12 member A (CLEC12A); C-type lectin-like molecule-1 (CLL-1 or CLECL1); Cyclin B1; Cytochrome P450 1B1 (CYP1B1); DLL3; EBV-EBNA3c; EGF-like module-containing mucin-like hormone receptor-like 2 (EMR2); Elongation factor 2 mutated (ELF2M); Ephrin B2; Ephrin type-A receptor 2 (EphA2); Epidermal growth factor receptor (EGFR); Epidermal growth factor receptor variant III (EGFRviii); Epithelial cell adhesion molecule (EPCAM); ERG; ETS translocation-variant gene 6 located on chromosome 12p (ETV6-AML); Fc fragment of IgA receptor (FCAR or CD89); Fc receptor-like 5 (FCRL5); Fibroblast activation protein alpha (FAP); FITC; Fms Like Tyrosine Kinase 3 (FLT3); Folate receptor alpha (FRa or FR1); Folate receptor beta (FRb); Follicle stimulating hormone receptor (FSHR); Fos-related antigen 1; Fucosyl-GM1; G protein coupled receptor class C group 5 member D (GPRC5D); G protein-coupled receptor 20 (GPR20); GAD; Ganglioside G2 (GD2); Ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlcp(1-

l)Cer); Ganglioside GM3 (aNeu5Ac(2-3)bDCIalp(1- 4)bDGlep(1-1)Cer); GD3; GFRalpha4;
 Glycoprotein 100 (gp100); Glypican-3 (GPC3); Gonadotropin Hormone receptor (CGHR or GR);
 GpA33; GpNMB; GPRC5D; Guanylyl cyclase C (GCC); Heat shock protein 70-2 mutated (mut
 hsp70-2); Hepatitis A virus cellular receptor 1 (HAVCR1); Hexasaccharide portion of globoH
 glycosphingolipid (GloboH); High molecular weight-melanoma associated antigen (HMWMAA);
 HIV1 envelope glycoprotein; HLA; HLA-DOA; HLA-A; HLA-A2; HLA-B; HLA-C; HLA-DM;
 HLA-DOB; HLA-DP; HLA-DQ; HLA-DR; HLA-G; HTLV1-Tax; Human papilloma virus E6
 (HPV E6); Human papilloma virus E7 (HPV E7); Human Telomerase reverse transcriptase
 (hTERT); IgE; IL13Ra2; IL13Ra; Immunoglobulin lambda-like polypeptide 1 (IGLL1); Influenza
 A hemagglutinin (HA); Insulin-like growth factor 1 receptor (IGF-I receptor); Interleukin 11
 receptor alpha (IL-11Ra); Interleukin- 13 receptor subunit alpha-2 (IL-13Ra2 or CD213A2);
 Intestinal carboxyl esterase; KIT (CD117); KSHV K8.1; KSHV-gH; LAMP1 ; Legumain;
 Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2); Leukocyte-
 associated immunoglobulin-like receptor 1 (LAIR1); Leutenizing hormone receptor (LHR);
 Lewis(Y) antigen; Lews Ag; Liv1; Locus K 9 (LY6K); Low conductance chloride channel;
 Lymphocyte antigen 6 complex; Lymphocyte antigen 75 (LY75); Lymphocyte-specific protein
 tyrosine kinase (LCK); Mammary gland differentiation antigen (NY-BR-1); Melanoma antigen
 recognized by T cells 1 (MelanA or MART1); Melanoma- associated antigen 1 (MAGE-A1);
 Melanoma cancer testis antigen-1 (MAD-CT-1); Melanoma cancer testis antigen-2 (MAD-CT-2);
 Melanoma inhibitor of apoptosis (ML-IAP); Mesothelin; MPL; Mucin 1 cell surface associated
 (MUC1); N-Acetyl glucosaminyl-transferase V (NA17); Nectin-4; Neural cell adhesion molecule
 (NCAM); NKG2D; NYBR1; O-acetyl-GD2 ganglioside (OAcGD2); Olfactory receptor 51E2
 (OR51E2); Oncogene fusion protein consisting of breakpoint cluster region (BCR) and Abelson
 murine leukemia viral oncogene homolog 1 (Abl) (bcr-abl); P53 mutant; Paired box protein Pax-
 3 (PAX3); Paired box protein Pax-5 (PAX5); Pannexin 3 (PANX3); PDL1; P-glycoprotein;
 Placenta-specific 1 (PLAC1); Platelet-derived growth factor receptor beta (PDGFR-beta);
 Polysialic acid; Proacrosin binding protein sp32 (OY-TES1); Prostase; Prostate carcinoma tumor
 antigen-1 (PCT A-1 or Galectin 8); Prostate stem cell antigen (PSCA); Prostate-specific membrane
 antigen (PSMA); Prostatic acid phosphatase (PAP); Prostein; Protease Serine 21 (Testisin or
 PRSS21); Proteasome (Prosome Macropain) Subunit Beta Type 9 (LMP2); PTK7; Ras G12V; Ras
 Homolog Family Member C (RhoC); Rat sarcoma (Ras) mutant; Receptor for Advanced Gly

cation Endproducts (RAGE-1); Receptor tyrosine kinase-like orphan receptor 1 (ROR1); Receptor tyrosine-protein kinase ERBB2 or Her-22/neu; Renal ubiquitous 1 (RU1); Renal ubiquitous 2 (RU2); Sarcoma translocation breakpoints; Serine 2 (TMPRSS2) ETS fusion gene; Sialyl Lewis adhesion molecule (sLe); SLAMF4; SLAMF6; Slea (CA19.9 or Sialyl Lewis Antigen); Sperm protein 17 (SPA17); Squamous Cell Carcinoma Antigen Recognized By T Cells 3 (SART3); Stage-specific embryonic antigen-4 (SSEA-4); STEAP1; Survivin; Synovial sarcoma X breakpoint 2 (SSX2); TCR Gamma Alternate Reading Frame Protein (TARP); TCR-beta1 chain; TCR-beta2 chain; TCR-delta chain; TCR-gamma chain; TCRgamma-delta; Telomerase; TGFbetaR2; The antigen recognized by TNT antibody; Thyroid stimulating hormone receptor (TSHR); Tim1/HVCR1; Tissue Factor 1 (TF1); Tn ag; Tn antigen ((Tn Ag) or (GalNAc-Ser/Thr)); TNF receptor family member B cell maturation (BCMA); Transglutaminase 5 (TGS5); Transmembrane protease; TROP2; Tumor endothelial marker 1 (TEM1/CD248); Tumor endothelial marker 7-related (TEM7R); Tumor protein p53 (p53); Tumor-associated glycoprotein 72 (TAG72); Tyrosinase; Tyrosinase-related protein 2 (TRP-2); Uroplakin 2 (UPK2); Vascular endothelial growth factor receptor 2 (VEGFR2); V-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN); Wilms tumor protein (WT1); or X Antigen Family Member 1A (XAGE1) on the target cell.

49. A method of delivering a heterologous molecule to a target cell in a subject, the method comprising administering to the subject a viral vector comprising:

- a) the VSV-G polypeptide of any one of embodiments 1-31;
- b) a targeting moiety that binds to the target cell; and
- c) a nucleic acid molecule encoding the heterologous molecule.

50. The method of embodiment 49, wherein the target cell is an immune cell, such as a T cell, B cell; NK cell, dendritic cell, neutrophils, macrophages, a cancer cell; or, for example, CD3+ T cell; CD4+ T cell; CD7+ T cell, CD8+ T cell; CD19+ B cell; CD19+ cancer cell; CD20+ B cell; CD20+ cancer cell; CD30+ lung epithelial cell; CD34+ haematopoietic stem cell; CD105+ endothelial cell; CD105+ haematopoietic stem cell; CD117+ haematopoietic stem cell; CD133+ cancer cell; EpCAM+ cancer cell; GluA2+ neuron; GluA4+ neuron; Haematopoietic stem cell; Hepatocyte; Her2/Neu+ cancer cell; NKG2D+ natural killer cell; SLC1A3+ astrocyte; SLC7A10+

adipocyte.

51. The method of embodiments 49 or 50, wherein the targeting moiety binds to CD7, CD8, cKit (CD117), CD4, CD3, CD5, CD6, CD2, TCR alpha, TCR beta, TCR gamma, TCR delta, CD10, CD34, CD110, CD33, CD14, CD68, CCR7, CD62L, CD25, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, or CXCR3, A glycosylated CD43 epitope expressed on acute leukemia or lymphoma but not on hematopoietic progenitors; A glycosylated CD43 epitope expressed on non-hematopoietic cancers; A kinase anchor protein 4 (AKAP-4); Adrenoceptor beta 3 (ADRB3); AFP; Anaplastic lymphoma kinase (ALK); Androgen receptor; Angiopoietin-binding cell surface receptor 2 (Tie 2); Auto antibody to desmoglein 1 (Dsg1); Auto antibody to desmoglein 3 (Dsg3); B7H3 (CD276); Biotin; Bone marrow stromal cell antigen 2 (BST2); BST1/CD157; Cancer/testis antigen 1 (NY-ESO-1); Cancer/testis antigen 2 (LAGE-la); Carbonic anhydrase IX (CA1X); Carcinoembryonic antigen (CEA); CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS or Brother of the Regulator of Imprinted Sites); CCR4; CD5; CD19; CD20; CD22; CD24; CD30; CD32 (FCGR2A); CD33; CD34; CD38; CD44v6; CD72; CD79a; CD79b; CD97; CD99; CD123; CD171; CD179a; CD179b-IGL1; CD200R; CD276/B7H3; CD300 molecule-like family member f (CD300LF); CDH1-CD324; CDH6; CDH17; CDH19; Chromosome X open reading frame 61 (CXORF61); Claudin 6 (CLDN6); Claudin18.2 (CLDN18A2 or CLDN18A.2); CMV pp65; C-MYC epitope Tag; Cripto; CS1 (also referred to as CD2 subset 1 or CRACC or SLAMF7 or CD319 or 19A24); CSF2RA (GM-CSFR-alpha); C-type lectin domain family 12 member A (CLEC12A); C-type lectin-like molecule-1 (CLL-1 or CLECL1); Cyclin B1; Cytochrome P450 1B1 (CYP1B1); DLL3; EBV-EBNA3c; EGF-like module-containing mucin-like hormone receptor-like 2 (EMR2); Elongation factor 2 mutated (ELF2M); Ephrin B2; Ephrin type-A receptor 2 (EphA2); Epidermal growth factor receptor (EGFR); Epidermal growth factor receptor variant III (EGFRviii); Epithelial cell adhesion molecule (EPCAM); ERG; ETS translocation-variant gene 6 located on chromosome 12p (ETV6-AML); Fc fragment of IgA receptor (FCAR or CD89); Fc receptor-like 5 (FCRL5); Fibroblast activation protein alpha (FAP); FITC; Fms Like Tyrosine Kinase 3 (FLT3); Folate receptor alpha (FRa or FR1); Folate receptor beta (FRb); Follicle stimulating hormone receptor (FSHR); Fos-related antigen 1; Fucosyl-GM1; G protein coupled receptor class C group 5 member D (GPRC5D); G protein-coupled receptor 20 (GPR20); GAD; Ganglioside G2 (GD2); Ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlcp(1-

l)Cer); Ganglioside GM3 (aNeu5Ac(2-3)bDCIalp(1- 4)bDGlep(1-1)Cer); GD3; GFRalpha4;
 Glycoprotein 100 (gp100); Glypican-3 (GPC3); Gonadotropin Hormone receptor (CGHR or GR);
 GpA33; GpNMB; GPRC5D; Guanylyl cyclase C (GCC); Heat shock protein 70-2 mutated (mut
 hsp70-2); Hepatitis A virus cellular receptor 1 (HAVCR1); Hexasaccharide portion of globoH
 glycosphingolipid (GloboH); High molecular weight-melanoma associated antigen (HMWMAA);
 HIV1 envelope glycoprotein; HLA; HLA-DOA; HLA-A; HLA-A2; HLA-B; HLA-C; HLA-DM;
 HLA-DOB; HLA-DP; HLA-DQ; HLA-DR; HLA-G; HTLV1-Tax; Human papilloma virus E6
 (HPV E6); Human papilloma virus E7 (HPV E7); Human Telomerase reverse transcriptase
 (hTERT); IgE; IL13Ra2; IL13Ra; Immunoglobulin lambda-like polypeptide 1 (IGLL1); Influenza
 A hemagglutinin (HA); Insulin-like growth factor 1 receptor (IGF-I receptor); Interleukin 11
 receptor alpha (IL-11Ra); Interleukin- 13 receptor subunit alpha-2 (IL-13Ra2 or CD213A2);
 Intestinal carboxyl esterase; KIT (CD117); KSHV K8.1; KSHV-gH; LAMP1 ; Legumain;
 Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2); Leukocyte-
 associated immunoglobulin-like receptor 1 (LAIR1); Leutenizing hormone receptor (LHR);
 Lewis(Y) antigen; Lews Ag; Liv1; Locus K 9 (LY6K); Low conductance chloride channel;
 Lymphocyte antigen 6 complex; Lymphocyte antigen 75 (LY75); Lymphocyte-specific protein
 tyrosine kinase (LCK); Mammary gland differentiation antigen (NY-BR-1); Melanoma antigen
 recognized by T cells 1 (MelanA or MART1); Melanoma- associated antigen 1 (MAGE-A1);
 Melanoma cancer testis antigen-1 (MAD-CT-1); Melanoma cancer testis antigen-2 (MAD-CT-2);
 Melanoma inhibitor of apoptosis (ML-IAP); Mesothelin; MPL; Mucin 1 cell surface associated
 (MUC1); N-Acetyl glucosaminyl-transferase V (NA17); Nectin-4; Neural cell adhesion molecule
 (NCAM); NKG2D; NYBR1; O-acetyl-GD2 ganglioside (OAcGD2); Olfactory receptor 51E2
 (OR51E2); Oncogene fusion protein consisting of breakpoint cluster region (BCR) and Abelson
 murine leukemia viral oncogene homolog 1 (Abl) (bcr-abl); P53 mutant; Paired box protein Pax-
 3 (PAX3); Paired box protein Pax-5 (PAX5); Pannexin 3 (PANX3); PDL1; P-glycoprotein;
 Placenta-specific 1 (PLAC1); Platelet-derived growth factor receptor beta (PDGFR-beta);
 Polysialic acid; Proacrosin binding protein sp32 (OY-TES1); Prostase; Prostate carcinoma tumor
 antigen-1 (PCT A-1 or Galectin 8); Prostate stem cell antigen (PSCA); Prostate-specific membrane
 antigen (PSMA); Prostatic acid phosphatase (PAP); Prostein; Protease Serine 21 (Testisin or
 PRSS21); Proteasome (Prosome Macropain) Subunit Beta Type 9 (LMP2); PTK7; Ras G12V; Ras
 Homolog Family Member C (RhoC); Rat sarcoma (Ras) mutant; Receptor for Advanced Gly

cation Endproducts (RAGE-1); Receptor tyrosine kinase-like orphan receptor 1 (ROR1); Receptor tyrosine-protein kinase ERBB2 or Her-22/neu; Renal ubiquitous 1 (RU1); Renal ubiquitous 2 (RU2); Sarcoma translocation breakpoints; Serine 2 (TMPRSS2) ETS fusion gene; Sialyl Lewis adhesion molecule (sLe); SLAMF4; SLAMF6; Slea (CA19.9 or Sialyl Lewis Antigen); Sperm protein 17 (SPA17); Squamous Cell Carcinoma Antigen Recognized By T Cells 3 (SART3); Stage-specific embryonic antigen-4 (SSEA-4); STEAP1; Survivin; Synovial sarcoma X breakpoint 2 (SSX2); TCR Gamma Alternate Reading Frame Protein (TARP); TCR-beta1 chain; TCR-beta2 chain; TCR-delta chain; TCR-gamma chain; TCRgamma-delta; Telomerase; TGFbetaR2; The antigen recognized by TNT antibody; Thyroid stimulating hormone receptor (TSHR); Tim1/HVCR1; Tissue Factor 1 (TF1); Tn ag; Tn antigen ((Tn Ag) or (GalNAc-Ser/Thr)); TNF receptor family member B cell maturation (BCMA); Transglutaminase 5 (TGS5); Transmembrane protease; TROP2; Tumor endothelial marker 1 (TEM1/CD248); Tumor endothelial marker 7-related (TEM7R); Tumor protein p53 (p53); Tumor-associated glycoprotein 72 (TAG72); Tyrosinase; Tyrosinase-related protein 2 (TRP-2); Uroplakin 2 (UPK2); Vascular endothelial growth factor receptor 2 (VEGFR2); V-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN); Wilms tumor protein (WT1); or X Antigen Family Member 1A (XAGE1) on the target cell.

52. The method of any one of embodiments 46-51, wherein the heterologous molecule is an siRNA, an shRNA, a non-coding RNA (e.g. a guide RNA for a CRISPR system), a peptide, a polypeptide, a protein, a viral payload, a viral genome, or a combination thereof, such as a chimeric antigen receptor ("CAR").

53. A method of treating cancer in a subject, the method comprising administering to the subject a viral vector comprising:

- a) the VSV-G polypeptide of any one of embodiments 1-31;
- b) a targeting moiety that binds to the target cell; and
- c) a nucleic acid molecule encoding the heterologous molecule.

54. The method of embodiment 53, wherein the heterologous molecule is a chimeric antigen receptor.

55. The method of embodiments 53 or 54, wherein the target cell is an immune cell, such as a T cell, B cell; NK cell, dendritic cell, neutrophils, macrophages, a cancer cell; or, for example, CD3+ T cell; CD4+ T cell; CD7+ T cell, CD8+ T cell; CD19+ B cell; CD19+ cancer cell; CD20+ B cell; CD30+ lung epithelial cell; CD34+ haematopoietic stem cell; CD105+ endothelial cell; CD105+ haematopoietic stem cell; CD117+ haematopoietic stem cell; CD133+ cancer cell; EpCAM+ cancer cell; GluA2+ neuron; GluA4+ neuron; Haematopoietic stem cell; Hepatocyte; Her2/Neu+ cancer cell; NKG2D+ natural killer cell; SLC1A3+ astrocyte; SLC7A10+ adipocyte.

56. The method of any one of embodiments 53-55, wherein the targeting moiety binds to CD7, CD8, cKit (CD117), CD4, CD3, CD5, CD6, CD2, TCR alpha, TCR beta, TCR gamma, TCR delta, CD10, CD34, CD110, CD33, CD14, CD68, CCR7, CD62L, CD25, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, or CXCR3, A glycosylated CD43 epitope expressed on acute leukemia or lymphoma but not on hematopoietic progenitors; A glycosylated CD43 epitope expressed on non-hematopoietic cancers; A kinase anchor protein 4 (AKAP-4); Adrenoceptor beta 3 (ADRB3); AFP; Anaplastic lymphoma kinase (ALK); Androgen receptor; Angiopoietin-binding cell surface receptor 2 (Tie 2); Auto antibody to desmoglein 1 (Dsg1); Auto antibody to desmoglein 3 (Dsg3); B7H3 (CD276); Biotin; Bone marrow stromal cell antigen 2 (BST2); BST1/CD157; Cancer/testis antigen 1 (NY-ESO-1); Cancer/testis antigen 2 (LAGE-la); Carbonic anhydrase IX (CA1X); Carcinoembryonic antigen (CEA); CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS or Brother of the Regulator of Imprinted Sites); CCR4; CD5; CD19; CD20; CD22; CD24; CD30; CD32 (FCGR2A); CD33; CD34; CD38; CD44v6; CD72; CD79a; CD79b; CD97; CD99; CD123; CD171; CD179a; CD179b-IGLII; CD200R; CD276/B7H3; CD300 molecule-like family member f (CD300LF); CDH1-CD324; CDH6; CDH17; CDH19; Chromosome X open reading frame 61 (CXORF61); Claudin 6 (CLDN6); Claudin18.2 (CLD18A2 or CLDN18A.2); CMV pp65; C-MYC epitope Tag; Cripto; CS1 (also referred to as CD2 subset 1 or CRACC or SLAMF7 or CD319 or 19A24); CSF2RA (GM-CSFR-alpha); C-type lectin domain family 12 member A (CLEC12A); C-type lectin-like molecule-1 (CLL-1 or CLECL1); Cyclin B1; Cytochrome P450 IB 1 (CYP1B 1); DLL3; EBV-EBNA3c; EGF-bke module- containing mucin-like hormone receptor-like 2 (EMR2); Elongation factor 2 mutated (ELF2M); Ephrin B2; Ephrin type-A receptor 2 (EphA2); Epidermal growth factor receptor (EGFR); Epidermal growth factor receptor variant III

(EGFRviii); Epithelial cell adhesion molecule (EPCAM); ERG; ETS translocation-variant gene 6 located on chromosome 12p (ETV6-AML); Fc fragment of IgA receptor (FCAR or CD89); Fc receptor-like 5 (FCRL5); Fibroblast activation protein alpha (FAP); FITC; Fms Like Tyrosine Kinase 3 (FLT3); Folate receptor alpha (FRa or FR1); Folate receptor beta (FRb); Follicle stimulating hormone receptor (FSHR); Fos-related antigen 1; Fucosyl-GM1; G protein coupled receptor class C group 5 member D (GPRC5D); G protein-coupled receptor 20 (GPR20); GAD; Ganglioside G2 (GD2) ; Ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlc(1-1)Cer); Ganglioside GM3 (aNeu5Ac(2-3)bDClalp(1- 4)bDGlc(1-1)Cer); GD3; GFRalpha4; Glycoprotein 100 (gp100); Glypican-3 (GPC3); Gonadotropin Hormone receptor (CGHR or GR); GpA33; GpNMB; GPRC5D; Guanylyl cyclase C (GCC); Heat shock protein 70-2 mutated (mut hsp70-2); Hepatitis A virus cellular receptor 1 (HAVCR1); Hexasaccharide portion of globoH glycosphingolipid (GloboH); High molecular weight-melanoma associated antigen (HMWMAA); HIV1 envelope glycoprotein; HLA; HLA-DOA; HLA-A; HLA-A2; HLA-B; HLA-C; HLA-DM; HLA-DOB; HLA-DP; HLA-DQ; HLA-DR; HLA-G; HTLV1-Tax; Human papilloma virus E6 (HPV E6); Human papilloma virus E7 (HPV E7); Human Telomerase reverse transcriptase (hTERT); IgE; IL13Ra2; IL1 lRa; Immunoglobulin lambda-like polypeptide 1 (IGLL1); Influenza A hemagglutinin (HA); Insulin-like growth factor 1 receptor (IGF-I receptor); Interleukin 11 receptor alpha (IL-11Ra); Interleukin- 13 receptor subunit alpha-2 (IL-13Ra2 or CD213A2); Intestinal carboxyl esterase; KIT (CD117); KSHV K8.1; KSHV-gH; LAMP1 ; Legumain; Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2); Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1); Leutenizing hormone receptor (LHR); Lewis(Y) antigen; Lews Ag; Liv1; Locus K 9 (LY6K); Low conductance chloride channel; Lymphocyte antigen 6 complex; Lymphocyte antigen 75 (LY75); Lymphocyte-specific protein tyrosine kinase (LCK); Mammary gland differentiation antigen (NY-BR-1); Melanoma antigen recognized by T cells 1 (MelanA or MART1); Melanoma- associated antigen 1 (MAGE-A1); Melanoma cancer testis antigen-1 (MAD-CT-1); Melanoma cancer testis antigen-2 (MAD-CT-2); Melanoma inhibitor of apoptosis (ML-IAP); Mesothelin; MPL; Mucin 1 cell surface associated (MUC1); N-Acetyl glucosaminyl-transferase V (NA17); Nectin-4; Neural cell adhesion molecule (NCAM); NKG2D; NYBR1; O-acetyl-GD2 ganglioside (OAcGD2); Olfactory receptor 51E2 (OR51E2); Oncogene fusion protein consisting of breakpoint cluster region (BCR) and Abelson murine leukemia viral oncogene homolog 1 (Abl) (bcr-abl); P53 mutant; Paired box protein Pax-

3 (PAX3); Paired box protein Pax-5 (PAX5); Pannexin 3 (PANX3); PDL1; P-glycoprotein; Placenta-specific 1 (PLAC1); Platelet-derived growth factor receptor beta (PDGFR-beta); Polysialic acid; Proacrosin binding protein sp32 (OY-TES1); Prostase; Prostate carcinoma tumor antigen-1 (PCT A-1 or Galectin 8); Prostate stem cell antigen (PSCA); Prostate-specific membrane antigen (PSMA); Prostatic acid phosphatase (PAP); Prostein; Protease Serine 21 (Testisin or PRSS21); Proteasome (Prosome Macropain) Subunit Beta Type 9 (LMP2); PTK7; Ras G12V; Ras Homolog Family Member C (RhoC); Rat sarcoma (Ras) mutant; Receptor for Advanced Glycation Endproducts (RAGE-1); Receptor tyrosine kinase-like orphan receptor 1 (ROR1); Receptor tyrosine-protein kinase ERBB2 or Her-22/neu; Renal ubiquitous 1 (RU1); Renal ubiquitous 2 (RU2); Sarcoma translocation breakpoints; Serine 2 (TMPRSS2) ETS fusion gene; Sialyl Lewis adhesion molecule (sLe); SLAMF4; SLAMF6; Slea (CA19.9 or Sialyl Lewis Antigen); Sperm protein 17 (SPA17); Squamous Cell Carcinoma Antigen Recognized By T Cells 3 (SART3); Stage-specific embryonic antigen-4 (SSEA-4); STEAP1; Survivin; Synovial sarcoma X breakpoint 2 (SSX2); TCR Gamma Alternate Reading Frame Protein (TARP); TCR-beta1 chain; TCR-beta2 chain; TCR-delta chain; TCR-gamma chain; TCRgamma-delta; Telomerase; TGFbetaR2; The antigen recognized by TNT antibody; Thyroid stimulating hormone receptor (TSHR); Tim1-/HVCR1; Tissue Factor 1 (TF1); Tn ag; Tn antigen ((Tn Ag) or (GalNAc-Ser/Thr)); TNF receptor family member B cell maturation (BCMA); Transglutaminase 5 (TGS5); Transmembrane protease; TROP2; Tumor endothelial marker 1 (TEM1/CD248); Tumor endothelial marker 7-related (TEM7R); Tumor protein p53 (p53); Tumor-associated glycoprotein 72 (TAG72); Tyrosinase; Tyrosinase-related protein 2 (TRP-2); Uroplakin 2 (UPK2); Vascular endothelial growth factor receptor 2 (VEGFR2); V-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN); Wilms tumor protein (WT1); or X Antigen Family Member 1A (XAGE1) on the target cell.

57. The method of any one of embodiments 53-56, wherein the cancer is a cancer as provided for herein, such as a T cell or B cell disorder.

58. The viral particle of any one of embodiments 35-45, wherein the targeting moiety binds to CD7.

59. The viral particle of embodiment 58, wherein the targeting moiety that binds to CD7 comprises a polypeptide comprising: (i) a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 35; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 36; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 37, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 38; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 39; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 40; or variants of any of the foregoing.

60. The viral particle of embodiment 59, wherein the heavy chain comprises a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 47, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; and HCDR3 as set forth in SEQ ID NO: 37.

61. The viral particle of embodiments 59 and 60, wherein the light chain comprises: a light chain variable region having at least 90% sequence identity to SEQ ID NO: 48, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 38; LCDR2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40.

62. The viral particle of any one of embodiments 59-61, wherein the polypeptide comprises a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; HCDR3 as set forth in SEQ ID NO: 37; LCDR1 as set forth in SEQ ID NO: 38; LCD2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40.

63. The viral particle of any one of embodiments 59-62, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 90% sequence

identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48.

64. The viral particle of any one of embodiments 59-63, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 48.

65. The viral particle of any one of embodiments 59-64, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 48.

66. The viral particle of any one of embodiments 59-65, wherein the light chain and a heavy chain comprise: a heavy chain variable region comprising SEQ ID NO: 47, and a light chain variable region comprising SEQ ID NO: 48.

67. The viral particle of any one of embodiments 59-66, wherein the heavy chain variable region and the light chain variable region are, or are not, linked by a linker, such as a peptide linker, which can be for example, a glycine/serine linker.

68. The viral particle of embodiment 67, wherein the peptide linker comprises a sequence of (GGGGS)_n (SEQ ID NO: 49), wherein n is independently 1-5.

69. The viral particle of any one of embodiments 58-68, wherein the targeting moiety that binds to CD7 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 51, at least 95% sequence identity to SEQ ID NO: 51, at least 99% sequence identity to SEQ ID NO: 51, or a sequence as set forth in SEQ ID NO: 51.

70. The viral particle of any one of embodiments 58-68, wherein the targeting moiety that binds to CD7 comprises a polypeptide comprising a sequence having at least 90% sequence

identity to SEQ ID NO: 52, having at least 95% sequence identity to SEQ ID NO: 52, having at least 99% sequence identity to SEQ ID NO: 52, or a sequence as set forth in SEQ ID NO: 52.

71. The viral particle of any one of embodiments 35-45, wherein the targeting moiety binds to CD8.

72. The viral particle of embodiment 71, wherein the targeting moiety comprises a polypeptide that comprises: a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 55; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 56; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 57, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 58; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 59; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 60; or variants of any of the foregoing.

73. The viral particle of embodiment 72, wherein the targeting moiety comprises a heavy chain that comprises a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 67, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; and HCDR3 as set forth in SEQ ID NO: 57.

74. The viral particle of embodiment 72 or 73, wherein the CD8 targeting moiety comprises a light chain that comprises a light chain variable region having at least 90% sequence identity to SEQ ID NO: 68, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 58; LCDR2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60.

75. The viral particle of any one of embodiments 72-74, wherein the CD8 targeting moiety comprises a polypeptide comprising a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID

NO: 68, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; HCDR3 as set forth in SEQ ID NO: 57; LCDR1 as set forth in SEQ ID NO: 58; LCD2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60.

76. The viral particle of any one of embodiments 72-75, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68.

77. The viral particle of any one of embodiments 72-75, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 68.

78. The viral particle of any one of embodiments 72-75, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 68.

79. The viral particle of embodiment 71, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: wherein the heavy chain variable region comprises SEQ ID NO: 67, and the light chain variable region comprises SEQ ID NO: 68.

80. The viral particle of any one of embodiments 72-79, wherein the heavy chain variable region and the light chain variable region are, or are not, linked by a linker, such as a peptide linker.

81. The viral particle of embodiment 80, wherein the peptide linker is a glycine/serine linker.
82. The viral particle of embodiments 80 or 81, wherein the peptide linker comprises a sequence of (GGGGS)_n (SEQ ID NO: 49), wherein each n is independently 1-5.
83. The viral particle of any one of embodiments 71-82, wherein the targeting moiety that binds to CD8 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 69, at least 95% sequence identity to SEQ ID NO: 69, at least 99% sequence identity to SEQ ID NO: 69, or a sequence as set forth in SEQ ID NO: 69.
84. The viral particle of any one of embodiments 71-82, wherein the targeting moiety that binds to CD8 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 70, having at least 95% sequence identity to SEQ ID NO: 70, having at least 99% sequence identity to SEQ ID NO: 70, or a sequence as set forth in SEQ ID NO: 70.
85. The method of any one of embodiments 46-57, wherein the targeting moiety that binds to the target cell binds to CD7.
86. The method of embodiment 85, wherein the targeting moiety comprises a polypeptide comprising: (i) a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 35; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 36; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 37, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 38; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 39; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 40; or variants of any of the foregoing.
87. The method of embodiment 86, wherein the heavy chain comprises a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 47, wherein the polypeptide maintains

the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; and HCDR3 as set forth in SEQ ID NO: 37.

88. The method of embodiments 86 and 87, wherein the light chain comprises: a light chain variable region having at least 90% sequence identity to SEQ ID NO: 48, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 38; LCDR2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40.

89. The method of any one of embodiments 86-88, wherein the polypeptide comprises a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; HCDR3 as set forth in SEQ ID NO: 37; LCDR1 as set forth in SEQ ID NO: 38; LCD2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40.

90. The method of any one of embodiments 86-89, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48.

91. The method of any one of embodiments 86-90, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 48.

92. The method of any one of embodiments 86-91, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 48.

93. The method of any one of embodiments 86-92, wherein the light chain and a heavy chain comprise: a heavy chain variable region comprising SEQ ID NO: 47, and a light chain variable region comprising SEQ ID NO: 48.

94. The method of any one of embodiments 86-93, wherein the heavy chain variable region and the light chain variable region are, or are not, linked by a linker, such as a peptide linker, which can be for example, a glycine/serine linker.

95. The method of embodiment 94, wherein the peptide linker comprises a sequence of (GGGGS)_n (SEQ ID NO: 49), wherein n is independently 1-5.

96. The method of any one of embodiments 85-95, wherein the targeting moiety that binds to CD7 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 51, at least 95% sequence identity to SEQ ID NO: 51, at least 99% sequence identity to SEQ ID NO: 51, or a sequence as set forth in SEQ ID NO: 51.

97. The method of any one of embodiments 85-95, wherein the targeting moiety that binds to CD7 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 52, having at least 95% sequence identity to SEQ ID NO: 52, having at least 99% sequence identity to SEQ ID NO: 52, or a sequence as set forth in SEQ ID NO: 52.

98. The method of any one of embodiments 46-57, wherein the targeting moiety that binds to the target cell binds to CD8.

99. The method of embodiment 98, wherein the targeting moiety comprises a polypeptide that comprises: a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 55; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 56; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 57, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO:

58; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 59; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 60; or variants of any of the foregoing.

100. The method of embodiment 99, wherein the targeting moiety comprises a heavy chain that comprises a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 67, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; and HCDR3 as set forth in SEQ ID NO: 57.

101. The method of embodiment 99 or 100, wherein the CD8 targeting moiety comprises a light chain that comprises a light chain variable region having at least 90% sequence identity to SEQ ID NO: 68, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 58; LCDR2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60.

102. The method of any one of embodiments 99-101, wherein the CD8 targeting moiety comprises a polypeptide comprising a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; HCDR3 as set forth in SEQ ID NO: 57; LCDR1 as set forth in SEQ ID NO: 58; LCD2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60.

103. The method of any one of embodiments 99-102, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68.

104. The method of any one of embodiments 99-102, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain

variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 67; and
a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 68.

105. The method of any one of embodiments 99-102, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 67; and
a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 68.

106. The method of embodiment 98, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: wherein the heavy chain variable region comprises SEQ ID NO: 67, and the light chain variable region comprises SEQ ID NO: 68.

107. The method of any one of embodiments 99-106, wherein the heavy chain variable region and the light chain variable region are, or are not, linked by a linker, such as a peptide linker.

108. The method of embodiment 107, wherein the peptide linker is a glycine/serine linker.

109. The method of embodiments 107 or 108, wherein the peptide linker comprises a sequence of (GGGS)_n (SEQ ID NO: 49), wherein each n is independently 1-5.

110. The method of any one of embodiments 98-109, wherein the targeting moiety that binds to CD8 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 69, at least 95% sequence identity to SEQ ID NO: 69, at least 99% sequence identity to SEQ ID NO: 69, or a sequence as set forth in SEQ ID NO: 69.

111. The method of any one of embodiments 98-109, wherein the targeting moiety that binds to CD8 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 70, having at least 95% sequence identity to SEQ ID NO: 70, having at least 99% sequence identity to SEQ ID NO: 70, or a sequence as set forth in SEQ ID NO: 70.

112. The viral particle of any one of embodiments 40-45 or 58-84, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 75.

113. The viral particle of embodiment 112, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 90% identity to SEQ ID NO: 75.

114. The viral particle of embodiment 112, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 95% identity to SEQ ID NO: 75.

115. The viral particle of embodiment 112, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 98% identity to SEQ ID NO: 75.

116. The viral particle of embodiment 112, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 99% identity to SEQ ID NO: 75.

117. The viral particle of embodiment 112, wherein the CAR comprises an antigen binding domain having an amino acid sequence having the amino acid sequence of SEQ ID NO: 75.

118. The viral particle of any one of embodiments 40-45 or 58-84, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 76.

119. The viral particle of embodiment 118, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 90% identity to SEQ ID NO: 76.

120. The viral particle of embodiment 118, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 95% identity to SEQ ID NO: 76.

121. The viral particle of embodiment 118, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 98% identity to SEQ ID NO: 76.

122. The viral particle of embodiment 118, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 99% identity to SEQ ID NO: 76.

123. The viral particle of embodiment 118, wherein the CAR comprises an antigen binding domain having an amino acid sequence having the amino acid sequence of SEQ ID NO: 76.

124. The method of any one of embodiments 46-57 or 85-111 wherein the heterologous molecule of interest is a chimeric antigen receptor (CAR), and wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 75.

125. The method of embodiment 124, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 90% identity to SEQ ID NO: 75.

126. The method of embodiment 124, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 95% identity to SEQ ID NO: 75.

127. The method of embodiment 124, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 98% identity to SEQ ID NO: 75.

128. The method of embodiment 124, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 99% identity to SEQ ID NO: 75.

129. The method of embodiment 124, wherein the CAR comprises an antigen binding domain having the amino acid sequence of SEQ ID NO: 75.

130. The method of any one of embodiments 46-57 or 85-111 wherein the heterologous

molecule of interest is a chimeric antigen receptor (CAR), and wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identity to SEQ ID NO: 76.

131. The method of embodiment 130, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 90% identity to SEQ ID NO: 76.

132. The method of embodiment 130, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 95% identity to SEQ ID NO: 76.

133. The method of embodiment 130, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 98% identity to SEQ ID NO: 76.

134. The method of embodiment 130, wherein the CAR comprises an antigen binding domain having an amino acid sequence having at least 99% identity to SEQ ID NO: 76.

135. The method of embodiment 130, wherein the CAR comprises an antigen binding domain having the amino acid sequence of SEQ ID NO: 76.

136. A viral particle comprising a targeting moiety that binds to CD7 or CD8, or a pharmaceutical composition comprising the same.

137. The viral particle of embodiment 136, or a pharmaceutical composition comprising the same, wherein the targeting moiety that binds to CD7 comprises a polypeptide comprising: (i) a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 35; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 36; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 37, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 38; the light

chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 39; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 40; or variants of any of the foregoing.

138. The viral particle of embodiment 137, or a pharmaceutical composition comprising the same, wherein the targeting moiety comprises a heavy chain comprising a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 47, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; and HCDR3 as set forth in SEQ ID NO: 37.

139. The viral particle of embodiment 137, or a pharmaceutical composition comprising the same, wherein the targeting moiety comprises a light chain comprising: a light chain variable region having at least 90% sequence identity to SEQ ID NO: 48, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 38; LCDR2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40.

140. The viral particle of embodiment 137, or a pharmaceutical composition comprising the same, wherein the targeting moiety that binds to CD7 comprises a polypeptide comprising a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; HCDR3 as set forth in SEQ ID NO: 37; LCDR1 as set forth in SEQ ID NO: 38; LCD2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40.

141. The viral particle of embodiment 137, or a pharmaceutical composition comprising the same, wherein the targeting moiety comprises a light chain and a heavy chain, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48.

141. The viral particle of embodiment 141, or a pharmaceutical composition comprising the

same, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 48.

142. The viral particle embodiment 141, or a pharmaceutical composition comprising the same, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 48.

143. The viral particle embodiment 141, or a pharmaceutical composition comprising the same, wherein the light chain and a heavy chain comprise: a heavy chain variable region comprising SEQ ID NO: 47, and a light chain variable region comprising SEQ ID NO: 48.

144. The viral particle of embodiment 136, or a pharmaceutical composition comprising the same, wherein the targeting moiety that binds to CD7 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 51, at least 95% sequence identity to SEQ ID NO: 51, at least 99% sequence identity to SEQ ID NO: 51, or a sequence as set forth in SEQ ID NO: 51.

145. The viral particle of embodiment 136, or a pharmaceutical composition comprising the same, wherein the targeting moiety that binds to CD7 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 52, having at least 95% sequence identity to SEQ ID NO: 52, having at least 99% sequence identity to SEQ ID NO: 52, or a sequence as set forth in SEQ ID NO: 52.

146. The viral particle of embodiment 136, or a pharmaceutical composition comprising the same, wherein the targeting moiety binds to CD8.

147. The viral particle of embodiment 146, or a pharmaceutical composition comprising the same, wherein the targeting moiety comprises a polypeptide that comprises: a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain

CDR1 sequence has the amino acid sequence of SEQ ID NO: 55; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 56; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 57, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 58; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 59; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 60; or variants of any of the foregoing.

148. The viral particle of embodiment 146, or a pharmaceutical composition comprising the same, wherein the targeting moiety comprises a heavy chain that comprises a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 67, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; and HCDR3 as set forth in SEQ ID NO: 57.

149. The viral particle of embodiment 146, or a pharmaceutical composition comprising the same, wherein the CD8 targeting moiety comprises a light chain that comprises a light chain variable region having at least 90% sequence identity to SEQ ID NO: 68, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 58; LCDR2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60.

150. The viral particle of embodiment 146, or a pharmaceutical composition comprising the same, wherein the CD8 targeting moiety comprises a polypeptide comprising a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; HCDR3 as set forth in SEQ ID NO: 57; LCDR1 as set forth in SEQ ID NO: 58; LCD2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60.

150. The viral particle of embodiment 146, or a pharmaceutical composition comprising the same, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a

heavy chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68.

151. The viral particle of embodiment 146, or a pharmaceutical composition comprising the same, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 68.

152. The viral particle of embodiment 146, or a pharmaceutical composition comprising the same, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 68.

153. The viral particle of embodiment 146, or a pharmaceutical composition comprising the same, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: wherein the heavy chain variable region comprises SEQ ID NO: 67, and the light chain variable region comprises SEQ ID NO: 68.

154. The viral particle of embodiment 146, or a pharmaceutical composition comprising the same, wherein the targeting moiety that binds to CD8 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 69, at least 95% sequence identity to SEQ ID NO: 69, at least 99% sequence identity to SEQ ID NO: 69, or a sequence as set forth in SEQ ID NO: 69.

155. The viral particle of embodiment 146, or a pharmaceutical composition comprising the same, wherein the targeting moiety that binds to CD8 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 70, having at least 95% sequence identity to SEQ ID NO: 70, having at least 99% sequence identity to SEQ ID NO: 70, or a sequence

as set forth in SEQ ID NO: 70.

156. The viral particle of any one of embodiments 136-155, or a pharmaceutical composition comprising the same, wherein the viral particle is a lentivirus derived viral particle.

157. The viral particle of any one of embodiments 136-156, or a pharmaceutical composition comprising the same, wherein the viral particle comprises a heterologous fusogenic protein, such as a glycoprotein from such as those provided for herein .

158. The viral particle of embodiment 157, or a pharmaceutical composition comprising the same, wherein the fusogenic protein is a VSV-G protein.

159. The viral particle of embodiment 158, or a pharmaceutical composition comprising the same, wherein the VSV-G protein is a mutant VSV-G protein.

160. The viral particle of embodiment 159, or a pharmaceutical composition comprising the same, wherein the mutant VSV-G protein abrogates binding to its natural co-receptor, such as LDL-R.

161. The viral particle of embodiment 159, or a pharmaceutical composition comprising the same, wherein the mutant VSV-G protein comprises a polypeptide of any one of embodiments 1-30.

162. The viral particle of embodiment 159, or a pharmaceutical composition comprising the same, wherein the viral particle comprises a nucleic acid molecule encoding a chimeric antigen receptor.

163. The viral particle of embodiment 162, or a pharmaceutical composition comprising the same, wherein the chimeric antigen receptor comprises an antigen binding domain that binds to CD20.

164. The viral particle of embodiment 163, or a pharmaceutical composition comprising the same, wherein the antigen binding domain that binds to CD20 comprises a VH and VL as provided for herein.

165. A method of delivering a heterologous molecule to a CD7+ or CD8+ target cell or a method of treating cancer in a subject, the method comprising contacting a cell with, or administering to the subject, a pseudotyped viral vector comprising a targeting moiety that specifically binds to CD7 or CD8 on the target cell and a nucleic acid molecule encoding the heterologous molecule, wherein:

i) the targeting moiety that binds to CD7 comprises:

a polypeptide comprising: (i) a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 35; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 36; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 37, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 38; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 39; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 40; or variants of any of the foregoing;

a heavy chain comprising a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 47, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; and HCDR3 as set forth in SEQ ID NO: 37;

a light chain comprising: a light chain variable region having at least 90% sequence identity to SEQ ID NO: 48, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 38; LCDR2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40;

a polypeptide comprises a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47, and a light chain variable region of the light chain having at least 90% sequence identity

to SEQ ID NO: 48, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; HCDR3 as set forth in SEQ ID NO: 37; LCDR1 as set forth in SEQ ID NO: 38; LCD2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40;

a light chain and a heavy chain, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48;

a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 48;

a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 48;

a heavy chain variable region comprising SEQ ID NO: 47, and a light chain variable region comprising SEQ ID NO: 48; a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 51, at least 95% sequence identity to SEQ ID NO: 51, at least 99% sequence identity to SEQ ID NO: 51, or a sequence as set forth in SEQ ID NO: 51; or

a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 52, having at least 95% sequence identity to SEQ ID NO: 52, having at least 99% sequence identity to SEQ ID NO: 52, or a sequence as set forth in SEQ ID NO: 52; or
ii) the targeting moiety that binds to CD8 comprises:

a polypeptide that comprises: a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 55; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 56; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 57, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 58; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 59; and the light chain CDR3

sequence has the amino acid sequence of SEQ ID NO: 60; or variants of any of the foregoing;

a heavy chain that comprises a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 67, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; and HCDR3 as set forth in SEQ ID NO: 57;

a light chain that comprises a light chain variable region having at least 90% sequence identity to SEQ ID NO: 68, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 58; LCDR2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60;

a polypeptide comprising a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; HCDR3 as set forth in SEQ ID NO: 57; LCDR1 as set forth in SEQ ID NO: 58; LCD2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60;

a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68;

a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 68;

a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 68;

a polypeptide comprising a light chain and a heavy chain comprising: wherein the heavy chain variable region comprises SEQ ID NO: 67, and the light chain variable region

comprises SEQ ID NO: 68;

a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 69, at least 95% sequence identity to SEQ ID NO: 69, at least 99% sequence identity to SEQ ID NO: 69, or a sequence as set forth in SEQ ID NO: 69; or

a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 70, having at least 95% sequence identity to SEQ ID NO: 70, having at least 99% sequence identity to SEQ ID NO: 70, or a sequence as set forth in SEQ ID NO: 70.

166. The method of embodiment 165, wherein the pseudotyped viral particle is a pseudotyped lentivirus derived viral particle.

167. The method of embodiments 165 or 166, wherein the pseudotyped viral particle comprises a heterologous fusogenic protein, such as a glycoprotein from those provided for herein .

168. The method of embodiment 167, wherein the fusogenic protein is a VSV-G protein.

169. The method of embodiment 168, wherein the VSV-G protein is a mutant VSV-G protein.

170. The method of embodiment 169, wherein the mutant VSV-G protein abrogates binding to its natural co-receptor, such as LDL-R.

171. The method of embodiment 169, wherein the mutant VSV-G protein comprises a polypeptide of any one of embodiments 1-30.

172. The method of any one of embodiments 165-171, wherein the heterologous nucleic acid molecule encodes for an siRNA, an shRNA, a non-coding RNA, a peptide, a polypeptide, a protein, a viral payload, a viral genome, a chimeric antigen receptor (“CAR”) or a combination thereof.

173. The method of embodiment 172, wherein the heterologous nucleic acid molecule encodes for a chimeric antigen receptor.

174. The method of embodiment 173, wherein the chimeric antigen receptor comprises an antigen binding domain that binds to CD20.

175. The method of embodiment 174, wherein the antigen binding domain that binds to CD20 comprises a VH and VL as provided for herein and linked polypeptides comprising the same.

176. A cell comprising one or more heterologous nucleic acid molecules encoding for a targeting moiety and a viral glycoprotein.

177. The cell of embodiment 176, wherein the heterologous nucleic acid molecules comprise a nucleic acid sequence encoding for a targeting moiety that binds to CD7 or CD8.

178. The cell of embodiment 177, wherein the nucleic acid sequence encoding for a targeting moiety that binds to CD7 encodes for a polypeptide comprising: (i) a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 35; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 36; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 37, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 38; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 39; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 40; or variants of any of the foregoing.

179. The cell of embodiment 177, wherein the nucleic acid molecule that encodes for a targeting moiety that binds CD7 encodes for a polypeptide comprising a heavy chain comprising a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 47, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; and HCDR3 as set forth in SEQ ID NO: 37.

180. The cell of embodiment 177, wherein the nucleic acid molecule that encodes for a targeting moiety that binds CD7 encodes for a polypeptide comprising a light chain comprising: a light chain

variable region having at least 90% sequence identity to SEQ ID NO: 48, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 38; LCDR2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40.

181. The cell of embodiment 166, wherein the nucleic acid molecule that encodes for a targeting moiety that binds CD7 encodes for a polypeptide comprising a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; HCDR3 as set forth in SEQ ID NO: 37; LCDR1 as set forth in SEQ ID NO: 38; LCD2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40.

182. The cell of embodiment 177, wherein the nucleic acid molecule that encodes for a targeting moiety that binds CD7 encodes for a polypeptide comprising a light chain and a heavy chain, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48.

183. The cell of embodiment 177, wherein the nucleic acid molecule that encodes for a targeting moiety that binds CD7 encodes for a polypeptide comprising a light chain and a heavy chain that comprise: a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 48.

184. The cell of embodiment 183, wherein the light chain and the heavy chain comprise: a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 48.

185. The cell of embodiment 183, wherein the light chain and the heavy chain comprise: a heavy

chain variable region comprising SEQ ID NO: 47, and a light chain variable region comprising SEQ ID NO: 48.

186. The cell of embodiment 177, wherein the nucleic acid sequence encoding for a targeting moiety that binds to CD7 encodes for a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 51, at least 95% sequence identity to SEQ ID NO: 51, at least 99% sequence identity to SEQ ID NO: 51, or a sequence as set forth in SEQ ID NO: 51.

187. The cell of embodiment 177, wherein the nucleic acid sequence encoding for a targeting moiety that binds to CD7 encodes for a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 52, having at least 95% sequence identity to SEQ ID NO: 52, having at least 99% sequence identity to SEQ ID NO: 52, or a sequence as set forth in SEQ ID NO: 52.

188. The cell of embodiment 177, wherein the nucleic acid sequence encoding for a targeting moiety that binds to CD8.

189. The cell of embodiment 188, wherein the targeting moiety comprises a polypeptide that comprises: a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 55; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 56; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 57, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 58; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 59; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 60; or variants of any of the foregoing.

190. The cell of embodiment 188, wherein the targeting moiety comprises a heavy chain that comprises a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 67, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55;

HCDR2 as set forth in SEQ ID NO: 56; and HCDR3 as set forth in SEQ ID NO: 57.

191. The cell of embodiment 188, wherein the CD8 targeting moiety comprises a light chain that comprises a light chain variable region having at least 90% sequence identity to SEQ ID NO: 68, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 58; LCDR2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60.

192. The cell of embodiment 188, wherein the CD8 targeting moiety comprises a polypeptide comprising a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; HCDR3 as set forth in SEQ ID NO: 57; LCDR1 as set forth in SEQ ID NO: 58; LCD2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60.

193. The cell of embodiment 188, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68.

194. The cell of embodiment 188, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 68.

195. The cell of embodiment 188, wherein the CD8 targeting moiety comprises a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 68.

196. The cell of embodiment 188, wherein the CD8 targeting moiety comprises a polypeptide

comprising a light chain and a heavy chain comprising: wherein the heavy chain variable region comprises SEQ ID NO: 67, and the light chain variable region comprises SEQ ID NO: 68.

197. The cell of embodiment 188, wherein the targeting moiety that binds to CD8 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 69, at least 95% sequence identity to SEQ ID NO: 69, at least 99% sequence identity to SEQ ID NO: 69, or a sequence as set forth in SEQ ID NO: 69.

198. The cell of embodiment 188, wherein the targeting moiety that binds to CD8 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 70, having at least 95% sequence identity to SEQ ID NO: 70, having at least 99% sequence identity to SEQ ID NO: 70, or a sequence as set forth in SEQ ID NO: 70.

199. The cells of any one of embodiments 176-198, wherein the heterologous nucleic acid molecule encoding the viral glycoprotein is a glycoprotein of any one of embodiments 1-30.

200. A cell comprising the nucleic acid molecule of embodiment 31, the vector of embodiment 32, or the plasmid of embodiment 33.

201. A method of making a viral particle comprising culturing the cell of any one of embodiments 176-198 under conditions sufficient to make the viral particle.

EXAMPLES

[0352] The following examples are illustrative, but not limiting, of the compounds, compositions, particles, polypeptids, and methods described herein. Other suitable modifications and adaptations known to those skilled in the art are within the scope of the following embodiments.

Example 1: Mutation at position 182 of VSV-G abrogates LDL-R interaction, but retains fusing properties.

[0353] Plasmids/Sequences. All VSV-G plasmids were derived from pCMV-VSV-G Envelope Vector (Cell Bio Labs, catalog RV-110). Point mutations and combinations thereof were

introduced using site directed mutagenesis (New England Biolabs). Individual mutations H8A and K47Q were previously shown to partially “blind” VSV-G, reducing its binding to LDL-R, the native cellular receptor for VSV (PMID: 29531262, DOI: 10.1038/s41467-018-03432-4). In this experiment, a single binder molecule consisting of a CD7-targeting scFv (clone MT701) fused to an IgG “stalk” bearing a CD28 transmembrane domain was used.

[0354] Cells. HEK293T cells were grown in DMEM with 10% FBS. SupT1 cells were maintained in RPMI media with 10% FBS. Human PBMCs were purchased from AllCells and cultured in X-Vivo 10 (Lonza) supplemented with 20ng/mL IL-2 (Peprotech). PBMCs were activated 48 hours prior to transduction using anti-CD3/CD28 Dynabeads (Cell Therapy Systems).

[0355] Generation of lentiviral particles. The recombinant lentiviral particles co-expressing VSV-G glycoprotein and binder molecules were generated by plasmid transection into HEK293T cells using Lipofectamine 3000 (ThermoFisher Scientific). A total of 5 plasmids were transfected: (1) plasmid expressing the VSV-G glycoprotein, (2) plasmid expressing the binder protein (3) plasmid expressing the lentiviral transfer genome encoding for eGFP, (4) plasmid expressing gag-pol, and (5) plasmid expressing rev. Transfected cell supernatant was harvested 48 hours later. Virus in the cell supernatant was concentrated by centrifugation through a sucrose cushion and resuspended in PBS. Lentiviral particle titer was determined using the Lenti-X p24 Rapid Titer Kit (Takara Bio, San Jose, CA).

[0356] Lentivirus transduction assay. A series of 10-fold dilutions (in cell culture media) of the concentrated lentivirus was performed and used to infect SupT1 and activated human PBMCs. Media was replaced 6 hours later, and the transduced cells were analyzed by flow cytometry on days 4 and 7 after transduction. Cells were stained with a viability stain and an anti-CD7 antibody to detect CD7 positive cells (PeCy7 mouse-anti-human CD7, clone CD7-6B7, BD Biosciences). Expression of eGFP was measured to calculate transduction efficiency.

[0357] Structure-guided design of novel blinding mutations. Using a published crystal structures of VSV-G bound to CR2 and CR3 of the LDL-R (pdb 50YL and 50Y9, respectively), we identified two putative positions in VSV-G with side chains oriented toward the binding interface on LDL-R (FIG. 1). Residue Q10 (SEQ ID NO: 2) appeared to form several interactions with residues in both CR2 and CR3. In CR3, this included interactions with a positively charged arginine residue. Thus, three substitutions were tested: Q10A to reduce side-chain interactions that potentially stabilize LDL-R binding and Q10R and Q10K to create electrostatic repulsion.

[0358] Residue I182 (SEQ ID NO: 2) appeared to contact several residues in both CR2 and CR3 as well. Three substitutions were tested: I182A to reduce side-chain interactions that potentially stabilize LDL-R binding and I182D and I182E to create electrostatic repulsion against the primary binding interfaces on LDL-R.

[0359] Addition of negative charges in the binding interface ablate native tropism without altering fusogenicity. Titration of viral supernatants on the CD7+ T cell line SupT1 validated the structural predictions for residue I182. In the absence of any compensatory binder molecule, WT VSV-G reached titers of 3.0×10^8 while both I182D and I182E were ~ 3 orders of magnitude lower (Fig 2). Substitutions at residue I182 preserved fusogenicity, as titers were restored to 1×10^8 in the presence of a binder redirecting the virions to CD7.

[0360] The data is illustrated in FIG. 2, which shows that the addition of negative charges in the binding interface ablate native tropism without altering fusogenicity. FIG. 2A shows the titration of VSV-G constructs on SupT1 cells. Plotted is the percentage of SupT1 cells expressing GFP at each amount of viral input in terms of p24 antigen. Dashed lines/open circles indicate VSV-G constructs alone, solid lines/filled circles indicate the same construct with a CD7 targeting molecule expressed in trans. FIG. 2B illustrates Functional titer of each construct calculated from the titration in A, expressed as transducing units per mL of concentrated virus supernatant (TU/mL).

[0361] Thus these examples demonstrate that a mutation at position 182 is sufficient to abrogate the LDL-R interaction, but retain fusogenic properties when combining with a targeting moiety that binds to a target on the target cell.

Example 2: Serum Stable VSV-G Protein

[0362] WT VSV-G is reported to be sensitive to inactivation by naïve human serum, with inactivation ranging from minimal change to about 100-fold decrease depending on study parameters. Accordingly, it was assessed i) whether substitution at position I182 produced a variant VSV-G with similar sensitivity to serum and ii) whether incorporation of known serum stabilizing VSV-G mutations into the I182 substituted constructs had an effect on the serum stability of the constructs.

[0363] WT VSV-G, Serum Stable VSV-G (T214N, T352A), LDL-R de-targeted VSV-G (I182E), and LDL-R de-targeted serum stable VSV-G (I182E, T214N, T352A) were tested for their ability to infect SupT1 cells in the presence or absence of naïve serum or heat inactivated

serum (HI) (FIG. 4). As expected WT VSV-G had a dramatic decrease in percent infectivity when administered in the presence of serum with a mild rescue of infectivity if serum was heat inactivated (HI) (FIG. 4, columns 1-3). The inclusion of the CD7 targeting moiety on an IgG1 Fc stalk (WT) did not rescue the serum inactivation (FIG. 4, columns 4-6). The VSV-G constructs harboring the T214N and T352A mutations showed a similar reduction in infectivity in the presence of serum with no distinction between naïve or HI serum (FIG. 4, columns 7-9). However, the inclusion of the targeting moiety was able to partially rescue the inactivation in either serum condition (FIG. 4, columns 10-12). Surprisingly, the LDL-R de-targeted VSV-G (I182E) showed no reduction in infectivity percentage in the presence of either serum condition (FIG. 4, columns 13-15), but rather showed increased infectivity in the presence of naïve serum as compared to serum free or HI serum conditions. Further inclusion of the known serum stable mutations into VSV-G (I182E, T214N, T352A) also demonstrated increased infectivity in the presence of serum as compared to serum free conditions (FIG. 4, columns 16-18). Further, the I182E, T214N, T352A construct also demonstrates increased infectivity in the presence of HI serum (FIG. 4, column 18), which indicates that the triple mutant construct has a higher degree of serum stability than the other constructs examined.

[0364] To further characterize the effect of including serum stabilizing mutations in the viral constructs of the present application, WT VSV-G, VSV-G (I182E), and VSV-G (I182E, T230N, T352A) were assessed at a lower dose of vector (FIG. 5). In agreement with the previous assessment, WT VSV-G showed decreased infectivity in serum as compared to no-serum conditions (FIG. 5, columns 1 and 2). However, heat inactivation (HI) of serum did appear to have a rescuing effect on WT VSV-G infectivity (FIG. 5, column 3). In further agreement with the previous assessment, VSV-G (I182E) demonstrated enhanced infectivity in the presence of serum as compared to no serum conditions (FIG. 5, columns 4 and 5). The presence of HI serum also resulted in an increased infectivity compared to no serum, with a smaller effect than naïve serum (FIG. 5, column 6). VSV-G (I182E, T214N, T352A) also demonstrated increased infectivity in serum and HI serum conditions as compared to no serum conditions (FIG. 5, columns 7-9), in agreement with the previous assessment. Further, when compared to the respective no serum condition, VSV-G (I182E, T230N, T352A) results in a greater percent increased infectivity compared to VSV-G (I182E), suggesting that inclusion of all three mutations results in greater serum stability than just I182E alone.

[0365] Thus, these examples demonstrate that i) the LDL-R detargeting mutation I182E affords an unexpected level of protection against the serum inactivation observed for WT VSV-G, ii) the LDL-R detargeting mutation I182E does not abrogate the serum stabilizing effect of the T230N + T352A mutations, and iii) the combined mutation construct VSV-G (I182E, T230N, T352A) appears to have a greater serum stabilizing effect than either I182E alone or T230N + T352A alone. Thus, VSV-G (I182E, T230N, T352A) is able to ablate native tropism without alternating fusogenicity and additionally demonstrates a more stable profile and is not neutralized in serum.

Example 3: CD7 Targeting Moiety can be linked to a Fc stalk to transduce PBMCs.

[0366] Generation of Plasmids and Sequences.

[0367] The amino acid sequence including CDRs of the CD7 binder was determined via mass spectrometry (Rapid Novor). CD7 binder sequences were synthesized by IDT or GenScript (Piscataway, NJ) using codon optimization for human expression and inserted onto viral glycoprotein or IgG-based stalks and flanked by a G4S linker. An example viral glycoprotein is a Nipah-G protein with the CD7 binder attached to the extracellular region. An example IgG-based stalk was a human IgG1 Fc dimer with a CD8 or CD28 transmembrane region and an envelope incorporating motif with the CD7 binder attached to the extracellular region. The resulting binders were expressed under the direction of a CMV promoter.

[0368] Generation of Engineered Lentiviral Particles.

[0369] The recombinant lentiviral particles expressing the CD7 binder incorporated on the surface were generated by plasmid transection into HEK293T cells using Lipofectamine 3000 (ThermoFisher Scientific). In some embodiments, a total of 5 plasmids were transfected: (1) plasmid expressing the CD7 binder with the IgG stalk or a plasmid expressing the CD7 binder on a Nipah-G protein, (2) plasmid expressing a detargeted-VSV-G protein, (3) lentiviral genome expressing eGFP, (4) plasmid expressing gag-pol, (5) plasmid expressing rev. In some embodiments, a total of 5 plasmids were transfected: (1) plasmid expressing the CD7 binder with the Nipah-G protein, (2) plasmid expressing Nipah-F protein, (3) lentiviral genome expressing eGFP, (4) plasmid expressing gag-pol, (5) plasmid expressing rev. Media was changed 6 hours after transfection and cells were harvested 48 hours later. Virus in the media was concentrated by centrifugation through a sucrose cushion and resuspended in media. Lentiviral particle titer was determined using the Lenti-X p24 Rapid Titer Kit (Takara Bio, San Jose, CA)

[0370] CD7 binders attach to both human and non-human primate PBMCs, leading to CD7+ cell transduction.

[0371] Human PBMC cells were maintained in X-Vivo10 media with 5% human serum and 20ng/mL of human IL-2. Non-human primate cells were maintained in RPMI media with 10% FBS, 1% Pen/Strep with 1mM Sodium Pyruvate and 100units/mL hIL-2. Concentrated lentivirus was used to infect human and non-human primate cells. Media was replaced 24 hours later, and the transduced cells were analyzed by flow cytometry on days 4 and 7 after transduction. Cells were stained with an anti-CD7 antibody to detect CD7 positive cells (PE-Cy7 mouse-anti-human CD7, clone MT-701, Biolegend) as well as GFP expression. Cell viability was also determined. CD7 binders attached to Paramyxoviridae glycoproteins, such as Nipah-G (FIG. 6A-D and FIG. 7A-D) or with Rhabdoviridae glycoproteins, such as VSV-G (FIG. 6I-L and FIG. 7I-L) successfully transduced human PBMCs (FIG. 6A-L) or NHP PBMCs (FIG. 7A-L) compared to the binder alone (FIG. 6E-H and FIG. 7E-H) which shows no transduction. When CD7 with an IgG stalk was tested, the binder alone (top row) again did not transduce human (FIG. 6M) or NHP PBMCs (FIG. 7M) but when administered with detargeted VSV-G, transduction was seen in both human (FIG. 6M, bottom panel) and NHP PBMCs (FIG. 7M, bottom panel). GFP transduction experiments were repeated as described above for detargeted VSV-G pseudotyped lentiviral constructs with CD7 binders. Results of the repeat experiments were in agreement with the data of FIGS 6 and 7 and are provided in Table 1 below.

TABLE 1

	WT VSVG	Detartgeted VSVG	CD7 Binder (SEQ ID NO. 51)	CD7 Binder (SEQ ID NO. 52)	CD8 Binder (SEQ ID NO. 69)	CD7 Binder (SEQ ID NO. 70)
Human PBMC Transduction (% positive)	52.76	0.00798	19.32	21.55	4.09	3.17
NHP PBMC Transduction (% positive)	48.4	0.026	2.61	5.64	1.8	0.76

[0372] These examples and embodiments demonstrate that the polypeptides can be used to bind to CD7 and target viral particles to cells expressing CD7 to transduce the cells expressing

CD7.

Example 4: CD8 Targeting Moiety can be linked to a Fc stalk to transduce PBMCs.

[0373] Generation of Plasmids and Sequences.

[0374] The amino acid sequence including CDRs of the CD8 binder was determined via mass spectrometry (Rapid Novor). CD8 binder sequences were synthesized by IDT or GenScript (Piscataway, NJ) using codon optimization for human expression and inserted onto viral glycoprotein or IgG-based stalks and flanked by a G4S linker. An example viral glycoprotein is a Nipah-G protein with the CD8 binder attached to the extracellular region. An example IgG-based stalk was a human IgG1 Fc dimer with a CD8 or CD28 transmembrane region and an envelope incorporating motif with the CD8 binder attached to the extracellular region. The resulting binders were expressed under the direction of a CMV promoter.

[0375] Generation of Engineered Lentiviral Particles.

[0376] The recombinant lentiviral particles expressing the CD8 binder incorporated on the surface were generated by plasmid transection into HEK293T cells using Lipofectamine 3000 (ThermoFisher Scientific). In some embodiments, a total of 5 plasmids were transfected: (1) plasmid expressing the CD8 binder with the IgG stalk, (2) plasmid expressing a detargeted-VSV-G protein, (3) lentiviral genome expressing eGFP, (4) plasmid expressing gag-pol, (5) plasmid expressing rev. In some embodiments, the lentiviral genome expressing eGFP was replaced with a CAR molecule. Media was changed 6 hours after transfection and cells were harvested 48 hours later. Virus in the media was concentrated by centrifugation through a sucrose cushion and resuspended in media. Lentiviral particle titer was determined using the Lenti-X p24 Rapid Titer Kit (Takara Bio, San Jose, CA).

[0377] CD8 binders attach to both human and non-human primate PBMCs, leading to CD8+ cell transduction.

[0378] Human PBMC cells were maintained in X-Vivo10 media with 5% human serum and 20ng/mL of human IL-2. Non-human primate cells were maintained in RPMI media with 10% FBS, 1% Pen/Strep with 1mM Sodium Pyruvate and 100units/mL hIL-2. Concentrated lentivirus was used to infect human and non-human primate cells. Media was replaced 24 hours later, and the transduced cells were analyzed by flow cytometry on day 6 after transduction. Cells were stained with an anti-CD8 antibody to detect CD8 positive cells (BV421 mouse-anti-human CD8, clone RPA-T8, Biolegend) as well as GFP or CAR20 expression. Cell viability was also

determined. CD8 binders with detargeted VSV-G successfully transduced human PBMCs (FIG. 8A) or NHP PBMCs (FIG. 8B) with GFP (FIG. 8A-8B, top row) and CAR20 (FIG. 8A-8B, bottom row). GFP transduction experiments were repeated as described above for detargeted VSV-G pseudotyped lentiviral constructs with CD8 binders. Results of the repeat experiments were in agreement with the data of FIG 8 and are provided in Table 1 above.

[0379] These examples and embodiments demonstrate that the polypeptides can be used to bind to CD8 and target viral particles to cells expressing CD8 to transduce the cells expressing CD8.

[0380] **Example 5: Binder construct affects transgene expression.**

[0381] Concentrated lentivirus was used to transduce SupT1 cells. Media was replaced 24 hours later, and the transduced cells were analyzed by flow cytometry. Cells were stained with an anti-CD8 antibody to detect CD8 positive cells or an anti-CD7 antibody to detect CD7 positive cells as well as GFP or CAR20 expression.

[0382] Lentiviral constructs utilizing CD7 or CD8 binders as provided for herein exhibited comparable GFP transduction (FIG. 9A). Binder orientation (VH-VL vs VL-VH) had minimal effect on transgene expression. WT VSV-G was used as a positive control, and detargeted VSV-G was used as a negative control. With respect to a CD20-CAR transgene, lentiviral constructs utilizing CD7 binders as provided for herein provided superior transgene transduction as compared to lentiviral constructs utilizing CD8 binders (FIG. 9B). As with the GFP transgene, binder orientation had minimal effect on transgene expression. Lentiviral constructs utilizing CD7 or CD8 binders both demonstrated transgene expression levels near the level of WT VSV-G when greater amounts of virus were utilized.

[0383] These examples and embodiments demonstrate that the CD7 and CD8 binders of the present disclosure can be incorporated into a lentiviral particle to direct successful transgene expression in a target cell. Further, these examples and embodiments demonstrate that different binders result in differential transgene expression, and that both the identity of the binder and the identity of the transgene are important variable in the efficacy of the viral particle.

[0384] **Example 6: VSV-G* pseudotyped lentiviruses harboring CD7 binder transduce PBMCs.**

[0385] Human PBMC cells were maintained in X-Vivo10 media with 5% human serum and 20ng/mL of human IL-2. Non-human primate cells were maintained in RPMI media with 10%

FBS, 1% Pen/Strep with 1mM Sodium Pyruvate and 100units/mL hIL-2. Concentrated lentivirus was used to infect human and non-human primate cells. The VSV-G protein utilized was a variant VSV-G protein harboring a mutation to prevent binding of VSV-G to the LDL-R. The variant VSV-G is denoted VSV-G* and corresponds to VSV-G (I182E, T214N, T352A) (e.g. SEQ ID NO: 23), as provided for herein. A CD7 binder construct was utilized in the present example. The CD7 binder was as provided for herein. Media was replaced 24 hours later, and the transduced cells were analyzed by flow cytometry on day 7 after transduction to determine CAR transduction. CAR expression was observed in all human and non-human primate PBMCs assessed (FIG. 10A) compared to the binder alone (FIG. 10B) which shows no transduction. Data are also presented as percent CAR expression vs multiplicity of infection (FIG. 10C).

[0386] These examples and embodiments demonstrate that the VSV-G* pseudotyped lentiviruses are able to successfully transduce a variety of human and non-human primate PMBCs.

[0387] Example 7: VSV-G* pseudotyped lentiviruses harboring a CD7 binder exhibit minimal off target transduction.

[0388] The off target transduction of B-cells via VSV-G* pseudotyped lentiviral constructs utilizing CD7 binders as provided for herein was assessed in several B-cell populations. The VSV-G protein utilized was a variant VSV-G protein harboring a mutation to prevent binding of VSV-G to the LDL-R. The variant VSV-G is denoted VSV-G* and corresponds to VSV-G (I182E, T214N, T352A) (e.g. SEQ ID NO: 23), as provided for herein. B-cells cells were maintained in X-Vivo10 media with 5% human serum and 20ng/mL of human IL-2. Non-human primate cells were maintained in RPMI media with 10% FBS, 1% Pen/Strep with 1mM Sodium Pyruvate and 100units/mL hIL-2. Concentrated lentivirus was used to infect the various B-cell populations. A CD7 binder construct was utilized in the present example. The CD7 binder was as provided for herein. Media was replaced 24 hours later, and the transduced cells were analyzed by flow cytometry on day 5 after transduction to determine GFP transduction. As a control, SupT1 cells were also infected with the same lentiviral constructs. None of the B-cell populations assessed demonstrated any significant level of GFP expression (FIG. 11A).

[0389] Lentiviral constructs delivering a CAR20-T2A-GFP transgene were also assessed in a similar manner. VSV-G* pseudotyped lentiviruses utilizing the CD7 binder still exhibited minimal off target transduction of B-cells in all cell populations tested (FIG. 11B). The percent of CAR-T2A-GFP positive cells only exceeded 1% in line DB at the highest lentiviral

concentrations utilized.

[0390] These examples and embodiments demonstrate that the i) VSV-G* pseudotyped lentiviruses utilizing the CD7 binders of the present disclosure have low off target transduction of B-cells and ii) delivery of a CAR20 transgene instead of a GFP transgene does not significantly increase off target transduction in B-cells.

Example 8: VSV-G* pseudotyped lentiviruses with CD20-CAR transgene payload kills CD20 positive lymphoma cells *in vitro*

[0391] The ability of VSV-G* pseudotyped lentiviral constructs with a CD20-CAR transgene to kill CD20 positive lymphoma cells *in vitro* was assessed. The VSV-G* pseudotyped lentiviral constructs utilized CD7 binders as provided for herein. The VSV-G protein utilized was a variant VSV-G protein harboring a mutation to prevent binding of VSV-G to the LDL-R. The variant VSV-G is denoted VSV-G* and corresponds to VSV-G (I182E, T214N, T352A) (e.g. SEQ ID NO: 23), as provided for herein. Human PBMCs were transduced with a lentiviral construct carrying a CD20-CAR (CAR20) transgene. The CAR20 transgenes comprised antigen binding domains of SEQ ID NO: 75 or SEQ ID NO: 76. CD20 positive lymphoma cells were then added to the CAR20 cells at a given effector to target ratio (E:T). All lentiviral constructs tested produced a dose dependent killing of CAR20 positive lymphoma cells (FIG 12)

[0392] These examples and embodiments demonstrate that the VSV-G* pseudotyped lentiviruses utilizing the CD7 binders of the present disclosure not only transduce the appropriate target cells as demonstrated by the previous examples, but also produce robust and dose dependent killing of CD20 positive lymphoma cells *in vivo*.

Example 9: VSV-G* pseudotyped lentiviruses deplete B-cells and eliminate established tumors in mice.

[0393] The ability of VSV-G* pseudotyped lentiviral constructs to deplete B-cell populations *in vivo* was assessed. The VSV-G* pseudotyped lentiviral constructs utilized CD7 binders as provided for herein. The VSV-G protein utilized was a variant VSV-G protein harboring a mutation to prevent binding of VSV-G to the LDL-R. The variant VSV-G is denoted VSV-G* and corresponds to VSV-G (I182E, T214N, T352A) (e.g. SEQ ID NO: 23, SEQ ID NO: 25), as provided for herein. Mice were injected with lentiviral particles expressing a GFP transgene, or a CD20-CAR transgene. The mice utilized were huCD34 NSG mice which have circulating human T and B cells. Mice receiving lentiviral constructs expressing GFP saw no loss

of CD20 (FIG. 13A) or CD19 (FIG. 13B) B cells. Mice receiving the CD20-CAR transgene saw a dramatic and sustained loss of B cells over three weeks as exhibited by a dramatic loss of both CD20 positive B cells (FIG. 13A) and CD19 positive B cells (FIG. 13B).

[0394] To further characterize the *in vivo* efficacy of the viral particles, the ability of VSV-G* pseudotyped lentiviral constructs to prevent tumor formation was assessed. Mice were injected with viral particles expressing a GFP transgene or a CD20-CAR transgene on protocol Day 1. On Day 9, mice were infused with Raji tumor cells and tumor progression was monitored. The experimental protocol illustrated in FIG. 14A. Mice receiving no treatment or lentiviral constructs expressing GFP demonstrated progressive tumor development and spread as determined by IVIS imaging. In contrast, mice receiving lentiviral constructs expressing the CD20-CAR transgene demonstrated no tumor development after Raji cell infusion (FIG. 14B).

[0395] The ability of the viral particles to affect established tumors *in vivo* was also assessed. Mice were infused with Raji tumor cells on protocol Day 0. Mice were humanized via human PBMC engraftment on protocol Day 5, and mice were either not injected with viral particles (control) or injected with viral particles expressing a CD20-CAR transgene on protocol Day 6. Tumor progression was monitored through protocol Day 30. The experimental protocol is illustrated in FIG. 15A. Control and virus treated mice demonstrated comparable tumor load on Day 5 (prior to virus administration), Day 7, and Day 8 as determined by IVIS imaging, indicating that Raji tumors were established in both groups. At Day 12, control mice continued to exhibit increased tumor load, whereas virus treated mice exhibited a dramatic decrease in tumor load. The decreased tumor load was sustained throughout the duration of the experiment (FIG. 15B).

[0396] These examples and embodiments demonstrate that the VSV-G* pseudotyped lentiviral constructs of the present disclosure are able to properly target B-cells *in vivo*, prevent tumor formation *in vivo*, and eliminate established tumors *in vivo*.

Example 10: VSV-G* pseudotyped lentiviruses deplete B-cells in non-human primates.

[0397] The ability of VSV-G* pseudotyped lentiviral constructs to deplete B-cell populations *in vivo* was also assessed in non-human primates (Macaques). The VSV-G* pseudotyped lentiviral constructs utilized CD7 binders as provided for herein. The VSV-G protein utilized was a variant VSV-G protein harboring a mutation to prevent binding of VSV-G to the LDL-R. The variant VSV-G is denoted VSV-G* and corresponds to VSV-G (I182E, T214N,

T352A) (e.g. SEQ ID NO: 23), as provided for herein. Macaques were injected with lentiviral particles on protocol Day 0 and their CD20+ cells were monitored. Injection of the lentiviral particles results in B-cell depletion in 6/6 Macaques tested.

[0398] This specification contains numerous citations to patents, patent applications, and publications. Each is hereby incorporated by reference for all purposes.

[0399] The specification also makes reference to various sequences, such as those provided herein and below.

VSV-G Indiana Full length WT:

MKCLLYLAFLFIGVNCKFTIVFPHNQGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYTGEWVDSQFINGKCSNYICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPKNPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISGTTTERELWDDWAPYEDVEIGPNGVLRRTSSGYKFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESFFGDTGLSKNPIELVEGWFSWKSSIASFFFFIIGLIIGLFLVLRVGIHLICIKLKHTKKRQIYTDIEMNRLGK (SEQ ID NO: 1)

VSV-G Indiana Ectodomain WT:

KFTIVFPHNQGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYTGEWVDSQFINGKCSNYICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPKNPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISGTTTERELWDDWAPYEDVEIGPNGVLRRTSSGYKFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESFFGDTGLSKNPIELVEGWFSWKSSIASFFFFIIGLIIGLFLVLRVGIHLICIKLKHTKKRQIYTDIEMNRLGK (SEQ ID NO: 2)

VSV-G Indiana ectodomain I182A:

KFTIVFPHNQGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYTGEWVDSQFINGKCSNYICPTVHNSTTWHSDYKVKGLCDSNLASMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPKNPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISGTTTERELWDDWAPYEDVEIGPNGVLRRTSSGYKFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESFFGDTGLSKNPIELVEGWFSWKSSIASFFFFIIGLIIGLFLVLRVGIHLICIKLKHTKKRQIYTDIEMNRLGK (SEQ ID NO: 3)

VSV-G Indiana ectodomain I182D:

KFTIVFPHNQGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYTGEWVDSQFINGKCSNYICPTVHNSTTWHSDYKVKGLCDSNLSMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRLPSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPKNPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISGTTTERELWDDWAPYEDVEIGPNGVLRRTSSGYKFPLYMIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESFFGDTGLSKNPIELVEGWFSWKSSIASFFFFIIGLIIGLFLVLRVGIHLICIKLKHTKKRQIYTDIEMNRLGK (SEQ ID NO: 4)

VSV-G Indiana ectodomain I182E:

KFTIVFPHNQGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPKYITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYTGEWVDSQFINGKCS

NYICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRL
PSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPK
NPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISGTTTERELWDDWAPYEDVEIGPNGVLRRTSSGYKFLY
MIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESFFGDTGLSKNP IELVEGWFSWKSSIASFFFIIGLIIGL
FLVLRVGIHLCIKLKHTKKRQIYTDIEMNRLGK (SEQ ID NO: 5)

VSV-G Indiana ectodomain H8A + K47Q:

KFTIVFPANQKGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPQSHKAIQADGWMCHASKWVTTCDFRWYGPY
ITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYTGWVDSQFINGKCS
NYICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRL
PSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPK
NPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISGTTTERELWDDWAPYEDVEIGPNGVLRRTSSGYKFLY
MIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESFFGDTGLSKNP IELVEGWFSWKSSIASFFFIIGLIIGL
FLVLRVGIHLCIKLKHTKKRQIYTDIEMNRLGK (SEQ ID NO: 6)

VSV-G Indiana ectodomain Q10A:

KFTIVFPNNAKGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPY
ITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYTGWVDSQFINGKCS
NYICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRL
PSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPK
NPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISGTTTERELWDDWAPYEDVEIGPNGVLRRTSSGYKFLY
MIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESFFGDTGLSKNP IELVEGWFSWKSSIASFFFIIGLIIGL
FLVLRVGIHLCIKLKHTKKRQIYTDIEMNRLGK (SEQ ID NO: 7)

VSV-G Indiana ectodomain Q10R:

KFTIVFPNNRKNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPY
ITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYTGWVDSQFINGKCS
NYICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRL
PSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPK
NPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISGTTTERELWDDWAPYEDVEIGPNGVLRRTSSGYKFLY
MIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESFFGDTGLSKNP IELVEGWFSWKSSIASFFFIIGLIIGL
FLVLRVGIHLCIKLKHTKKRQIYTDIEMNRLGK (SEQ ID NO: 8)

VSV-G Indiana ectodomain Q10K:

KFTIVFPNNKGNWKNVPSNYHYCPSSSDLNWHNDLIGTALQVKMPKSHKAIQADGWMCHASKWVTTCDFRWYGPY
ITHSIRSFTPSVEQCKESIEQTKQGTWLNPGFPPQSCGYATVTDAAEAVIVQVTPHHVLVDEYTGWVDSQFINGKCS
NYICPTVHNSTTWHSDYKVKGLCDSNLISMDITFFSEDGELSSLGKEGTGFRSNYFAYETGGKACKMQYCKHWGVRL
PSGVWFEMADKDLFAAARFPECPEGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAGLPISPVDLSYLAPK
NPGTGPAFTIINGTLKYFETRYIRVDIAAPILSRMVGMISGTTTERELWDDWAPYEDVEIGPNGVLRRTSSGYKFLY
MIGHGMLDSDLHLSSKAQVFEHPHIQDAASQLPDDESFFGDTGLSKNP IELVEGWFSWKSSIASFFFIIGLIIGL
FLVLRVGIHLCIKLKHTKKRQIYTDIEMNRLGK (SEQ ID NO: 9)

VSV-G New Jersey Full length WT:

MLSYLIFALVSPILGKIEIVFPQHTTGDWKRVPHEYNYCPTSADKNSHGTQTGIPVELTMPKGLTTHQVDGFMCHS
ALWMTTCDFRWYGPYITHSIHNEEPTDYQCLEAIKAYKDGVSFNPFGPPQSCGYGTVDAAEHIVTVTPHSVKVDE
YTGEWIDPHFIGRCKGQICETVHNSTKWFTSSDGESVCSQLFTLVGGTFFSDSEEITSMGLPETGIRSNYFPYVST
EGICKMPFCRKPGYKLNLDLWFQITDPLDKTVRDLPHIKDCDLSSSIVTPGEHATDISLSDVERILDYALCQNTW
SKIEAGEPITPVDSLGLPKNPAGPVFTIINGSLHYFMSKYLRVELESPIPRMEGKVAGTRIVRQLWDQWFFPGE
VEIGPNGVLKTKQGYKFLPHIIGTGEVDNDIKMERIVKHWEHPHIEAAQTFLKKDDTEEVLYYGDGTGVSKNPVOLVE
GWFSGWRSSIMGVLAVIIGFVILIFLIRLIGVLSSLFQRKRRPIYKSDVEMAHFR (SEQ ID NO: 10)

VSV-G New Jersey ectodomain WT:

KIEIVFPQHTTGDWKRVPHEYNYCPTSADKNSHGTQTGIPVELTMPKGLTTHQVDGFMCHSALWMTTCDFRWYGPY

ITHSIHNEEPTDYQCLEAIKAYKDGVSFNPGFPPQSCGYGTVDAAEHIVTVTPHSVKVDEYTG EWIDPHFIGGRCK
GQICETVHNSTKWFTSSDGE SVCSQLFTLVGGTFFSDSEEITSMGLPETGIRSNYFPYVSTEGICKMPFCRKP GYKL
KNDLWFQITDPDLDKTVRDLPHIKDCDLSSSIVTPGEHATDISLISDVERILDYALCQNTWSKIEAGEPITPVDLSY
LGPKNPGAGPVFTIINGS LHYFMSKYLRVELESPVIPRMEGKVAGTRIVRQLWDQWFPFGEVEIGPNGVLKTKQGYK
FPLHIIGTGEVDNDIKMERIVKHWEHPHIEAAQTFLKKDDTEEVLYYGD TGVSKNPVELVEGWFSGWRSSIMGVLAV
IIGFVILIFLIRLIGVLSLFRQKRRPIYKSDVEMAFHR (SEQ ID NO: 11)

VSV-G Marraba Full length WT:

MLRLFLFCFLALGAHSKFTIVFP HHQKGNWKNVPSTYHYCPSSSDQNWHDLTGVSLHV KIPKSHKAIQADGWMCHA
AKWVTTCDFRWYGPKYITHSIHSMSP TLEQCKTSIEQTKQGVWINPGFPPQSCGYATVTD AEV VVVQATPHHVLVDE
YTGEWIDSQVLGGKCSKEVCQTVHNSTVWHADYKITGLCESNLASVDITFFSEDGQKTS LGKPNTGFRSNHFAYESG
EKACRMQYCTQWGIRLPSGVWFELVDKDLFQA AKLPECPRGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIR
AKLPVSPVDLSYLAPKNPGSGPAFTIINGTLKYFETRYIRVDISNPIIPH MVGTMSGTTTERELWNDWYPYEDVEIG
PNGVLKTPGTGKFPLYMIGHGMLDSDLHKSSQAQVFEH PHAKDAASQLPDDETLFFGDTGLSKNPVELVEGWFSWK
STLASFFLIIGLVALIFIIRIIVAIRYKYKGRKTQKIYNDVEMSRLGNK (SEQ ID NO: 12)

VSV-G Marraba ectodomain WT:

KFTIVFP HHQKGNWKNVPSTYHYCPSSSDQNWHDLTGVSLHV KIPKSHKAIQADGWMCHAAKWVTTCDFRWYGPKY
ITHSIHSMSP TLEQCKTSIEQTKQGVWINPGFPPQSCGYATVTD AEV VVVQATPHHVLVDEYTG EWIDSQVLGGKCS
KEVCQTVHNSTVWHADYKITGLCESNLASVDITFFSEDGQKTS LGKPNTGFRSNHFAYESGEKACRMQYCTQWGIRL
PSGVWFELVDKDLFQA AKLPECPRGSSISAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAKLPVSPVDLSYLAPK
NPGSGPAFTIINGTLKYFETRYIRVDISNPIIPH MVGTMSGTTTERELWNDWYPYEDVEIGPNGVLKTPGTGKFPLY
MIGHGMLDSDLHKSSQAQVFEH PHAKDAASQLPDDETLFFGDTGLSKNPVELVEGWFSWKSTLASFFLIIGLVAL
IFIIRIIVAIRYKYKGRKTQKIYNDVEMSRLGNK (SEQ ID NO: 13)

VSV-G Carajas Full length WT:

MKMKMVIAGLILCIGILPAIGKITISFPQSLKGDWRPV PKGYNYCPTSADKNLHGD LIDIGLRLRAPKSFKGISADG
WMCHAARWITTCDFRWYGPKYITHSIHSFRPSNDQCKE AIRLTNEGNWINPGFPPQSCGYASVTDSESVVTVTKHQ
VLVDEYSGSWIDSQFP GGSTSPICDTVHNSTLWHADHTLDSICDQEFVAMDAVLFTESGKFEEFGKPN SGIRSNYF
PYESLKDVCQMDFCKRKGFKLPSGVWFEIEDAEKSHKAQVELKIKRCPHGAVISAPNQNAADINLIMDVERILDYSL
CQATWSKIQNKEALTPIDISYLGPKNPGPGPAFTIINGTLHYFNTRYIRVDIAGPVTKEITGFVSGTSTSRVLWDQW
FPYGENSIGPNGLLKTASGYKPLFMVGTGVLDADIHKLGEATVIEHPHAKEAQKVDDSEVIFFGDTGVSKNPVEV
VEGWFSGWRSSSLMSIFGIILLIVCLVLIVRILIALKYCCVRHKKRTIYKEDLEMGRIPRA (SEQ ID NO: 14)

VSV-G Carajas ectodomain WT:

KITISFPQSLKGDWRPV PKGYNYCPTSADKNLHGD LIDIGLRLRAPKSFKGISADGWMCHAARWITTCDFRWYGPKY
ITHSIHSFRPSNDQCKE AIRLTNEGNWINPGFPPQSCGYASVTDSESVVTVTKHQVLVDEYSGSWIDSQFP GGST
SPICDTVHNSTLWHADHTLDSICDQEFVAMDAVLFTESGKFEEFGKPN SGIRSNYFPYESLKDVCQMDFCKRKGFKL
PSGVWFEIEDAEKSHKAQVELKIKRCPHGAVISAPNQNAADINLIMDVERILDYSLCQATWSKIQNKEALTPIDISY
LGPKNPGPGPAFTIINGTLHYFNTRYIRVDIAGPVTKEITGFVSGTSTSRVLWDQWFPYGENSIGPNGLLKTASGYK
YPLFMVGTGVLDADIHKLGEATVIEHPHAKEAQKVDDSEVIFFGDTGVSKNPVEVVEGWFSGWRSSSLMSIFGIILL
IVCLVLIVRILIALKYCCVRHKKRTIYKEDLEMGRIPRA (SEQ ID NO: 15)

VSV-G Alagoa Full length WT:

MTPAFILCMLLAGSSWAKFTIVFPQS QKGDWVDVPPNYRYCPSSADQNWHDLLGVNIRAKMPKVHKA IKADGWMCH
AAKWVTTCDFRWYGPKYITHSIHSFIPTKAQCEESIKQTKEGVWINPGFPPKNCGYASVSDAESIIVQATAHSVMID
EYSGDWLDSQFPTGRCTGSTCETIHNSTLWYADYQVTGLCDSALVSTEVTFYSEDGLMTSIGRQNTGYRSNYFPY EK
GAAACRMKYCTHEGIRLPSGVWFEMVDKELLESVQMPECPAGLTISAPTQTSVDVSLILDVERMLDYS LCQETWSKV
HSGLPISPVDLGYIAPKNPGAGPAFTIIVNGTLKYFDTRYLRIDIEGPVLKMTGKVS GTPTKRELWTEWFFPYDDVEI
GPNGVLKTPEGYKFP LYMIGHGLDSDLQKTSQA EVFHHQPQIAEAVQKLPDDETLFFGDTGISKNPVEVIEGWFSNW
RSSVMAIVFAILLLVITVLMVRLCVAFRHFCCQKRHKIYNDLEMNQLRR (SEQ ID NO: 16)

VSV-G Alagoa ectodomain WT:

KFTIVFPQSQKGDWKDVPPNYRYCPSSADQNWHGDLGVDNIRAKMPKVHKAIKADGWMCHAAKWVTTCDYRWYGPQY
 ITHSIHSFIPTKAQCEESIKQTKEGVWINPGFPPKNCGYASVSDAESIIVQATAHSVMIDEYSQDWLDSQFPTGRCT
 GSTCETIHNSTLWYADYQVTGLCDSALVSTEVTFYSEDGLMTSIGRQNTGYRSNYFFYEKGAAACRMKYCTHEGIRL
 PSGVWFEMVDKELLESVQMPCEPAGLTISAPTQTSVDVSLILDVERMLDYSLCQETWSKVHSGPLISPVDLGYIAPK
 NPGAGPAFTIIVNGTLKYFDTRYLRIDIEGPVLKMTGKVSPTKRELWTEWFFPYDDVEIGPNGVLKTPEGYKFPLY
 MIGHGLDSDLQKTSQAEVFFHPQIAEAVQKLPDDETLFFGDTGISKNPVEVIEGWFSNWRSSVMAIVFAILLVIT
 VLMVRLCVAFRHFCCQKRHKIYNLDLEMNQLRR (SEQ ID NO: 17)

VSV-G Cocal Full length WT:

MNLLLLTFIVLPLCSHAKFSIVFPQSQKGNWKNVPSSYHYCPSSSDQNWHNDLLGITMKVKMPKTHKAIQADGWMCH
 AAKWITTCDFRWYGPKYITHSIHSIQPTSEQCKESIKQTKQGTWMSPGFPQNCGYATVTDVAVVVQATPHHVLVD
 EYTGEWIDSQFPNGKCEETECETVHNSTVWYSYKVTGLCDATLVDTEITFFSEDGKKEISIGKPNTGYRSNYFAYEK
 GDKVCKMNYCKHAGVRLPSGVWFEFVDQDVYAAAKLPECPVGATISAPTQTSVDVSLILDVERILDYSLCQETWSKI
 RSKQPVSPVDSLAPKNPGTGPAFTIINGTLKYFETRYIRIDIDNPIISKMVGKISGSQTERELWTEWFFPYEGVEI
 GPNGILKTPTGYKFLPMIGHGMLDSDLHKTSQAEVFEHPLAEAPKQLPEEETLFFGDTGISKNPVELIEGWFSW
 KSTVVTFFFAIGVFILLYVVARIVIAVRYRYQGSNNKRIYNDIEMSRFRK (SEQ ID NO: 18)

VSV-G Cocal ectodomain WT:

KFSIVFPQSQKGNWKNVPSSYHYCPSSSDQNWHNDLLGITMKVKMPKTHKAIQADGWMCHAAKWITTCDFRWYGPKY
 ITHSIHSIQPTSEQCKESIKQTKQGTWMSPGFPQNCGYATVTDVAVVVQATPHHVLVDEYTGEWIDSQFPNGKCE
 TEECETVHNSTVWYSYKVTGLCDATLVDTEITFFSEDGKKEISIGKPNTGYRSNYFAYEKGDVKCKMNYCKHAGVRL
 PSGVWFEFVDQDVYAAAKLPECPVGATISAPTQTSVDVSLILDVERILDYSLCQETWSKIRSKQPVSPVDSLAPK
 NPGTGPAFTIINGTLKYFETRYIRIDIDNPIISKMVGKISGSQTERELWTEWFFPYEGVEIGPNGILKTPTGYKFLP
 MIGHGMLDSDLHKTSQAEVFEHPLAEAPKQLPEEETLFFGDTGISKNPVELIEGWFSWVSTVVTFFFAIGVFILL
 YVVARIVIAVRYRYQGSNNKRIYNDIEMSRFRK (SEQ ID NO: 19)

VSV-G Morreton Full length WT:

MLVLYLLLSLLALGAQCKFTIVFPHNQKGNWKNVPANYQYCPSSSDLNWHNGLIGTSLQVKMPKSHKAIQADGWMCH
 AAKWVTTCDFRWYGPKYVTHSIKSMIPTVDQCKESIAQTKQGTWLNPGFPPQSCGYASVTDAAEAVIVKATPHQVLVD
 EYTGEWVDSQFPTGKCNKDICTVHNSTTWHSDYKVTGLCDANLISMDITFFSEDGKLTSLGKEGTGFRSNYFAYEN
 GDKACRMQYCKHWGVRPLPSGVWFEMADKDIYNDAKFPDCPEGSSIAAPSQTSVDVSLIQDVERILDYSLCQETWSKI
 RAHLPISPVDLSYLSKPNPGTGPAFTIINGTLKYFETRYIRVDIAGPIIPQMRGVISGTTTERELWTDWYPYEDVEI
 GPNGVLKTATGYKFLPYMIGHGMLDSDLHISSKAQVFEHPHIQDAASQLPDDDETLFFGDTGLSKNPIELVEGWFSGW
 KSTIASFFFIIGLVIGLYLVLRIALCIKCRVQEKRPKIYTDVEMNRLDR (SEQ ID NO: 20)

VSV-G Morreton ectodomain WT:

KFTIVFPHNQKGNWKNVPANYQYCPSSSDLNWHNGLIGTSLQVKMPKSHKAIQADGWMCHAAKWVTTCDFRWYGPKY
 VTHSIKSMIPTVDQCKESIAQTKQGTWLNPGFPPQSCGYASVTDAAEAVIVKATPHQVLVDEYTGEWVDSQFPTGKCN
 KDICTVHNSTTWHSDYKVTGLCDANLISMDITFFSEDGKLTSLGKEGTGFRSNYFAYENGDKACRMQYCKHWGVRPL
 PSGVWFEMADKDIYNDAKFPDCPEGSSIAAPSQTSVDVSLIQDVERILDYSLCQETWSKIRAHLPISPVDLSYLSK
 NPGTGPAFTIINGTLKYFETRYIRVDIAGPIIPQMRGVISGTTTERELWTDWYPYEDVEIGPNGVLKTATGYKFLPY
 MIGHGMLDSDLHISSKAQVFEHPHIQDAASQLPDDDETLFFGDTGLSKNPIELVEGWFSGWKSTIASFFFIIGLVIGL
 YLVLRIALCIKCRVQEKRPKIYTDVEMNRLDR (SEQ ID NO: 21)

Fc IgG1 (Accession No. P01857)

ASTKGPSVFPFLAPSSKSTSGGTAAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVTVTPSSSLG
 TQTYICNVNHKPSNTKVDKKVEPKSCDKHTCTPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVDVSHEDP
 EVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQV

YTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFC
SVMHEALHNHYTQKSLSLSPGK (SEQ ID NO: 26)

Fc IgG2 (Accession No. AAN76043)

STKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTISWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGT
QTYTCNVDPKPSNTKVDKVERKCCVECPPEPPVAGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVQFN
WYVDGVEVHNAKTKPREEQFNSTFRVVSFLTIVHQQDWLNGKEYKCKVSNKGLPAPIEKTISKTKGQPREPQVYTLPP
SREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPMLDSDGSFFLYSKLTVDKSRWQQGNVFC
SVMHEALHNHYTQKSLSLSPGK (SEQ ID NO: 27)

Fc IgG4 (Accession No. AAB59394)

STKGPSVFPLAPCSRSTSESTAALGCLVKDYFPEPVTISWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGT
KTYTCNVDPKPSNTKVDKRVESKYGPPCPSCPAPEFLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVQF
NWKVDGVEVHNAKTKPREEQFNSTYRVVSFLTIVHQQDWLNGKEYKCKVSNKGLPSSIEKTISKAKGQPREPQVYTLPP
PSQEEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQEGNVFC
SVMHEALHNHYTQKSLSLSPGK (SEQ ID NO: 28)

What is Claimed:

1. A VSV-G polypeptide comprising a mutation that corresponds to a mutation at position 182 of SEQ ID NO: 2.
2. The VSV-G polypeptide of claim 1, wherein the protein comprises an amino acid sequence of SEQ ID NO: 2 with a mutation at position 182 and has at least 95% identity to SEQ ID NO: 2.
3. The VSV-G polypeptide of claims 1 or 2, wherein the polypeptide comprises a I182E or I182D mutation as compared to SEQ ID NO: 2.
4. The VSV-G polypeptide of any one of claims 1-3, wherein the polypeptide comprises an amino acid sequence of SEQ ID NO: 1 with a mutation at position 198 and has at least 70% identity to SEQ ID NO: 2.
5. The VSV-G polypeptide of any one of claims 1-4, wherein the polypeptide comprises a mutation that corresponds to I182D or I182E as compared to a sequence of SEQ ID NO: 2.
6. The VSV-G polypeptide of any one of claims 1-5, wherein the polypeptide comprises an amino acid sequence at least 95% identical to a sequence of SEQ ID NO: 4.
7. The VSV-G polypeptide of any one of claims 1-5, wherein the polypeptide comprises an amino acid sequence of SEQ ID NO: 4.
8. The VSV-polypeptide of any one of claims 1-5, wherein the polypeptide comprises an amino acid sequence at least 95% identical to a sequence of SEQ ID NO: 5.
9. The VSV-G polypeptide of any one of claims 1-5, wherein the polypeptide comprises an amino acid sequence of SEQ ID NO: 5.
10. The VSV-G polypeptide of any one of claims 1-9, further comprising a mutation in the VSV-G protein that corresponds to a position of 8, 10, 47, 209 and/or 354 as compared to SEQ ID

NO: 2.

11. The VSV-G polypeptide of any one of claims 1-10, further comprising a mutation that corresponds to position 8 in SEQ ID NO: 2, wherein the mutation is any amino acid different from the amino acid indicated at that position in SEQ ID NO: 2, except Y.

12. The VSV-G polypeptide of any one of claims 1-11, further comprising a mutation that corresponds to position 209 in SEQ ID NO: 2, wherein the mutation is any amino acid different from the amino acid indicated at that position in SEQ ID NO: 2, except H.

13. The VSV-G polypeptide of any one of claims 1-12, further comprising a mutation that corresponds to position 47 in SEQ ID NO: 2, wherein the mutation is any amino acid different from the amino acid indicated at that position in SEQ ID NO: 2, except K or R.

14. The VSV-G polypeptide of any one of claims 1-13, further comprising a mutation that corresponds to position 354 in SEQ ID NO: 2, wherein the mutation is any amino acid different from the amino acid indicated at that position in SEQ ID NO: 2, except K or R.

15. The VSV-G polypeptide of any one of claims 1-14, further comprising a mutation that corresponds to position 10 in SEQ ID NO: 2, wherein the mutation is any amino acid different from the amino acid indicated at that position in SEQ ID NO: 2, except Q or N.

16. The VSV-G polypeptide of any one of claims 1-15, wherein the protein comprises a substitution at position 47 or at position 354, or at both positions 47 and 354, wherein each position is, independently, substituted by A, G, F, Q, or N.

17. The VSV-G polypeptide of any one of claims 1-16, wherein the protein comprises a substitution at position 8, wherein the substitution is H8A, H8I, H8V, H8L, and the like.

18. The VSV-G polypeptide of any one of claims 1-17, wherein the protein comprises a substitution at position 47, wherein the substitution is K47Q or K47N.

19. The VSV-G polypeptide of any one of claims 1-18, wherein the protein comprises a substitution H8A and/or K47Q mutation.
20. The VSV-G polypeptide of any one of claims 1-19, wherein the protein comprises a Q10A, Q10R, or Q10K substitution.
21. The VSV-G polypeptide of any one of claims 1-20, further comprising a mutation that corresponds to a mutation at positions 214 and/or 352 of SEQ ID NO: 2.
22. The VSV-G polypeptide of claim 21, wherein the polypeptide comprises a T214N and/or T352A mutation.
23. A VSV-G polypeptide comprising a substitution at positions I182 and at least one of T214, and T352 of SEQ ID NO: 2.
24. The VSV-G polypeptide of claim 23, wherein the polypeptide comprises substitutions at positions I182, T214, and T352 of SEQ ID NO: 2.
25. The VSV-G polypeptide of claims 23 or 24, wherein the substitution at position 182 is I182D or I182E, the substitution at position 214 is T214N, and the substitution at position 352 is T352A.
26. The VSV-G polypeptide of any one of claims 23-25, wherein the polypeptide comprises a sequence of SEQ ID NO: 23, SEQ ID NO: 22, SEQ ID NO: 24 or SEQ ID NO: 25.
27. The VSV-G polypeptide of claim 1, wherein the polypeptide comprises a sequence of SEQ ID NO: 23.
28. The VSV-G polypeptide of any one of claims 23-25, wherein the polypeptide comprises a sequence of SEQ ID NO: 22.

29. The VSV-G polypeptide of any one of claims 23-25, wherein the polypeptide comprises a sequence of SEQ ID NO: 24.
30. The VSV-G polypeptide of any one of claims 23-25, wherein the polypeptide comprises a sequence of SEQ ID NO: 25.
31. A nucleic acid molecule encoding the VSV-G polypeptide of any one of claims 1-30.
32. A vector comprising the nucleic acid molecule of claim 31.
33. A plasmid comprising the nucleic acid molecule of claim 31.
34. A viral particle comprising the polypeptide of any one of claims 1-30.
35. The viral particle of claim 34, further comprising a targeting moiety.
36. The viral particle of claims 34 or 35, wherein the viral particle is a pseudotyped lentivirus.
37. The viral particle of any one of claims 34-36, wherein the viral particle further comprises a nucleic acid molecule encoding a heterologous molecule of interest.
38. The viral particle of claim 37, wherein the heterologous molecule of interest is an siRNA, an shRNA, a non-coding RNA (e.g. a guide RNA for a CRISPR system), a peptide, a polypeptide, a protein, a viral payload, a viral genome, or a combination thereof.
39. The viral particle of claims 37 or 38, wherein the heterologous molecule of interest is a chimeric antigen receptor ("CAR").
40. The viral particle of claim 39, wherein the chimeric antigen receptor comprises antigen binding domain that binds to CD20.

41. The viral particle of any one of claims 34-40, wherein the targeting moiety binds to an immune cell, T cell, B cell, NK cell, dendritic cell, neutrophils, macrophages, cancer cell, CD3+ T cell, CD4+ T cell, CD7+ T cell, CD8+ T cell, CD19+ B cell, CD19+ cancer cell, CD20+ B cell, CD20+ cancer cell, CD30+ lung epithelial cell, CD34+ haematopoietic stem cell, CD105+ endothelial cell, CD105+ haematopoietic stem cell, CD117+ haematopoietic stem cell, CD133+ cancer cell, EpCAM+ cancer cell, GluA2+ neuron, GluA4+ neuron, Haematopoietic stem cell, Hepatocyte, Her2/Neu+ cancer cell, NKG2D+ natural killer cell, SLC1A3+ astrocyte, or SLC7A10+ adipocyte.

42. The viral particle of any one of claims 34-41, wherein the targeting moiety binds to CD7, CD8, cKit (CD117), CD4, CD3, CD5, CD6, CD2, TCR alpha, TCR beta, TCR gamma, TCR delta, CD10, CD34, CD110, CD33, CD14, CD68, CCR7, CD62L, CD25, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, or CXCR3, A glycosylated CD43 epitope expressed on acute leukemia or lymphoma but not on hematopoietic progenitors, A glycosylated CD43 epitope expressed on non-hematopoietic cancers, A kinase anchor protein 4 (AKAP-4), Adrenoceptor beta 3 (ADRB3), AFP, Anaplastic lymphoma kinase (ALK), Androgen receptor, Angiopoietin-binding cell surface receptor 2 (Tie 2), Auto antibody to desmoglein 1 (Dsg1), Auto antibody to desmoglein 3 (Dsg3), B7H3 (CD276), Biotin, Bone marrow stromal cell antigen 2 (BST2), BST1/CD157, Cancer/testis antigen 1 (NY-ESO-1), Cancer/testis antigen 2 (LAGE-la), Carbonic anhydrase IX (CA1X), Carcinoembryonic antigen (CEA), CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS or Brother of the Regulator of Imprinted Sites), CCR4, CD5, CD19, CD20, CD22, CD24, CD30, CD32 (FCGR2A), CD33, CD34, CD38, CD44v6, CD72, CD79a, CD79b, CD97, CD99, CD123, CD171, CD179a, CD179b-IGLII, CD200R, CD276/B7H3, CD300 molecule-like family member f (CD300LF), CDH1-CD324, CDH6, CDH17, CDH19, Chromosome X open reading frame 61 (CXORF61), Claudin 6 (CLDN6), Claudin18.2 (CLD18A2 or CLDN18A.2), CMV pp65, C-MYC epitope Tag, Cripto, CS1 (also referred to as CD2 subset 1 or CRACC or SLAMF7 or CD319 or 19A24), CSF2RA (GM-CSFR-alpha), C-type lectin domain family 12 member A (CLEC12A), C-type lectin-like molecule-1 (CLL-1 or CLECL1), Cyclin B1, Cytochrome P450 IB 1 (CYP1B 1), DLL3, EBV-EBNA3c, EGF-bke module- containing mucin-like hormone receptor-like 2 (EMR2), Elongation factor 2 mutated (ELF2M), Ephrin B2, Ephrin type-A receptor 2 (EphA2), Epidermal

growth factor receptor (EGFR), Epidermal growth factor receptor variant III (EGFRviii), Epithelial cell adhesion molecule (EPCAM), ERG, ETS translocation-variant gene 6 located on chromosome 12p (ETV6-AML), Fc fragment of IgA receptor (FCAR or CD89), Fc receptor-like 5 (FCRL5), Fibroblast activation protein alpha (FAP), FITC, Fms Like Tyrosine Kinase 3 (FLT3), Folate receptor alpha (FRa or FR1), Folate receptor beta (FRb), Follicle stimulating hormone receptor (FSHR), Fos-related antigen 1, Fucosyl-GM1, G protein coupled receptor class C group 5 member D (GPRC5D), G protein-coupled receptor 20 (GPR20), GAD, Ganglioside G2 (GD2), Ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlc(1-1)Cer), Ganglioside GM3 (aNeu5Ac(2-3)bDGalp(1-4)bDGlc(1-1)Cer), GD3, GFRalpha4, Glycoprotein 100 (gp100), Glypican-3 (GPC3), Gonadotropin Hormone receptor (CGHR or GR), GpA33, GpNMB, GPRC5D, Guanylyl cyclase C (GCC), Heat shock protein 70-2 mutated (mut hsp70-2), Hepatitis A virus cellular receptor 1 (HAVCR1), Hexasaccharide portion of globoH glycosphingolipid (GloboH), High molecular weight-melanoma associated antigen (HMWMAA), HIV1 envelope glycoprotein, HLA, HLA-DOA, HLA-A, HLA-A2, HLA-B, HLA-C, HLA-DM, HLA-DOB, HLA-DP, HLA-DQ, HLA-DR, HLA-G, HTLV1-Tax, Human papilloma virus E6 (HPV E6), Human papilloma virus E7 (HPV E7), Human Telomerase reverse transcriptase (hTERT), IgE, IL13Ra2, IL13Ra, Immunoglobulin lambda-like polypeptide 1 (IGLL1), Influenza A hemagglutinin (HA), Insulin-like growth factor 1 receptor (IGF-I receptor), Interleukin 11 receptor alpha (IL-11Ra), Interleukin-13 receptor subunit alpha-2 (IL-13Ra2 or CD213A2), Intestinal carboxyl esterase, KIT (CD117), KSHV K8.1, KSHV-gH, LAMP1, Legumain, Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2), Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1), Leutenizing hormone receptor (LHR), Lewis(Y) antigen, Lewis Ag, Liv1, Locus K 9 (LY6K), Low conductance chloride channel, Lymphocyte antigen 6 complex, Lymphocyte antigen 75 (LY75), Lymphocyte-specific protein tyrosine kinase (LCK), Mammary gland differentiation antigen (NY-BR-1), Melanoma antigen recognized by T cells 1 (MelanA or MART1), Melanoma-associated antigen 1 (MAGE-A1), Melanoma cancer testis antigen-1 (MAD-CT-1), Melanoma cancer testis antigen-2 (MAD-CT-2), Melanoma inhibitor of apoptosis (ML-IAP), Mesothelin, MPL, Mucin 1 cell surface associated (MUC1), N-Acetyl glucosaminyl-transferase V (NA17), Nectin-4, Neural cell adhesion molecule (NCAM), NKG2D, NYBR1, O-acetyl-GD2 ganglioside (OAcGD2), Olfactory receptor 51E2 (OR51E2), Oncogene fusion protein consisting of breakpoint cluster region (BCR) and Abelson murine

leukemia viral oncogene homolog 1 (Abl) (bcr-abl), P53 mutant, Paired box protein Pax-3 (PAX3), Paired box protein Pax-5 (PAX5), Pannexin 3 (PANX3), PDL1, P-glycoprotein, Placenta-specific 1 (PLAC1), Platelet-derived growth factor receptor beta (PDGFR-beta), Polysialic acid, Proacrosin binding protein sp32 (OY-TES1), Prostase, Prostate carcinoma tumor antigen-1 (PCT A-1 or Galectin 8), Prostate stem cell antigen (PSCA), Prostate-specific membrane antigen (PSMA), Prostatic acid phosphatase (PAP), Prostein, Protease Serine 21 (Testisin or PRSS21), Proteasome (Prosome Macropain) Subunit Beta Type 9 (LMP2), PTK7, Ras G12V, Ras Homolog Family Member C (RhoC), Rat sarcoma (Ras) mutant, Receptor for Advanced Glycation Endproducts (RAGE-1), Receptor tyrosine kinase-like orphan receptor 1 (ROR1), Receptor tyrosine-protein kinase ERBB2 or Her-22/neu, Renal ubiquitous 1 (RU1), Renal ubiquitous 2 (RU2), Sarcoma translocation breakpoints, Serine 2 (TMPRSS2) ETS fusion gene, Sialyl Lewis adhesion molecule (sLe), SLAMF4, SLAMF6, Slea (CA19.9 or Sialyl Lewis Antigen), Sperm protein 17 (SPA17), Squamous Cell Carcinoma Antigen Recognized By T Cells 3 (SART3), Stage-specific embryonic antigen-4 (SSEA-4), STEAP1, Survivin, Synovial sarcoma X breakpoint 2 (SSX2), TCR Gamma Alternate Reading Frame Protein (TARP), TCR-beta1 chain, TCR-beta2 chain, TCR-delta chain, TCR-gamma chain, TCRgamma-delta, Telomerase, TGFbetaR2, The antigen recognized by TNT antibody, Thyroid stimulating hormone receptor (TSHR), Timl/HVCR1, Tissue Factor 1 (TF1), Tn ag, Tn antigen ((Tn Ag) or (GalNAc-Ser/Thr)), TNF receptor family member B cell maturation (BCMA), Transglutaminase 5 (TGS5), Transmembrane protease, TROP2, Tumor endothelial marker 1 (TEM1/CD248), Tumor endothelial marker 7-related (TEM7R), Tumor protein p53 (p53), Tumor-associated glycoprotein 72 (TAG72), Tyrosinase, Tyrosinase-related protein 2 (TRP-2), Uroplakin 2 (UPK2), Vascular endothelial growth factor receptor 2 (VEGFR2), V-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN), Wilms tumor protein (WT1), or X Antigen Family Member 1A (XAGE1).

43. The viral particle of claim 42, wherein the targeting moiety binds to CD7.

44. The viral particle of claim 43, wherein the targeting moiety that binds to CD7 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 52, having at least 95% sequence identity to SEQ ID NO: 52, having at least 99% sequence identity

to SEQ ID NO: 52, or a sequence as set forth in SEQ ID NO: 52.

45. The viral particle of claim 42, wherein the targeting moiety binds to CD8.

46. The viral particle of claim 45, wherein the targeting moiety that binds to CD8 comprises a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 69, at least 95% sequence identity to SEQ ID NO: 69, at least 99% sequence identity to SEQ ID NO: 69, or a sequence as set forth in SEQ ID NO: 69.

47. The viral particle of any one of claims 34-46, wherein the targeting moiety is an antibody, a scFv antibody, an antigen binding domain, an ankyrin repeat (e.g., DARPIN), a VHH domain antibody, a nanobody, single domain antibody, a FN3 domain, or any combination thereof.

48. The viral particle of any one of claims 34-47, wherein the targeting moiety is attached to the viral surface through

an IgG Fc stalk;

an envelope glycoprotein G or H of a virus of the Paramyxoviridae family, such as a morbillivirus, such as Measles virus, or a henipavirus, such as Nipah virus, Cedar virus, or Hendra virus;

a glycoprotein of a virus of the Rhabdoviridae family, such as a vesicular stomatitis New Jersey virus, a vesicular stomatitis Indiana virus, a vesicular stomatitis Alagoas virus, a vesicular stomatitis Maraba virus, a vesicular stomatitis Carajas virus, Parainfluenza virus, Spodoptera frugiperda rhabdovirus isolate Sf G, Drosophila obscura sigmavirus 10A, Wuhan insect virus 7, Perch virus, or Spring viremia of carp virus;

a glycoprotein of a virus of the Filoviridae family, such as Ebola virus; or

a glycoprotein of a virus of the Arenaviridae family, such as Machupo virus.

49. The viral particle of claim 48, wherein the stalk comprises a CD8 or CD28 transmembrane domain.

50. A method of delivering a heterologous molecule to a target cell, the method comprising contacting the cell with a viral vector comprising:

- a) the VSV-G protein of any one of claims 1-30;
- b) a targeting moiety that binds to the target cell; and
- c) a nucleic acid molecule encoding the heterologous molecule.

51. The method of claim 45, wherein the target cell is an immune cell, T cell, B cell, NK cell, dendritic cell, neutrophils, macrophages, a cancer cell, CD3+ T cell, CD4+ T cell, CD7+ T cell, CD8+ T cell, CD19+ B cell, CD19+ cancer cell, CD20+ B cell, CD20+ cancer cell, CD30+ lung epithelial cell, CD34+ haematopoietic stem cell, CD105+ endothelial cell, CD105+ haematopoietic stem cell, CD117+ haematopoietic stem cell, CD133+ cancer cell, EpCAM+ cancer cell, GluA2+ neuron, GluA4+ neuron, Haematopoietic stem cell, Hepatocyte, Her2/Neu+ cancer cell, NKG2D+ natural killer cell, SLC1A3+ astrocyte, or SLC7A10+ adipocyte.

52. The method of claims 50 or 51, wherein the targeting moiety binds to CD7, CD8, cKit (CD117), CD4, CD3, CD5, CD6, CD2, TCR alpha, TCR beta, TCR gamma, TCR delta, CD10, CD34, CD110, CD33, CD14, CD68, CCR7, CD62L, CD25, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, or CXCR3, A glycosylated CD43 epitope expressed on acute leukemia or lymphoma but not on hematopoietic progenitors, A glycosylated CD43 epitope expressed on non-hematopoietic cancers, A kinase anchor protein 4 (AKAP-4), Adrenoceptor beta 3 (ADRB3), AFP, Anaplastic lymphoma kinase (ALK), Androgen receptor, Angiopoietin-binding cell surface receptor 2 (Tie 2), Auto antibody to desmoglein 1 (Dsg1), Auto antibody to desmoglein 3 (Dsg3), B7H3 (CD276), Biotin, Bone marrow stromal cell antigen 2 (BST2), BST1/CD157, Cancer/testis antigen 1 (NY-ESO-1), Cancer/testis antigen 2 (LAGE-la), Carbonic anhydrase IX (CAIX), Carcinoembryonic antigen (CEA), CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS or Brother of the Regulator of Imprinted Sites), CCR4, CD5, CD19, CD20, CD22, CD24, CD30, CD32 (FCGR2A), CD33, CD34, CD38, CD44v6, CD72, CD79a, CD79b, CD97, CD99, CD123, CD171, CD179a, CD179b-IGL1l, CD200R, CD276/B7H3, CD300 molecule-like family member f (CD300LF), CDH1-CD324, CDH6, CDH17, CDH19, Chromosome X open reading frame 61 (CXORF61), Claudin 6 (CLDN6), Claudin18.2 (CLD18A2 or CLDN18A.2), CMV pp65, C-MYC epitope Tag, Cripto, CS1 (also referred to as CD2 subset 1 or CRACC or SLAMF7 or CD319 or 19A24),

CSF2RA (GM-CSFR-alpha), C-type lectin domain family 12 member A (CLEC12A), C-type lectin-like molecule-1 (CLL-1 or CLECL1), Cyclin B1, Cytochrome P450 IB 1 (CYP1B 1), DLL3, EBV-EBNA3c, EGF-bke module- containing mucin-like hormone receptor-like 2 (EMR2), Elongation factor 2 mutated (ELF2M), Ephrin B2, Ephrin type-A receptor 2 (EphA2), Epidermal growth factor receptor (EGFR), Epidermal growth factor receptor variant III (EGFRviii), Epithelial cell adhesion molecule (EPCAM), ERG, ETS translocation-variant gene 6 located on chromosome 12p (ETV6-AML), Fc fragment of IgA receptor (FCAR or CD89), Fc receptor-like 5 (FCRL5), Fibroblast activation protein alpha (FAP), FITC, Fms Like Tyrosine Kinase 3 (FLT3), Folate receptor alpha (FRa or FR1), Folate receptor beta (FRb), Follicle stimulating hormone receptor (FSHR), Fos-related antigen 1, Fucosyl-GM1, G protein coupled receptor class C group 5 member D (GPRC5D), G protein-coupled receptor 20 (GPR20), GAD, Ganglioside G2 (GD2) , Ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlc(1-l)Cer), Ganglioside GM3 (aNeu5Ac(2-3)bDCIalp(1- 4)bDGlc(1-l)Cer), GD3, GFRalpha4, Glycoprotein 100 (gp100), Glypican-3 (GPC3), Gonadotropin Hormone receptor (CGHR or GR), GpA33, GpNMB, GPRC5D, Guanylyl cyclase C (GCC), Heat shock protein 70-2 mutated (mut hsp70-2), Hepatitis A virus cellular receptor 1 (HAVCR1), Hexasaccharide portion of globoH glycosphingolipid (GloboH), High molecular weight-melanoma associated antigen (HMWMAA), HIV1 envelope glycoprotein, HLA, HLA-DOA, HLA-A, HLA-A2, HLA-B, HLA-C, HLA-DM, HLA-DOB, HLA-DP, HLA-DQ, HLA-DR, HLA-G, HTLV1-Tax, Human papilloma virus E6 (HPV E6), Human papilloma virus E7 (HPV E7), Human Telomerase reverse transcriptase (hTERT), IgE, IL13Ra2, IL1 lRa, Immunoglobulin lambda-like polypeptide 1 (IGLL1), Influenza A hemagglutinin (HA), Insulin-like growth factor 1 receptor (IGF-I receptor), Interleukin 11 receptor alpha (IL-11Ra), Interleukin- 13 receptor subunit alpha-2 (IL-13Ra2 or CD213A2), Intestinal carboxyl esterase, KIT (CD117), KSHV K8.1, KSHV-gH, LAMP1 , Legumain, Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2), Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1), Leutenizing hormone receptor (LHR), Lewis(Y) antigen, Lews Ag, Livl, Locus K 9 (LY6K), Low conductance chloride channel, Lymphocyte antigen 6 complex, Lymphocyte antigen 75 (LY75), Lymphocyte-specific protein tyrosine kinase (LCK), Mammary gland differentiation antigen (NY-BR-1), Melanoma antigen recognized by T cells 1 (MelanA or MART1), Melanoma- associated antigen 1 (MAGE-A1), Melanoma cancer testis antigen-1 (MAD-CT-1), Melanoma cancer testis antigen-2 (MAD-CT-2), Melanoma

inhibitor of apoptosis (ML-IAP), Mesothelin, MPL, Mucin 1 cell surface associated (MUC1), N-Acetyl glucosaminyl-transferase V (NA17), Nectin-4, Neural cell adhesion molecule (NCAM), NKG2D, NYBR1, O-acetyl-GD2 ganglioside (OAcGD2), Olfactory receptor 51E2 (OR51E2), Oncogene fusion protein consisting of breakpoint cluster region (BCR) and Abelson murine leukemia viral oncogene homolog 1 (Abl) (bcr-abl), P53 mutant, Paired box protein Pax-3 (PAX3), Paired box protein Pax-5 (PAX5), Pannexin 3 (PANX3), PDL1, P-glycoprotein, Placenta-specific 1 (PLAC1), Platelet-derived growth factor receptor beta (PDGFR-beta), Polysialic acid, Proacrosin binding protein sp32 (OY-TES1), Prostase, Prostate carcinoma tumor antigen-1 (PCT A-1 or Galectin 8), Prostate stem cell antigen (PSCA), Prostate-specific membrane antigen (PSMA), Prostatic acid phosphatase (PAP), Prostein, Protease Serine 21 (Testisin or PRSS21), Proteasome (Prosome Macropain) Subunit Beta Type 9 (LMP2), PTK7, Ras G12V, Ras Homolog Family Member C (RhoC), Rat sarcoma (Ras) mutant, Receptor for Advanced Glycation Endproducts (RAGE-1), Receptor tyrosine kinase-like orphan receptor 1 (ROR1), Receptor tyrosine-protein kinase ERBB2 or Her-22/neu, Renal ubiquitous 1 (RU1), Renal ubiquitous 2 (RU2), Sarcoma translocation breakpoints, Serine 2 (TMPRSS2) ETS fusion gene, Sialyl Lewis adhesion molecule (sLe), SLAMF4, SLAMF6, Slea (CA19.9 or Sialyl Lewis Antigen), Sperm protein 17 (SPA17), Squamous Cell Carcinoma Antigen Recognized By T Cells 3 (SART3), Stage-specific embryonic antigen-4 (SSEA-4), STEAP1, Survivin, Synovial sarcoma X breakpoint 2 (SSX2), TCR Gamma Alternate Reading Frame Protein (TARP), TCR-beta1 chain, TCR-beta2 chain, TCR-delta chain, TCR-gamma chain, TCRgamma-delta, Telomerase, TGFbetaR2, The antigen recognized by TNT antibody, Thyroid stimulating hormone receptor (TSHR), Timl/HVCR1, Tissue Factor 1 (TF1), Tn ag, Tn antigen ((Tn Ag) or (GalNAc-Ser/Thr)), TNF receptor family member B cell maturation (BCMA), Transglutaminase 5 (TGS5), Transmembrane protease, TROP2, Tumor endothelial marker 1 (TEM1/CD248), Tumor endothelial marker 7-related (TEM7R), Tumor protein p53 (p53), Tumor-associated glycoprotein 72 (TAG72), Tyrosinase, Tyrosinase-related protein 2 (TRP-2), Uroplakin 2 (UPK2), Vascular endothelial growth factor receptor 2 (VEGFR2), V-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN), Wilms tumor protein (WT1), or X Antigen Family Member 1A (XAGE1) on the target cell.

53. A method of delivering a heterologous molecule to a target cell in a subject, the method

comprising administering to the subject a viral vector comprising:

- a) the VSV-G polypeptide of any one of claims 1-30;
- b) a targeting moiety that binds to the target cell; and
- c) a nucleic acid molecule encoding the heterologous molecule.

54. The method of claim 53, wherein the target cell is an immune cell, T cell, B cell, NK cell, dendritic cell, neutrophils, macrophages, a cancer cell, CD3+ T cell, CD4+ T cell, CD7+ T cell, CD8+ T cell, CD19+ B cell, CD19+ cancer cell, CD20+ B cell, CD20+ cancer cell, CD30+ lung epithelial cell, CD34+ haematopoietic stem cell, CD105+ endothelial cell, CD105+ haematopoietic stem cell, CD117+ haematopoietic stem cell, CD133+ cancer cell, EpCAM+ cancer cell, GluA2+ neuron, GluA4+ neuron, Haematopoietic stem cell, Hepatocyte, Her2/Neu+ cancer cell, NKG2D+ natural killer cell, SLC1A3+ astrocyte, or + adipocyte.

55. The method of claims 53 or 54, wherein the targeting moiety binds to CD7, CD8, cKit (CD117), CD4, CD3, CD5, CD6, CD2, TCR alpha, TCR beta, TCR gamma, TCR delta, CD10, CD34, CD110, CD33, CD14, CD68, CCR7, CD62L, CD25, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, or CXCR3, A glycosylated CD43 epitope expressed on acute leukemia or lymphoma but not on hematopoietic progenitors, A glycosylated CD43 epitope expressed on non-hematopoietic cancers, A kinase anchor protein 4 (AKAP-4), Adrenoceptor beta 3 (ADRB3), AFP, Anaplastic lymphoma kinase (ALK), Androgen receptor, Angiopoietin-binding cell surface receptor 2 (Tie 2), Auto antibody to desmoglein 1 (Dsg1), Auto antibody to desmoglein 3 (Dsg3), B7H3 (CD276), Biotin, Bone marrow stromal cell antigen 2 (BST2), BST1/CD157, Cancer/testis antigen 1 (NY-ESO-1), Cancer/testis antigen 2 (LAGE-la), Carbonic anhydrase IX (CAIX), Carcinoembryonic antigen (CEA), CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS or Brother of the Regulator of Imprinted Sites), CCR4, CD5, CD19, CD20, CD22, CD24, CD30, CD32 (FCGR2A), CD33, CD34, CD38, CD44v6, CD72, CD79a, CD79b, CD97, CD99, CD123, CD171, CD179a, CD179b-IGLII, CD200R, CD276/B7H3, CD300 molecule-like family member f (CD300LF), CDH1-CD324, CDH6, CDH17, CDH19, Chromosome X open reading frame 61 (CXORF61), Claudin 6 (CLDN6), Claudin8.2 (CLD18A2 or CLDN18A.2), CMV pp65, C-MYC epitope Tag, Cripto, CS1 (also referred to as CD2 subset 1 or CRACC or SLAMF7 or CD319 or 19A24), CSF2RA (GM-CSFR-alpha), C-type lectin domain family 12 member A (CLEC12A), C-type

lectin-like molecule-1 (CLL-1 or CLECL1), Cyclin B1, Cytochrome P450 1B 1 (CYP1B 1), DLL3, EBV-EBNA3c, EGF-bke module- containing mucin-like hormone receptor-like 2 (EMR2), Elongation factor 2 mutated (ELF2M), Ephrin B2, Ephrin type-A receptor 2 (EphA2), Epidermal growth factor receptor (EGFR), Epidermal growth factor receptor variant III (EGFRviii), Epithelial cell adhesion molecule (EPCAM), ERG, ETS translocation-variant gene 6 located on chromosome 12p (ETV6-AML), Fc fragment of IgA receptor (FCAR or CD89), Fc receptor-like 5 (FCRL5), Fibroblast activation protein alpha (FAP), FITC, Fms Like Tyrosine Kinase 3 (FLT3), Folate receptor alpha (FRa or FR1), Folate receptor beta (FRb), Follicle stimulating hormone receptor (FSHR), Fos-related antigen 1, Fucosyl-GM1, G protein coupled receptor class C group 5 member D (GPRC5D), G protein-coupled receptor 20 (GPR20), GAD, Ganglioside G2 (GD2) , Ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlc(1-1)Cer), Ganglioside GM3 (aNeu5Ac(2-3)bDCIalp(1- 4)bDGlc(1-1)Cer), GD3, GFRalpha4, Glycoprotein 100 (gp100), Glypican-3 (GPC3), Gonadotropin Hormone receptor (CGHR or GR), GpA33, GpNMB, GPRC5D, Guanylyl cyclase C (GCC), Heat shock protein 70-2 mutated (mut hsp70-2), Hepatitis A virus cellular receptor 1 (HAVCR1), Hexasaccharide portion of globoH glycosphingolipid (GloboH), High molecular weight-melanoma associated antigen (HMWMAA), HIV1 envelope glycoprotein, HLA, HLA-DOA, HLA-A, HLA-A2, HLA-B, HLA-C, HLA-DM, HLA-DOB, HLA-DP, HLA-DQ, HLA-DR, HLA-G, HTLV1-Tax, Human papilloma virus E6 (HPV E6), Human papilloma virus E7 (HPV E7), Human Telomerase reverse transcriptase (hTERT), IgE, IL13Ra2, IL1 lRa, Immunoglobulin lambda-like polypeptide 1 (IGLL1), Influenza A hemagglutinin (HA), Insulin-like growth factor 1 receptor (IGF-I receptor), Interleukin 11 receptor alpha (IL-11Ra), Interleukin- 13 receptor subunit alpha-2 (IL-13Ra2 or CD213A2), Intestinal carboxyl esterase, KIT (CD117), KSHV K8.1, KSHV-gH, LAMP1 , Legumain, Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2), Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1), Leutenizing hormone receptor (LHR), Lewis(Y) antigen, Lews Ag, Liv1, Locus K 9 (LY6K), Low conductance chloride channel, Lymphocyte antigen 6 complex, Lymphocyte antigen 75 (LY75), Lymphocyte-specific protein tyrosine kinase (LCK), Mammary gland differentiation antigen (NY-BR-1), Melanoma antigen recognized by T cells 1 (MelanA or MART1), Melanoma- associated antigen 1 (MAGE-A1), Melanoma cancer testis antigen-1 (MAD-CT-1), Melanoma cancer testis antigen-2 (MAD-CT-2), Melanoma inhibitor of apoptosis (ML-IAP), Mesothelin, MPL, Mucin 1 cell surface associated (MUC1), N-

Acetyl glucosaminyl-transferase V (NA17), Nectin-4, Neural cell adhesion molecule (NCAM), NKG2D, NYBR1, O-acetyl-GD2 ganglioside (OAcGD2), Olfactory receptor 51E2 (OR51E2), Oncogene fusion protein consisting of breakpoint cluster region (BCR) and Abelson murine leukemia viral oncogene homolog 1 (Abl) (bcr-abl), P53 mutant, Paired box protein Pax-3 (PAX3), Paired box protein Pax-5 (PAX5), Pannexin 3 (PANX3), PDL1, P-glycoprotein, Placenta-specific 1 (PLAC1), Platelet-derived growth factor receptor beta (PDGFR-beta), Polysialic acid, Proacrosin binding protein sp32 (OY-TES1), Prostase, Prostate carcinoma tumor antigen-1 (PCT A-1 or Galectin 8), Prostate stem cell antigen (PSCA), Prostate-specific membrane antigen (PSMA), Prostatic acid phosphatase (PAP), Prostein, Protease Serine 21 (Testisin or PRSS21), Proteasome (Prosome Macropain) Subunit Beta Type 9 (LMP2), PTK7, Ras G12V, Ras Homolog Family Member C (RhoC), Rat sarcoma (Ras) mutant, Receptor for Advanced Glycation Endproducts (RAGE-1), Receptor tyrosine kinase-like orphan receptor 1 (ROR1), Receptor tyrosine-protein kinase ERBB2 or Her-22/neu, Renal ubiquitous 1 (RU1), Renal ubiquitous 2 (RU2), Sarcoma translocation breakpoints, Serine 2 (TMPRSS2) ETS fusion gene, Sialyl Lewis adhesion molecule (sLe), SLAMF4, SLAMF6, Slea (CA19.9 or Sialyl Lewis Antigen), Sperm protein 17 (SPA17), Squamous Cell Carcinoma Antigen Recognized By T Cells 3 (SART3), Stage-specific embryonic antigen-4 (SSEA-4), STEAP1, Survivin, Synovial sarcoma X breakpoint 2 (SSX2), TCR Gamma Alternate Reading Frame Protein (TARP), TCR-beta1 chain, TCR-beta2 chain, TCR-delta chain, TCR-gamma chain, TCRgamma-delta, Telomerase, TGFbetaR2, The antigen recognized by TNT antibody, Thyroid stimulating hormone receptor (TSHR), Timl-/HVCR1, Tissue Factor 1 (TF1), Tn ag, Tn antigen ((Tn Ag) or (GalNAc-Ser/Thr)), TNF receptor family member B cell maturation (BCMA), Transglutaminase 5 (TGS5), Transmembrane protease, TROP2, Tumor endothelial marker 1 (TEM1/CD248), Tumor endothelial marker 7-related (TEM7R), Tumor protein p53 (p53), Tumor-associated glycoprotein 72 (TAG72), Tyrosinase, Tyrosinase-related protein 2 (TRP-2), Uroplakin 2 (UPK2), Vascular endothelial growth factor receptor 2 (VEGFR2), V-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN), Wilms tumor protein (WT1), or X Antigen Family Member 1A (XAGE1) on the target cell.

56. The method of any one of claims 50-55, wherein the heterologous molecule is an siRNA, an shRNA, a non-coding RNA, a peptide, a polypeptide, a protein, a viral payload, a viral genome,

a chimeric antigen receptor (“CAR”) or a combination thereof.

57. A method of treating cancer in a subject, the method comprising administering to the subject a viral vector comprising:

- a) the VSV-G polypeptide of any one of claims 1-30;
- b) a targeting moiety that binds to the target cell; and
- c) a nucleic acid molecule encoding the heterologous molecule.

58. The method of claim 57, wherein the heterologous molecule is a chimeric antigen receptor.

59. The method of claims 57 or 58, wherein the target cell is an immune cell, T cell, B cell, NK cell, dendritic cell, neutrophils, macrophages, a cancer cell, or, for example, CD3+ T cell, CD4+ T cell, CD7+ T cell, CD8+ T cell, CD19+ B cell, CD19+ cancer cell, CD20+ B cell, CD30+ lung epithelial cell, CD34+ haematopoietic stem cell, CD105+ endothelial cell, CD105+ haematopoietic stem cell, CD117+ haematopoietic stem cell, CD133+ cancer cell, EpCAM+ cancer cell, GluA2+ neuron, GluA4+ neuron, Haematopoietic stem cell, Hepatocyte, Her2/Neu+ cancer cell, NKG2D+ natural killer cell, SLC1A3+ astrocyte, or SLC7A10+ adipocyte.

60. The method of any one of claims 57-59, wherein the targeting moiety binds to CD7, CD8, cKit (CD117), CD4, CD3, CD5, CD6, CD2, TCR alpha, TCR beta, TCR gamma, TCR delta, CD10, CD34, CD110, CD33, CD14, CD68, CCR7, CD62L, CD25, CCR2, CCR3, CCR4, CCR5, CCR6, CCR7, or CXCR3, A glycosylated CD43 epitope expressed on acute leukemia or lymphoma but not on hematopoietic progenitors, A glycosylated CD43 epitope expressed on non-hematopoietic cancers, A kinase anchor protein 4 (AKAP-4), Adrenoceptor beta 3 (ADRB3), AFP, Anaplastic lymphoma kinase (ALK), Androgen receptor, Angiopoietin-binding cell surface receptor 2 (Tie 2), Auto antibody to desmoglein 1 (Dsgl), Auto antibody to desmoglein 3 (Dsg3), B7H3 (CD276), Biotin, Bone marrow stromal cell antigen 2 (BST2), BST1/CD157, Cancer/testis antigen 1 (NY-ESO-1), Cancer/testis antigen 2 (LAGE-Ia), Carbonic anhydrase IX (CAIX), Carcinoembryonic antigen (CEA), CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS or Brother of the Regulator of Imprinted Sites), CCR4, CD5, CD19, CD20, CD22, CD24, CD30, CD32 (FCGR2A), CD33, CD34, CD38, CD44v6, CD72, CD79a, CD79b, CD97, CD99, CD123,

CD171, CD179a, CD179b-IGLII, CD200R, CD276/B7H3, CD300 molecule-like family member f (CD300LF), CDH1-CD324, CDH6, CDH17, CDH19, Chromosome X open reading frame 61 (CXORF61), Claudin 6 (CLDN6), Claudin18.2 (CLD18A2 or CLDN18A.2), CMV pp65, C-MYC epitope Tag, Cripto, CS1 (also referred to as CD2 subset 1 or CRACC or SLAMF7 or CD319 or 19A24), CSF2RA (GM-CSFR-alpha), C-type lectin domain family 12 member A (CLEC12A), C-type lectin-like molecule-1 (CLL-1 or CLECL1), Cyclin B1, Cytochrome P450 IB 1 (CYP1B 1), DLL3, EBV-EBNA3c, EGF-bke module- containing mucin-like hormone receptor-like 2 (EMR2), Elongation factor 2 mutated (ELF2M), Ephrin B2, Ephrin type-A receptor 2 (EphA2), Epidermal growth factor receptor (EGFR), Epidermal growth factor receptor variant III (EGFRviii), Epithelial cell adhesion molecule (EPCAM), ERG, ETS translocation-variant gene 6 located on chromosome 12p (ETV6-AML), Fc fragment of IgA receptor (FCAR or CD89), Fc receptor-like 5 (FCRL5), Fibroblast activation protein alpha (FAP), FITC, Fms Like Tyrosine Kinase 3 (FLT3), Folate receptor alpha (FRa or FR1), Folate receptor beta (FRb), Follicle stimulating hormone receptor (FSHR), Fos-related antigen 1, Fucosyl-GM1, G protein coupled receptor class C group 5 member D (GPRC5D), G protein-coupled receptor 20 (GPR20), GAD, Ganglioside G2 (GD2) , Ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlc(1-1)Cer), Ganglioside GM3 (aNeu5Ac(2-3)bDCIalp(1- 4)bDGlc(1-1)Cer), GD3, GFRalpha4, Glycoprotein 100 (gp100), Glypican-3 (GPC3), Gonadotropin Hormone receptor (CGHR or GR), GpA33, GpNMB, GPRC5D, Guanylyl cyclase C (GCC), Heat shock protein 70-2 mutated (mut hsp70-2), Hepatitis A virus cellular receptor 1 (HAVCR1), Hexasaccharide portion of globoH glycosphingolipid (GloboH), High molecular weight-melanoma associated antigen (HMWMAA), HIV1 envelope glycoprotein, HLA, HLA-DOA, HLA-A, HLA-A2, HLA-B, HLA-C, HLA-DM, HLA-DOB, HLA-DP, HLA-DQ, HLA-DR, HLA-G, HTLV1-Tax, Human papilloma virus E6 (HPV E6), Human papilloma virus E7 (HPV E7), Human Telomerase reverse transcriptase (hTERT), IgE, IL13Ra2, IL1 lRa, Immunoglobulin lambda-like polypeptide 1 (IGLL1), Influenza A hemagglutinin (HA), Insulin-like growth factor 1 receptor (IGF-I receptor), Interleukin 11 receptor alpha (IL-11Ra), Interleukin- 13 receptor subunit alpha-2 (IL-13Ra2 or CD213A2), Intestinal carboxyl esterase, KIT (CD117), KSHV K8.1, KSHV-gH, LAMP1 , Legumain, Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2), Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1), Leutenizing hormone receptor (LHR), Lewis(Y) antigen, Lews Ag, Livl, Locus K 9 (LY6K), Low conductance chloride channel, Lymphocyte

antigen 6 complex, Lymphocyte antigen 75 (LY75), Lymphocyte-specific protein tyrosine kinase (LCK), Mammary gland differentiation antigen (NY-BR-1), Melanoma antigen recognized by T cells 1 (MelanA or MART1), Melanoma-associated antigen 1 (MAGE-A1), Melanoma cancer testis antigen-1 (MAD-CT-1), Melanoma cancer testis antigen-2 (MAD-CT-2), Melanoma inhibitor of apoptosis (ML-IAP), Mesothelin, MPL, Mucin 1 cell surface associated (MUC1), N-Acetyl glucosaminyl-transferase V (NA17), Nectin-4, Neural cell adhesion molecule (NCAM), NKG2D, NYBR1, O-acetyl-GD2 ganglioside (OAcGD2), Olfactory receptor 51E2 (OR51E2), Oncogene fusion protein consisting of breakpoint cluster region (BCR) and Abelson murine leukemia viral oncogene homolog 1 (Abl) (bcr-abl), P53 mutant, Paired box protein Pax-3 (PAX3), Paired box protein Pax-5 (PAX5), Pannexin 3 (PANX3), PDL1, P-glycoprotein, Placenta-specific 1 (PLAC1), Platelet-derived growth factor receptor beta (PDGFR-beta), Polysialic acid, Proacrosin binding protein sp32 (OY-TES1), Prostase, Prostate carcinoma tumor antigen-1 (PCT A-1 or Galectin 8), Prostate stem cell antigen (PSCA), Prostate-specific membrane antigen (PSMA), Prostatic acid phosphatase (PAP), Prostein, Protease Serine 21 (Testisin or PRSS21), Proteasome (Prosome Macropain) Subunit Beta Type 9 (LMP2), PTK7, Ras G12V, Ras Homolog Family Member C (RhoC), Rat sarcoma (Ras) mutant, Receptor for Advanced Glycation Endproducts (RAGE-1), Receptor tyrosine kinase-like orphan receptor 1 (ROR1), Receptor tyrosine-protein kinase ERBB2 or Her-22/neu, Renal ubiquitous 1 (RU1), Renal ubiquitous 2 (RU2), Sarcoma translocation breakpoints, Serine 2 (TMPRSS2) ETS fusion gene, Sialyl Lewis adhesion molecule (sLe), SLAMF4, SLAMF6, Slea (CA19.9 or Sialyl Lewis Antigen), Sperm protein 17 (SPA17), Squamous Cell Carcinoma Antigen Recognized By T Cells 3 (SART3), Stage-specific embryonic antigen-4 (SSEA-4), STEAP1, Survivin, Synovial sarcoma X breakpoint 2 (SSX2), TCR Gamma Alternate Reading Frame Protein (TARP), TCR-beta1 chain, TCR-beta2 chain, TCR-delta chain, TCR-gamma chain, TCRgamma-delta, Telomerase, TGFbetaR2, The antigen recognized by TNT antibody, Thyroid stimulating hormone receptor (TSHR), Timl-/HVCR1, Tissue Factor 1 (TF1), Tn ag, Tn antigen ((Tn Ag) or (GalNAc-Ser/Thr)), TNF receptor family member B cell maturation (BCMA), Transglutaminase 5 (TGS5), Transmembrane protease, TROP2, Tumor endothelial marker 1 (TEM1/CD248), Tumor endothelial marker 7-related (TEM7R), Tumor protein p53 (p53), Tumor-associated glycoprotein 72 (TAG72), Tyrosinase, Tyrosinase-related protein 2 (TRP-2), Uroplakin 2 (UPK2), Vascular endothelial growth factor receptor 2 (VEGFR2), V-myc avian myelocytomatosis viral oncogene

neuroblastoma derived homolog (MYCN), Wilms tumor protein (WT1), or X Antigen Family Member 1A (XAGE1) on the target cell.

61. The method of any one of claims 57-60, wherein the cancer is carcinoma, blastoma, sarcoma, leukemia, lymphoid malignancies, benign tumors, malignant tumors, sarcoma, carcinoma, melanoma, breast cancer, prostate cancer, ovarian cancer, cervical cancer, skin cancer, pancreatic cancer, colorectal cancer, renal cancer, liver cancer, brain cancer, lymphoma, leukemia, lung cancer, thyroid cancer, B cell related cancers or T cell related cancers.

62. A viral particle comprising a targeting moiety that binds to CD7 or CD8, or a pharmaceutical composition comprising the same, wherein:

i) the targeting moiety that binds to CD7 comprises:

a polypeptide comprising: (i) a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 35; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 36; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 37, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 38; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 39; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 40; or variants of any of the foregoing;

a heavy chain comprising a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 47, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; and HCDR3 as set forth in SEQ ID NO: 37;

a light chain comprising: a light chain variable region having at least 90% sequence identity to SEQ ID NO: 48, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 38; LCDR2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40;

a polypeptide comprises a heavy chain and a light chain comprising: a heavy chain

variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; HCDR3 as set forth in SEQ ID NO: 37; LCDR1 as set forth in SEQ ID NO: 38; LCD2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40;

a light chain and a heavy chain, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48;

a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 48;

a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 48;

a heavy chain variable region comprising SEQ ID NO: 47, and a light chain variable region comprising SEQ ID NO: 48; a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 51, at least 95% sequence identity to SEQ ID NO: 51, at least 99% sequence identity to SEQ ID NO: 51, or a sequence as set forth in SEQ ID NO: 51; or

a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 52, having at least 95% sequence identity to SEQ ID NO: 52, having at least 99% sequence identity to SEQ ID NO: 52, or a sequence as set forth in SEQ ID NO: 52; or
ii) the targeting moiety that binds to CD8 comprises:

a polypeptide that comprises: a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 55; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 56; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 57, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the

light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 58; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 59; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 60; or variants of any of the foregoing;

a heavy chain that comprises a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 67, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; and HCDR3 as set forth in SEQ ID NO: 57;

a light chain that comprises a light chain variable region having at least 90% sequence identity to SEQ ID NO: 68, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 58; LCDR2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60;

a polypeptide comprising a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; HCDR3 as set forth in SEQ ID NO: 57; LCDR1 as set forth in SEQ ID NO: 58; LCD2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60;

a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68;

a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 68;

a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 68;

a polypeptide comprising a light chain and a heavy chain comprising: wherein the heavy chain variable region comprises SEQ ID NO: 67, and the light chain variable region comprises SEQ ID NO: 68;

a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 69, at least 95% sequence identity to SEQ ID NO: 69, at least 99% sequence identity to SEQ ID NO: 69, or a sequence as set forth in SEQ ID NO: 69; or

a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 70, having at least 95% sequence identity to SEQ ID NO: 70, having at least 99% sequence identity to SEQ ID NO: 70, or a sequence as set forth in SEQ ID NO: 70.

63. The viral particle of claim 62, or a pharmaceutical composition comprising the same, wherein the viral particle comprises a heterologous fusogenic protein, such as a glycoprotein from such as those provided for herein .

64. The viral particle of claim 63, or a pharmaceutical composition comprising the same, wherein the fusogenic protein is a VSV-G protein or a mutant VSV-G protein or other viral glycoprotein.

65. The viral particle of claim 64, or a pharmaceutical composition comprising the same, wherein the mutant VSV-G protein comprises a polypeptide of any one of claims 1-30.

66. The viral particle of claim 62, or a pharmaceutical composition comprising the same, wherein the viral particle comprises a nucleic acid molecule encoding a chimeric antigen receptor.

67. A method of delivering a heterologous molecule to a CD7+ or CD8+ target cell or a method of treating cancer in a subject, the method comprising contacting a cell with, or administering to the subject, a pseudotyped viral vector comprising a targeting moiety that specifically binds to CD7 or CD8 on the target cell and a nucleic acid molecule encoding the heterologous molecule, wherein:

i) the targeting moiety that binds to CD7 comprises:

a polypeptide comprising: (i) a heavy chain variable region comprising heavy

chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 35; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 36; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 37, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 38; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 39; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 40; or variants of any of the foregoing;

a heavy chain comprising a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 47, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; and HCDR3 as set forth in SEQ ID NO: 37;

a light chain comprising: a light chain variable region having at least 90% sequence identity to SEQ ID NO: 48, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 38; LCDR2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40;

a polypeptide comprises a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; HCDR3 as set forth in SEQ ID NO: 37; LCDR1 as set forth in SEQ ID NO: 38; LCD2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40;

a light chain and a heavy chain, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48;

a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 48;

a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 48;

a heavy chain variable region comprising SEQ ID NO: 47, and a light chain variable region comprising SEQ ID NO: 48; a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 51, at least 95% sequence identity to SEQ ID NO: 51, at least 99% sequence identity to SEQ ID NO: 51, or a sequence as set forth in SEQ ID NO: 51; or

a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 52, having at least 95% sequence identity to SEQ ID NO: 52, having at least 99% sequence identity to SEQ ID NO: 52, or a sequence as set forth in SEQ ID NO: 52; or

ii) the targeting moiety that binds to CD8 comprises:

a polypeptide that comprises: a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 55; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 56; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 57, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 58; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 59; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 60; or variants of any of the foregoing;

a heavy chain that comprises a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 67, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; and HCDR3 as set forth in SEQ ID NO: 57;

a light chain that comprises a light chain variable region having at least 90% sequence identity to SEQ ID NO: 68, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 58; LCDR2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60;

a polypeptide comprising a heavy chain and a light chain comprising: a heavy chain

variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; HCDR3 as set forth in SEQ ID NO: 57; LCDR1 as set forth in SEQ ID NO: 58; LCD2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60;

a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68;

a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 68;

a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 68;

a polypeptide comprising a light chain and a heavy chain comprising: wherein the heavy chain variable region comprises SEQ ID NO: 67, and the light chain variable region comprises SEQ ID NO: 68;

a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 69, at least 95% sequence identity to SEQ ID NO: 69, at least 99% sequence identity to SEQ ID NO: 69, or a sequence as set forth in SEQ ID NO: 69; or

a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 70, having at least 95% sequence identity to SEQ ID NO: 70, having at least 99% sequence identity to SEQ ID NO: 70, or a sequence as set forth in SEQ ID NO: 70.

68. A cell comprising one or more heterologous nucleic acid molecules encoding for a targeting moiety and a viral glycoprotein, wherein the heterologous nucleic acid molecules

comprise a nucleic acid sequence encoding for a targeting moiety that binds to CD7 or CD8, wherein:

i) the targeting moiety that binds to CD7 comprises:

a polypeptide comprising: (i) a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 35; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 36; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 37, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 38; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 39; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 40; or variants of any of the foregoing;

a heavy chain comprising a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 47, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; and HCDR3 as set forth in SEQ ID NO: 37;

a light chain comprising: a light chain variable region having at least 90% sequence identity to SEQ ID NO: 48, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 38; LCDR2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40;

a polypeptide comprises a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 48, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 35; HCDR2 as set forth in SEQ ID NO: 36; HCDR3 as set forth in SEQ ID NO: 37; LCDR1 as set forth in SEQ ID NO: 38; LCD2 as set forth in SEQ ID NO: 39; and LCDR3 as set forth in SEQ ID NO: 40;

a light chain and a heavy chain, wherein the light chain and a heavy chain comprise: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 90%

sequence identity to SEQ ID NO: 48;

a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 48;

a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 47; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 48;

a heavy chain variable region comprising SEQ ID NO: 47, and a light chain variable region comprising SEQ ID NO: 48; a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 51, at least 95% sequence identity to SEQ ID NO: 51, at least 99% sequence identity to SEQ ID NO: 51, or a sequence as set forth in SEQ ID NO: 51; or

a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 52, having at least 95% sequence identity to SEQ ID NO: 52, having at least 99% sequence identity to SEQ ID NO: 52, or a sequence as set forth in SEQ ID NO: 52; or

ii) the targeting moiety that binds to CD8 comprises:

a polypeptide that comprises: a heavy chain variable region comprising heavy chain CDR1, CDR2, and CDR3 sequences, wherein the heavy chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 55; the heavy chain CDR2 has the amino acid sequence of SEQ ID NO: 56; and the heavy chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 57, or variants of any of the foregoing; and (ii) a light chain variable region comprising light chain CDR1, CDR2, and CDR3 sequences, wherein the light chain CDR1 sequence has the amino acid sequence of SEQ ID NO: 58; the light chain CDR2 sequence has the amino acid sequence of SEQ ID NO: 59; and the light chain CDR3 sequence has the amino acid sequence of SEQ ID NO: 60; or variants of any of the foregoing;

a heavy chain that comprises a heavy chain variable region having at least 90% sequence identity to SEQ ID NO: 67, wherein the polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; and HCDR3 as set forth in SEQ ID NO: 57;

a light chain that comprises a light chain variable region having at least 90%

sequence identity to SEQ ID NO: 68, wherein the polypeptide maintains the sequences of LCDR1 as set forth in SEQ ID NO: 58; LCDR2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60;

a polypeptide comprising a heavy chain and a light chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67, and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68, wherein polypeptide maintains the sequences of HCDR1 as set forth in SEQ ID NO: 55; HCDR2 as set forth in SEQ ID NO: 56; HCDR3 as set forth in SEQ ID NO: 57; LCDR1 as set forth in SEQ ID NO: 58; LCD2 as set forth in SEQ ID NO: 59; and LCDR3 as set forth in SEQ ID NO: 60;

a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 90% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 90% sequence identity to SEQ ID NO: 68;

a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 95% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 95% sequence identity to SEQ ID NO: 68;

a polypeptide comprising a light chain and a heavy chain comprising: a heavy chain variable region of the heavy chain having at least 99% sequence identity to SEQ ID NO: 67; and a light chain variable region of the light chain having at least 99% sequence identity to SEQ ID NO: 68;

a polypeptide comprising a light chain and a heavy chain comprising: wherein the heavy chain variable region comprises SEQ ID NO: 67, and the light chain variable region comprises SEQ ID NO: 68;

a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 69, at least 95% sequence identity to SEQ ID NO: 69, at least 99% sequence identity to SEQ ID NO: 69, or a sequence as set forth in SEQ ID NO: 69; or

a polypeptide comprising a sequence having at least 90% sequence identity to SEQ ID NO: 70, having at least 95% sequence identity to SEQ ID NO: 70, having at least 99% sequence identity to SEQ ID NO: 70, or a sequence as set forth in SEQ ID NO: 70.

69. A cell comprising the nucleic acid molecule of claim 31, the vector of claim 32, or the plasmid of claim 33.

70. A method of making a viral particle comprising culturing the cell of any one of the preceding claims under conditions sufficient to make the viral particle.

ABSTRACT

Provided for herein are mutant VSV-G polypeptides, compositions comprising the same, and methods of using the same. Also provided for herein are polypeptides and compositions that bind to CD7 and uses thereof. Also provided for herein are polypeptides and compositions that bind to CD8 and uses thereof.

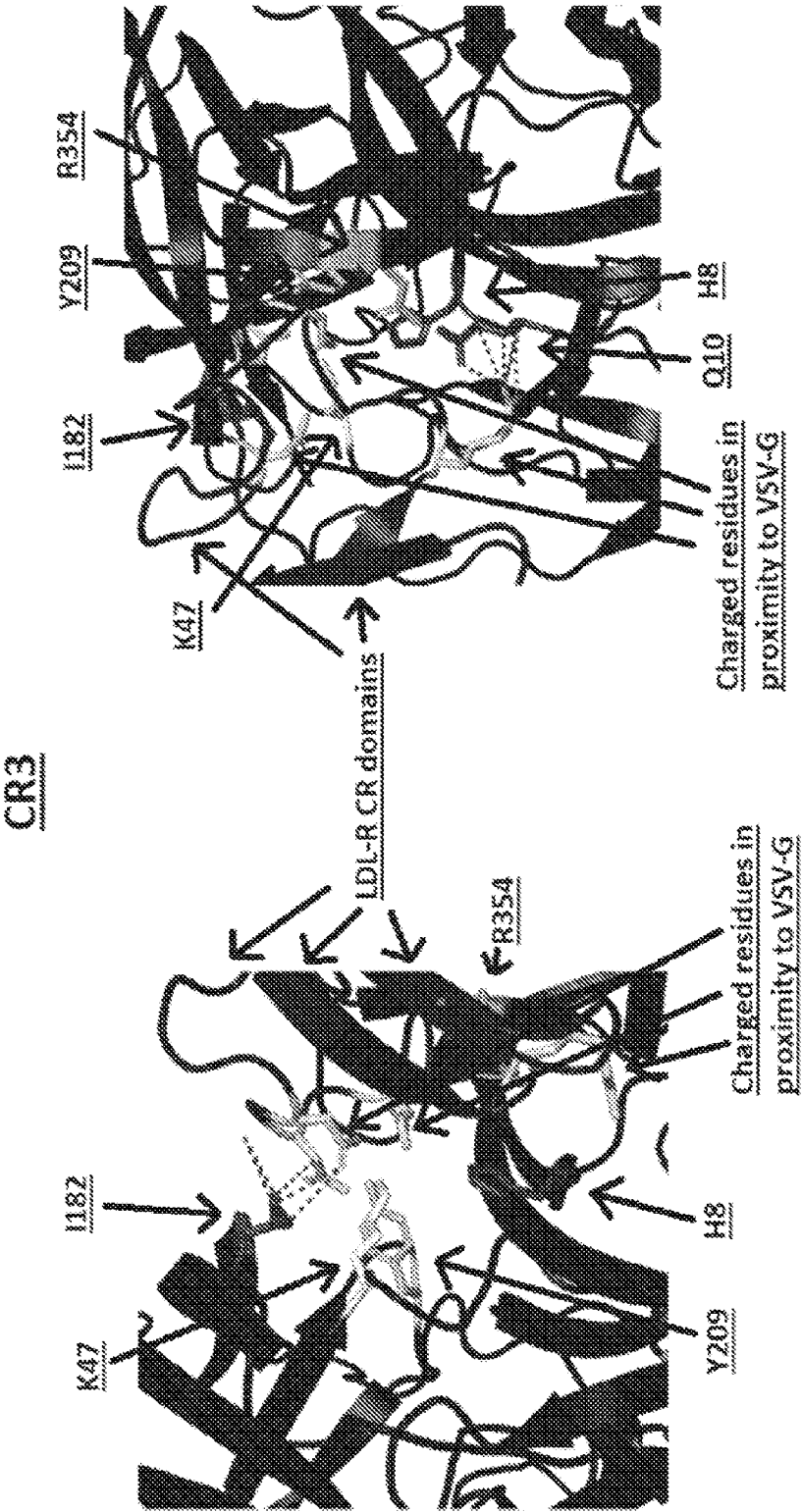


FIG. 1A

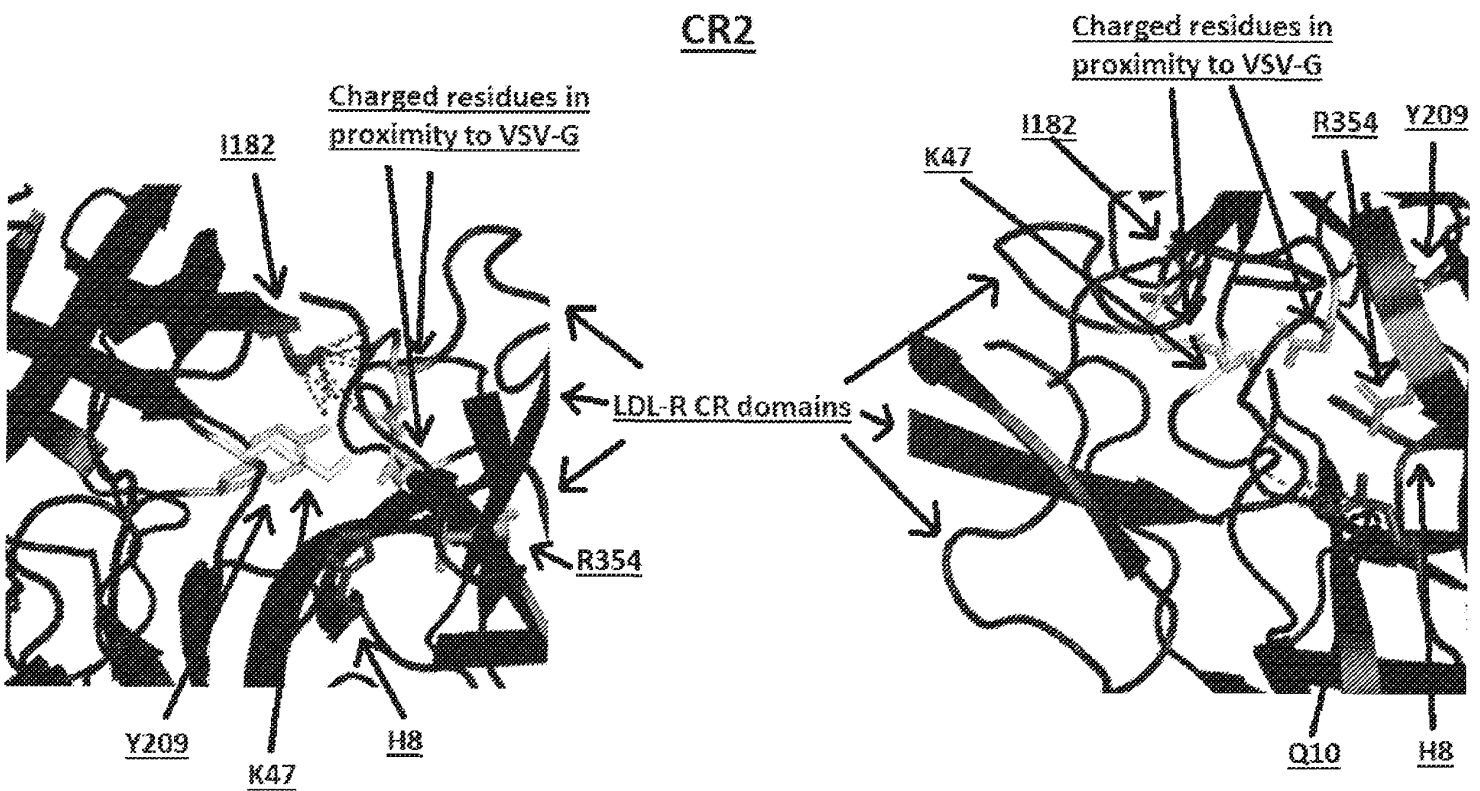


FIG. 1B

3/26

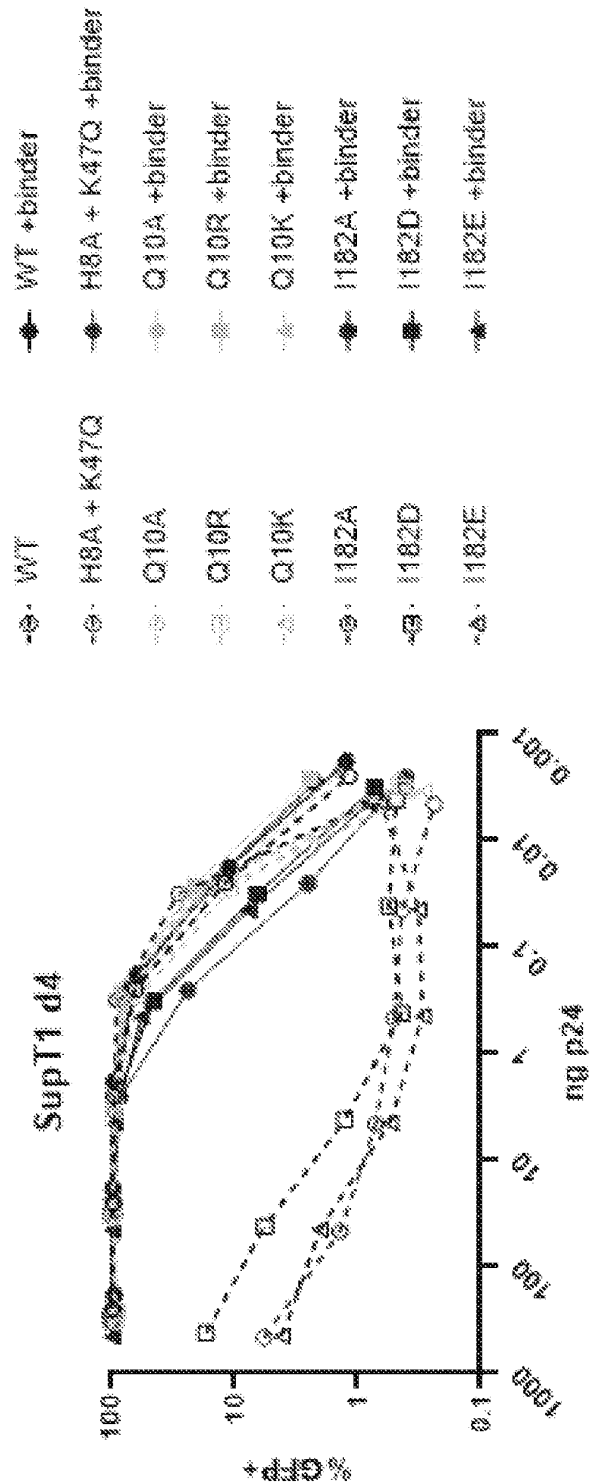


FIG. 2A

4/26

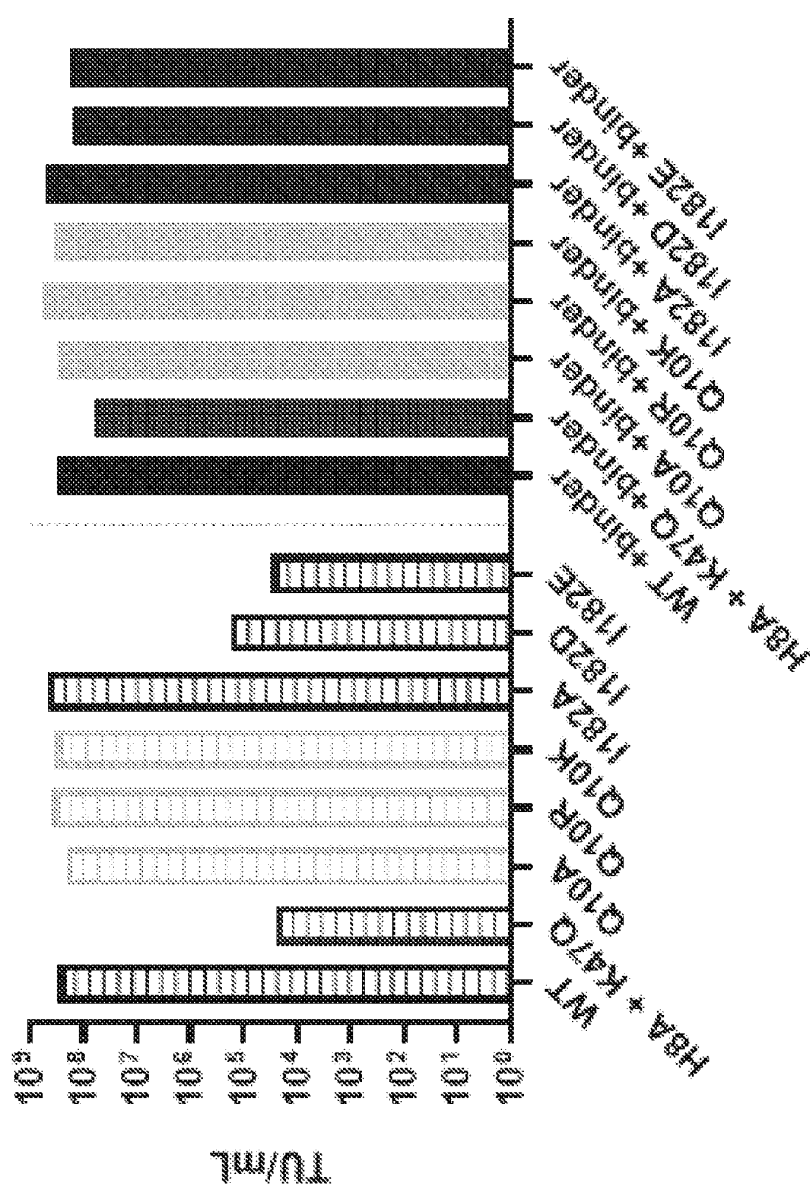


FIG. 2B

SEQ_ID_NO_2	KSTFVFPFHUKXGKNWYVPSNTHYCPSSDILWNNHDLIGTALQVNMKFSKKAQADGWMC	69
SEQ_ID_NO_11	NTEIVFPQHTFSDWKPVEHE/YNCPFHADKNHSHGQTGTGPEYELNMKFKGLTHGQVDPGPMH	69
SEQ_ID_NO_13	NTEIVFPKHGKSNWYVPSNTHYCPSSDQNNHDLITGVSILHVFIKKHNATGADGWMC	69
SEQ_ID_NO_15	NTEIVFPQSLGSDWKPVEKGYVCPSSADKNHDLIGTALQVNMKFSKKAQADGWMC	69
SEQ_ID_NO_17	KPTIVFPQSQGKSDWKPVEPNYKCPSSADQNNHDLIGTALQVNMKFSKKAQADGWMC	69
SEQ_ID_NO_19	KPTIVFPQSQGKSDWKNVPSYK/CPSSDQNNHDLIGTALQVNMKFSKKAQADGWMC	69
SEQ_ID_NO_21	KPTIVFPHDKGHNWYVANYQVCPSSDQNNHDLIGTALQVNMKFSKKAQADGWMC	69
	*: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *	
SEQ_ID_NO_2	ASKNVWTCDFPWHYGHYITHSIRSFPSVEQCKESIEGTNQCOTWNLGPPFQSGGYATVT	120
SEQ_ID_NO_11	SATWNTCCDFPWHYGHYITHSIRSFPSVEQCKESIEGTNQCOTWNLGPPFQSGGYATVT	120
SEQ_ID_NO_13	AAKWVWTCDFPWHYGHYITHSIRSFPSVEQCKESIEGTNQCOTWNLGPPFQSGGYATVT	120
SEQ_ID_NO_15	AAKWVWTCDFPWHYGHYITHSIRSFPSVEQCKESIEGTNQCOTWNLGPPFQSGGYATVT	120
SEQ_ID_NO_17	AAKWVWTCDFPWHYGHYITHSIRSFPSVEQCKESIEGTNQCOTWNLGPPFQSGGYATVT	120
SEQ_ID_NO_19	AAKWVWTCDFPWHYGHYITHSIRSFPSVEQCKESIEGTNQCOTWNLGPPFQSGGYATVT	120
SEQ_ID_NO_21	AAKWVWTCDFPWHYGHYITHSIRSFPSVEQCKESIEGTNQCOTWNLGPPFQSGGYATVT	120
	: : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :	
SEQ_ID_NO_2	DAAAVVQVTFHHVIVDEYTGWNVDSQPIGKSNHYICPTVHNSTHWSHYKVEGLCSN	180
SEQ_ID_NO_11	DAAAVVQVTFHHVIVDEYTGWNVDSQPIGKSNHYICPTVHNSTHWSHYKVEGLCSN	180
SEQ_ID_NO_13	DAAAVVQVTFHHVIVDEYTGWNVDSQPIGKSNHYICPTVHNSTHWSHYKVEGLCSN	180
SEQ_ID_NO_15	DAAAVVQVTFHHVIVDEYTGWNVDSQPIGKSNHYICPTVHNSTHWSHYKVEGLCSN	180
SEQ_ID_NO_17	DAAAVVQVTFHHVIVDEYTGWNVDSQPIGKSNHYICPTVHNSTHWSHYKVEGLCSN	180
SEQ_ID_NO_19	DAAAVVQVTFHHVIVDEYTGWNVDSQPIGKSNHYICPTVHNSTHWSHYKVEGLCSN	180
SEQ_ID_NO_21	DAAAVVQVTFHHVIVDEYTGWNVDSQPIGKSNHYICPTVHNSTHWSHYKVEGLCSN	180
	*: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *	
SEQ_ID_NO_2	LISMSTTFESDGGKILSLGKESGTFRNHYFAYETGGKACNMQYCKHMGVFLPSQWFEA	240
SEQ_ID_NO_11	PTLVAGTTFESDGGKILSLGKESGTFRNHYFAYETGGKACNMQYCKHMGVFLPSQWFEA	240
SEQ_ID_NO_13	LAIVAGTTFESDGGKILSLGKESGTFRNHYFAYETGGKACNMQYCKHMGVFLPSQWFEA	240
SEQ_ID_NO_15	LAIVAGTTFESDGGKILSLGKESGTFRNHYFAYETGGKACNMQYCKHMGVFLPSQWFEA	240
SEQ_ID_NO_17	LAIVAGTTFESDGGKILSLGKESGTFRNHYFAYETGGKACNMQYCKHMGVFLPSQWFEA	240
SEQ_ID_NO_19	LAIVAGTTFESDGGKILSLGKESGTFRNHYFAYETGGKACNMQYCKHMGVFLPSQWFEA	240
SEQ_ID_NO_21	LAIVAGTTFESDGGKILSLGKESGTFRNHYFAYETGGKACNMQYCKHMGVFLPSQWFEA	240
	: : : : : : : : : : * : * : * : * : * : * : * : * : * : * : * : * : * :	
SEQ_ID_NO_2	DKLEA-----AAFFPECHGSSISABSGQTVGVVSLGQDVHPIILVSLGQETWSKIPAGLE	296
SEQ_ID_NO_11	DKLEA-----AAFFPECHGSSISABSGQTVGVVSLGQDVHPIILVSLGQETWSKIPAGLE	296
SEQ_ID_NO_13	DKLEA-----AAFFPECHGSSISABSGQTVGVVSLGQDVHPIILVSLGQETWSKIPAGLE	296
SEQ_ID_NO_15	DKLEA-----AAFFPECHGSSISABSGQTVGVVSLGQDVHPIILVSLGQETWSKIPAGLE	296
SEQ_ID_NO_17	DKLEA-----AAFFPECHGSSISABSGQTVGVVSLGQDVHPIILVSLGQETWSKIPAGLE	296
SEQ_ID_NO_19	DKLEA-----AAFFPECHGSSISABSGQTVGVVSLGQDVHPIILVSLGQETWSKIPAGLE	296
SEQ_ID_NO_21	DKLEA-----AAFFPECHGSSISABSGQTVGVVSLGQDVHPIILVSLGQETWSKIPAGLE	296
	: : : : : : : : : : * : * : * : * : * : * : * : * : * : * : * : * : *	
SEQ_ID_NO_2	ISPVDSVLAPKNPPTGPAFTIINGTLKYFETRYTPVOIAPFISPHAVMSGTTERELL	356
SEQ_ID_NO_11	ITPVDSVLAPKNPPTGPAFTIINGTLKYFETRYTPVOIAPFISPHAVMSGTTERELL	356
SEQ_ID_NO_13	ISPVDSVLAPKNPPTGPAFTIINGTLKYFETRYTPVOIAPFISPHAVMSGTTERELL	356
SEQ_ID_NO_15	ISPVDSVLAPKNPPTGPAFTIINGTLKYFETRYTPVOIAPFISPHAVMSGTTERELL	356
SEQ_ID_NO_17	ISPVDSVLAPKNPPTGPAFTIINGTLKYFETRYTPVOIAPFISPHAVMSGTTERELL	356
SEQ_ID_NO_19	ISPVDSVLAPKNPPTGPAFTIINGTLKYFETRYTPVOIAPFISPHAVMSGTTERELL	356
SEQ_ID_NO_21	ISPVDSVLAPKNPPTGPAFTIINGTLKYFETRYTPVOIAPFISPHAVMSGTTERELL	356
	: : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *	
SEQ_ID_NO_2	WDWAPYEDVETGPHVIRTSQYKPELYMIGHGMLSDLLHSSKAQVFEHPHIQDAASQ	416
SEQ_ID_NO_11	WDWAPYEDVETGPHVIRTSQYKPELYMIGHGMLSDLLHSSKAQVFEHPHIQDAASQ	416
SEQ_ID_NO_13	WDWAPYEDVETGPHVIRTSQYKPELYMIGHGMLSDLLHSSKAQVFEHPHIQDAASQ	416
SEQ_ID_NO_15	WDWAPYEDVETGPHVIRTSQYKPELYMIGHGMLSDLLHSSKAQVFEHPHIQDAASQ	416
SEQ_ID_NO_17	WDWAPYEDVETGPHVIRTSQYKPELYMIGHGMLSDLLHSSKAQVFEHPHIQDAASQ	416
SEQ_ID_NO_19	WDWAPYEDVETGPHVIRTSQYKPELYMIGHGMLSDLLHSSKAQVFEHPHIQDAASQ	416
SEQ_ID_NO_21	WDWAPYEDVETGPHVIRTSQYKPELYMIGHGMLSDLLHSSKAQVFEHPHIQDAASQ	416
	* : * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *: * *	
SEQ_ID_NO_2	LP-----DDESLFPPTGGLSKNPIELVEGWSSSSISLSPFFITGLIIGLFLVLPVGIHAC	473
SEQ_ID_NO_11	LPKKDESLFPPTGGLSKNPIELVEGWSSSSISLSPFFITGLIIGLFLVLPVGIHAC	473
SEQ_ID_NO_13	LP-----DDESLFPPTGGLSKNPIELVEGWSSSSISLSPFFITGLIIGLFLVLPVGIHAC	473
SEQ_ID_NO_15	LP-----DDESLFPPTGGLSKNPIELVEGWSSSSISLSPFFITGLIIGLFLVLPVGIHAC	473
SEQ_ID_NO_17	LP-----DDESLFPPTGGLSKNPIELVEGWSSSSISLSPFFITGLIIGLFLVLPVGIHAC	473
SEQ_ID_NO_19	LP-----DDESLFPPTGGLSKNPIELVEGWSSSSISLSPFFITGLIIGLFLVLPVGIHAC	473
SEQ_ID_NO_21	LP-----DDESLFPPTGGLSKNPIELVEGWSSSSISLSPFFITGLIIGLFLVLPVGIHAC	473
	: : : : : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :	
SEQ_ID_NO_2	IKIKHTK-KRQYTDLENNRIGH-- 495	
SEQ_ID_NO_11	-SLFGK-PRTQKTYNDVEMSRIGH-- 501	
SEQ_ID_NO_13	YKNG-KRQYTDLENNRIGH-- 496	
SEQ_ID_NO_15	YKNG-KRQYTDLENNRIGH-- 496	</

FIG. 3

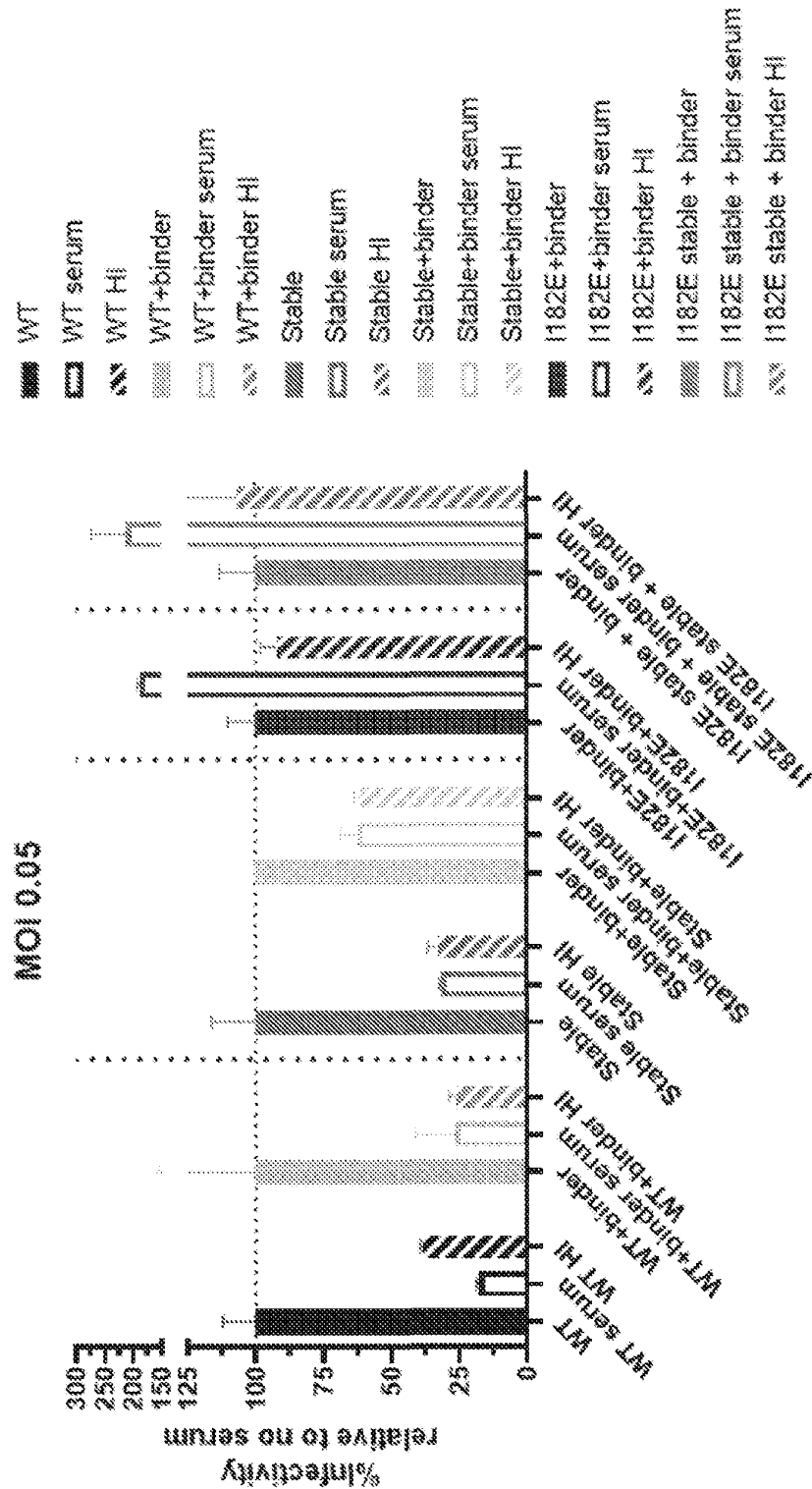


FIG. 4

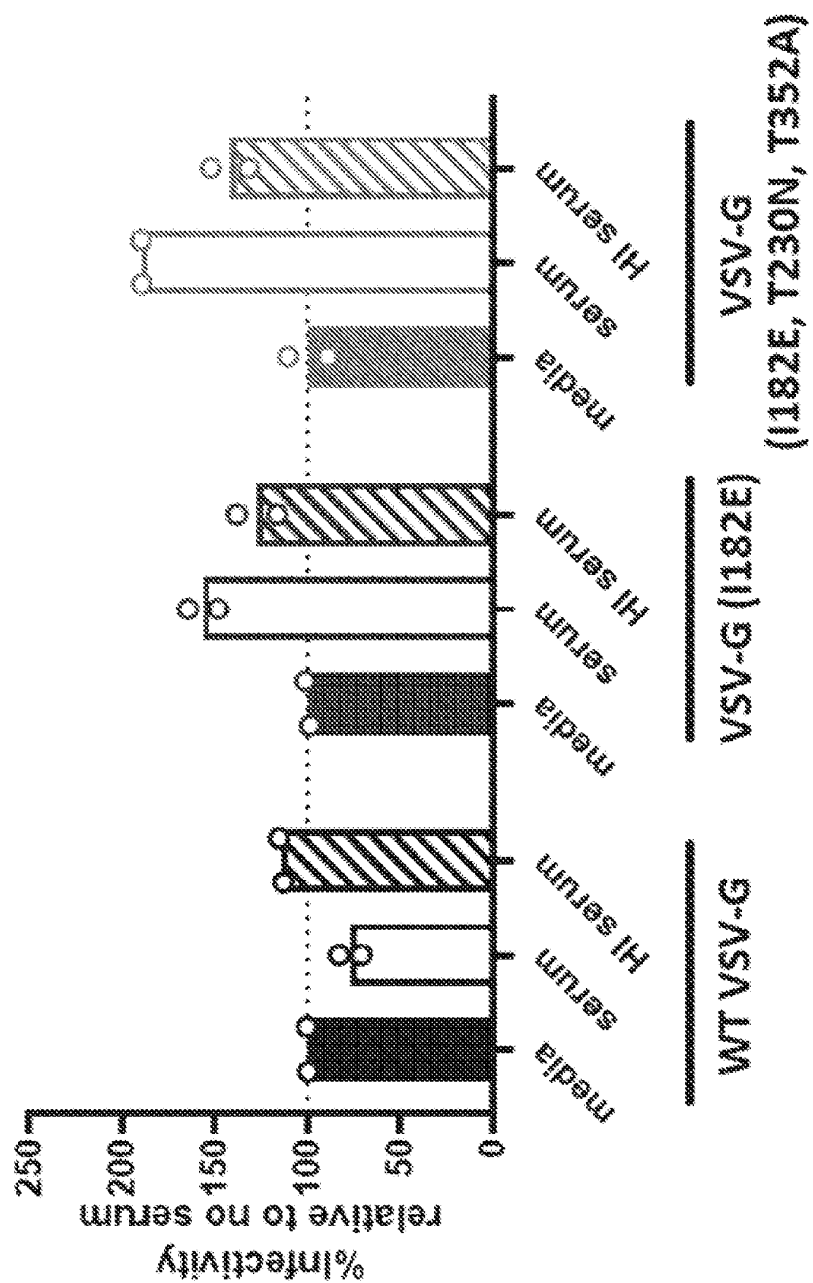


FIG. 5

With Paramyxoviridae glycoprotein

Example
Construct 1:
L-H

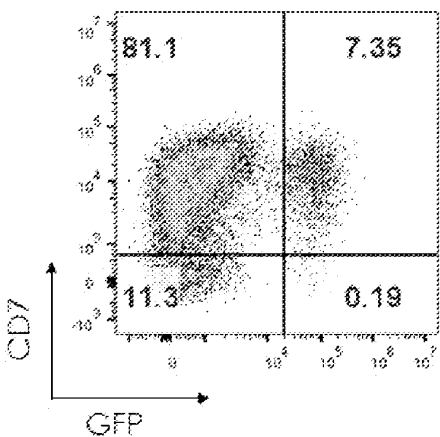


FIG. 6A

Example
Construct 2:
L-H

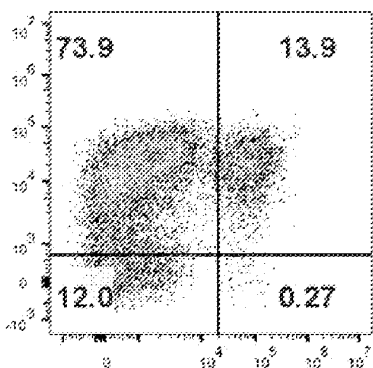


FIG. 6B

Example
Construct 3:
H-L

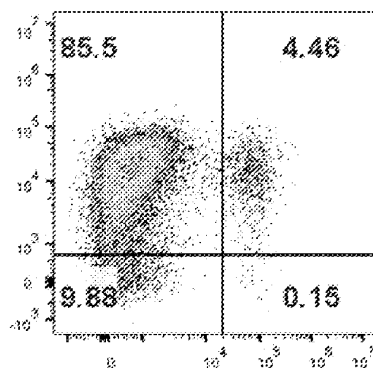


FIG. 6C

Example
Construct 4:
H-L

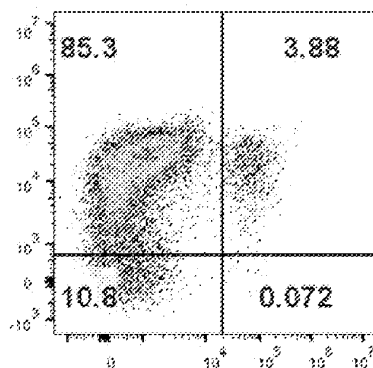


FIG. 6D

Binder Alone

Example
Construct 1:
L-H

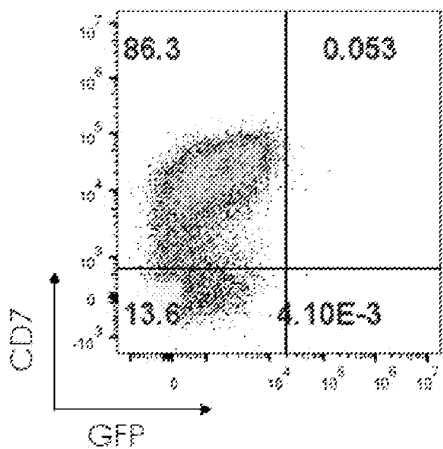


FIG. 6E

Example
Construct 2:
L-H

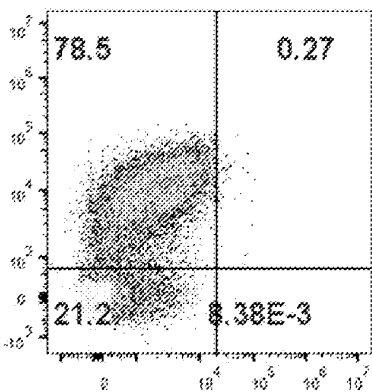


FIG. 6F

Example
Construct 3:
H-L

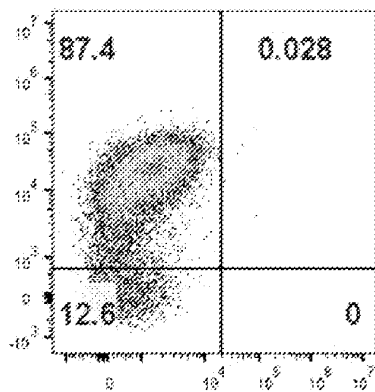


FIG. 6G

Example
Construct 4:
H-L

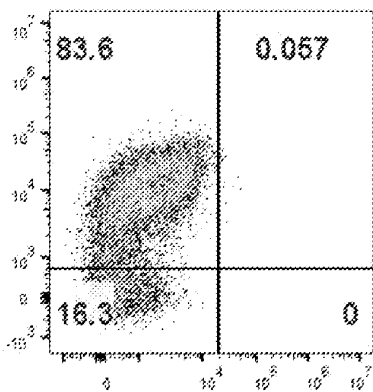


FIG. 6H

10/26

PATENT

With Rhabdoviridae glycoprotein

Example
Construct 1:
L-H

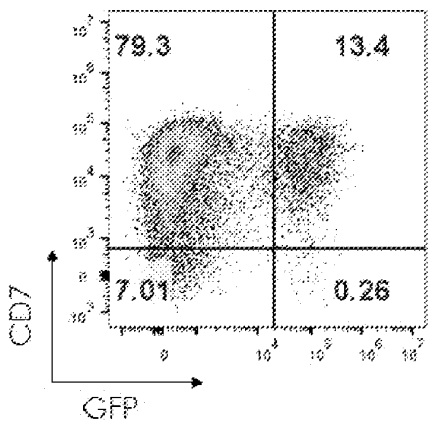


FIG. 6I

Example
Construct 2:
L-H

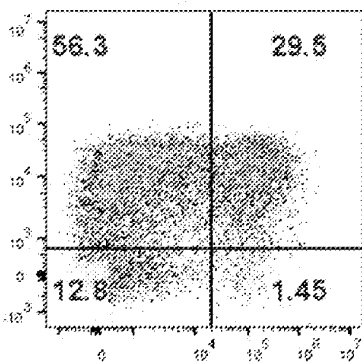


FIG. 6J

Example
Construct 3:
H-L

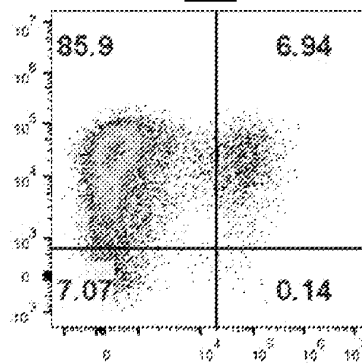


FIG. 6K

Example
Construct 4:
H-L

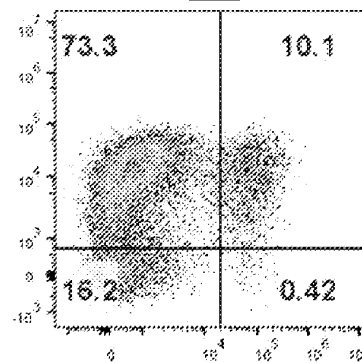


FIG. 6L

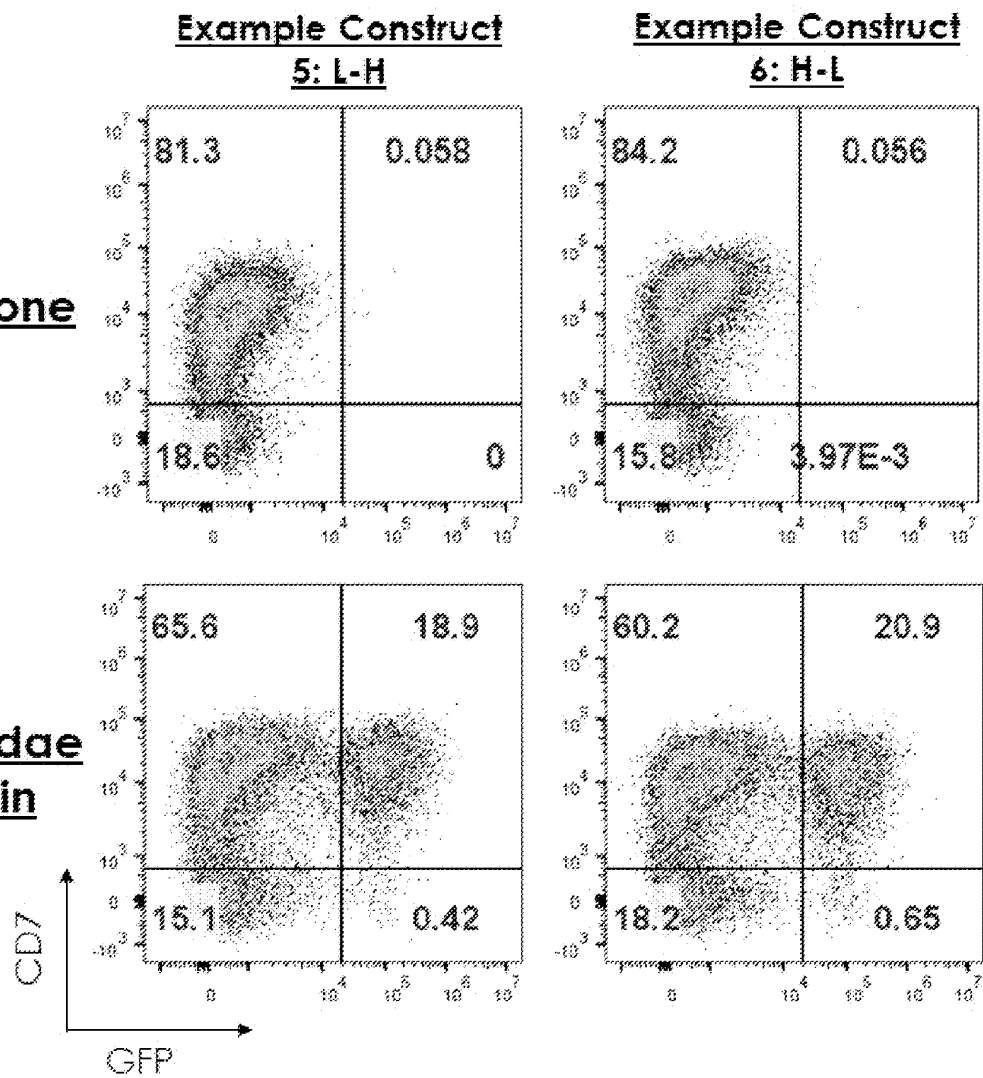


FIG. 6M

With Paramyxoviridae glycoprotein

Example
Construct 1:
L-H

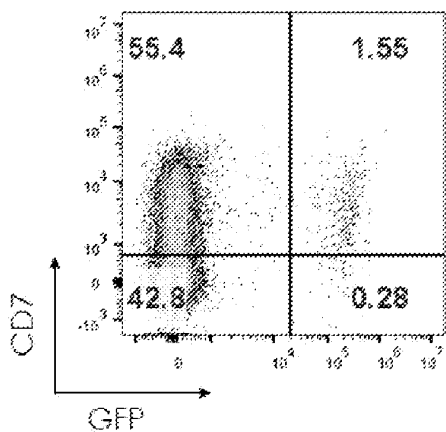


FIG. 7A

Example
Construct 2:
L-H

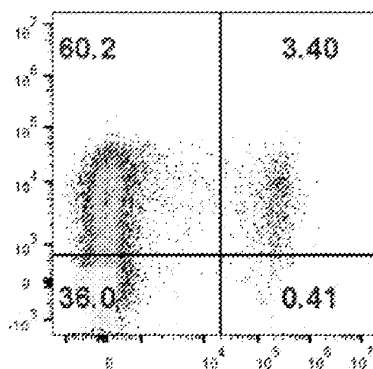


FIG. 7B

Example
Construct 3:
H-L

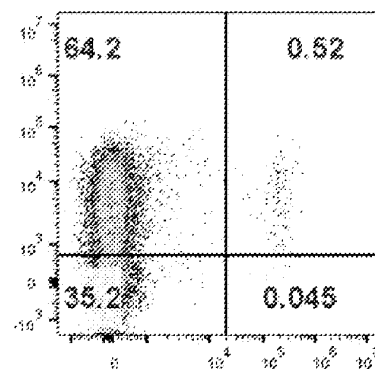


FIG. 7C

Example
Construct 4:
H-L

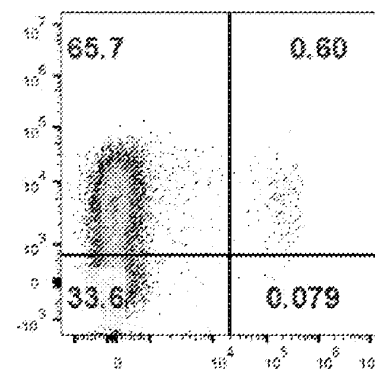


FIG. 7D

Binder Alone

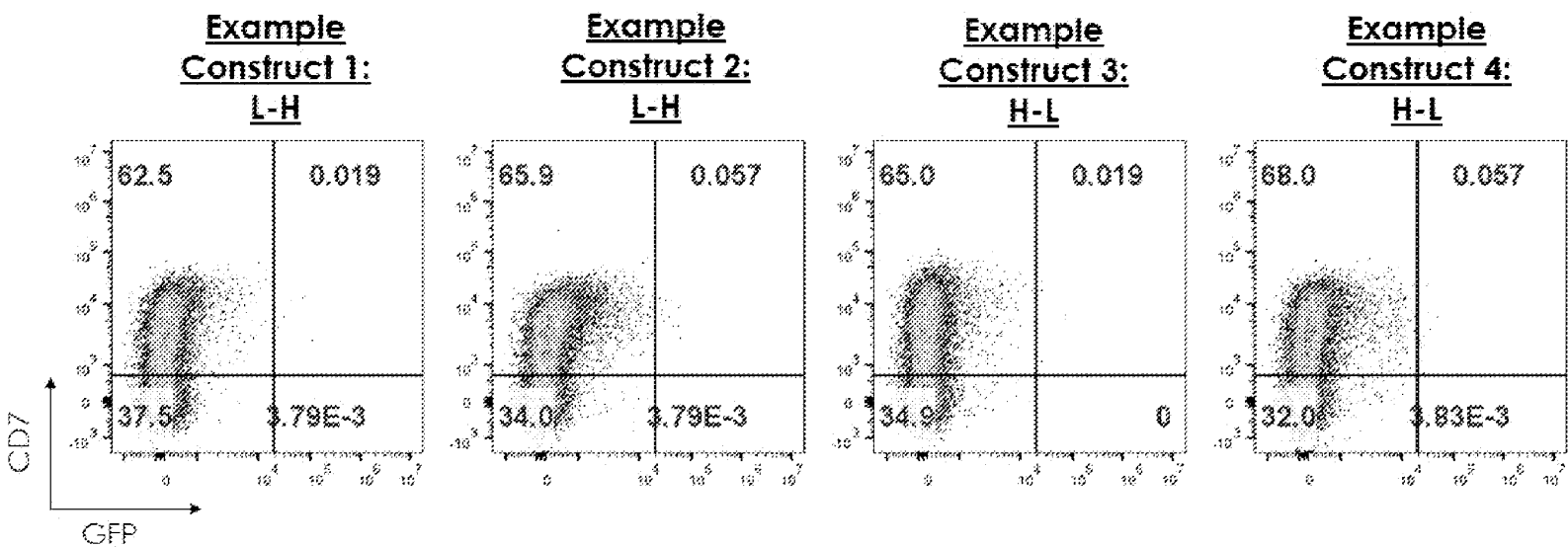


FIG. 7E

FIG. 7F

FIG. 7G

FIG. 7H

With Rhabdoviridae glycoprotein

Example
Construct 1:
L-H

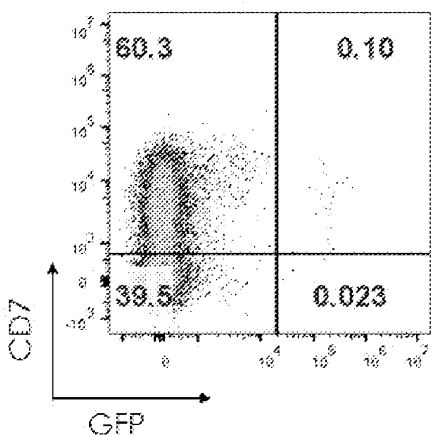


FIG. 7I

Example
Construct 2:
L-H

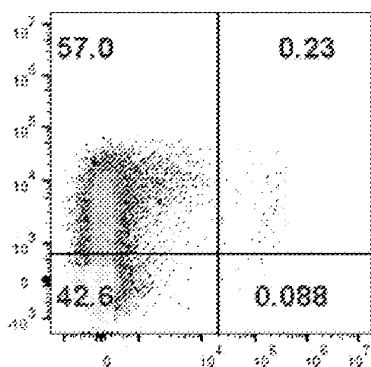


FIG. 7J

Example
Construct 3:
H-L

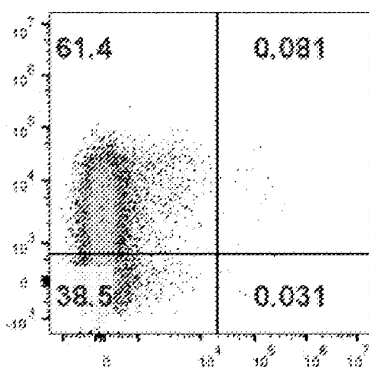


FIG. 7K

Example
Construct 4:
H-L

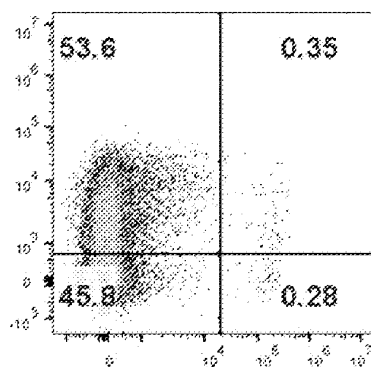


FIG. 7L

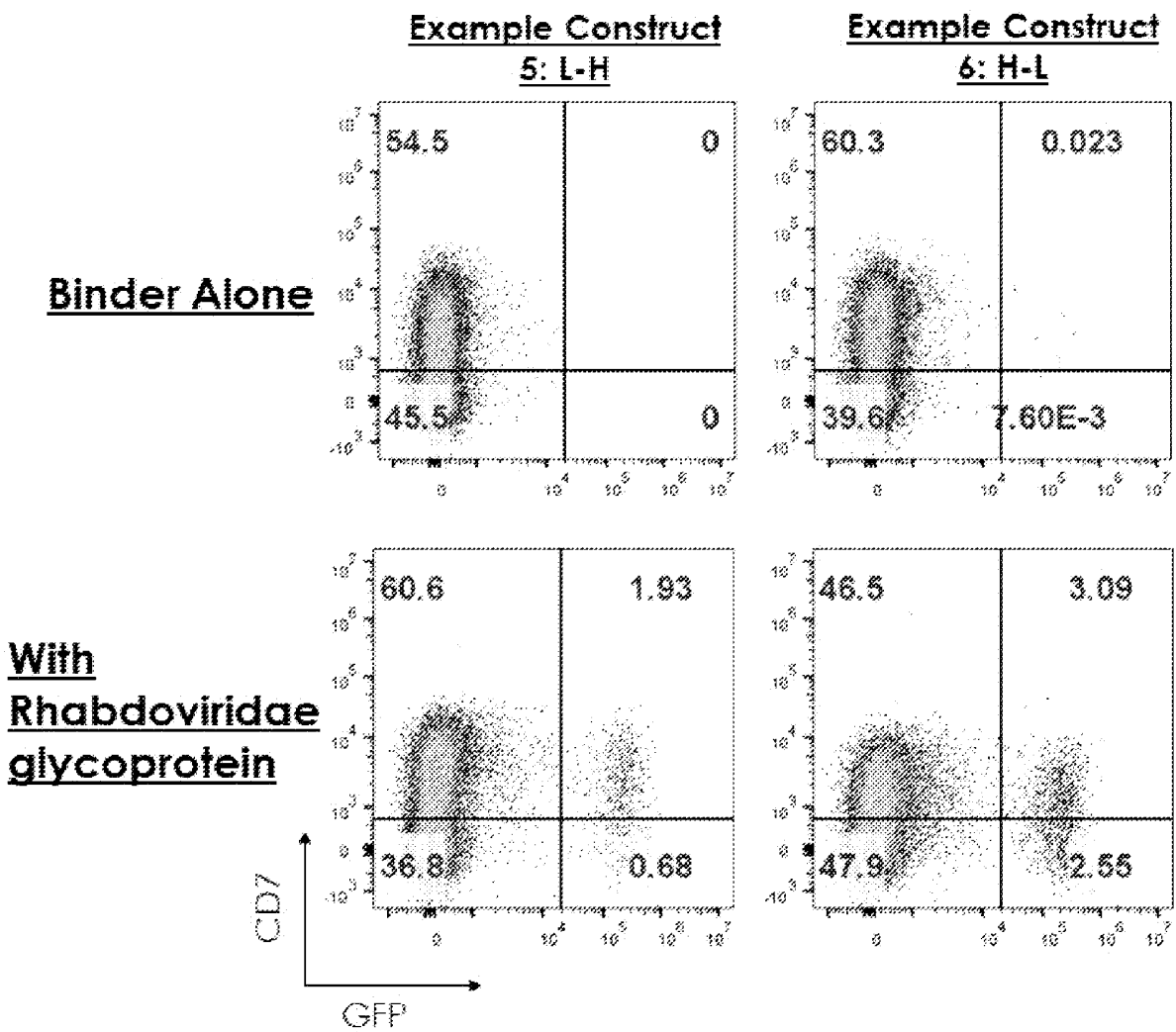


FIG. 7M

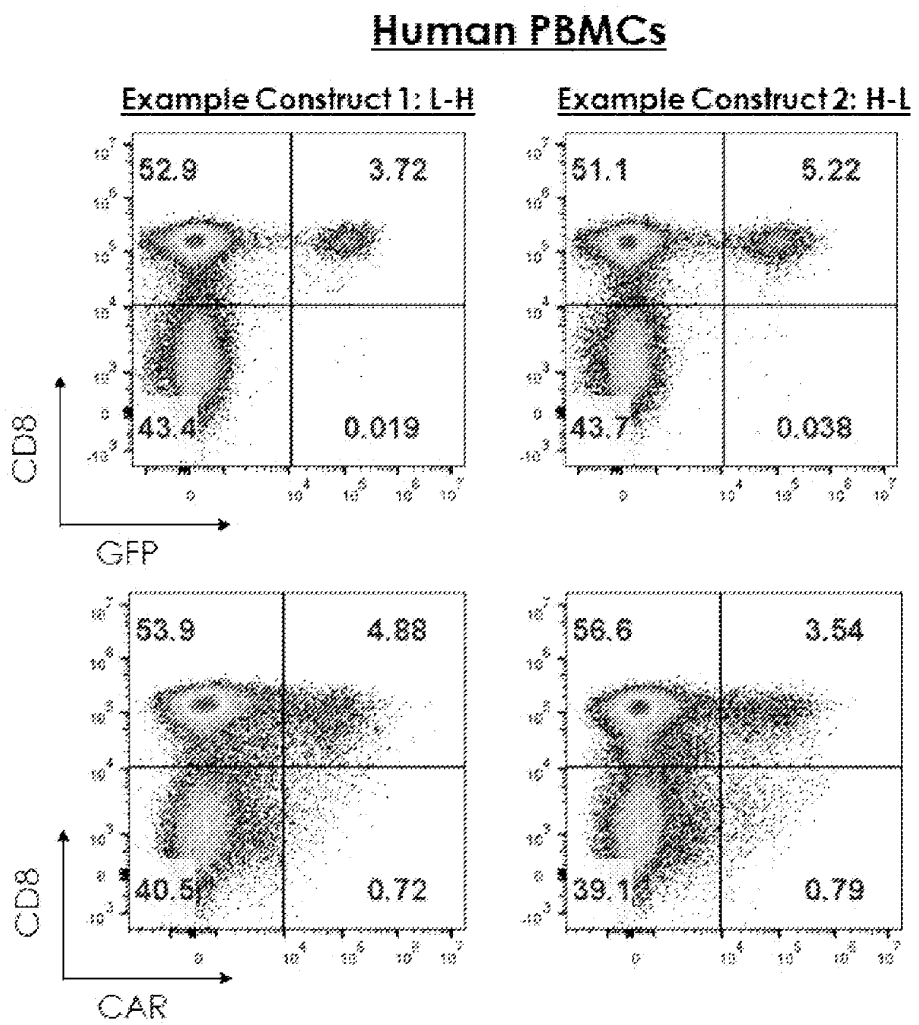


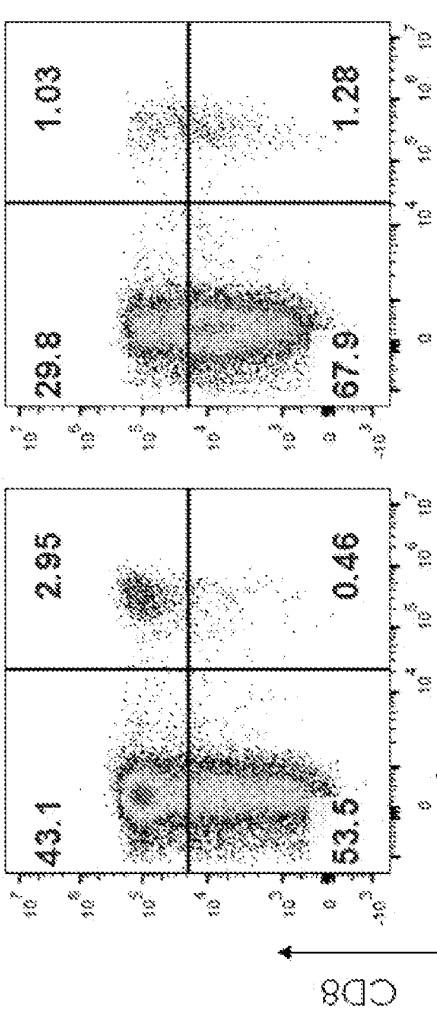
FIG. 8A

17/26

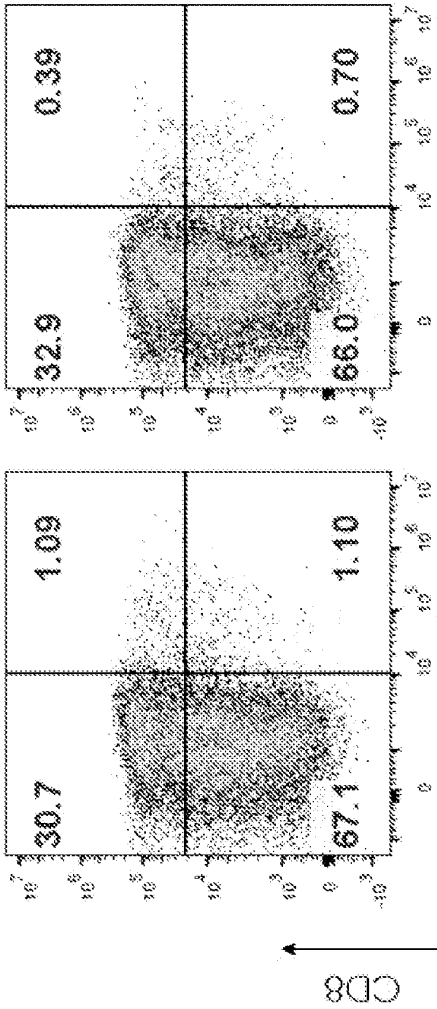
NHP PBMCs

Example Construct 1: L-H

Example Construct 2: H-L



GFP



CAR

FIG. 8B

18/26

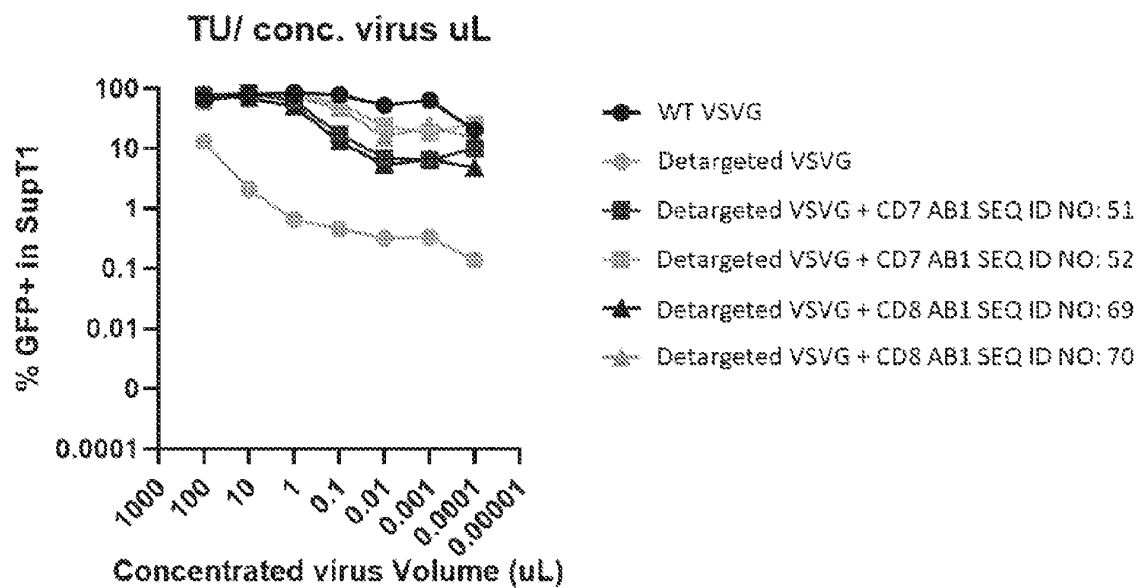


FIG. 9A

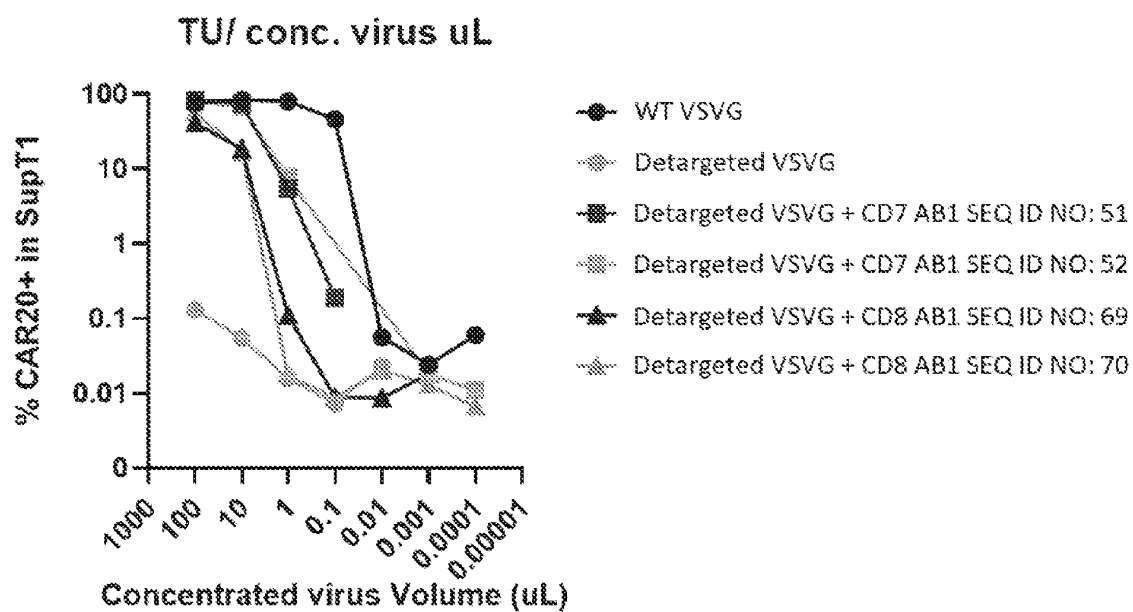


FIG. 9B

19/26

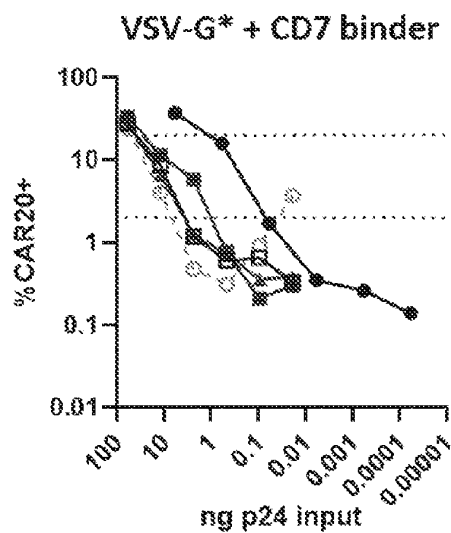


FIG. 10A

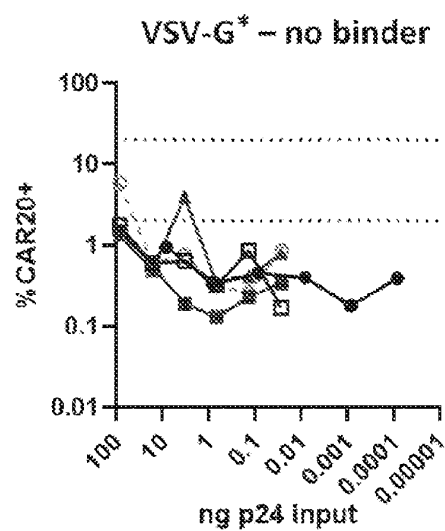


FIG. 10B

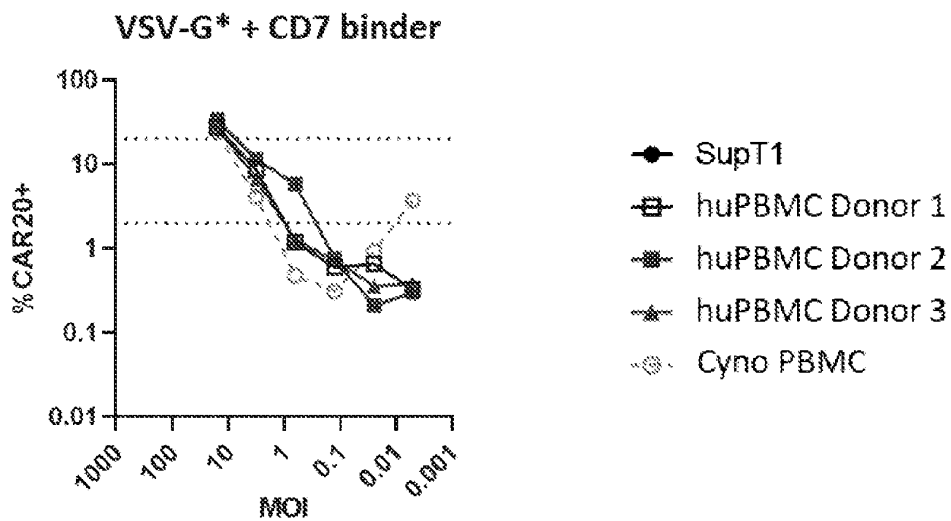


FIG. 10C

20/26

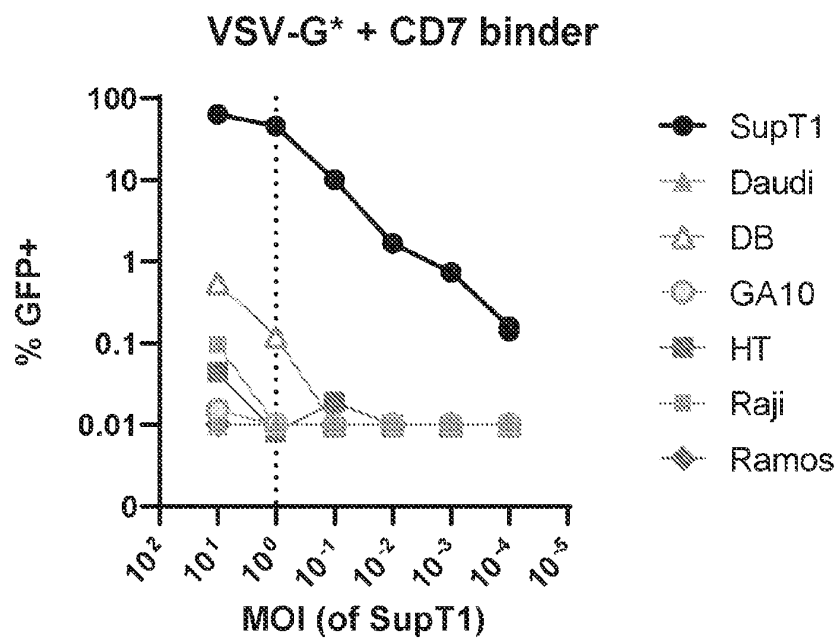


FIG. 11A

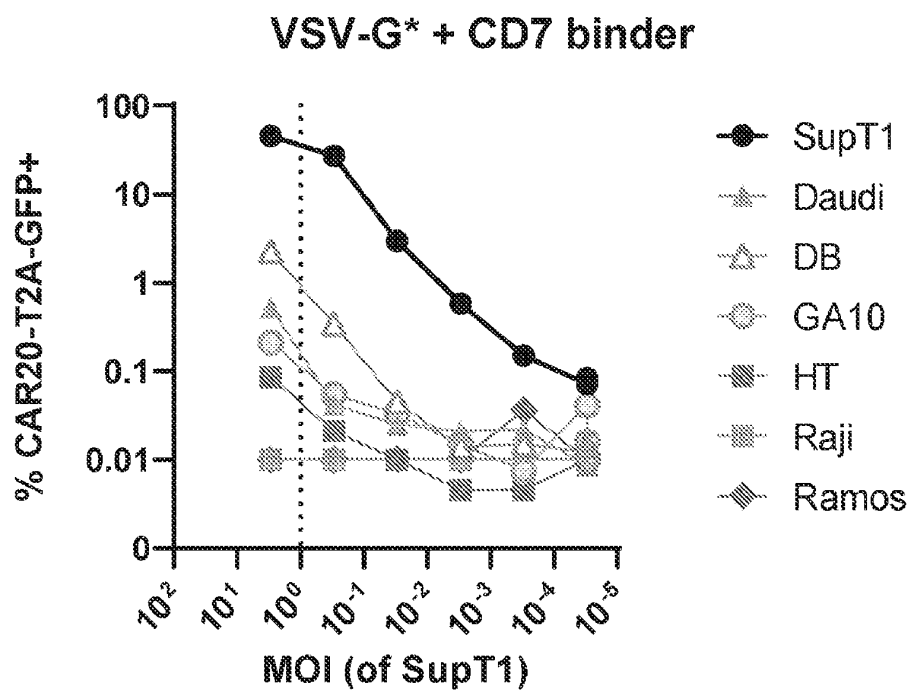


FIG. 11B

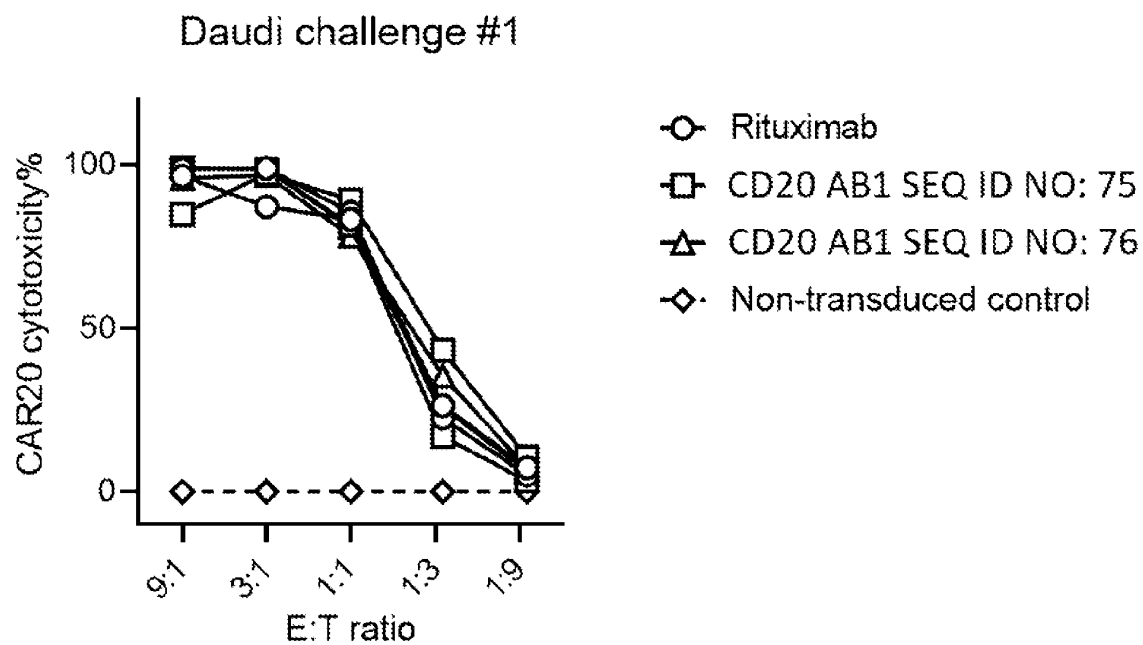


FIG. 12A

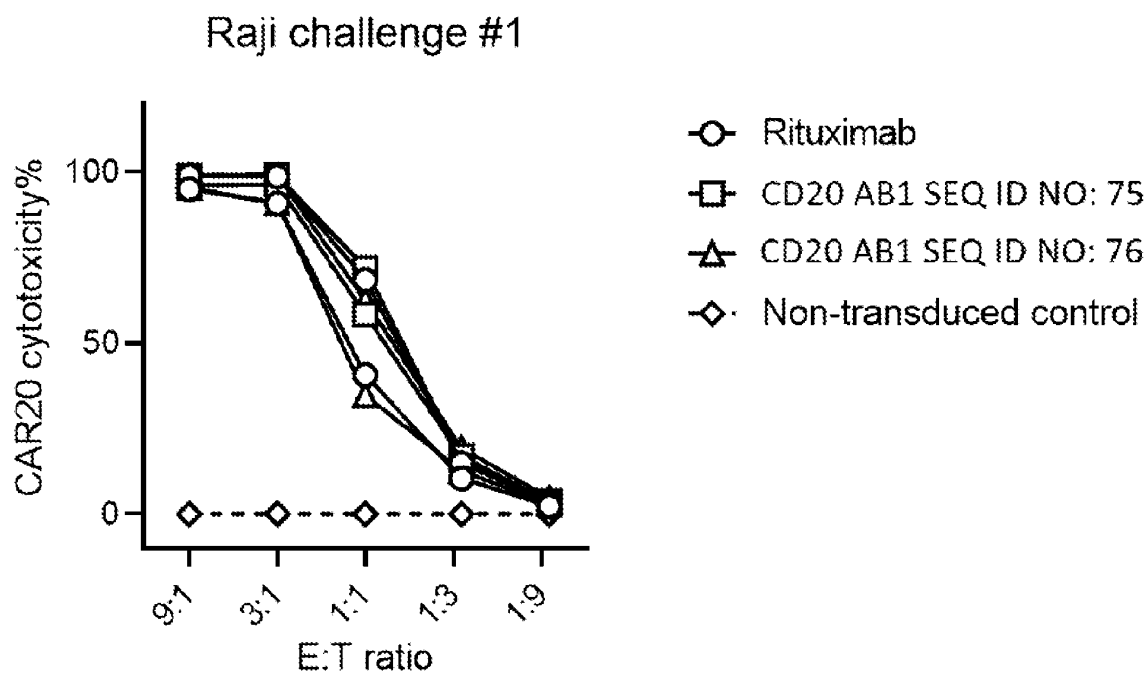


FIG. 12B

22/26

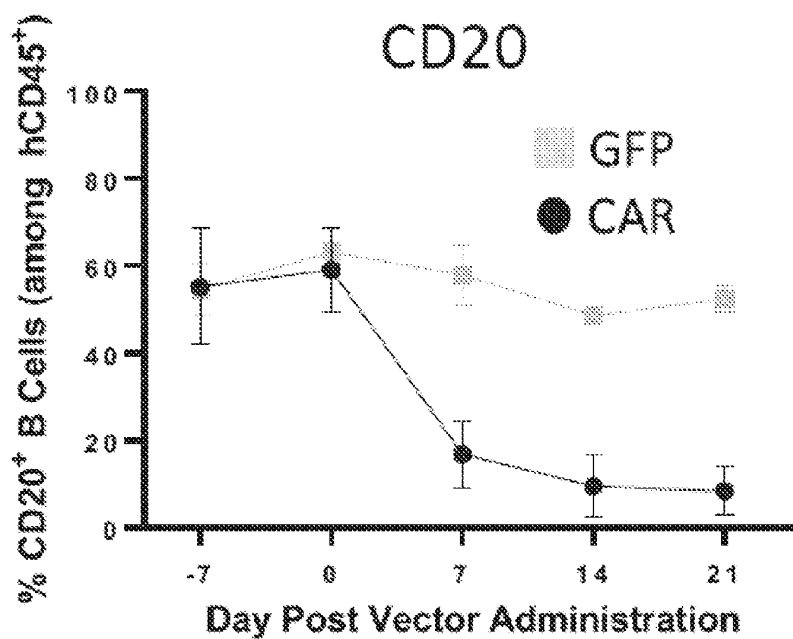


FIG. 13A

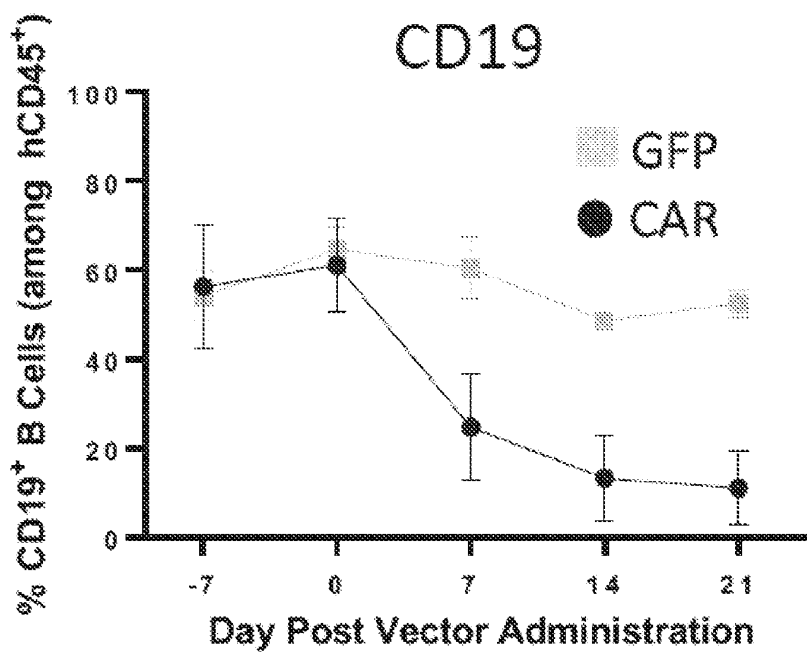


FIG. 13B

23/26

Experimental Design

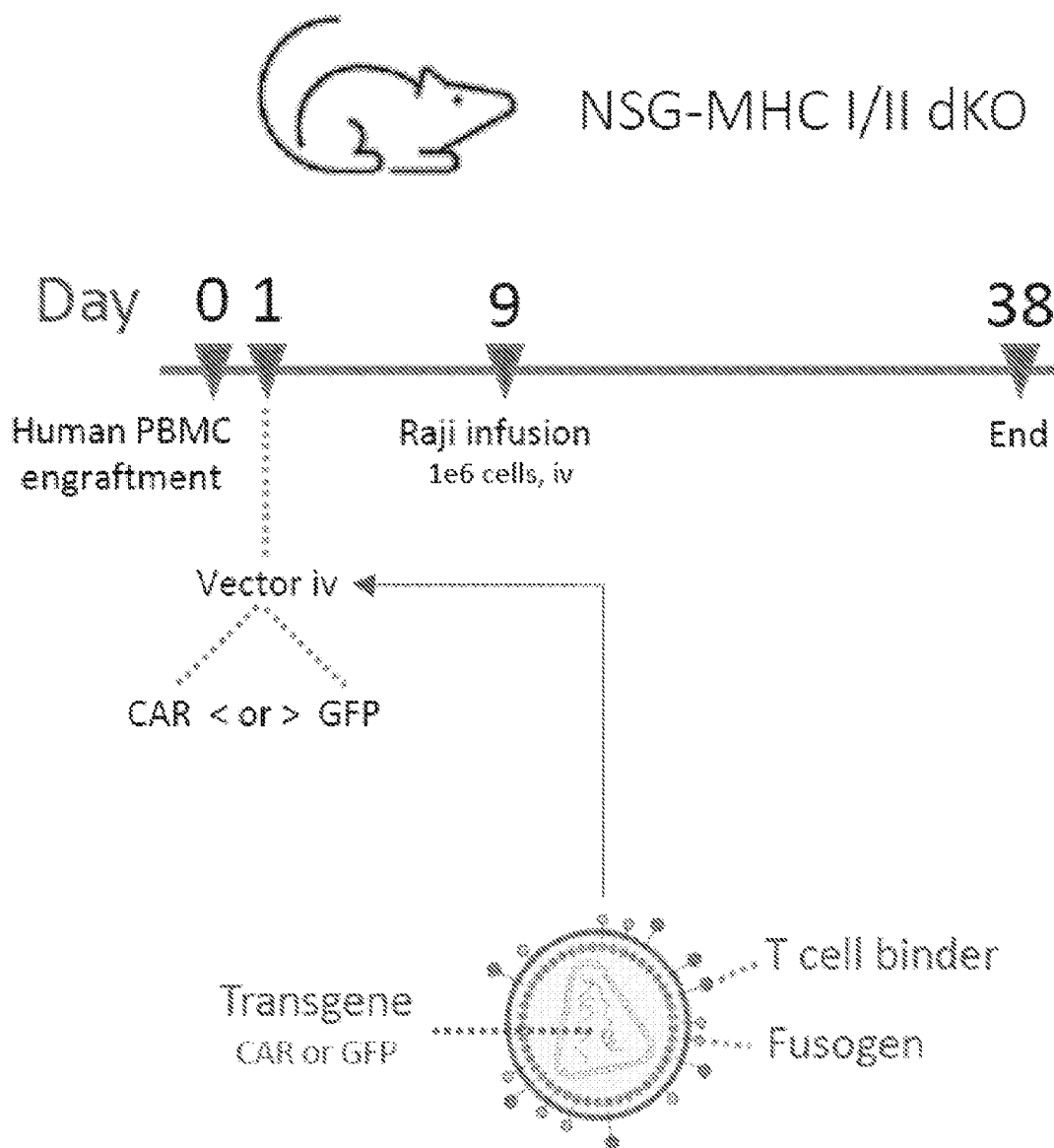


FIG. 14A

24/26

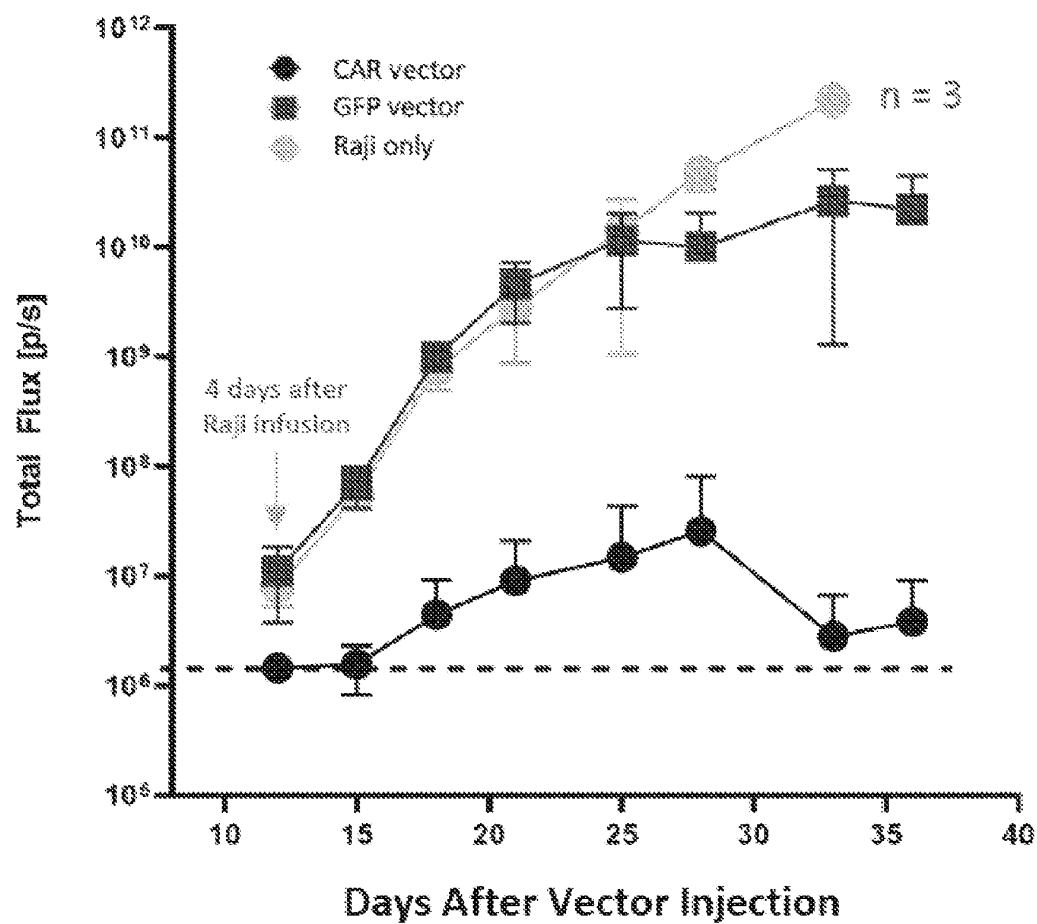
Raji tumor load (IVIS image)

FIG. 14B

25/26

Experimental Design

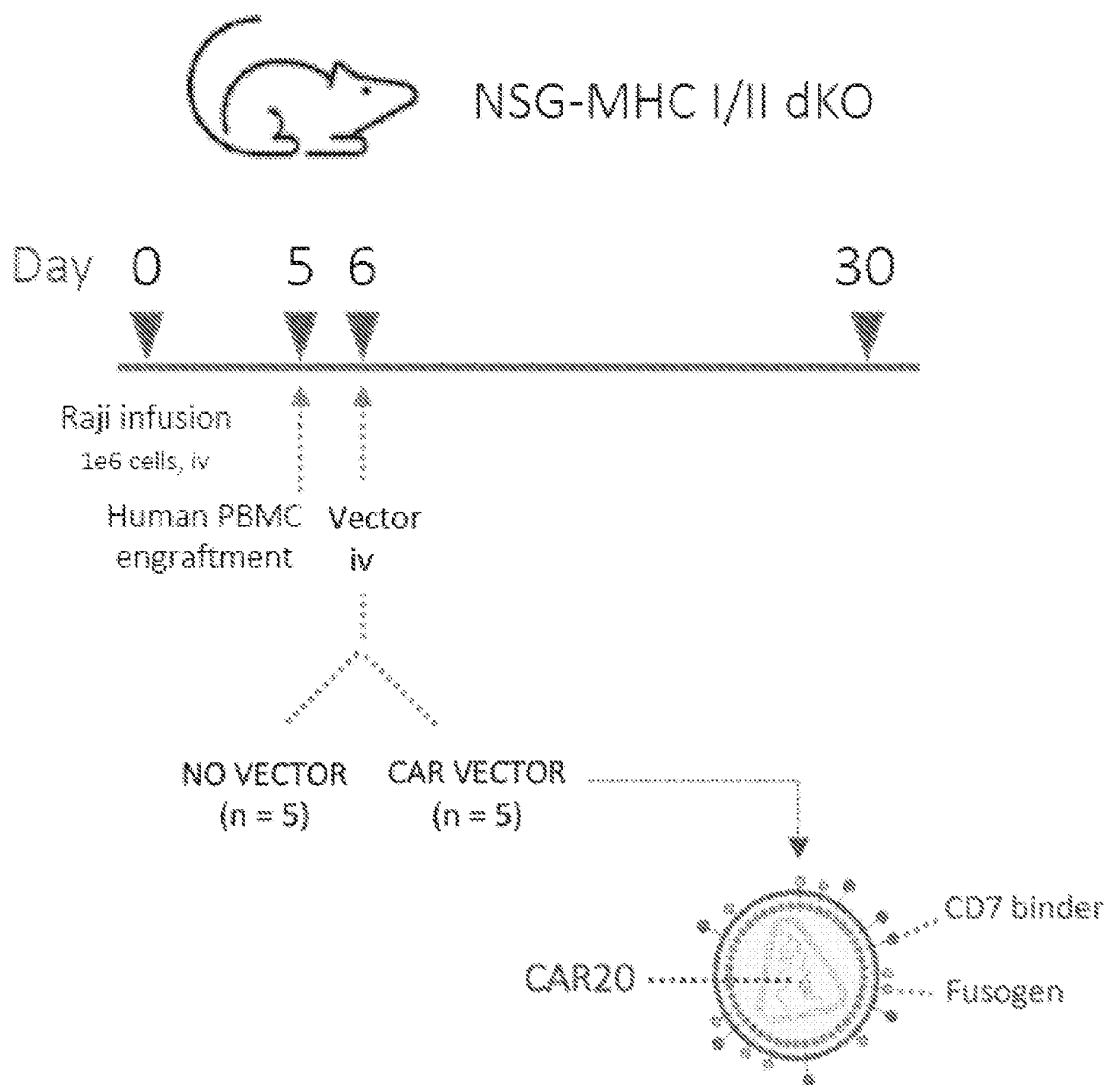


FIG. 15A

26/26

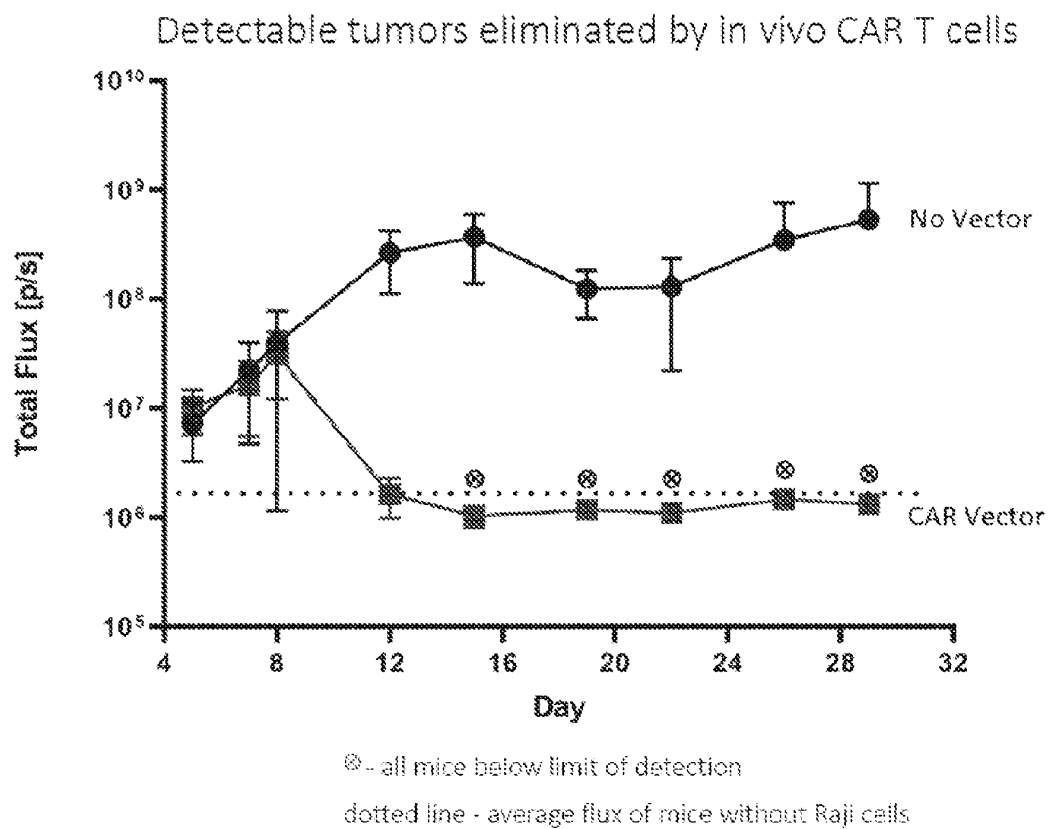


FIG. 15B

**CERTIFICATION AND REQUEST FOR PRIORITIZED EXAMINATION
UNDER 37 CFR 1.102(e) (Page 1 of 1)**

First Named Inventor:	Ronnie M. RUSSELL	Nonprovisional Application Number (if known):	
Title of Invention:	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF		

APPLICANT HEREBY CERTIFIES THE FOLLOWING AND REQUESTS PRIORITIZED EXAMINATION FOR THE ABOVE-IDENTIFIED APPLICATION.

1. The processing fee set forth in 37 CFR 1.17(i)(1) and the prioritized examination fee set forth in 37 CFR 1.17(c) have been filed with the request. The publication fee requirement is met because that fee, set forth in 37 CFR 1.18(d), is currently \$0. The basic filing fee, search fee, and examination fee are filed with the request or have been already been paid. I understand that any required excess claims fees or application size fee must be paid for the application.
2. I understand that the application may not contain, or be amended to contain, more than four independent claims, more than thirty total claims, or any multiple dependent claims, and that any request for an extension of time will cause an outstanding Track I request to be dismissed.
3. The applicable box is checked below:

I. ☒ Original Application (Track One) - Prioritized Examination under § 1.102(e)(1)

- i. (a) The application is an original nonprovisional utility application filed under 35 U.S.C. 111(a). This certification and request is being filed with the utility application via EFS-Web.
---OR---
(b) The application is an original nonprovisional plant application filed under 35 U.S.C. 111(a). This certification and request is being filed with the plant application in paper.
- ii. An executed inventor's oath or declaration under 37 CFR 1.63 or 37 CFR 1.64 for each inventor, or the application data sheet meeting the conditions specified in 37 CFR 1.53(f)(3)(i) is filed with the application.

II. ☐ Request for Continued Examination - Prioritized Examination under § 1.102(e)(2)

- i. A request for continued examination has been filed with, or prior to, this form.
- ii. If the application is a utility application, this certification and request is being filed via EFS-Web.
- iii. The application is an original nonprovisional utility application filed under 35 U.S.C. 111(a), or is a national stage entry under 35 U.S.C. 371.
- iv. This certification and request is being filed prior to the mailing of a first Office action responsive to the request for continued examination.
- v. No prior request for continued examination has been granted prioritized examination status under 37 CFR 1.102(e)(2).

Signature <u>/Christopher R. Bodle/</u>	Date <u>15-December-2022</u>
Name (Print/Typed) <u>Christopher R. Bodle</u>	Practitioner Registration Number <u>80,876</u>
Note: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. Submit multiple forms if more than one signature is required.*	
☒ *Total of <u>1</u> forms are submitted.	

Privacy Act Statement

The **Privacy Act of 1974 (P.L. 93-579)** requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary, and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 213(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (*i.e.*, GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	148165.001401
		Application Number	
Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF		
<p>The application data sheet is part of the provisional or nonprovisional application for which it is being submitted. The following form contains the bibliographic data arranged in a format specified by the United States Patent and Trademark Office as outlined in 37 CFR 1.76.</p> <p>This document may be completed electronically and submitted to the Office in electronic format using the Electronic Filing System (EFS) or the document may be printed and included in a paper filed application.</p>			

Secrecy Order 37 CFR 5.2:

☐ Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2 (Paper filers only. Applications that fall under Secrecy Order may not be filed electronically.)

Inventor Information:

Inventor 1		Remove	
Legal Name			
Prefix	Given Name	Middle Name	Family Name
	Ronnie	M.	Russell
Residence Information (Select One) ☒ US Residency ☐ Non US Residency ☐ Active US Military Service			
City	Philadelphia	State/Province	PA
Country of Residence		US	
Mailing Address of Inventor:			
Address 1		3401 Grays Ferry Avenue	
Address 2		Building 200, Suite 300	
City	Philadelphia	State/Province	PA
Postal Code	19146	Country	US
Inventor 2		Remove	
Legal Name			
Prefix	Given Name	Middle Name	Family Name
	Philip	R.	Johnson
Residence Information (Select One) ☒ US Residency ☐ Non US Residency ☐ Active US Military Service			
City	Philadelphia	State/Province	PA
Country of Residence		US	
Mailing Address of Inventor:			
Address 1		3401 Grays Ferry Avenue	
Address 2		Building 200, Suite 300	
City	Philadelphia	State/Province	PA
Postal Code	19146	Country	US
Inventor 1		Remove	
Legal Name			
Prefix	Given Name	Middle Name	Family Name
	James	I.	Andorko
Residence Information (Select One) ☒ US Residency ☐ Non US Residency ☐ Active US Military Service			

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number		148165.001401	
		Application Number			
Title of Invention		PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF			

City	Philadelphia	State/Province	PA	Country of Residence	US
------	--------------	----------------	----	----------------------	----

Mailing Address of Inventor:

Address 1	3401 Grays Ferry Avenue				
Address 2	Building 200, Suite 300				
City	Philadelphia	State/Province	PA		
Postal Code	19146	Country i	US		
Inventor	4				Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Mohit		Sachdeva	

Residence Information (Select One) ☒ US Residency ☐ Non US Residency ☐ Active US Military Service

City	Philadelphia	State/Province	PA	Country of Residence	US
------	--------------	----------------	----	----------------------	----

Mailing Address of Inventor:

Address 1	3401 Grays Ferry Avenue				
Address 2	Building 200, Suite 300				
City	Philadelphia	State/Province	PA		
Postal Code	19146	Country i	US		
Inventor	4				Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Bruce		Schnepp	

Residence Information (Select One) ☒ US Residency ☐ Non US Residency ☐ Active US Military Service

City	Havertown	State/Province	PA	Country of Residence	US
------	-----------	----------------	----	----------------------	----

Mailing Address of Inventor:

Address 1	3401 Grays Ferry Avenue				
Address 2	Building 200, Suite 300				
City	Philadelphia	State/Province	PA		
Postal Code	19146	Country i	US		
Inventor	4				Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Michael		Marinelli	

Residence Information (Select One) ☒ US Residency ☐ Non US Residency ☐ Active US Military Service

City	Philadelphia	State/Province	PA	Country of Residence	US
------	--------------	----------------	----	----------------------	----

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	148165.001401
		Application Number	
Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF		

Mailing Address of Inventor:				
Address 1	3401 Grays Ferry Avenue			
Address 2	Building 200, Suite 300			
City	Philadelphia	State/Province	PA	
Postal Code	19146	Country i	US	
Inventor	7			Remove
Legal Name				
Prefix	Given Name	Middle Name	Family Name	Suffix
	Jeffrey		Foss	
Residence Information (Select One) ☒ US Residency ☐ Non US Residency ☐ Active US Military Service				
City	Philadelphia	State/Province	PA	Country of Residence
				US

Mailing Address of Inventor:				
Address 1	3401 Grays Ferry Avenue			
Address 2	Building 200, Suite 300			
City	Philadelphia	State/Province	PA	
Postal Code	19146	Country i	US	
Inventor	8			Remove
Legal Name				
Prefix	Given Name	Middle Name	Family Name	Suffix
	Debasish		Boral	
Residence Information (Select One) ☒ US Residency ☐ Non US Residency ☐ Active US Military Service				
City	Philadelphia	State/Province	PA	Country of Residence
				US

Mailing Address of Inventor:				
Address 1	3401 Grays Ferry Avenue			
Address 2	Building 200, Suite 300			
City	Philadelphia	State/Province	PA	
Postal Code	19146	Country i	US	
All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the Add button.				

Correspondence Information:

Enter either Customer Number or complete the Correspondence Information section below. For further information see 37 CFR 1.33(a).	
☐ An Address is being provided for the correspondence information of this application.	
Customer Number	78442

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	148165.001401
		Application Number	
Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF		
Email Address	docketingTP@troutman.com	<input type="button" value="Add Email"/>	<input type="button" value="Remove Email"/>

Application Information:

Title of the Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF		
Attorney Docket Number	148165.001401	Small Entity Status Claimed	☒
Application Type	Nonprovisional		
Subject Matter	Utility		
Total Number of Drawing Sheets (if any)	26	Suggested Figure for Publication (if any)	

Filing By Reference:

Only complete this section when filing an application by reference under 35 U.S.C. 111(c) and 37 CFR 1.57(a). Do not complete this section if application papers including a specification and any drawings are being filed. Any domestic benefit or foreign priority information must be provided in the appropriate section(s) below (i.e., "Domestic Benefit/National Stage Information" and "Foreign Priority Information").

For the purposes of a filing date under 37 CFR 1.53(b), the description and any drawings of the present application are replaced by this reference to the previously filed application, subject to conditions and requirements of 37 CFR 1.57(a).

Application number of the previously filed application	Filing date (YYYY-MM-DD)	Intellectual Property Authority or Country

Publication Information:

☐ Request Early Publication (Fee required at time of Request 37 CFR 1.219)
☐ Request Not to Publish. I hereby request that the attached application not be published under 35 U.S.C. 122(b) and certify that the invention disclosed in the attached application has not and will not be the subject of an application filed in another country, or under a multilateral international agreement, that requires publication at eighteen months after filing.

Representative Information:

Representative information should be provided for all practitioners having a power of attorney in the application. Providing this information in the Application Data Sheet does not constitute a power of attorney in the application (see 37 CFR 1.32). Either enter Customer Number or complete the Representative Name section below. If both sections are completed the customer Number will be used for the Representative Information during processing.			
Please Select One:	☒ Customer Number	☐ US Patent Practitioner	☐ Limited Recognition (37 CFR 11.9)
Customer Number	78442		

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	148165.001401
		Application Number	
Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF		

Domestic Benefit/National Stage Information:

This section allows for the applicant to either claim benefit under 35 U.S.C. 119(e), 120, 121, 365(c), or 386(c) or indicate National Stage entry from a PCT application. Providing benefit claim information in the Application Data Sheet constitutes the specific reference required by 35 U.S.C. 119(e) or 120, and 37 CFR 1.78.

When referring to the current application, please leave the "Application Number" field blank.

Prior Application Status	Pending	Remove	
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
	Claims benefit of provisional	63/289888	2021-12-15
Prior Application Status	Pending	Remove	
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
	Claims benefit of provisional	63/289977	2021-12-15
Prior Application Status	Pending	Remove	
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
	Claims benefit of provisional	63/266044	2021-12-27
Prior Application Status	Pending	Remove	
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
	Claims benefit of provisional	63/267039	2022-01-21
Additional Domestic Benefit/National Stage Data may be generated within this form by selecting the Add button.			Add

Foreign Priority Information:

This section allows for the applicant to claim priority to a foreign application. Providing this information in the application data sheet constitutes the claim for priority as required by 35 U.S.C. 119(b) and 37 CFR 1.55. When priority is claimed to a foreign application that is eligible for retrieval under the priority document exchange program (PDX)ⁱ the information will be used by the Office to automatically attempt retrieval pursuant to 37 CFR 1.55(i)(1) and (2). Under the PDX program, applicant bears the ultimate responsibility for ensuring that a copy of the foreign application is received by the Office from the participating foreign intellectual property office, or a certified copy of the foreign priority application is filed, within the time period specified in 37 CFR 1.55(g)(1).

Application Number	Country ^j	Filing Date (YYYY-MM-DD)	Remove
			Access Code ^j (if applicable)
Additional Foreign Priority Data may be generated within this form by selecting the Add button.			Add

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	148165.001401
		Application Number	
Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF		

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications

<p>This application (1) claims priority to or the benefit of an application filed before March 16, 2013 and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013.</p> <p>☐ NOTE: By providing this statement under 37 CFR 1.55 or 1.78, this application, with a filing date on or after March 16, 2013, will be examined under the first inventor to file provisions of the AIA.</p>
--

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	148165.001401
		Application Number	
Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF		

Authorization or Opt-Out of Authorization to Permit Access:

When this Application Data Sheet is properly signed and filed with the application, applicant has provided written authority to permit a participating foreign intellectual property (IP) office access to the instant application-as-filed (see paragraph A in subsection 1 below) and the European Patent Office (EPO) access to any search results from the instant application (see paragraph B in subsection 1 below).

Should applicant choose not to provide an authorization identified in subsection 1 below, applicant **must opt-out** of the authorization by checking the corresponding box A or B or both in subsection 2 below.

NOTE: This section of the Application Data Sheet is **ONLY** reviewed and processed with the **INITIAL** filing of an application. After the initial filing of an application, an Application Data Sheet cannot be used to provide or rescind authorization for access by a foreign IP office(s). Instead, Form PTO/SB/39 or PTO/SB/69 must be used as appropriate.

1. Authorization to Permit Access by a Foreign Intellectual Property Office(s)

A. Priority Document Exchange (PDX) - Unless box A in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office of the People's Republic of China (SIPO), the World Intellectual Property Organization (WIPO), and any other foreign intellectual property office participating with the USPTO in a bilateral or multilateral priority document exchange agreement in which a foreign application claiming priority to the instant patent application is filed, access to: (1) the instant patent application-as-filed and its related bibliographic data, (2) any foreign or domestic application to which priority or benefit is claimed by the instant application and its related bibliographic data, and (3) the date of filing of this Authorization. See 37 CFR 1.14(h)(1).

B. Search Results from U.S. Application to EPO - Unless box B in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the EPO access to the bibliographic data and search results from the instant patent application when a European patent application claiming priority to the instant patent application is filed. See 37 CFR 1.14(h)(2).

The applicant is reminded that the EPO's Rule 141(1) EPC (European Patent Convention) requires applicants to submit a copy of search results from the instant application without delay in a European patent application that claims priority to the instant application.

2. Opt-Out of Authorizations to Permit Access by a Foreign Intellectual Property Office(s)

☐ A. Applicant **DOES NOT** authorize the USPTO to permit a participating foreign IP office access to the instant application-as-filed. If this box is checked, the USPTO will not be providing a participating foreign IP office with any documents and information identified in subsection 1A above.

☐ B. Applicant **DOES NOT** authorize the USPTO to transmit to the EPO any search results from the instant patent application. If this box is checked, the USPTO will not be providing the EPO with search results from the instant application.

NOTE: Once the application has published or is otherwise publicly available, the USPTO may provide access to the application in accordance with 37 CFR 1.14.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	148165.001401
		Application Number	
Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF		

Applicant Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.			
Applicant	1		<input type="button" value="Remove"/>
<p>If the applicant is the inventor (or the remaining joint inventor or inventors under 37 CFR 1.45), this section should not be completed. The information to be provided in this section is the name and address of the legal representative who is the applicant under 37 CFR 1.43; or the name and address of the assignee, person to whom the inventor is under an obligation to assign the invention, or person who otherwise shows sufficient proprietary interest in the matter who is the applicant under 37 CFR 1.46. If the applicant is an applicant under 37 CFR 1.46 (assignee, person to whom the inventor is obligated to assign, or person who otherwise shows sufficient proprietary interest) together with one or more joint inventors, then the joint inventor or inventors who are also the applicant should be identified in this section.</p>			
<input type="button" value="Clear"/>			
☒ Assignee	☐ Legal Representative under 35 U.S.C. 117		☐ Joint Inventor
Person to whom the inventor is obligated to assign.		Person who shows sufficient proprietary interest	
If applicant is the legal representative, indicate the authority to file the patent application, the inventor is:			
<div style="border: 1px solid black; height: 20px; width: 100%;"></div>			
Name of the Deceased or Legally Incapacitated Inventor: <div style="border: 1px solid black; height: 20px; width: 100%;"></div>			
If the Applicant is an Organization check here. ☒			
Organization Name	INTERIUS BIOTHERAPEUTICS, INC.		
Mailing Address Information For Applicant:			
Address 1	3401 Grays Ferry Avenue		
Address 2	Building 200, Suite 300		
City	Philadelphia	State/Province	PA
Country	US	Postal Code	19146
Phone Number		Fax Number	
Email Address			
Additional Applicant Data may be generated within this form by selecting the Add button. <input type="button" value="Add"/>			

Assignee Information including Non-Applicant Assignee Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	148165.001401
		Application Number	
Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF		

Assignee	1			
Complete this section if assignee information, including non-applicant assignee information, is desired to be included on the patent application publication. An assignee-applicant identified in the "Applicant Information" section will appear on the patent application publication as an applicant. For an assignee-applicant, complete this section only if identification as an assignee is also desired on the patent application publication.				
Remove				
If the Assignee or Non-Applicant Assignee is an Organization check here. ☐				
Prefix	Given Name	Middle Name	Family Name	Suffix
Mailing Address Information For Assignee including Non-Applicant Assignee:				
Address 1				
Address 2				
City		State/Province		
Country i		Postal Code		
Phone Number		Fax Number		
Email Address				
Additional Assignee or Non-Applicant Assignee Data may be generated within this form by selecting the Add button. Add				

Signature:	Remove			
NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b). However, if this Application Data Sheet is submitted with the INITIAL filing of the application and either box A or B is not checked in subsection 2 of the "Authorization or Opt-Out of Authorization to Permit Access" section, then this form must also be signed in accordance with 37 CFR 1.14(c).				
This Application Data Sheet must be signed by a patent practitioner if one or more of the applicants is a juristic entity (e.g., corporation or association). If the applicant is two or more joint inventors, this form must be signed by a patent practitioner, all joint inventors who are the applicant, or one or more joint inventor-applicants who have been given power of attorney (e.g., see USPTO Form PTO/AIA/81) on behalf of all joint inventor-applicants.				
See 37 CFR 1.4(d) for the manner of making signatures and certifications.				
Signature	/Christopher R. Bodle/		Date (YYYY-MM-DD)	2022-12-15
First Name	Christopher R.	Last Name	Bodle	Registration Number
				80,876
Additional Signature may be generated within this form by selecting the Add button. Add				

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76		Attorney Docket Number	148165.001401
		Application Number	
Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF		

This collection of information is required by 37 CFR 1.76. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 23 minutes to complete, including gathering, preparing, and submitting the completed application data sheet form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.



UNITED STATES
PATENT AND TRADEMARK OFFICE

P.O. Box 1450
Alexandria, VA 22313 - 1450
www.uspto.gov

ELECTRONIC PAYMENT RECEIPT

APPLICATION #
18/066,420

RECEIPT DATE / TIME
12/15/2022 12:02:21 PM ET

ATTORNEY DOCKET #
148165.001401

Title of Invention

PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF

Application Information

APPLICATION TYPE Utility - Nonprovisional Application
under 35 USC 111(a)

PATENT # -

CONFIRMATION # 1909

FILED BY Melissa Seebaran

PATENT CENTER # 61303902

AUTHORIZED BY Christopher Bodie

CUSTOMER # 78442

FILING DATE -

CORRESPONDENCE ADDRESS -

FIRST NAMED INVENTOR Ronnie M. Russell

Payment Information

PAYMENT METHOD
DA / 500436

PAYMENT TRANSACTION ID
E2022BEC03262358

PAYMENT AUTHORIZED BY
Melissa Seebaran

PRE-AUTHORIZED ACCOUNT
500436

PRE-AUTHORIZED CATEGORY
37 CFR 1.16 (National application filing, search, and examination fees); 37 CFR 1.17 (Patent application and reexamination processing fees); 37 CFR 1.19 (Document supply fees); 37 CFR 1.20 (Post Issuance fees); 37 CFR 1.21 (Miscellaneous fees and charges)

FEE CODE	DESCRIPTION	ITEM PRICE(\$)	QUANTITY	ITEM TOTAL(\$)
2830	PROCESSING FEE, EXCEPT IN PROVISIONAL APPLICATIONS	70.00	1	70.00
2817	REQUEST FOR PRIORITIZED EXAMINATION	2100.00	1	2100.00
2111	UTILITY PATENT APPL. SEARCH FEE	350.00	1	350.00
4011	BASIC FILING FEE- UTILITY	80.00	1	80.00

2051	SURCHARGE- LATE FILING FEE, SEARCH FEE, EXAMINATION FEE, INVENTOR'S OATH OR DECLARATION, OR APPLICATION FILED WITHOUT AT LEAST ONE CLAIM OR BY REFERENCE	80.00	1	80.00
2311	EXAMINATION OF ORIGINAL PATENT APPLICATION	400.00	1	400.00
2081	UTILITY APPL. SIZE FEE > THAN 100 PAGES (FEE X PAGES >= 101 IN 50 PAGE INCREMENTS)	210.00	2	420.00
2202	EACH CLAIM IN EXCESS OF 20	50.00	2	100.00
TOTAL AMOUNT:				\$3,600.00

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.



ELECTRONIC ACKNOWLEDGEMENT RECEIPT

APPLICATION #
18/066,420

RECEIPT DATE / TIME
12/15/2022 12:02:21 PM ET

ATTORNEY DOCKET #
148165.001401

Title of Invention

PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF

Application Information

APPLICATION TYPE Utility - Nonprovisional Application
under 35 USC 111(a)

PATENT # -

CONFIRMATION # 1909

FILED BY Melissa Seebaran

PATENT CENTER # 61303902

FILING DATE -

CUSTOMER # 78442

FIRST NAMED INVENTOR Ronnie M. Russell

CORRESPONDENCE ADDRESS -

AUTHORIZED BY Christopher Bodle

Documents

TOTAL DOCUMENTS: 10

DOCUMENT	PAGES	DESCRIPTION	SIZE (KB)
148165_001401_ADS_as-filed-15-Dec-2022.pdf	11	Application Data Sheet	1227 KB
148165.001401_SL.XML	-	Sequence Listing (ST.26 XML File)	86 KB
148165_001401_Track1_Request_as-filed-15-Dec-2022.pdf	2	Track One Request	112 KB
148165_001401_Preliminary__Amendment_as_filed_15-Dec-2022.pdf	6	-	37 KB

148165_001401_Preliminary_Amendment_as_filed_15-Dec-2022-A.PE.pdf	(1-1)	1	Preliminary Amendment	21 KB
148165_001401_Preliminary_Amendment_as_filed_15-Dec-2022-CLM.pdf	(2-4)	3	Claims	22 KB
148165_001401_Preliminary_Amendment_as_filed_15-Dec-2022-REM.pdf	(5-6)	2	Applicant Arguments/Remarks Made in an Amendment	23 KB
148165_001401_Drawings_as_filed_15-Dec-2022.pdf		26	Drawings-only black and white line drawings	4257 KB
148165_001401_US_NP_Application_as-filed_15-Dec-2022.pdf		188	-	938 KB
148165_001401_US_NP_Application_as-filed_15-Dec-2022-SPEC.pdf	(1-159)	159	Specification	796 KB
148165_001401_US_NP_Application_as-filed_15-Dec-2022-CLM.pdf	(160-187)	28	Claims	136 KB
148165_001401_US_NP_Application_as-filed_15-Dec-2022-ABST.pdf	(188-188)	1	Abstract	19 KB

Digest

DOCUMENT	MESSAGE DIGEST(SHA-512)
148165_001401_ADS_as-filed-15-Dec-2022.pdf	F9EC11E3A83D1F3340A02220C4F925080721AE83D3EF3FFBC 6FCE3483B790663E2C1692A80707ECD2D88BDB5E1B1895543 6323AFD846EC19E983C5D9C8BC7F33

148165.001401_SL.XML	7071DE06722AF4DCA9091F17DD3A9290C06A3AA68710D9B50 2F0BA0350BEDB8D1E96071AFF65981AB23F2AF8EF931ACFB2 7B99E9BB6A4EAC8394951976EA68D9
148165_001401_Track1_Reque st_as-filed-15-Dec-2022.pdf	5595EAED70DF87437131EC3945F15E6B99AAD66B7166CAADF 6983973A7A1A16976DB1BEFD89C77EE7450A9911B76EC3A0A 9DC5FC0798A5FB33CC570E17D4A002
148165_001401_Preliminary_A mendment_as_filed_15-Dec- 2022.pdf	883F4F967A06C62A10315DD5B74742B42CEF66BC9F006C5186 2454569C57319BA2587F97D63242D91623E8E550711FBE48E0 E03A1C748FC2390C6540BF4557A5
148165_001401_Preliminary_A mendment_as_filed_15-Dec- 2022-A.PE.pdf	FA9B7B33DF027AA5E9D9B813C6951A00B102752381866604D7 8443BDAFBCD75E8123D79B2C53423B9A957B6893E428D9789 8F6CC07A246259F9F6AAF8316D036
148165_001401_Preliminary_A mendment_as_filed_15-Dec- 2022-CLM.pdf	8C2234E1BC2B51C0F68E2F1BA5E23512533AB374D6C5E2A2B DDFE86A39244483F7A397A604E1B11DB05307701DE9D6B9DD BD24A6665F80A39D58FC09DB167215
148165_001401_Preliminary_A mendment_as_filed_15-Dec- 2022-REM.pdf	726FA5DD8B91A33733F530C9A9A9FD498FB7298920F993628C 67C0A796F8B49C024220E7110B71444BA802D58184A3DE4DC 35E29266B3C4D48F2B1A4F2168333
148165_001401_Drawings_as_ filed_15-Dec-2022.pdf	F3599E8F524A0C4D48605A3D9CBAEC6CEFE0424BC194DE85 B97C575CAEFF086F971B8A612EFCA981116B4A602886FB479 A0A8CC2809DB25EE1127518729BC448
148165_001401_US_NP_Appli cation_as-filed_15-Dec- 2022.pdf	9B70819C6881B1A06FA7C3138E702B2C0E8ADB56918A526733 378F04BF9BCF3C2CC524C2FD3752600B6E967DB1122ADB43A 45117E136458AC92E58E262E64F86
148165_001401_US_NP_Appli cation_as-filed_15-Dec-2022- SPEC.pdf	2D8F7551C771F6AAD755FBB6CE0A10361E440C197A4E6741E 6BA6B94D4DCD83504DD62ABE6DE0BD31B1AE3FB39B5B0A9 7BB9513F6BCFD68249779EA0E41790DC
148165_001401_US_NP_Appli cation_as-filed_15-Dec-2022- CLM.pdf	37195346DAEF40688BF289A159E85D3C2C7F45857668424190 B89E9DBCE27236638341A55C6BBF74C6565E62F9602FEA828 81E2F7B99F92588DF6ED58851BE3A

148165_001401_US_NP_Appli	B5A3DCAAC658A9EAB8827144FC88B0E801A1E61B4E73AE7A
cation_as-filed_15-Dec-2022-	08C26BF8D69FE5973836A1DBE721378F4DE7B60C50613D239
ABST.pdf	1115BC3AE52A4CCC1C565C0BE8B2778

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

TRANSMITTAL FOR POWER OF ATTORNEY TO ONE OR MORE REGISTERED PRACTITIONERS

NOTE: This form is to be submitted with the Power of Attorney by Applicant form (PTO/AIA/82B) to identify the application to which the Power of Attorney is directed, in accordance with 37 CFR 1.5, unless the application number and filing date are identified in the Power of Attorney by Applicant form. If neither form PTO/AIA/82A nor form PTO/AIA82B identifies the application to which the Power of Attorney is directed, the Power of Attorney will not be recognized in the application.

Application Number	18/066,420
Filing Date	12/15/2022
First Named Inventor	Ronnie M. Russell
Title	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF
Art Unit	Not yet assigned
Examiner Name	Not yet assigned
Attorney Docket Number	148165.001401

SIGNATURE of Applicant or Patent Practitioner

Signature	/Christopher Bodle, Reg. #80,876/	Date (Optional)	January 19, 2023
Name	Christopher Bodle	Registration Number	80,876
Title (if Applicant is a juristic entity)			
Applicant Name (if Applicant is a juristic entity)			

NOTE: This form must be signed in accordance with 37 CFR 1.33. See 37 CFR 1.4(d) for signature requirements and certifications. If more than one applicant, use multiple forms.



*Total of 1 forms are submitted.

This collection of information is required by 37 CFR 1.131, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

POWER OF ATTORNEY BY APPLICANT

I hereby revoke all previous powers of attorney given in the application identified in either the attached transmittal letter or the boxes below.

Application Number

Filing Date

(Note: The boxes above may be left blank if information is provided on form PTO/AIA/82A.)



I hereby appoint the Patent Practitioner(s) associated with the following Customer Number as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above:

78442

OR



I hereby appoint Practitioner(s) named in the attached list (form PTO/AIA/82C) as my/our attorney(s) or agent(s), and to transact all business in the United States Patent and Trademark Office connected therewith for the patent application referenced in the attached transmittal letter (form PTO/AIA/82A) or identified above. (Note: Complete form PTO/AIA/82C.)

Please recognize or change the correspondence address for the application identified in the attached transmittal letter or the boxes above to:



The address associated with the above-mentioned Customer Number

OR



The address associated with Customer Number:

78442

OR



Firm or
Individual Name

Address

City

State

Zip

Country

Telephone

Email

I am the Applicant (if the Applicant is a juristic entity, list the Applicant name in the box):

Interius Biotherapeutics, Inc.



Inventor or Joint Inventor (title not required below)



Legal Representative of a Deceased or Legally Incapacitated Inventor (title not required below)



Assignee or Person to Whom the Inventor is Under an Obligation to Assign (provide signer's title if applicant is a juristic entity)



Person Who Otherwise Shows Sufficient Proprietary Interest (e.g., a petition under 37 CFR 1.46(b)(2) was granted in the application or is concurrently being filed with this document) (provide signer's title if applicant is a juristic entity)

SIGNATURE of Applicant for Patent

The undersigned (whose title is supplied below) is authorized to act on behalf of the applicant (e.g., where the applicant is a juristic entity).

Signature

Date (Optional)

Name

Title

NOTE: Signature - This form must be signed by the applicant in accordance with 37 CFR 1.33. See 37 CFR 1.4 for signature requirements and certifications. If more than one applicant, use multiple forms.



Total of _____ forms are submitted.

This collection of information is required by 37 CFR 1.131, 1.32, and 1.33. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 3 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.



UNITED STATES
PATENT AND TRADEMARK OFFICE

P.O. Box 1450
Alexandria, VA 22313 - 1450
www.uspto.gov

ELECTRONIC ACKNOWLEDGEMENT RECEIPT

APPLICATION #
18/066,420

RECEIPT DATE / TIME
01/19/2023 05:32:44 PM ET

ATTORNEY DOCKET #
148165.001401

Title of Invention

PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF

Application Information

APPLICATION TYPE Utility - Nonprovisional Application
under 35 USC 111(a)

PATENT # -

CONFIRMATION # 1909

FILED BY TAMMY COLLINS

PATENT CENTER # 61463864

FILING DATE -

CUSTOMER # 78442

FIRST NAMED INVENTOR Ronnie M. Russell

CORRESPONDENCE ADDRESS -

AUTHORIZED BY Christopher Bodle

Documents

TOTAL DOCUMENTS: 1

DOCUMENT	PAGES	DESCRIPTION	SIZE (KB)
148165_001401_POA.pdf	2	Power of Attorney	632 KB

Digest

DOCUMENT	MESSAGE DIGEST(SHA-512)
148165_001401_POA.pdf	D8F6061741E7DF754342D030689F6C651D4A57936384C2AC2A 3675AA171257EDF449144BDE7467557FCCDD9BFE8384F943 5B47035C06C2A6A771B66C359E1DC

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized

by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF
<p>As the below named inventor, I hereby declare that:</p> <p>This declaration is directed to: ☐ The attached application, or ☒ United States application or PCT international application number <u>16/066,420</u> filed on <u>December 15, 2022</u></p> <p>The above-identified application was made or authorized to be made by me.</p> <p>I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.</p> <p>I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.</p> <p style="text-align: center;">WARNING:</p> <p>Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identify theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.</p>	
<p>LEGAL NAME OF INVENTOR</p> <p>Inventor: <u>Ronnie M. Russell</u> Date (Optional): <u>January 5, 2023</u></p> <p>Signature: <u><i>Ronnie M. Russell</i></u></p>	
<p><small>Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.</small></p>	

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-6199 and select option 2.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF
--------------------	---

As the below named inventor, I hereby declare that:

This declaration is directed to: ☐ The attached application, or ☒ United States application or PCT international application number 18/066,420 filed on December 15, 2022

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

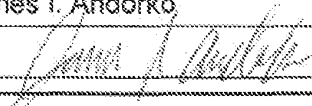
I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2036 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2036 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: James I. Andorko Date (Optional): 1/5/23

Signature: 

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF
--------------------	---

As the below named inventor, I hereby declare that:

This declaration is directed to: ☐ The attached application, or ☒ United States application or PCT international application number 18/066,420 filed on December 15, 2022

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Bruce Schnepf Date (Optional): Jan 5, 2023

Signature: Bruce Schnepf

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 118 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF
--------------------	---

As the below named inventor, I hereby declare that:

This declaration is directed to: ☐ The attached application, or ☒ United States application or PCT international application number 18/066,420 filed on December 15, 2022

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Mohit Sachdeva Date (Optional): Jan 20 2023

Signature: M Sachdeva

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF
--------------------	---

As the below named inventor, I hereby declare that:

This declaration is directed to: ☐ The attached application, or ☒ United States application or PCT international application number 18/066,420 filed on December 15, 2022

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Michael Marinelli Date (Optional): 1/10/23

Signature: [Signature]

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9129 and select option 2.



ELECTRONIC ACKNOWLEDGEMENT RECEIPT

APPLICATION #
18/066,420

RECEIPT DATE / TIME
01/23/2023 04:17:58 PM ET

ATTORNEY DOCKET #
148165.001401

Title of Invention

PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF

Application Information

APPLICATION TYPE Utility - Nonprovisional Application
under 35 USC 111(a)

PATENT # -

CONFIRMATION # 1909

FILED BY TAMMY COLLINS

PATENT CENTER # 61477939

FILING DATE -

CUSTOMER # 78442

FIRST NAMED Ronnie M. Russell
INVENTOR

CORRESPONDENCE -
ADDRESS

AUTHORIZED BY Christopher Bodle

Documents

TOTAL DOCUMENTS: 8

DOCUMENT	PAGES	DESCRIPTION	SIZE (KB)
Declaration_Andorko.pdf	1	Oath or Declaration filed	458 KB
Declaration_Boral.pdf	1	Oath or Declaration filed	447 KB
Declaration_Foss.pdf	1	Oath or Declaration filed	446 KB
Declaration_Johnson.pdf	1	Oath or Declaration filed	448 KB
Declaration_Marinelli.pdf	1	Oath or Declaration filed	449 KB
Declaration_Russell.pdf	1	Oath or Declaration filed	448 KB
Declaration_Sachdeva.pdf	1	Oath or Declaration filed	130 KB

Declaration_Schnepp.pdf	1	Oath or Declaration filed	447 KB
-------------------------	---	---------------------------	--------

Digest

DOCUMENT	MESSAGE DIGEST(SHA-512)
Declaration_Andorko.pdf	6C0A729CCC8A81BE323E5B34F250263A1C409923A388E3B1D 5E3F5AB74B4A95171D0DDCB60E4D66CF08E5A71CCFC6BD13 8BC5CF97A4318A67B953A47B6DFAB61
Declaration_Boral.pdf	E3D845865C48BD9EBB42DE4662134FB84F6221F78478786236 0BE85DEE93D8E3B95C3CC21D600A47C1C47A158C10947A67 F5D94BE3D7DB455FCEBC5B2D795320
Declaration_Foss.pdf	E9EEDCD5222B5F501EEA76F79B88054FF1260FE2F5377C346 1A3C93A6FF9639E960FEE82F4D2D6AAD13F1FFA1742F6DF8C 2B88635E4543C30DF08A135496E97C
Declaration_Johnson.pdf	89AD71498400FCE98A76E719D888721C45180E2375C454350A A33D6980FD5D59CD99B6482C2E7FD626E123531C68EA57CD3 4C9FB8EE1795B78A988E7E18C533D
Declaration_Marinelli.pdf	93EC059B5DC9452E1F78D583C6C6097BE690722A0209A42299 CA22FEFACB5E94F6782A5479DD4EE8737384FB1D3AF147A45 F3DBA81046F35796C8AE055F84E92
Declaration_Russell.pdf	D33527AF0DF68EF4D5035BAC7CF2C99A9DA1BE8FD79243A4 2645452E98E8B94718489E4CFEEE36C32240ACD6102E32D74 D04A46F12F3D2214FC1AAD1608ABA1B
Declaration_Sachdeva.pdf	46A907B792699B9285A0F66582C4912572074EA8811A1CAC82 A67318CD37AEBE187AE6428F58806551C7F659046020E418E0 7B6B898D3A58A736742A8829C42B
Declaration_Schnepp.pdf	FC9B89BBD432AC334DA0577A18EAC697AAD8ED19FF75F8A7 C6774A31537CA5BF2D2477BD9AFE92AEAD161B9903D050CB 90D407ABF881A4C87E26E78AFF34D657

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

**DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN
APPLICATION DATA SHEET (37 CFR 1.76)**

Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF
-------------------------------	--

As the below named inventor, I hereby declare that:

This declaration is directed to: ☐ The attached application, or
☒ United States application or PCT international application number 18/066,420
filed on December 15, 2022

The above-identified application was made or authorized to be made by me.

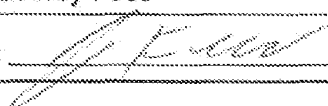
I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor Jeffrey Foss Date (Optional) 1/10/23
Signature 

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1460, Alexandria, VA 22313-1460. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1460, Alexandria, VA 22313-1460.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

**DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN
APPLICATION DATA SHEET (37 CFR 1.76)**

Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF
-------------------------------	--

As the below named inventor, I hereby declare that:

This declaration
is directed to:

☐

The attached application, or

☒

United States application or PCT international application number 18/066,420

filed on December 15, 2022

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

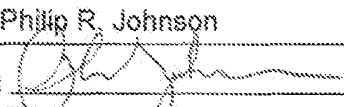
WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Philip R. Johnson

Date (Optional): Jan 17, 2023

Signature: 

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9198 and select option 2.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF
--------------------	---

As the below named inventor, I hereby declare that:

This declaration is directed to: ☐ The attached application, or ☒ United States application or PCT international application number 18/066,420 filed on December 15, 2022

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Debasish Boral Date (Optional): Jan 5, 2023
Signature: Debasish Boral

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 1 minute to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2.

Doc code: IDS

Doc description: Information Disclosure Statement (IDS) Filed

PTO/SB/05a (02-18)

Approved for use through 11/30/2020. OMB 0651-0031

U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)	Application Number	18066420
	Filing Date	2022-12-15
	First Named Inventor	Ronnie M. Russell
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	148165.001401

U.S.PATENTS						Remove
Examiner Initial*	Cite No	Patent Number	Kind Code ¹	Issue Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines where Relevant Passages or Relevant Figures Appear
	1	4439196	A	1984-03-27	Higuchi	
	1	4447224	A	1984-05-08	DeCant, Jr. et al.	
	1	4447224	A	1984-05-08	Mayfield	
	1	4439196	A	1984-03-27	Higuchi	
	1	4596556	A	1984-05-08	Morrow et al.	
	1	4447224	A	1988-12-13	Morrow et al.	
	1	4447224	A	1984-05-08	Cabilly et al.	
	1	4447224	A	1988-12-13	Burns	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)	Application Number		18066420	
	Filing Date		2022-12-15	
	First Named Inventor	Ronnie M. Russell		
	Art Unit	N/A		
	Examiner Name	Not Yet Assigned		
	Attorney Docket Number	148165.001401		

9	4946778	A	1990-08-07	Ladner et al.	
10	5064413	A	1991-11-12	McKinnon et al.	
9	4946778	A	1993-11-09	Ladner et al.	
9	4946778	A	1990-08-07	McKinnon et al.	
9	5064413	A	1995-01-24	McKinnon et al.	
10	4946778	A	1990-08-07	Peterson et al.	
9	6005079	A	1999-12-21	Casterman et al.	
10	6096002	A	2000-08-01	Landau	
9	4946778	B1	2003-09-16	Weston et al.	

If you wish to add additional U.S. Patent citation information please click the Add button.

Add

U.S.PATENT APPLICATION PUBLICATIONS

Remove

Examiner Initial*	Cite No	Publication Number	Kind Code ¹	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines where Relevant Passages or Relevant Figures Appear
-------------------	---------	--------------------	------------------------	------------------	---	--

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)	Application Number	18066420
	Filing Date	2022-12-15
	First Named Inventor	Ronnie M. Russell
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	148165.001401

	1	20200216502	A1	2020-07-09	Albertini et al.	
--	---	-------------	----	------------	------------------	--

If you wish to add additional U.S. Published Application citation information please click the Add button. [Add](#)

FOREIGN PATENT DOCUMENTS								Remove
Examiner Initial*	Cite No	Foreign Document Number ³	Country Code ² i	Kind Code ⁴	Publication Date	Name of Patentee or Applicant of cited Document	Pages, Columns, Lines where Relevant Passages or Relevant Figures Appear	T ⁵
	1	0404097	EP	A1	1990-12-27	Behringwerke AG		
	1	1988001649	EP	A1	1988-03-10	Genex Corp.		
	1	1993011161	WO	A1	1993-03-10	Enzon, Inc.		
	1	1994001649	EP	A1	1994-03-03	Casterman		
	1	1994001649	EP	A1	1994-12-27	Unilever PLC		
	1	2004041862	WO	A1	2004-05-21	Ablynx N.V.		
	1	2016065323	EP	A2	2016-04-28	Singh Biotechnology, LLC		
	1	2022183072	EP	A1	2024-05-21	Kelonia Therapeutics, Inc.		

If you wish to add additional Foreign Patent Document citation information please click the Add button. [Add](#)

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)	Application Number		18066420
	Filing Date		2022-12-15
	First Named Inventor	Ronnie M. Russell	
	Art Unit	N/A	
	Examiner Name	Not Yet Assigned	
	Attorney Docket Number	148165.001401	

NON-PATENT LITERATURE DOCUMENTS				Remove
Examiner Initials*	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc), date, pages(s), volume-issue number(s), publisher, city and/or country where published.		T ⁵
	1	Albertini et al., "Molecular and Cellular Aspects of Rhabdovirus Entry," Viruses (2012) 4:117-129.		
	2	Baert et al., "Influence of Immunogenicity on the Long-Term Efficacy of Infliximab in Crohn's Disease," (2003) New Engl. J. Med. 348:601-608.		
	1	Benjaminovitz et al., "Prevention of Rejection in Cardiac Transplantation by Blockade of the Interleukin-2 Receptor with a Monoclonal Antibody," (2000) New Engl. J. Med. 342(9):613-619.		
	2	Chothia et al., "Canonical Structures for the Hypervariable Regions of Immunoglobulins," (1987) J. Mol. Biol. 196:901-917.		
	1	Chothia et al., "Conformations of Immunoglobulin Hypervariable Regions," (1989) Nature 342: 877-883.		
	1	Clackson et al., "Making Antibody Fragments Using Phage Display Libraries," (1991) Nature 352:624-628. (Abstract Only)		
	2	Finkelshtein et al., "LDL Receptor and its Family Members Serve as the Cellular Receptors for Vesicular Stomatitis Virus," PNAS (2013) 110(18): 7306-7311.		
	3	Ghosh et al., "Natalizumab for Active Crohn's Disease," (2003) New Engl. J. Med. 348:24-32.		
	1	Herold et al., "Anti-CD3 Monoclonal Antibody in New-Onset Type 1 Diabetes Mellitus," (2002) New Engl. J. Med. 346 (22):1692-1698.		

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)	Application Number	18066420
	Filing Date	2022-12-15
	First Named Inventor	Ronnie M. Russell
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	148165.001401

10	Holliger et al., "Diabodies: Small Bivalent and Bispecific Antibody Fragments," (1993) Proc. Natl. Acad. Sci. USA 90: 6444-6448.
10	Holliger et al., "Engineered Antibody Fragments and the Rise of Single Domains," (2005) Nat. Biotechnol. 23(9): 1126-1136.
10	Hwang et al., "Engineering a Serum-Resistant and Thermostable Vesicular Stomatitis Virus G Glycoprotein for Pseudotyping Retroviral and Lentiviral Vectors," Gene Ther. (Aug. 2013) 20(8):807-815.
10	Kabat, "The Structural Basis for Antibody Complementarity," Adv. Prot. Chem. (1978) 32:1-75. (Abstract Only)
10	Kabat et al., "Unusual Distributions of Amino Acids in Complementarity-Determining (Hypervariable) Segments of Heavy and Light Chains of Immunoglobulins and Their Possible Roles in Specificity of Antibody-Combining Sites," (1977) J. Biol. Chem. 252(19):6609-6616.
10	Lalhe, "Synthetic Oligonucleotide Probes Deduced from Amino Acid Sequence Data," J. Molec. Biol. (1985) 183:1-12.
10	Lee et al., "ASTCT Consensus Grading for Cytokine Release Syndrome and Neurologic Toxicity Associated with Immune Effector Cells," Biol. Blood Marrow Transplant (2019) 25:625-638.
10	Lipsky et al., "Infliximab and Methotrexate in the Treatment of Rheumatoid Arthritis," (2000) New Engl. J. Med. 343 (22):1594-1602.
10	Liu et al., "Chimeric Mouse-Human IgG1 Antibody that can Mediate Lysis of Cancer Cells," Proc Natl. Acad. Sci., USA (1987) 84:3439-3443.
10	Liu et al., "Production of a Mouse-Human Chimeric Monoclonal Antibody to CD20 with Potent Fc-Dependent Biologic Activity," J. Immunology (1987) 139(10):3521-3526.
10	Liu et al., "Randomised, Double Blind, Placebo Controlled Study of Interferon Beta-1a in Relapsing-Remitting Multiple Sclerosis Analysed by Area under Disability/Time Curves," J. Neurol. Neurosurg. Psych. (1999) 67:451-456.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)	Application Number	18066420
	Filing Date	2022-12-15
	First Named Inventor	Ronnie M. Russell
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	148165.001401

	21	Marks et al., "By-Passing Immunization," J. Mol. Biol. (1991) 222: 581-597.	
	21	Milgrom et al., "Treatment of Allergic Asthma with Monoclonal Anti-IgE Antibody," New Engl. J. Med. (1999) 341:1966-1973.	
	23	Muller, "[43] Determination of Affinity and Specificity of Anti-Hapten Antibodies by Competitive Radioimmunoassay," Meth. Enzymol. (1983) 92:589-601. (Abstract Only)	
	21	Muyldermans et al., "Recognition of Antigens by Single-Domain Antibody Fragments: the Superfluous Luxury of Paired Domains," (2001) Trends Biochem. Sci. 26(4):230-235.	
	21	Neelapu et al., "Chimeric Antigen Receptor T-Cell Therapy – Assessment and Management of Toxicities," Nat. Rev. Clin. Oncology (Jan. 2018) 15(1):47-62.	
	23	Nikolic et al., "Structural Basis for the Recognition of LDL-Receptor Family Members by VSV Glycoprotein," Nature Comm. (2018) 9(1029):1-12.	
	21	Pluckthun, "Antibodies from Escherichia Coli," Nature (4 October 1990) Vol. 347, No. 6292, pgs. 497-498.	
	21	Portielji et al., "IL-12: A Promising Adjuvant for Cancer Vaccination," Cancer Immunol. Immunother. (2003) 52:133-144.	
	23	Presta, "Selection, Design, and Engineering of Therapeutic Antibodies," J. Allergy Clin. Immunol. (2005) 116 (4):731-736.	
	30	Reichmann et al., "Reshaping Human Antibodies for Therapy," Nature (1988) 332(6162):323-327.	
	21	Roche et al., "Crystal Structure of the Low-pH Form of the Vesicular Stomatitis Virus Glycoprotein G," Science (14 July 2006) 313:187-191.	

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)	Application Number		18066420
	Filing Date		2022-12-15
	First Named Inventor	Ronnie M. Russell	
	Art Unit	N/A	
	Examiner Name	Not Yet Assigned	
	Attorney Docket Number	148165.001401	

32	Roche et al., "Structure of the Prefusion Form of the Vesicular Stomatitis Virus Glycoprotein G," Science (9 Feb. 2007) 315:843-848.
32	Slamon et al., "Use of Chemotherapy Plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer that Overexpresses HER2," (2001) New Engl. J. Med. 344(11):783-792.
34	Teachey et al., "Identification of Predictive Biomarkers for Cytokine Release Syndrome after Chimeric Antigen Receptor T-cell Therapy for Acute Lymphoblastic Leukemia," Cancer Discov. (June 2016) 6(6):664-679.
32	Tsurushita et al., "Humanization of a Chicken Anti-IL-12 Monoclonal Antibody," J. Immuno. Methods (2004) Vol. 295, pgs. 9-19.
32	Yang et al., "A Randomized Trial of Bevacizumab, an Anti-Vascular Endothelial Growth Factor Antibody, for Metastatic Renal Cancer," (2003) New Engl. J. Med. 349(5):427-434.

If you wish to add additional non-patent literature document citation information please click the Add button

EXAMINER SIGNATURE

Examiner Signature		Date Considered	
--------------------	--	-----------------	--

*EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609. Draw line through a citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

¹ See Kind Codes of USPTO Patent Documents at www.USPTO.GOV or MPEP 901.04. ² Enter office that issued the document, by the two-letter code (WIPO Standard ST.3). ³ For Japanese patent documents, the indication of the year of the reign of the Emperor must precede the serial number of the patent document. ⁴ Kind of document by the appropriate symbols as indicated on the document under WIPO Standard ST.16 if possible. ⁵ Applicant is to place a check mark here if English language translation is attached.

INFORMATION DISCLOSURE STATEMENT BY APPLICANT (Not for submission under 37 CFR 1.99)	Application Number	18066420
	Filing Date	2022-12-15
	First Named Inventor	Ronnie M. Russell
	Art Unit	N/A
	Examiner Name	Not Yet Assigned
	Attorney Docket Number	148165.001401

CERTIFICATION STATEMENT

Please see 37 CFR 1.97 and 1.98 to make the appropriate selection(s):

That each item of information contained in the information disclosure statement was first cited in any communication from a foreign patent office in a counterpart foreign application not more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(1).

OR

☐ That no item of information contained in the information disclosure statement was cited in a communication from a foreign patent office in a counterpart foreign application, and, to the knowledge of the person signing the certification after making reasonable inquiry, no item of information contained in the information disclosure statement was known to any individual designated in 37 CFR 1.56(c) more than three months prior to the filing of the information disclosure statement. See 37 CFR 1.97(e)(2).

See attached certification statement.

The fee set forth in 37 CFR 1.17 (p) has been submitted herewith.

☒ A certification statement is not submitted herewith.

SIGNATURE

A signature of the applicant or representative is required in accordance with CFR 1.33, 10.18. Please see CFR 1.4(d) for the form of the signature.

Signature	/Christopher R. Bodle/	Date (YYYY-MM-DD)	2023-01-30
Name/Print	Christopher R. Bodle	Registration Number	80,876

This collection of information is required by 37 CFR 1.97 and 1.98. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 1 hour to complete, including gathering, preparing and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.



UNITED STATES
PATENT AND TRADEMARK OFFICE

P.O. Box 1450
Alexandria, VA 22313 - 1450
www.uspto.gov

ELECTRONIC ACKNOWLEDGEMENT RECEIPT

APPLICATION #
18/066,420

RECEIPT DATE / TIME
01/30/2023 05:35:31 PM ET

ATTORNEY DOCKET #
148165.001401

Title of Invention

PSEUDOTYPED VIRAL PARTICLES, COMPOSITIONS COMPRISING THE SAME, AND USES THEREOF

Application Information

APPLICATION TYPE Utility - Nonprovisional Application
under 35 USC 111(a)

PATENT # -

CONFIRMATION # 1909

FILED BY TAMMY COLLINS

PATENT CENTER # 61514619

FILING DATE -

CUSTOMER # 78442

FIRST NAMED INVENTOR Ronnie M. Russell

CORRESPONDENCE ADDRESS -

AUTHORIZED BY Christopher Bodle

Documents

TOTAL DOCUMENTS: 45

DOCUMENT	PAGES	DESCRIPTION	SIZE (KB)
EP0404097B1.pdf	39	Foreign Reference	977 KB
WO1988001649A1.pdf	124	Foreign Reference	4644 KB
WO1993011161A1.pdf	119	Foreign Reference	4196 KB
WO1994004678A1.pdf	88	Foreign Reference	4404 KB
WO1994025591A1.pdf	100	Foreign Reference	4102 KB
WO2004041862A2.pdf	92	Foreign Reference	4908 KB
WO2016065323A2.pdf	95	Foreign Reference	4141 KB

WO2022183072A1.pdf	162	Foreign Reference	1260 KB
Albertini117.pdf	23	Non Patent Literature	1978 KB
Baert601.pdf	8	Non Patent Literature	84 KB
Beniaminovitz613.pdf	7	Non Patent Literature	99 KB
Chothia901.pdf	18	Non Patent Literature	5296 KB
Chothia877_.pdf	8	Non Patent Literature	10696 KB
Clackson624__AbstractOnly.pdf	1	Non Patent Literature	79 KB
Finkelshtein7306.pdf	6	Non Patent Literature	1135 KB
Ghosh24.pdf	9	Non Patent Literature	102 KB
Herold1692.pdf	7	Non Patent Literature	100 KB
Holliger6444.pdf	5	Non Patent Literature	1487 KB
Holliger1126_.pdf	12	Non Patent Literature	1758 KB
Hwang807_.pdf	26	Non Patent Literature	4136 KB
Kabat1__AbstractOnly.pdf	1	Non Patent Literature	108 KB
Kabat6609.pdf	8	Non Patent Literature	1248 KB
Lathe1_.pdf	13	Non Patent Literature	6826 KB
Lee625.pdf	14	Non Patent Literature	467 KB
Lipsky1594.pdf	9	Non Patent Literature	120 KB
Liu3439.pdf	5	Non Patent Literature	937 KB

Liu3521_.pdf	7	Non Patent Literature	8302 KB
Liu451.pdf	6	Non Patent Literature	177 KB
Marks581_.pdf	17	Non Patent Literature	3448 KB
Milgrom1966.pdf	8	Non Patent Literature	105 KB
Muller589_AbstractOnly.pdf	1	Non Patent Literature	113 KB
Muyldermans230.pdf	6	Non Patent Literature	495 KB
Neelapu47_.pdf	37	Non Patent Literature	1657 KB
Nikolic1.pdf	12	Non Patent Literature	5215 KB
Pluckthun497_.pdf	2	Non Patent Literature	4312 KB
Portielje133_.pdf	12	Non Patent Literature	127 KB
Presta731.pdf	6	Non Patent Literature	120 KB
Reichmann323_.pdf	5	Non Patent Literature	1004 KB
Roche187.pdf	5	Non Patent Literature	716 KB
Roche843.pdf	6	Non Patent Literature	399 KB
Slamon783.pdf	10	Non Patent Literature	188 KB
Teachey664.pdf	16	Non Patent Literature	7271 KB
Tsurushita9.pdf	11	Non Patent Literature	368 KB
Yang427.pdf	8	Non Patent Literature	157 KB
148165_001401_IDS_30Jan23.pdf	9	Information Disclosure Statement (IDS) Form (SB08)	1014 KB

Digest

DOCUMENT	MESSAGE DIGEST(SHA-512)
EP0404097B1.pdf	56026166702BB9DF9E8A6FC471F6F8F378CEFC5967AB4CB7F 471B99B55130483A28537ED907326D87D9C88E1D2D8445EC00 1279559ECAF88C8DAADB19C324715
WO1988001649A1.pdf	935BBA1BF44632196EF18FD384A4CF9E2F2770352BAB4DC10 2DAE41E254BE1D4A207761F5FCA04796B5697ED6109AE8139 C5C0D8671292BA29FD01FA614A0FE4
WO1993011161A1.pdf	8DBAFA9E08218A61A4F46C3EEBF681176A1B5A3D29ECE91F EB48BE4ED40A9A43E0A13EDF52C8E88ACCCEAB36173D5B94 BAEAA537933C15DBD904A1142EA5AEA1
WO1994004678A1.pdf	C6262ED1F7F2F8661EBC7D9A7C2E7A6627C39E30AFA251B5B 7EE62A533576F32449F776156387D425BC61D157A7EC5E7C4A FF89B881E789F6C2908E14A799D7F
WO1994025591A1.pdf	69F955EDE170EBF45A36C2DCF7063EBAD726EDF280AE493D 5ABDEB01E55C3EB65F8661C3F1BB41FAC907E1C1F38DED15 4D1836E9AB6109726D880509A7879C04
WO2004041862A2.pdf	CE6874C6CBA2FE7CEBCAD459E2AA951EBFEC70EA1FA22CF 57EC46E5D19FE23EEDECEDFAF09D4B001DBAE80AB772797B 1519B076328BE560E13FBFCE3BF8EE8C8
WO2016065323A2.pdf	B42E6DFA27C3E26A7302B8A1EA3BEA31D5609491F65E685DB 35A1EAA05AC32CBB76703C8852C9027F069BC812EFCBBB26 B493597BE881207991D1B2F85D5B79B
WO2022183072A1.pdf	5D22FE11E39DA8143A0FC78041053F9C2BB911FD8FB1082208 24A746BE1B08843A69CAC90FEEA3545BE54A0E6418632E69E EBAF8534C476113870B5F7E6118D6
Albertini f 17.pdf	155DEB711A273731CD25EFCCDCD7D00754F322D15F808381D 8DBFD3DAE5B3D3C0F642786183C8DE76DD2BE623294F574A 0D58A1E434B129BE58E01C04281FA5B

Baert601.pdf	11BBF8118476ACC613DDB2D2718B2D85682C7C23537F9C2D3 9D50D52C30F81B6644B6DAAAC5244374CAA368B1FE275F89B 68E926C15262B54B5022847336F0C2
Beniaminovitz613.pdf	D4BC7CA4364E2631E994B899AD4C590889569D8B43CA9F918 89A118FFF05FC6AFCB8CAEE19045595F8A2E69A886E9F220B 727BCA2D70E23B222C4DB222F14424
Chothia901.pdf	AA02E18459933040ACDBAA9F2B7BCF64989347ADB883CB5F7 24571605438FD01FA6839535589C284C9711E2D91DD79F9D8C EC7AB3BDB5A44C9C60AAC04D0846E
Chothia877_.pdf	D2BC95E44C91672E5870DFE61E3CD46DEC02DFD43C91E12C 80849B158F81E3975955E0BC93137A363AD6338A70E931FE4E 13BB6B9DA5D4343419B7F898894139
Clackson624__AbstractOnly.pdf	B8AB3D10580E6A48AB4544345953BF617D1C088FD950152BB1 4373B13CA1F7D824096E7F6C388F5D8DB2BB605C2D4CF1EF6 0E986DECF9BA203995804D230BA0C
Finkelshtein7306.pdf	8B18B67CDD86CC1D0DD703E87EC425A56C0E39D7CFC31564 7D6E52A9CECA892603766840F3FF9AA8B636D1C863A68781D D7844745A2ECEE54F6C3ADFE5228774
Ghosh24.pdf	1045C6D900ADF83FA0C21286DC46BE5CF4E636F7A094EAF1B F735A8E5D31F5D26FEF12D3075E15B3DA0B60B64FDF3140AB B590B16810BF0D5E36C4378864F8DE
Herold1692.pdf	7D6108C1A00CE8511E30B7BC80EA485D4848E5ED2646102F7 6DF02043EC3B8AD6A91403352B49561C87E9528D55A2A5A0B B83F42386223301B23CFD2A0D9DE8E
Holliger6444.pdf	B9BC3C9CC4A09AFF4E83C33DEDF542E63541D1381A1E6F61 DDA79604373813A7974C5064B18221EEB0EB1841DD091CB74 7893CF52A8271BC23ACA7B19BFEF185
Holliger1126_.pdf	79F1938B98596D5E849F9D6F438EE7D87BEF99723B80A27331 EF6C304F7F87089501CC20D777FD744970E1014C1F7DFB3450 9084706DDCFD483F3E9AAE10F2ED

Hwang807_.pdf	387EF121C7BB7B9710E8E6887442E07134A769EAE0B9B7B0F6 DA48C62A643AE8CB7878E77656A61C0244FB679D53F3E9796 4A7177ED7F97AA39591FA76B2B5EA
Kabat1_AbtractOnly.pdf	8801AA6873BDB4CAAFBCEBADC5C01B0D822DE36165AFCC8 731DAAAE67B6C2BF68DE424221D2B371522FFF3987EBC4C25 C401497C79DF2B3C6D575C5E4C65387C
Kabat6609.pdf	763AFF335B179010E5B44DA94B14743E3E9C8E9D6E099C97C F9AAFAFDB271F67A51679C2778DB5E8E3C9103E71ABDAF780 699AD4CF7ABF926EB1F09C433B12A9
Lathe1_.pdf	C71C7F55450A1FEEBE6821ECD57BF1ED4ACAE46F55ADBDE 2C4A9AE9B9C164C509D98DAFAD5640F15D0B73147137E8504 99FA4207D92EFF7B63F135E03031E81F
Lee625.pdf	33943E164C58EED01FFC2B4A1A3253214385F9E03346D37D8E 523F56C53DFA6A4F4662C584B8B12EC7459E26F34EAFDC221 373EBE73E4F42D6C06B8C83CB24BF
Lipsky1594.pdf	41136DB949E7DF4008556A80898B2D281561EBAF8AF21AF82F 684CECE05935FBFE3E918A23656E77E10F29CB39915B9DA74 CFD018C4B3B08A518D1A8A9007B84
Liu3439.pdf	9E2419FE417F807D74420DE73221A559F0F6735E236FA68A21 5603167F9EE62B49AB1A314858B5EE35DDA3868564BD94741F 92D2E4CDA9221A3349F283543A3E
Liu3521_.pdf	D5FD7FA86A688A057EB7932E1BC961B85864FC6D4FEE11DD 2B49E928AE71BFD0DA23E5EFAC6AED1214E7E04F443B75125 EFDC67ED95EDD18C1A5B07CC5863E26
Liu451.pdf	3D94F6B32E22DFECB3D80A46E1D7A0256158B8640E676E621 EFE1BA41C6C20245FF2D83178D08D294F460D1295A4E2F096 0BFC76BD6938E1C9EFBFF25B23E22A
Marks581_.pdf	0B8F219243E563DB925F85C9788E61385A7B186B2EC196EA53 07021C4FF3153FDCEFA80BA1C1E6FAE18AA8B0E72B63CA9 6CFFA27952E0F323DBD813E8E10CB1

Milgrom1966.pdf	23B7B76CBC540A8E20318241760605A530453BE5A91BAE24A8 3E3F6A5769491F384516AE90F68AD635AA11C87E0B10715DC4 898B69D7F588DBAF6C9FA3AAF32F
Muller589_AbstractOnly.pdf	BDD88C0326B7C121281D4A8EA3FA4B1F11B3B7AC25682DFA FBC1A3B084701BCB2E018773B6EB41EAC7C5589F1BF218265 25BC840317DDB0E999BD7E3DB0B672F
Muyldermans230.pdf	B70BB4C0729D65B0165744B4B31390F97A2442F0E867C7B903 4F2A22E2243D45AEA4AD105DB5A3418AFA069CDF2EA403EB 1C0402140F8355DA1578BE4F2F8124
Neelapu47_.pdf	39BC8D639180BC1CE07A7C3A660D2698A566936390A40559D 7B3BD7EA03D3E70A4AC6D84330F268D272B9832C049B5661C FB892E74132F7965D0A9D3E24DA30F
Nikolic1.pdf	ED24D464A61DD03A8CCBA92CD4BCFD0ED0585BE814A9D3A 42A7D53B7E5EF9DA556A76330394CD7DBC0CD9CF6D9032D6 01A983287D91EFCE417A732DEDC315A02
Pluckihun497_.pdf	12BCBBCF9150FF768F3D3B42EBAB9EB318FE71BADB6F85BC 5937E23C64C3B5EB20DEA3EA0B10E25AEBB47985DEF653E9 9613D4A8E9E6968EA85486501A7BCA57
Portielje133_.pdf	334D2A454C9FBDF122DF2F9A1AD2AC6FB5BC00C8DDFF2C38 93DE0C266C881535C1D4BCB69DA4B921513DE356AED83615F 99FFBE59295E2FCC1EB744419644FDB
Presta731.pdf	7C09E9E6FC5E34E1C07146DA76CF04F58296B68EBCD910087 457FC715AB34695C5B4A68F6985FC69571FB8A36798FC01D6 DB49BB522B6A14B7C9F74382469E59
Reichmann323_.pdf	1536D7AFDE18C06BC441FA5BFA0405190DFBD881CBFAE4CA 0C42DF4796AC32A1164E5FA946B70589C8D7C354CE969B568 678E219087EAB30450E9DEC337A7396
Roche187.pdf	3AD97228F0DF17B8FEFD7DFF280CE396DAB98B7D17D3DEA7 805AD221F420DAB151F671802C9FA03A6D041F5398E487F304 04C6AA3280FB5983A4A129687DDFC4

Roche843.pdf	187B21ECE3F01B328582ECAE904ABCC95EDFD54E14DB8B6B 984A1D1DF630A2F36955599F058B29ABB88EACE42E26C654D 0022A5D4430640BA8DB2A75D33F84BC
Slamon783.pdf	6E330E0FDCB2E5FADCB023F1511B15E64F1519F4177ACAF0E AAD7429A77A4FD661247A45F064A4800B14B900F2E0A2EBB6 903894B2C46AA7ABD9CF95B2582019
Teachey664.pdf	5FCC6EB13369E04693BDC513ABACD7C338A3D2329B8A8261 F54309CC945293F24C1C4DA3FA249104F9051D891E8173B648 41876337F7E7DEAF2F9993FFF0948D
Tsurushita9.pdf	068FB3E03772EC2743BE37A6CB4EED2612B620CDE10031C26 575944D6B967EB460911E01FD4BEB2071BC7FE472BC5994EC A6F67F6500563B53EBAED1E3405AC3
Yang427.pdf	9A95AAD2E0B7FABEA5545068A21FBBE45035DB3CFA11AF55 09485DB14ADAE7BD71D3DD0012F062ECCE6167FA3D210DAF A619ACD7C7702300011475323FF238C6
148165__001401_IDS__30Jan23. pdf	070578C9209CAB291C6AB3B16B96FECA0F8F9A52040247A11 7117C87859443C66501914F60B789D6EE39B142E1F31042D56 F012244C71E06BB1C073B2F47261E

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

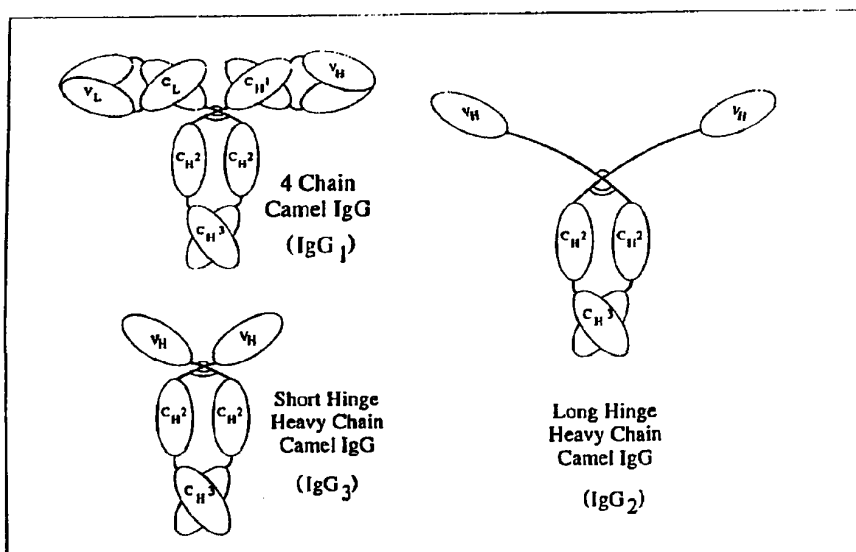
New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : C12N 15/13, C07K 15/06 C12P 21/08, C12N 5/10 A61K 39/395, C12N 15/62	A1	(11) International Publication Number: WO 94/04678 (43) International Publication Date: 3 March 1994 (03.03.94)
(21) International Application Number: PCT/EP93/02214 (22) International Filing Date: 18 August 1993 (18.08.93) (30) Priority data: 92402326.0 21 August 1992 (21.08.92) EP <i>(34) Countries for which the regional or international application was filed:</i> 93401310.3 21 May 1993 (21.05.93) AT et al. EP <i>(34) Countries for which the regional or international application was filed:</i> AT et al.		(74) Agents: GUTMANN, Ernest et al.; Ernest Gutmann - Yves Plasseraud S.A., 67, boulevard Haussmann, F-75008 Paris (FR). (81) Designated States: AU, BB, BG, BR, CA, CZ, FI, HU, JP, KP, KR, LK, MG, MN, MW, NO, NZ, PL, RO, RU, SD, SK, UA, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(71)(72) Applicants and Inventors: CASTERMAN, Cécile [BE/BE]; HAMERS, Raymond [BE/BE]; Vijversweg 15, B-1640 Sint-Genesius-Rode (BE).		

(54) Title: IMMUNOGLOBULINS DEVOID OF LIGHT CHAINS

A schematic representation of Camel immunoglobulins IgG₁, the putative IgG₂ and IgG₃.
 The large (Pro-X)₁₂ of the putative IgG₂ molecule can be modelled into a 6 aa repeat

(57) Abstract

The invention relates to isolated immunoglobulin, characterized in that it comprises two heavy polypeptide chains sufficient for the formation of a complete antigen binding site or several antigen binding sites, this immunoglobulin being further devoid of light polypeptide chains.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NE	Niger
BE	Belgium	GN	Guinea	NL	Netherlands
BF	Burkina Faso	GR	Greece	NO	Norway
BG	Bulgaria	HU	Hungary	NZ	New Zealand
BJ	Benin	IE	Ireland	PL	Poland
BR	Brazil	IT	Italy	PT	Portugal
BY	Belarus	JP	Japan	RO	Romania
CA	Canada	KP	Democratic People's Republic of Korea	RU	Russian Federation
CF	Central African Republic	KR	Republic of Korea	SD	Sudan
CG	Congo	KZ	Kazakhstan	SE	Sweden
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovak Republic
CM	Cameroon	LU	Luxembourg	SN	Senegal
CN	China	LV	Latvia	TD	Chad
CS	Czechoslovakia	MC	Monaco	TG	Togo
CZ	Czech Republic	MG	Madagascar	UA	Ukraine
DE	Germany	ML	Mali	US	United States of America
DK	Denmark	MN	Mongolia	UZ	Uzbekistan
ES	Spain			VN	Viet Nam
FI	Finland				

IMMUNOGLOBULINS DEVOID OF LIGHT CHAINS

The invention relates to new isolated immunoglobulins which are devoid of light polypeptide chains. These immunoglobulins do not consist in the degradation products of immunoglobulins composed of both heavy polypeptide and light polypeptide chains but to the contrary, the invention defines a new member of the family of the immunoglobulins, especially a new type of molecules capable of being involved in the immune recognition. Such immunoglobulins can be used for several purposes, especially for diagnosis or therapeutical purposes including protection against pathological agents or regulation of the expression or activity of proteins.

Up to now the structure proposed for immunoglobulins consists of a four-chain model referring to the presence of two identical light polypeptide chains (light chains) and two identical heavy polypeptide chains (heavy chains) linked together by disulfide bonds to form a y- or T-shaped macromolecules. These chains are composed of a constant region and a variable region, the constant region being subdivided in several domains. The two heavy polypeptide chains are usually linked by disulphide bounds in a so-called "hinge region" situated between the first and second domains of the constant region.

Among the proteins forming the class of the immunoglobulins, most of them are antibodies and accordingly present an antigen binding site or several antigen binding sites.

According to the four-chain model, the antigen binding site of an antibody is located in the variable domains of each of the heavy and light chains, and

requires the association of the heavy and the light chains variable domains.

For the definition of these four-chain model immunoglobulins, reference is made to Roitt. I et al (Immunology-second-Edition Gower Medical Publishing USA, 1989). Reference is especially made to the part concerning the definition of the four-chain immunoglobulins, their polypeptidic and genetic structures, the definition of their variable and constant regions and the obtention of the fragments produced by enzymatic degradation according to well known techniques.

The inventors have surprisingly established that different molecules can be isolated from animals which naturally produce them, which molecules have functional properties of immunoglobulins these functions being in some cases related to structural elements which are distinct from those involved in the function of four-chain immunoglobulins due for instance to the absence of light chains.

The invention relates to two-chain model immunoglobulins which neither correspond to fragments obtained for instance by the degradation in particular the enzymatic degradation of a natural four-chain model immunoglobulin, nor correspond to the expression in host cells, of DNA coding for the constant or the variable region of a natural four-chain model immunoglobulin or a part of these regions, nor correspond to antibodies produced in lymphopaties for example in mice, rats or human.

E.S. Ward et al (1) have described some experiments performed on variable domains of heavy polypeptide chains (V_H) or/and light polypeptide chains (V_K/F_V) to test the ability of these variable domains, to bind specific antigens. For this purpose, a library of V_H genes was prepared from the spleen

genomic DNA of mice previously immunized with these specific antigens.

Ward et al have described in their publication that V_H domains are relatively sticky, presumably due to the exposed hydrophobic surface normally capped by the V_K or V_L domains. They consequently envisage that it should be possible to design V_H domains having improved properties and further that V_H domains with binding activities could serve as the building blocks for making variable fragments (Fv fragments) or complete antibodies.

The invention does not start from the idea that the different fragments (light and heavy chains) and the different domains of these fragments of four-chain model immunoglobulin can be modified to define new or improved antigen binding sites or a four-chain model immunoglobulin.

The inventors have determined that immunoglobulins can have a different structure than the known four-chain model and that such different immunoglobulins offer new means for the preparation of diagnosis reagents, therapeutical agents or any other reagent for use in research or industrial purposes.

Thus the invention provides new immunoglobulins which are capable of showing functional properties of four-chain model immunoglobulins although their structure appears to be more appropriate in many circumstances for their use, their preparation and in some cases for their modification. Moreover these molecules can be considered as lead structures for the modification of other immunoglobulins. The advantages which are provided by these immunoglobulins comprise the possibility to prepare them with an increased facility.

The invention accordingly relates to immunoglobulins characterized in that they comprise two heavy polypeptide chains sufficient for the

formation of a complete antigen binding site or several antigen binding sites, these immunoglobulins being further devoid of light polypeptide chains. In a particular embodiment of the invention, these immunoglobulins are further characterized by the fact that they are the product of the expression in a prokaryotic or in a eukaryotic host cell, of a DNA or of a cDNA having the sequence of an immunoglobulin devoid of light chains as obtainable from lymphocytes or other cells of Camelids.

The immunoglobulins of the invention can be obtained for example from the sequences which are described in figure 7.

The immunoglobulins of the invention, which are devoid of light chains are such that the variable domains of their heavy chains have properties differing from those of the four-chain immunoglobulin V_H . The variable domain of a heavy-chain immunoglobulin of the invention has no normal interaction sites with the V_L or with the C_H1 domain which do not exist in the heavy chain immunoglobulins. it is hence a novel fragment in many of its properties such as solubility and position of the binding site. For clarity reasons we will call it V_{HH} in this text to distinguish it from the classical V_H of four-chain immunoglobulins.

By "a complete antigen binding site" it is meant according to the invention, a site which will alone allow the recognition and complete binding of an antigen. This could be verified by any known method regarding the testing of the binding affinity.

These immunoglobulins which can be prepared by the technique of recombinant DNA, or isolated from animals, will be sometimes called "heavy-chain immunoglobulins" in the following pages. In a preferred embodiment of the invention, these immunoglobulins are in a pure form.

In a first embodiment, the immunoglobulins of the invention are obtainable in prokaryotic cells, especially in E.coli cells by a process comprising the steps of :

- a) cloning in a Bluecript vector of a DNA or cDNA sequence coding for the V_{HH} domain of an immunoglobulin devoid of light chain obtainable for instance from lymphocytes of Camelids,
- b) recovering the cloned fragment after amplification using a 5' primer containing an Xho site and a 3' primer containing the Spe site having the following sequence
TC TTA ACT AGT GAG GAG ACG GTG ACC TG,
- c) cloning the recovered fragment in phase in the immuno PBS vector after digestion of the vector with Xho and Spe restriction enzymes,
- d) transforming host cells, especially E.coli by transfection with the recombinant immuno PBS vector of step c,
- e) recovering the expression product of the V_{HH} coding sequence, for instance by using antibodies raised against the dromadary V_{HH} domain.

In another embodiment the immunoglobulins are hetero-specific immunoglobulins obtainable by a process comprising the steps of:

- obtaining a first DNA or cDNA sequence coding for a V_{HH} domain or part thereof having a determined specificity against a given antigen and comprised between Xho and Spe sites,
- obtaining a second DNA or cDNA sequence coding for a V_{HH} domain or part thereof, having a determined specificity different from the specificity of the first DNA or cDNA sequence and comprised between the Spe and EcoRI sites,
- digesting an immuno PBS vector with EcoRI and XhoI restriction enzymes,

- ligating the obtained DNA or cDNA sequences coding for V_{HH} domains, so that the DNA or cDNA sequences are serially cloned in the vector,
- transforming a host cell, especially E.coli cell by transfection, and recovering the obtained immunoglobulins.

In another embodiment, the immunoglobulins are obtainable by a process comprising the steps of:

- obtaining a DNA or cDNA sequence coding for a V_{HH} domain or part thereof, having a determined specific antigen binding site,
- amplifying the obtained DNA or cDNA, using a 5' primer containing an initiation codon and a HindIII site, and a 3' primer containing a termination codon having a XhoI site,
- recombining the amplified DNA or cDNA into the HindIII (position 2650) and XhoI (position 4067) sites of a plasmid pMM984,
- transfecting permissive cells especially NB-E cells with the recombinant plasmid,
- recovering the obtained products.

Successful expression can be verified with antibodies directed against a region of a V_{HH} domain, especially by an ELISA assay.

According to another particular embodiment of this process, the immunoglobulins are cloned in a parvovirus.

In another example these immunoglobulins are obtainable by a process comprising the further cloning of a second DNA or cDNA sequence having another determined antigen binding site, in the pMM984 plasmid.

Such an Immunoglobulin can be further characterized in that it is obtainable by a process wherein the vector is Yep 52 and the transformed recombinant cell is a yeast especially S.cerevisiae.

A particular Immunoglobulin is characterized in that it has a catalytic activity, especially in that it is directed against an antigen mimicking an activated state of a given substrate. These catalytic antibodies can be modified at the level of their binding site, by random or directed mutagenesis in order to increase or modify their catalytic function. Reference may be made to the publication of Lerner et al (TIBS November 1987. 427-430) for the general technique for the preparation of such catalytic immunoglobulins.

According to a preferred embodiment, the immunoglobulins of the invention are characterized in that their variable regions contain in position 45, an amino-acid which is different from leucine, proline or glutamine residue.

Moreover the heavy-chain immunoglobulins are not products characteristic of lymphocytes of animals nor from lymphocytes of a human patient suffering from lymphopathies. Such immunoglobulins produced in lymphopathies are monoclonal in origin and result from pathogenic mutations at the genomic level. They have apparently no antigen binding site.

The two heavy polypeptide chains of these immunoglobulins can be linked by a hinge region according to the definition of Roitt et al.

In a particular embodiment of the invention, immunoglobulins corresponding to the above-defined molecules are capable of acting as antibodies.

The antigen binding site(s) of the immunoglobulins of the invention are located in the variable region of the heavy chain.

In a particular group of these immunoglobulins each heavy polypeptide chain contains one antigen binding site on its variable region, and these sites correspond to the same amino-acid sequence.

In a further embodiment of the invention the immunoglobulins are characterized in that their heavy polypeptide chains contain a variable region (V_{HH}) and a constant region (C_H) according to the definition of Roitt et al, but are devoid of the first domain of their constant region. This first domain of the constant region is called C_{H1} .

These immunoglobulins having no C_{H1} domain are such that the variable region of their chains is directly linked to the hinge region at the C-terminal part of the variable region.

The immunoglobulins of the type described hereabove can comprise type G immunoglobulins and especially immunoglobulins which are defined as immunoglobulins of class 2 (IgG2) or immunoglobulins of class 3 (IgG3).

The absence of the light chain and of the first constant domain lead to a modification of the nomenclature of the immunoglobulin fragments obtained by enzymatic digestion, according to Roitt et al.

The terms Fc and pFc on the one hand, Fc' and pFc' on the other hand corresponding respectively to the papain and pepsin digestion fragments are maintained.

The terms Fab $F(ab)_2$ $F(ab')_2$ Fabc, Fd and Fv are no longer applicable in their original sense as these fragments have either a light chain, the variable part of the light chain or the C_{H1} domain.

The fragments obtained by papain digestion and composed of the V_{HH} domain and the hinge region will be called FV_{HHh} or $F(V_{HHh})_2$ depending upon whether or not they remain linked by the disulphide bonds.

In another embodiment of the invention, immunoglobulins replying to the hereabove given definitions can be originating from animals especially from animals of the camelid family. The inventors have found out that the heavy-chain immunoglobulins which

are present in camelids are not associated with a pathological situation which would induce the production of abnormal antibodies with respect to the four-chain immunoglobulins. On the basis of a comparative study of old world camelids (Camelus bactrianus and Camelus dromaderius) and new world camelids (for example Lama Paccos, Lama Glama, and Lama Vicugna) the inventors have shown that the immunoglobulins of the invention, which are devoid of light polypeptide chains are found in all species. Nevertheless differences may be apparent in molecular weight of these immunoglobulins depending on the animals. Especially the molecular weight of a heavy chain contained in these immunoglobulins can be from approximately 43 kd to approximately 47 kd, in particular 45 kd.

Advantageously the heavy-chain immunoglobulins of the invention are secreted in blood of camelids.

Immunoglobulins according to this particular embodiment of the invention are obtainable by purification from serum of camelids and a process for the purification is described in details in the examples. In the case where the immunoglobulins are obtained from Camelids, the invention relates to immunoglobulins which are not in their natural biological environment.

According to the invention immunoglobulin IgG2 as obtainable by purification from the serum of camelids can be characterized in that :

- it is not adsorbed by chromatography on Protein G Sepharose column,
- it is adsorbed by chromatography on Protein A Sepharose column,
- it has a molecular weight of around 100 kd after elution with a pH 4.5 buffer (0.15 M NaCl, 0.58% acetic acid adjusted to pH 4.5 by NaOH),

- it consists of heavy $\gamma 2$ polypeptide chains of a molecular weight of around 46 kd preferably 45 after reduction.

According to a further embodiment of the invention another group of immunoglobulins corresponding to IgG3, as obtainable by purification from the serum of Camelids is characterized in that the immunoglobulin :

- is adsorbed by chromatography on a Protein A Sepharose column,
- has a molecular weight of around 100 kd after elution with a pH 3.5 buffer (0.15 M NaCl, 0.58% acetic acid),
- is adsorbed by chromatography on a Protein G Sepharose column and eluted with pH 3.5 buffer (0.15 M NaCl, 0.58% acetic acid).
- consists of heavy $\gamma 3$ polypeptide chains of a molecular weight of around 45 Kd in particular between 43 and 47 kd after reduction.

The immunoglobulins of the invention which are devoid of light chains, nevertheless comprise on their heavy chains a constant region and a variable region. The constant region comprises different domains.

The variable region of immunoglobulins of the invention comprises frameworks (FW) and complementarity determining regions (CDR), especially 4 frameworks and 3 complementarity regions. It is distinguished from the four-chain immunoglobulins especially by the fact that this variable region can itself contain an antigen binding site or several, without contribution of the variable region of a light chain which is absent.

The amino-acid sequences of frameworks 1 and 4 comprise among others respectively amino-acid sequences which can be selected from the following :

for the framework 1 domain

11

G G S V Q T G G S L R L S C E I S G L T F D
G G S V Q T G G S L R L S C A V S G F S F S
G G S E Q G G G S L R L S C A I S G Y T Y G
G G S V Q P G G S L T L S C T V S G A T Y S
G G S V Q A G G S L R L S C T G S G F P Y S
G G S V Q A G G S L R L S C V A G F G T S
G G S V Q A G G S L R L S C V S F S P S S

for the framework 4 domain

W G Q G T Q V T V S S
W G Q G T L V T V S S
W G Q G A Q V T V S S
W G Q G T Q V T A S S
R G Q G T Q V T V S L

for the CDR3 domain

A L Q P G G Y C G Y G X - - - - - C L
V S L M D R I S Q H - - - - - G C
V P A H L G P G A I L D L K K Y - - - - - K Y
F C Y S T A G D G G S G E - - - - - M Y
E L S G G S C E L P L L F - - - - - D Y
D W K Y W T C G A Q T G G Y F - - - - - G Q
R L T E M G A C D A R W A T L A T R T F A Y N Y
Q K K D R T R W A E P R E W - - - - - N N
G S R F S S P V G S T S R L E S - S D Y - - N Y
A D P S I Y Y S I L X I E Y - - - - - K Y
D S P C Y M P T M P A P P I R D S F G W - - D D
T S S F Y W Y C T T A P Y - - - - - N V
T E I E W Y G C N L R T T F - - - - - T R
N Q L A G G W Y L D P N Y W L S V G A Y - - A I
R L T E M G A C D A R W A T L A T R T F A Y N Y
D G W T R K E G G I G L P W S V Q C E D G Y N Y
D S Y P C H L L - - - - - D V
V E Y P I A D M C S - - - - - R Y

As stated above, the immunoglobulins of the invention are preferably devoid of the totality of their C_H1 domain.

Such immunoglobulins comprise C_H2 and C_H3 domains in the C-terminal region with respect to the hinge region.

According to a particular embodiment of the invention the constant region of the immunoglobulins comprises C_H2 and C_H3 domains comprising an amino-acid sequence selected from the following :

for the C_H2 domain:

APELLGGPTVFIFPPKPKDVLSITLTP
 APELPGGPSVFVFPTKPKDVLSISGRP
 APELPGGPSVFVFPPKPKDVLSISGRP
 APELLGGPSVFIFPPKPKDVLSISGRP

for the C_H3 domain:

GQTREPQVYTLA
 GQTREPQVYTLAPXRLEL
 GQPREPQVYTLPPSRDEL
 GQPREPQVYTLPPSREEM
 GQPREPQVYTLPPSQEEM

Interestingly the inventors have shown that the hinge region of the immunoglobulins of the invention can present variable lengths. When these immunoglobulins act as antibodies, the length of the hinge region will participate to the determination of the distance separating the antigen binding sites.

Preferably an immunoglobulin according to the invention is characterized in that its hinge region comprises from 0 to 50 amino-acids.

Particular sequences of hinge region of the immunoglobulins of the invention are the following.

GTNEVCKCPKCP

or,

EPKIPQFQPKPQFQFQFQPKFQPKPEPECTCPKCP

The short hinge region corresponds to an IgG3 molecule and the long hinge sequence corresponds to an IgG2 molecule.

Isolated V_{HH} derived from heavy chain immunoglobulins or V_{HH} libraries corresponding to the heavy chain immunoglobulins can be distinguished from V_{HH} cloning of four-chain model immunoglobulins on the basis of sequence features characterizing heavy chain immunoglobulins.

The camel heavy - chain immunoglobulin V_{HH} region shows a number of differences with the V_{HH} regions derived from 4-chain immunoglobulins from all species examined. At the levels of the residues involved in the V_{HH}/V_L interactions, an important difference is noted at the level of position 45 (FW) which is practically always leucine in the 4-chain immunoglobulins (98%), the other amino acids at this position being proline (1%) or glutamine (1%).

In the camel heavy-chain immunoglobulin, in the sequences examined at present, leucine at position 45 is only found once. It could originate from a four-chain immunoglobulin. In the other cases, it is replaced by arginine, cysteine or glutamic acid residue. The presence of charged amino acids at this position should contribute to making the V_{HH} more soluble.

The replacement by camelid specific residues such as those of position 45 appears to be interesting for the construction of engineered V_{HH} regions derived from the V_{HH} repertoire of 4-chain immunoglobulins.

A second feature specific of the camelid V_{HH} domain is the frequent presence of a cysteine in the CDR₃ region associated with a cysteine in the CDR₁ position 31 or 33 or FW₂ region at position 45. The possibility of establishing a disulphide bond between the CDR₃ region and the rest of the variable domain

would contribute to the stability and positioning of the binding site.

With the exception of a single pathogenic myeloma protein (DAW) such a disulphide bond has never been encountered in immunoglobulin V regions derived from 4 chain immunoglobulins.

The heavy-chain immunoglobulins of the invention have further the particular advantage of being not sticky. Accordingly these immunoglobulins being present in the serum, aggregate much less than isolated heavy chains of a four-chain immunoglobulins. The immunoglobulins of the invention are soluble to a concentration above 0.5 mg/ml, preferably above 1 mg/ml and more advantageously above 2 mg/ml.

These immunoglobulins further bear an extensive antigen binding repertoire and undergo affinity and specificity maturation in vivo. Accordingly they allow the isolation and the preparation of antibodies having defined specificity, regarding determined antigens.

Another interesting property of the immunoglobulins of the invention is that they can be modified and especially humanized. Especially it is possible to replace all or part of the constant region of these immunoglobulins by all or part of a constant region of a human antibody. For example the C_H2 and/or C_H3 domains of the immunoglobulin could be replaced by the C_H2 and/or C_H3 domains of the IgG γ 3 human immunoglobulin.

In such humanized antibodies it is also possible to replace a part of the variable sequence, namely one or more of the framework residues which do not intervene in the binding site by human framework residues, or by a part of a human antibody.

Conversely features (especially peptide fragments) of heavy-chain immunoglobulin V_{HH} regions, could be introduced into the V_H or V_L regions derived from four-chain immunoglobulins with for instance the

aim of achieving greater solubility of the immunoglobulins.

The invention further relates to a fragment of an immunoglobulin which has been described hereabove and especially to a fragment selected from the following group :

- a fragment corresponding to one heavy polypeptide chain of an immunoglobulin devoid of light chains,
- fragments obtained by enzymatic digestion of the immunoglobulins of the invention, especially those obtained by partial digestion with papain leading to the Fc fragment (constant fragment) and leading to FV_{HH}h fragment (containing the antigen binding sites of the heavy chains) or its dimer F(V_{HH}h)₂, or a fragment obtained by further digestion with papain of the Fc fragment, leading to the pFc fragment corresponding to the C-terminal part of the Fc fragment,
- homologous fragments obtained with other proteolytic enzymes,
- a fragment of at least 10 preferably 20 amino acids of the variable region of the immunoglobulin, or the complete variable region, especially a fragment corresponding to the isolated V_{HH} domains or to the V_{HH} dimers linked to the hinge disulphide,
- a fragment corresponding to the hinge region of the immunoglobulin, or to at least 6 amino acids of this hinge region,
- a fragment of the hinge region comprising a repeated sequence of Pro-X,
- a fragment corresponding to at least 10 preferably 20 amino acids of the constant region or to the complete constant region of the immunoglobulin.

The invention also relates to a fragment comprising a repeated sequence, Pro-X which repeated sequence contains at least 3 repeats of Pro-X, X being any amino-acid and preferably Gln (glutamine), Lys (lysine) or Glu (acide glutamique); a particular repeated fragment is composed of a 12-fold repeat of the sequence Pro-X.

Such a fragment can be advantageously used as a link between different types of molecules.

The amino-acids of the Pro-X sequence are chosen among any natural or non natural amino-acids.

The fragments can be obtained by enzymatic degradation of the immunoglobulins. They can also be obtained by expression in cells or organisms, of nucleotide sequence coding for the immunoglobulins, or they can be chemically synthesized.

The invention also relates to anti-idiotypes antibodies belonging to the heavy chain immunoglobulin classes. Such anti-idiotypes can be produced against human or animal idiotypes. A property of these anti-idiotypes is that they can be used as idiotypic vaccines, in particular for vaccination against glycoproteins or glycolipids and where the carbohydrate determines the epitope.

The invention also relates to anti-idiotypes capable of recognizing idiotypes of heavy-chain immunoglobulins.

Such anti-idiotypic antibodies can be either syngeneic antibodies or allogenic or xenogeneic antibodies.

The invention also concerns nucleotide sequences coding for all or part of a protein which amino-acid sequence comprises a peptide sequence selected from the following :

G G S V Q T G G S L R L S C E I S G L T F D
G G S V Q T G G S L R L S C A V S G F S F S

17

G G S E Q G G G S L R L S C A I S G Y T Y G
G G S V Q P G G S L T L S C T V S G A T Y S
G G S V Q A G G S L R L S C T G S G F P Y S
G G S V Q A G G S L R L S C V A G F G T S
G G S V Q A G G S L R L S C V S F S P S S

W G Q G T Q V T V S S
W G Q G T L V T V S S
W G Q G A Q V T V S S
W G Q G T Q V T A S S
R G Q G T Q V T V S L

A L Q P G G Y C G Y G X - - - - - C L
V S L M D R I S Q H - - - - - G C
V P A H L G P G A I L D L K K Y - - - - - K Y
F C Y S T A G D G G S G E - - - - - M Y
E L S G G S C E L P L L F - - - - - D Y
D W K Y W T C G A Q T G G Y F - - - - - G Q
R L T E M G A C D A R W A T L A T R T F A Y N Y
Q K K D R T R W A E P R E W - - - - - N N
G S R F S S P V G S T S R L E S - S D Y - - N Y
A D P S I Y Y S I L X I E Y - - - - - K Y
D S P C Y M P T M P A P P I R D S F G W - - D D
T S S F Y W Y C T T A P Y - - - - - N V
T E I E W Y G C N L R T T F - - - - - T R
N Q L A G G W Y L D P N Y W L S V G A Y - - A I
R L T E M G A C D A R W A T L A T R T F A Y N Y
D G W T R K E G G I G L P W S V Q C E D G Y N Y
D S Y P C H L L - - - - - D V
V E Y P I A D M C S - - - - - R Y

APELLGGPSVFVFPPKPKDVLSISGXPK
APELLGGPSVFVFPTKPKDVLSISGRPK
APELLGGPSVFVFPPKPKDVLSISGRPK
APELLGGPSVFIFPPKPKDVLSISGRPK
GQTREPQVYTLAPXRLEL
GQPREPQVYTLPPSRDEL
GQPREPQVYTLPPSREEM

GQPREPQVYTLPPSQEEM

VTVSSGTNEVCKCPKCPAPELPGGPSVFVFP

or,

VTVSSEPKIPQPKPQPKPQPKPQPKPEPECTCPKCPAPPELLGGPSVFIFP

GTNEVCKCPKCP

APELPGGPSVFVFP

EPKIPQPKPQPKPQPKPQPKPEPECTCPKCP

APPELLGGPSVFIFP

Such nucleotide sequences can be deduced from the amino-acid sequences taking into account the denegeracy of the genetic code. They can be synthesized or isolated from cells producing immunoglobulins of the invention.

A procedure for the obtention of such DNA sequences is described in the examples.

The invention also contemplates RNA, especially mRNA sequences corresponding to these DNA sequences, and also corresponding cDNA sequences.

The nucleotide sequences of the invention can further be used for the preparation of primers appropriate for the detection in cells or screening of DNA or cDNA libraries to isolate nucleotide sequences coding for immunoglobulins of the invention.

Such nucleotide sequences can be used for the preparation of recombinant vectors and the expression of these sequences contained in the vectors by host cells especially prokaryotic cells like bacteria or also eukaryotic cells and for example CHO cells, insect cells, simian cells like Vero cells, or any other mammalian cells. Especially the fact that the immunoglobulins of the invention are devoid of light chains permits to secrete them in eukaryotic cells since there is no need to have recourse to the step consisting in the formation of the BIP protein which is required in the four-chain immunoglobulins.

The inadequacies of the known methods for producing monoclonal antibodies or immunoglobulins by recombinant DNA technology comes from the necessity in the vast majority of cases to clone simultaneously the V_H and V_L domains corresponding to the specific binding site of 4 chain immunoglobulins. The animals and especially camelids which produce heavy-chain immunoglobulins according to the invention, and possibly other vertebrate species are capable of producing heavy-chain immunoglobulins of which the binding site is located exclusively in the V_{HH} domain. Unlike the few heavy-chain immunoglobulins produced in other species by chain separation or by direct cloning, the camelid heavy-chain immunoglobulins have undergone extensive maturation in vivo. Moreover their V region has naturally evolved to function in absence of the V_L . They are therefore ideal for producing monoclonal antibodies by recombinant DNA technology. As the obtention of specific antigen binding clones does not depend on a stochastic process necessitating a very large number of recombinant cells, this allows also a much more extensive examination of the repertoire.

This can be done at the level of the non rearranged V_{HH} repertoire using DNA derived from an arbitrarily chosen tissue or cell type or at the level of the rearranged V_{HH} repertoire, using DNA obtained from B lymphocytes. More interesting however is to transcribe the mRNA from antibody producing cells and to clone the cDNA with or without prior amplification into an adequate vector. This will result in the obtention of antibodies which have already undergone affinity maturation.

The examination of a large repertoire should prove to be particularly useful in the search for antibodies with catalytic activities.

The invention thus provides libraries which can be generated in a way which includes part of the hinge sequence, the identification is simple as the hinge is directly attached to the V_{HH} domain.

These libraries can be obtained by cloning cDNA from lymphoid cells with or without prior PCR amplification. The PCR primers are located in the promoter, leader or framework sequences of the V_{HH} for the 5' primer and in the hinge, CH₂, CH₃, 3' untranslated region or polyA tail for the 3' primer. A size selection of amplified material allows the construction of a library limited to heavy chain immunoglobulins.

In a particular example, the following 3' primer in which a KpnI site has been constructed and which corresponds to amino-acids 313 to 319 (CGC CAT CAA GGT AAC AGT TGA) is used in conjunction with mouse V_{HH} primers described by Sestry et al and containing a Xho site

AG GTC CAG CTG CTC GAG TCT GG

AG CTC CAG CTG CTC GAG TCT GG

AG GTC CAG CTT CTC GAG TCT GG

XhoI site

These primers yield a library of camelid heavy chain immunoglobulins comprising the V_{HH} region (related to mouse or human subgroup III), the hinge and a section of CH₂.

In another example, the cDNA is polyadenylated at its 5' end and the mouse specific V_{HH} primers are replaced by a poly T primer with an inbuilt XhoI site, at the level of nucleotide 12.

CTCGAGT₁₂.

The same 3' primer with a KpnI site is used.

This method generates a library containing all subgroups of immunoglobulins.

Part of the interest in cloning a region encompassing the hinge-CH₂ link is that in both $\gamma 2$ and

$\gamma 3$, a Sac site is present immediately after the hinge. This site allows the grafting of the sequence coding for the V_{HH} and the hinge onto the Fc region of other immunoglobulins, in particular the human IgG₁ and IgG₃ which have the same amino acid sequence at this site (Glu₂₄₆ Leu₂₄₇).

As an example, the invention contemplates a cDNA library composed of nucleotide sequences coding for a heavy-chain immunoglobulin, such as obtained by performing the following steps:

- a) treating a sample containing lymphoid cells, especially periferal, lymphocytes, spleen cells, lymph nodes or another lymphoid tissue from a healthy animal, especially selected among the Camelids, in order to separate the lymphoid cells,
- b) separating polyadenylated RNA from the other nucleic acids and components of the cells,
- c) reacting the obtained RNA with a reverse transcriptase in order to obtain the corresponding cDNA,
- d) contacting the cDNA of step c) with 5' primers corresponding to mouse V_H domain of four-chain immunoglobulins, which primer contains a determined restriction site, for example an XhoI site and with 3' primers corresponding to the N-terminal part of a C_H2 domain containing a KpnI site,
- e) amplifying the DNA,
- f) cloning the amplified sequence in a vector, especially in a bluescript vector,
- g) recovering the clones hybridizing with a probe corresponding to the sequence coding for a constant domain from an isolated heavy-chain immunoglobulin.

This cloning gives rise to clones containing DNA sequences including the sequence coding for the hinge. It thus permits the characterization of the subclass of the immunoglobulin and the SacI site useful for grafting the FV_{HH} to the Fc region.

The recovery of the sequences coding for the heavy-chain immunoglobulins can also be achieved by the selection of clones containing DNA sequences having a size compatible with the lack of the C_H1 domain.

It is possible according to another embodiment of the invention, to add the following steps between steps c) and d) of the above process:

- in the presence of a DNA polymerase and of deoxyribonucleotide triphosphates, contacting said cDNA with oligonucleotide degenerated primers, which sequences are capable of coding for the hinge region and N-terminal V_{HH} domain of an immunoglobulin, the primers being capable of hybridizing with the cDNA and capable of initiating the extension of a DNA sequence complementary to the cDNA used as template,

- recovering the amplified DNA.

The clones can be expressed in several types of expression vectors. As an example using a commercially available vector Immuno PBS (Huse et al : Science (1989) 246, 1275), clones produced in Bluescript [®] according to the above described procedure, are recovered by PCR using the same XhoI containing 5' primer and a new 3' primer, corresponding to residues 113-103 in the framework of the immunoglobulins, in which an Spe site has been constructed : TC TTA ACT AGT GAG GAG ACG GTG ACC TG. This procedure allows the cloning of the V_{HH} in the Xho/Spe site of the Immuno PBS vector. However, the 3' end of the gene is not in phase with the identification "tag" and the stop codon of the vector. To achieve this, the construct is cut with Spe and the 4 base overhangs are filled in, using the Klenow fragment after which the vector is religated. A further refinement consists in replacing the marker ("tag") with a poly histidine so that metal purification of the cloned V_{HH} can be performed. To

achieve this a Spe/EcoRI double stranded oligo-nucleotide coding for 6 histidines and a termination codon is first constructed by synthesis of both strands followed by heating and annealing :

CTA GTG CAC CAC CAT CAC CAT CAC TAA* TAG*

AC GTG GTG GTA GTG GTA GTG ATT ATC TTA A

The vector containing the insert is then digested with SpeI and EcoRI to remove the resident "tag" sequence which can be replaced by the poly-His/termination sequence. The produced V_{HH} can equally be detected by using antibodies raised against the dromedary V_{HH} regions. Under laboratory conditions, V_{HH} regions are produced in the Immuno PBS vector in mg amounts per liter.

The invention also relates to a DNA library composed of nucleotide sequences coding for a heavy-chain immunoglobulin, such as obtained from cells with rearranged immunoglobulin genes.

In a preferred embodiment of the invention, the library is prepared from cells from an animal previously immunized against a determined antigen. This allows the selection of antibodies having a preselected specificity for the antigen used for immunization.

In another embodiment of the invention, the amplification of the cDNA is not performed prior to the cloning of the cDNA.

The heavy-chain of the four-chain immunoglobulins remains sequestered in the cell by a chaperon protein (BIP) until it has combined with a light chain. The binding site for the chaperon protein is the C_{H1} domain. As this domain is absent from the heavy chain immunoglobulins, their secretion is independent of the presence of the BIP protein or of the light chain. Moreover the inventors have shown that the obtained immunoglobulins are not sticky and accordingly will not abnormally aggregate.

The invention also relates to a process for the preparation of a monoclonal antibody directed against a determined antigen, the antigen binding site of the antibody consisting of heavy polypeptide chains and which antibody is further devoid of light polypeptide chains, which process comprises :

- immortalizing lymphocytes, obtained for example from the peripheral blood of Camelids previously immunized with a determined antigen, with an immortal cell and preferably with myeloma cells, in order to form a hybridoma,
- culturing the immortalized cells (hybridoma) formed and recovering the cells producing the antibodies having the desired specificity.

The preparation of antibodies can also be performed without a previous immunization of Camelids.

According to another process for the preparation of antibodies, the recourse to the technique of the hybridoma cell is not required.

According to such process, antibodies are prepared in vitro and they can be obtained by a process comprising the steps of :

- cloning into vectors, especially into phages and more particularly filamentous bacteriophages, DNA or cDNA sequences obtained from lymphocytes especially PBLs of Camelids previously immunized with determined antigens,
- transforming prokaryotic cells with the above vectors in conditions allowing the production of the antibodies,
- selecting the antibodies for their heavy-chain structure and further by subjecting them to antigen-affinity selection,
- recovering the antibodies having the desired specificity,

In another embodiment of the invention the cloning is performed in vectors, especially into

plasmids coding for bacterial membrane proteins. Prokaryotic cells are then transformed with the above vectors in conditions allowing the expression of antibodies in their membrane.

The positive cells are further selected by antigen affinity selection.

The heavy chain antibodies which do not contain the C_H1 domain present a distinct advantage in this respect. Indeed, the C_H1 domain binds to BIP type chaperone proteins present within eukaryotic vectors and the heavy chains are not transported out of the endocytoplasmic reticulum unless light chains are present. This means that in eukaryotic cells, efficient cloning of 4-chain immunoglobulins in non mammalian cells such as yeast cells can depend on the properties of the resident BIP type chaperone and can hence be very difficult to achieve. In this respect the heavy chain antibodies of the invention which lack the CH₁ domain present a distinctive advantage.

In a preferred embodiment of the invention the cloning can be performed in yeast either for the production of antibodies or for the modification of the metabolism of the yeast. As example, Yep 52 vector can be used. This vector has the origin of replication (ORI) 2 μ of the yeast together with a selection marker Leu 2.

The cloned gene is under the control of gall promoter and accordingly is inducible by galactose. Moreover, the expression can be repressed by glucose which allows the obtention of very high concentration of cells before the induction.

The cloning between BamHI and SalI sites using the same strategy of production of genes by PCR as the one described above, allows the cloning of camelid immunoglobulin genes in E. coli. As example of metabolic modulation which can be obtained by

antibodies and proposed for the yeast, one can site the cloning of antibodies directed against cyclins, that is proteins involved in the regulation of the cellular cycle of the yeast (TIBS 16 430 J.D. McKinney, N. Heintz 1991). Another example is the introduction by genetic engineering of an antibody directed against CD₂₈, which antibody would be inducible (for instance by gall), within the genome of the yeast. The CD₂₈ is involved at the level of the initiation of cell division, and therefore the expression of antibodies against this molecule would allow an efficient control of multiplication of the cells and the optimization of methods for the production in bioreactors or by means of immobilized cells.

In yet another embodiment of the invention, the cloning vector is a plasmid or a eukaryotic virus vector and the cells to be transformed are eukaryotic cells, especially yeast cells, mammalian cells for example CHO cells or simian cells such as Vero cells, insect cells, plant cells, or protozoan cells.

For more details concerning the procedure to be applied in such a case, reference is made to the publication of Marks et al, J. Mol. Biol. 1991, 222:581-597.

Furthermore, starting from the immunoglobulins of the invention, or from fragments thereof, new immunoglobulins or derivatives can be prepared.

Accordingly immunoglobulins replying to the above given definitions can be prepared against determined antigens. Especially the invention provides monoclonal or polyclonal antibodies devoid of light polypeptide chains or antisera containing such antibodies and directed against determined antigens and for example against antigens of pathological agents such as bacteria, viruses or parasites. As example of antigens or antigenic determinants against which antibodies

could be prepared, one can cite the envelope glycoproteins of viruses or peptides thereof, such as the external envelope glycoprotein of a HIV virus, the surface antigen of the hepatitis B virus.

Immunoglobulins of the invention can also be directed against a protein, hapten, carbohydrate or nucleic acid.

Particular antibodies according to the invention are directed against the galactosyl α -1-3-galactose epitope.

The immunoglobulins of the invention allow further the preparation of combined products such as the combination of the heavy-chain immunoglobulin or a fragment thereof with a toxin, an enzyme, a drug, a hormone.

As example one can prepare the combination of a heavy-chain immunoglobulin bearing an antigen binding site recognizing a myeloma immunoglobulin epitope with the abrin or mistletoe lectin toxin. Such a construct would have its uses in patient specific therapy.

Another advantageous combination is that one can prepare between a heavy-chain immunoglobulins recognizing an insect gut antigen with a toxin specific for insects such as the toxins of the different serotypes of *Bacillus thuringiensis* or *Bacillus sphaericus*. Such a construct cloned into plants can be used to increase the specificity or the host range of existing bacterial toxins.

The invention also proposes antibodies having different specificities on each heavy polypeptide chains. These multifunctional, especially bifunctional antibodies could be prepared by combining two heavy chains of immunoglobulins of the invention or one heavy chain of an immunoglobulin of the invention with a fragment of a four-chain model immunoglobulin.

The invention also provides hetero-specific antibodies which can be used for the targetting of

drugs or any biological substance like hormones. In particular they can be used to selectively target hormones or cytokines to a limited category of cells. Examples are a combination of a murine or human antibody raised against interleukin 2 (IL₂) and a heavy-chain antibody raised against CD₄ cells. This could be used to reactivate CD₄ cells which have lost their IL₂ receptor.

The heavy-chain immunoglobulins of the invention can also be used for the preparation of hetero-specific antibodies. These can be achieved either according to the above described method by reduction of the bridges between the different chains and reoxydation, according to the usual techniques, of two antibodies having different specificities, but it can also be achieved by serial cloning of two antibodies for instance in the Immuno pBS vector.

In such a case, a first gene corresponding to the V_{HH} domain comprised between Xho site and a Spe site is prepared as described above. A second gene is then prepared through an analogous way by using as 5' extremity a primer containing a Spe site, and as 3' extremity a primer containing a termination codon and an EcoRI site. The vector is then digested with EcoRI and XhoI and further both V_{HH} genes are digested respectively by Xho/Spe and by Spe/EcoRI.

After ligation, both immunoglobulin genes are serially cloned. The spacing between both genes can be increased by the introduction of addition codons within the 5' SpeI primer.

In a particular embodiment of the invention, the hinge region of IgG2 immunoglobulins according to the invention is semi-rigid and is thus appropriate for coupling proteins. In such an application proteins or peptides can be linked to various substances, especially to ligands through the hinge region used as

spacer. Advantageously the fragment comprises at least 6 amino acids.

According to the invention it is interesting to use a sequence comprising a repeated sequence Pro-X, X being any amino-acid and preferably Gln, Lys or Glu, especially a fragment composed of at least a 3-fold repeat and preferably of a 12-fold repeat, for coupling proteins to ligand, or for assembling different protein domains.

The hinge region or a fragment thereof can also be used for coupling proteins to ligands or for assembling different protein domains.

Usual techniques for the coupling are appropriate and especially reference may be made to the technique of protein engineering by assembling cloned sequences.

The antibodies according to this invention could be used as reagents for the diagnosis in vitro or by imaging techniques. The immunoglobulins of the invention could be labelled with radio-isotopes, chemical or enzymatic markers or chemiluminescent markers.

As example and especially in the case of detection or observation with the immunoglobulins by imaging techniques, a label like technetium, especially technitium 99 is advantageous. This label can be used for direct labelling by a coupling procedure with the immunoglobulins or fragments thereof or for indirect labelling after a step of preparation of a complex with the technitium.

Other interesting radioactive labels are for instance indium and especially indium 111, or iodine, especially I^{131} , I^{125} and I^{123} .

For the description of these techniques reference is made to the FR patent application published under number 2649488.

In these applications the small size of the V_{HH} fragment is a definitive advantage for penetration into tissue.

The invention also concerns monoclonal antibodies reacting with anti-idiotypes of the above-described antibodies.

The invention also concerns cells or organisms in which heavy-chain immunoglobulins have been cloned. Such cells or organisms can be used for the purpose of producing heavy-chain immunoglobulins having a desired preselected specificity, or corresponding to a particular repertoire. They can also be produced for the purpose of modifying the metabolism of the cell which expresses them. In the case of modification of the metabolism of cells transformed with the sequences coding for heavy-chain immunoglobulins, these produced heavy-chain immunoglobulins are used like antisense DNA. Antisense DNA is usually involved in blocking the expression of certain genes such as for instance the variable surface antigen of trypanosomes or other pathogens. Likewise, the production or the activity of certain proteins or enzymes could be inhibited by expressing antibodies against this protein or enzyme within the same cell.

The invention also relates to a modified 4-chain immunoglobulin or fragments thereof, the V_H regions of which has been partially replaced by specific sequences or amino acids of heavy chain immunoglobulins, especially by sequences of the V_{HH} domain. A particular modified V_H domain of a four-chain immunoglobulin, is characterized in that the leucine, proline or glutamine in position 45 of the V_H regions has been replaced by other amino acids and preferably by arginine, glutamic acid or cysteine.

A further modified V_H or V_L domain of a four-chain immunoglobulin, is characterized by linking of CDR loops together or to FW regions by the

introduction of paired cysteines, the CDR region being selected between the CDR₁ and the CDR₃, the FW region being the FW₂ region, and especially in which one of the cysteines introduced is in position 31, 33 of FR₂ or 45 of CDR₂ and the other in CDR₃.

Especially the introduction of paired cysteines is such that the CDR₃ loop is linked to the FW₂ or CDR₁ domain and more especially the cysteine of the CDR₃ of the V_H is linked to a cysteine in position 31 or 33 of FW₂ or in position 45 of CDR₂.

In another embodiment of the invention, plant cells can be modified by the heavy-chain immunoglobulins according to the invention, in order that they acquire new properties or increased properties.

The heavy-chain immunoglobulins of the invention can be used for gene therapy of cancer for instance by using antibodies directed against proteins present on the tumor cells.

In such a case, the expression of one or two V_{HH} genes can be obtained by using vectors derived from parvo or adeno viruses. The parvo viruses are characterized by the fact that they are devoid of pathogenicity or almost not pathogenic for normal human cells and by the fact that they are capable of easily multiplying in cancer cells (Russel S.J. 1990, Immunol. Today II. 196-200).

The heavy-chain immunoglobulins are for instance cloned within HindIII/XbaI sites of the infectious plasmid of the murine MVM virus (pMM984). (Merchinsky et al, 1983, J. Virol. 47, 227-232) and then placed under the control of the MVM38 promoter.

The gene of the V_{HH} domain is amplified by PCR by using a 5' primer containing an initiation codon and a HindIII site, the 3' primer containing a termination codon and a XbaI site.

This construct is then inserted between positions 2650 (HindIII) and 4067 (XbaI) within the plasmid.

The efficiency of the cloning can be checked by transfection. The vector containing the antibody is then introduced in permissive cells (NB-E) by transfection.

The cells are recovered after two days and the presence of V_{HH} regions is determined with an ELISA assay by using rabbit antiserum reacting with the V_{HH} part.

The invention further allows the preparation of catalytic antibodies through different ways. The production of antibodies directed against components mimicking activated states of substrates (as example vanadate as component mimicking the activated state of phosphate in order to produce their phosphoesterase activities, phosphonate as compound mimicking the peptidic binding in order to produce proteases) permits to obtain antibodies having a catalytic function. Another way to obtain such antibodies consists in performing a random mutagenesis in clones of antibodies for example by PCR, in introducing abnormal bases during the amplification of clones. These amplified fragments obtained by PCR are then introduced within an appropriate vector for cloning. Their expression at the surface of the bacteria permits the detection by the substrate of clones having the enzymatic activity. These two approaches can of course be combined. Finally, on the basis of the data available on the structure, for example the data obtained by XRay crystallography or NMR, the modifications can be directed. These modifications can be performed by usual techniques of genetic engineering or by complete synthesis. One advantage of the V_{HH} of the heavy chain immunoglobulins of the invention is the fact that they are sufficiently soluble.

The heavy chain immunoglobulins of the invention can further be produced in plant cells, especially in transgenics plants. As example the heavy chain immunoglobulins can be produced in plants using the pMon530 plasmid (Roger et al. Meth Enzym 153 1566 1987) constitutive plant expression vector as has been described for classical four chain antibodies (Hiat et al. Nature 342 76-78, 1989) once again using the appropriate PCR primers as described above, to generate a DNA fragment in the right phase.

Other advantages and characteristics of the invention will become apparent in the examples and figures which follow.

F I G U R E S

Figure 1 : Characterisation and purification of camel IgG by affinity chromatography on Protein A and Protein G sepharose (Pharmacia)

(A) shows, after reduction, the SDS-PAGE protein profile of the adsorbed and non adsorbed fractions of Camelus dromedarius serum. The fraction adsorbed on Protein A and eluted with NaCl 0.15 M acetic acid 0.58% show upon reduction (lane c) three heavy chain components of respectively 50, 46 and 43 Kd and light chain (rabbit IgG in lane a). The fractions adsorbed on a Protein G Sepharose (Pharmacia) derivative which has been engineered to delete the albumin binding region (lane e) and eluted with 0.1 M gly HCl pH 2.7 lacks the 46 Kd heavy chain which is recovered in the non adsorbed fraction (lane f). None of these components are present in the fraction non adsorbed on Protein A (lane d), lane b contains the molecular weight markers.

(B) and (C) By differential elution, immunoglobulin fractions containing the 50 and 43 Kd heavy chain can be separated. 5 ml of C. dromedarius serum is adsorbed onto a 5 ml Protein G sepharose column and the column is extensively washed with 20mM phosphate buffer, pH 7.0. Upon elution with pH 3.5 buffer (0.15 M NaCl, 0.58% acetic acid) a 100 Kd component is eluted which upon reduction yields a 43 Kd heavy chain, (lane 1). After column eluant absorbance has fallen to background level a second immunoglobulin component of 170 Kd can be eluted with pH 2.7 buffer (0.1 M glycine HC). This fraction upon reduction yields a 50 Kd heavy chain and a board light chain band (lane 2).

The fraction non adsorbed on Protein G is then brought on a 5 ml Protein A Sepharose column. After washing

and elution with pH 3.5 buffer (0.15 M NaCl, 0.58% acetic acid) a third immunoglobulin of 100 Kd is obtained which consists solely of 46 Kd heavy chains (lane 3).

Figure 2 : Immunoglobulins of Camelus bactrianus, Lama vicugna, Lama glama and Lama pacos to Protein A (A lanes) and to Protein G (G lanes) analyzed on SDS-PAGE before (A) and after reduction (B)

10 μ l of serum obtained from the different species were added to Eppendorf^R tubes containing 10 mg of Protein A or Protein G sepharose suspended in 400 μ l of pH 8.3 immunoprecipitation buffer (NaCl 0.2 M, Tris 0.01 M; EDTA 0.01 M, Triton X100 1%, ovalbumin 0.1%). The tubes were slowly rotated for 2 hours at 4°C. After centrifugation the pellets were washed 3 times in buffer and once in buffer in which the Triton and ovalbumin had been omitted. The pellets were then resuspended in the SDS-PAGE sample solution 70 μ l per pellet with or without dithiotreitol as reductant. After boiling for 3 min at 100°C, the tubes were centrifuged and the supernatants analysed.

In all species examined the unreduced fractions (A) contain in addition to molecules of approximately 170 Kd also smaller major components of approximately 100 Kd. In the reduced sample (B) the constituent heavy and light chains are detected. In all species a heavy chain component (marked by an asterisk *) is present in the material eluted from the Protein A but absent in the material eluted from the Protein G.

Figure 3 : IgG₁, IgG₂ and IgG₃ were prepared from serum obtained from healthy or Trypanosoma evansi infected Camelus dromedarius (CATT titer 1/160 (3) and analysed by radioimmunoprecipitation

or Western Blotting for anti
trypanosome activity

(A) ^{35}S methionine labelled Trypanosome evansi antigens lysate (500.000 counts) was added to Eppendorf tubes containing 10 μl of serum or, 20 μg of IgG_1 , IgG_2 or IgG_3 in 200 μl of pH 8.3 immunoprecipitation buffer containing 0.1 M TLCK as proteinase inhibitor and slowly rotated at 4°C during one hour. The tubes were then supplemented with 10 mg of Protein A Sepharose suspended in 200 μl of the same pH 8.3 buffer and incubated at 4°C for an additional hour.

After washing and centrifugation at 15000 rpm for 12 s, each pellet was resuspended in 75 μl SDS-PAGE sample solution containing DTT and heated for 3 min. at 100°C. After centrifugation in an Eppendorf minifuge at 15000 rpm for 30 s, 5 μl of the supernatant was saved for radioactivity determination and the remainder analysed by SDS-PAGE and fluorography. The counts/5 μl sample are inscribed on for each line.

(B) 20 μg of IgG_1 , IgG_2 and IgG_3 from healthy and trypanosome infected animals were separated by SDS-PAGE without prior reduction or heating. The separated samples were then electro transferred to a nitrocellulose membrane, one part of the membrane was stained with Ponceau Red to localise the protein material and the remainder incubated with 1% ovalbumin in TST buffer (Tris 10 mM, NaCl 150 mM, Tween 0.05%) to block protein binding sites.

After blocking, the membrane was extensively washed with TST buffer and incubated for 2 hours with ^{35}S -labelled trypanosome antigen. After extensive washing, the membrane was dried and analysed by autoradiography. To avoid background and unspecific binding, the labelled trypanosome lysate was filtered

through a 45 μ millipore filter and incubated with healthy camel immunoglobulin and ovalbumin adsorbed on a nitrocellulose membrane.

figure 4 : Purified IgG3 of the camel, by affinity chromatography on Protein A Sepharose are partially digested with papain and separated on Protein A sepharose.

14mg of purified IgG3 were dissolved in 0.1M phosphate buffer pH 7.0 containing 2mM EDTA. They were digested by 1 hour incubation at 37°C with mercurypapain (1% enzyme to protein ratio) activated by $5 \cdot 10^{-4}$ M cysteine. The digestion was blocked by the addition of excess iodoacetamide ($4 \cdot 10^{-2}$ M) (13). After centrifugation of the digest in an ependorf centrifuge for 5min at 15000 rpm, the papain fragments were separated on a protein A Sepharose column into binding (B) and non binding (NB) fractions. The binding fraction was eluted from the column with 0.1M glycine HCl buffer pH 1.7.

Figure 5 : Schematic presentation of a model for IgG3 molecules devoid of light chains.

Figure 6 : . Schematic representation of immunoglobulins having heavy polypeptide chains and devoid of light chains, regarding conventional four-chain model immunoglobulin
. Representation of a hinge region.

Figure 7: Alignement of 17 V_{HH} DNA sequences of Camel heavy chain immunoglobulins

Figure 8: Expression and purification of the camel V_{HH}21 protein from E.coli

I HEAVY CHAIN ANTIBODIES IN CAMELIDS

When Camelus dromedarius serum is adsorbed on Protein G sepharose, an appreciable amount (25-35%) of immunoglobulins (Ig) remains in solution which can then be recovered by affinity chromatography on Protein A sepharose (fig. 1A). The fraction adsorbed on Protein G can be differentially eluted into a tightly bound fraction (25%) consisting of molecules of an unreduced apparent molecular weight (MW) of 170 Kd and a more weakly bound fraction (30-45%) having an apparent molecular weight of 100 Kd (fig. 1B). The 170 Kd component when reduced yields 50 Kd heavy chains and large 30 Kd light chains. The 100 Kd fraction is totally devoid of light chains and appears to be solely composed of heavy chains which after reduction have an apparent MW of 43 Kd (Fig. 1C). The fraction which does not bind to Protein G can be affinity purified and eluted from a Protein A column as a second 100 Kd component which after reduction appears to be composed solely of 46 Kd heavy chains.

The heavy chain immunoglobulins devoid of light chains total up to 75% of the molecules binding to Protein A.

As all three immunoglobulins bind to Protein A we refer to them as IgG : namely IgG₁ (light chain and heavy chain γ_1 (50 Kd) binding to Protein G, IgG₂ (heavy chain γ_2 (46 Kd) non binding to Protein G and IgG₃ (heavy chain γ_3 (43 Kd) binding to Protein G. There is a possibility that these three sub(classes) can be further subdivided.

A comparative study of old world camelids (Camelus bactrianus and Camelus dromedarius) and new world camelids (Lama pacos, Lama glama, Lama vicugna) showed that heavy chain immunoglobulins are found in all species examined, albeit with minor differences in apparent molecular weight and proportion. The new

world camelids differs from the old world camelids in having a larger IgG₃ molecule (heavy chain immunoglobulin binding to Protein G) in which the constituent heavy chains have an apparent molecular weight of 47 Kd (fig. 2).

The abundance of the heavy chain immunoglobulins in the serum of camelids raises the question of what their role is in the immune response and in particular whether they bear antigen binding specificity and if so how extensive is the repertoire. This question could be answered by examining the immunoglobulins from Trypanosoma evansi infected camels (Camelus dromedarius).

For this purpose, the corresponding fractions of IgG₁, IgG₂, IgG₃ were prepared from the serum of a healthy camel and from the serum of camels with a high antitrypanosome titer, measured by the Card Agglutination Test (3). In radio-immunoprecipitation, IgG₁, IgG₂ and IgG₃ derived from infected camel indicating extensive repertoire heterogeneity and complexity (Fig. 3A) were shown to bind a large number of antigens present in a ³⁵S methionine labelled trypanosome lysate.

In blotting experiments ³⁵S methionine labelled trypanosome lysate binds to SDS PAGE separated IgG₁, IgG₂ and IgG₃ obtained from infected animals (Fig. 3B).

This leads us to conclude that the camelid heavy chain IgG₂ and IgG₃ are bona fide antigen binding antibodies.

An immunological paradigm states that an extensive antibody repertoire is generated by the combination of the light and heavy chain variable V region repertoires (6). The heavy chain immunoglobulins of the camel seem to contradict this paradigm.

Immunoglobulins are characterized by a complex I.E.F. (isoelectric focussing) pattern reflecting their extreme heterogeneity. To determine whether the two heavy chains constituting the IgG₂ and IgG₃ are identical or not, the isoelectric focussing (I.E.F.) pattern were observed before and after chain separation by reduction and alkylation using iodoacetamide as alkylating agent.

As this alkylating agent does not introduce additional charges in the molecule, the monomers resulting from the reduction and alkylation of a heavy chain homodimer will have practically the same isoelectric point as the dimer, whereas if they are derived from a heavy chain heterodimer, the monomers will in most cases differ sufficiently in isoelectric point to generate a different pattern in I.E.F.

Upon reduction, and alkylation by iodoacetamide the observed pattern is not modified for the Camelus dromedarius IgG₂ and IgG₃ indicating that these molecules are each composed of two identical heavy chains which migrate to the same position as the unreduced molecule they originated from.

In contrast, the I.E.F. pattern of IgG₁ is completely modified after reduction as the isoelectric point of each molecule is determined by the combination of the isoelectric points of the light and heavy chains which after separation will each migrate to a different position.

These findings indicate that the heavy chains alone can generate an extensive repertoire and question the contribution of the light chain to the useful antibody repertoire. If this necessity be negated, what other role does the light chain play.

Normally, isolated heavy chain from mammalian immunoglobulins tend to aggregate considerably but are only solubilized by light chains (8, 9) which bind to the C_H1 domain of the heavy chain.

In humans and in mice a number of spontaneous or induced myelomas produce a pathological immunoglobulin solely composed of heavy chains (heavy chain disease). These myeloma protein heavy chains carry deletions in the C_{H1} and V_{HH} domains (10). The reason why full length heavy chains do not give rise to secreted heavy chain in such pathological immunoglobulins seems to stem from the fact that the synthesis of Ig involves a chaperoning protein, the immunoglobulin heavy chain binding protein or BIP (11), which normally is replaced by the light chain (12). It is possible that the primordial role of the light chain in the four-chain model immunoglobulins is that of a committed heavy chain chaperon and that the emergence of light chain repertoires has just been an evolutionary bonus.

The camelid $\gamma 2$ and $\gamma 3$ chains are considerably shorter than the normal mammalian γ chain. This would suggest that deletions have occurred in the C_{H1} domain. Differences in sizes of the $\gamma 2$ and $\gamma 3$ immunoglobulins of old and new world camelids suggests that deletions occurred in several evolutionary steps especially in the C_{H1} domain.

II THE HEAVY CHAIN IMMUNOGLOBULINS OF THE CAMELIDS LACK THE C_{H1} DOMAIN.

The strategy followed for investigating the heavy chain immunoglobulin primary structure is a combination of protein and cDNA sequencing ; the protein sequencing is necessary to identify sequence stretches characteristic of each immunoglobulin. The N-terminal of the immunoglobulin being derived from the heavy chain variable region repertoire only yields information on the V_{HH} subgroups (variable region of the heavy chain) and cannot be used for class or subclass identification. This means that sequence data

had to be obtained from internal enzymatic or chemical cleavage sites.

A combination of papain digestion and Protein A affinity chromatography allowed the separation of various fragments yielding information on the general structure of IgG3.

The IgG3 of the camel (Camelus dromedarius) purified by affinity chromatography on Protein A Sepharose were partially digested with papain and the digest was separated on Protein A Sepharose into binding and non binding fractions. These fractions were analysed by SDS PAGE under reducing and non reducing conditions (fig 4).

The bound fraction contained two components, one of 28 Kd and one of 14.4 Kd, in addition to uncleaved or partially cleaved material. They were well separated by gel electrophoresis (from preparative 19% SDS-PAGE gels) under non reducing conditions and were further purified by electroelution(in 50mM ammonium bicarbonate, 0.1% (w/v) SDS using a BioRad electro-eluter). After lyophilization of these electroeluted fractions, the remaining SDS was eliminated by precipitating the protein by the addition of 90% ethanol, mixing and incubating the mixture overnight at -20°C (14). The precipitated protein was collected in a pellet by centrifuging (15000 rpm, 5min) and was used for protein sequencing. N-terminal sequencing was performed using the automated Edman chemistry of an Applied Biosystem 477A pulsed liquid protein sequencer. Amino acids were identified as their phenylthiohydantoin (PTH) derivatives using an Applied Biosystem 120 PTH analyser. All chemical and reagents were purchased from Applied Biosystems. Analysis of the chromatographic data was performed using Applied Biosystems software version 1.61. In every case the computer aided sequence analysis was confirmed by

direct inspection of the chromatograms from the PTH analyser. Samples for protein sequencing were dissolved in either 50% (v/v) trifluoroacetic acid (TFA) (28Kd fragment) or 100% TFA (14Kd fragment). Samples of dissolved protein equivalent to 2000 pmol (28Kd fragment) or 500 pmol (14Kd fragment) were applied to TFA-treated glass fibre discs. The glass fibre discs were coated with BioBrene (3mg) and precycled once before use.

N-terminal sequencing of the 28 Kd fragment yields a sequence homologous to the N-terminal part of γ C_H2 domain and hence to the N-terminal end of the Fc fragment. The N-terminal sequence of the 14.4 Kd fragment corresponds to the last lysine of a γ C_H2 and the N-terminal end of a γ C_H3 domain (Table 1). The molecular weight (MW) of the papain fragments and the identification of their N-terminal sequences led us to conclude that the C_H2 and C_H3 domains of the γ 3 heavy chains are normal in size and that the deletion must occur either in the C_H1 or in the V_{HH} domain to generate the shorted γ 3 chain. The fractions which do not bind to Protein A Sepharose contain two bands of 34 and 17 Kd which are more diffuse in SDS PAGE indicating that they originate from the variable N-terminal part of the molecule (fig 4).

Upon reduction, a single diffuse band of 17 Kd is found indicating that the 34 Kd is a disulfide bonded dimer of the 17 Kd component. The 34 Kd fragment apparently contains the hinge and the N-terminal domain V_{HH}.

The protein sequence data can be used to construct degenerate oligonucleotide primers allowing PCR amplification of cDNA or genomic DNA.

It has been shown that the cells from camel spleen imprint cells reacted with rabbit and anti camel immunoglobulin sera and that the spleen was hence a site of synthesis of at least one

immunoglobulin class. cDNA was therefore synthesised from camel spleen mRNA. The conditions for the isolation of RNA were the following: total RNA was isolated from the dromedary spleen by the guanidium isothiocyanate method (15). mRNA was purified with oligo T-paramagnetic beads.

cDNA synthesis is obtained using 1 μ g mRNA template, an oligodT primer and reverse transcriptase (BOEHRINGER MAN). Second strand cDNA is obtained using RNase H and E coli DNA polymerase I according to the condition given by the supplier.

Relevant sequences were amplified by PCR: 5ng of cDNA was amplified by PCR in a 100 μ l reaction mixture (10mM Tris-HCl pH 8.3, 50mM KCl, 15mM MgCl₂, 0.01% (w/v) gelatine, 200 μ M of each dNTP and 25 pmoles of each primer) overlaid with mineral oil (Sigma).

Degenerate primers containing EcoRI and KpnI sites and further cloned into pUC 18. After a round of denaturing and annealing (94°C for 5 min and 54°C for 5 min), 2 units of Taq DNA polymerase were added to the reaction mixture before subjecting it to 35 cycles of amplification: 1 min at 94°C (denature) 1min at 54°C (anneal) , 2 min at 72°C (elongate). To amplify DNA sequences between V_{HH} and C_{H2} domains , (# 72 clones), the PCR was performed in the same conditions with the exception that the annealing temperature was increased to 60°C.

One clone examined (#56/36) had a sequence corresponding to the N-terminal part of a C_{H2} domain identical to the sequence of the 28 Kd fragment. The availability of this sequence data allowed the construction of an exact 3' primer and the cloning of the region between the N-terminal end of the V_{HH} and the C_{H2} domain.

5' primers corresponding to the mouse V_{HH} (16) and containing a XhoI restriction site were used in conjunction with the 3' primer in which a KpnI site

had been inserted and the amplified sequences were cloned into pBluescript^R. Clone #56/36 which displayed two internal HaeIII sites was digested with this enzyme to produce a probe to identify PCR positive clones.

After amplification the PCR products were checked on a 1.2% (w/v) agarose gel. Cleaning up of the PCR products included a phenol-chloroform extraction followed by further purification by HPLC (GEN-PAC FAX column, Waters) and finally by using the MERMAID or GENE CLEAN II kit, BIO 101, Inc) as appropriate. After these purification steps, the amplified cDNA was then digested with EcoRI and KpnI for series #56 clones and with XhoI and KpnI for series #72 clones. A final phenol-chloroform extraction preceded the ligation into pUC 18(series #56 clones) or into pBluescript^R (series #72 clones).

All the clones obtained were smaller than the 860 base pairs to be expected if they possessed a complete V_{HH} and C_{H1} region. Partial sequence data corresponding to the N-terminal of the V_{HH} region reveals that out of 20 clones, 3 were identical and possibly not independent. The sequences obtained resemble the human subgroup III and the murine subgroups IIIa and IIIb (Table 2).

Clones corresponding to two different sets of C_{H2} protein sequences were obtained. A first set of sequences (#72/41) had a N-terminal C_{H2} region identical to the one obtained by protein sequencing of the 28 Kd papain fragments of the $\gamma 3$ heavy chain, a short hinge region containing 3 cysteines and a variable region corresponding to the framework (FR4) residues encoded by the J minigenes adjoining the hinge. The C_{H1} domain is entirely lacking. This cDNA corresponds to the $\gamma 3$ chain (Table 4).

In one closely related sequence (#72/1) the proline in position 259 is replaced by threonine.

The sequence corresponding to the C_H3 and the remaining part of the C_H2 was obtained by PCR of the cDNA using as KpnI primer a poly T in which a KpnI restriction site had been inserted at the 5' end. The total sequence of the γ 3 chain corresponds to a molecular weight (MW) which is in good agreement with the data obtained from SDS PAGE electrophoresis.

The sequence of this γ 3 chain presents similarities with other γ chains except that it lacks the C_H1 domain, the V_{HH} domain being adjacent to the hinge.

One or all three of the cysteines could be probably responsible for holding the two γ 3 chains together.

These results have allowed us to define a model for the IgG3 molecule based on sequence and papain cleavage (fig. 5).

Papain can cleave the molecule on each side of the hinge disulfides and also between C_H2 and C_H3. Under non reducing conditions the V_{HH} domains of IgG3 can be isolated as disulfide linked dimer or as monomer depending on the site of papain cleavage.

A second set of clones #72/29 had a slightly different sequence for the C_H2 and was characterized by a very long hinge immediately preceded by the variable domain. This hinge region has 3 cysteines at its C-terminal end in a sequence homologous to the γ 3 hinge. Such second set of clones could represent the IgG2 subclass. For the constant part of the γ 3 and also for the putative γ 2, most clones are identical showing the γ 2 or γ 3 specific sequences. A few clones such as #72/1 however show minor differences. For instance in the case of clones #72/1 two nucleotide differences are detected.

Several V_{HH} regions cDNA's have now been totally or partially sequenced with the exception of a short stretch at the N-terminal end which is primer derived.

Upon translation the majority shows the characteristic heavy chain Ser₂₁ Cys₂₂ and Tyr₉₀ Tyr₉₁ Cys₉₂ sequences, of the intra V_{HH} region disulfide bridge linking residues 22 and 92. All these clones have a sequence corresponding to the framework 4 (FR4) residues of the variable region immediately preceding the postulated hinge sequence (Table 3). This sequence is generated by the J minigenes and is in the majority of cases similar to the sequence encoded by the human and murine J minigenes. The sequence length between region Cys₉₂ and the C-terminal end of the V_{HH} regions is variable and, in the sequences determined, range from 25 to 37 amino-acids as one might expect from the rearrangements of J and D minigenes varying in length.

Several important questions are raised by the sole existence of these heavy chain immunoglobulins in a non pathological situation. First of all, are they bonafide antibodies ? The heavy chain immunoglobulins obtained from trypanosome infected camels react with a large number of parasite antigens as shown in part I of these examples. This implies that the camelid immune system generates an extensive number of binding sites composed of single V_{HH} domains. This is confirmed by the diversity of the V_{HH} regions of the heavy chain immunoglobulins obtained by PCR.

The second question is "how are they secreted ?". The secretion of immunoglobulin heavy chains composing four-chain model immunoglobulins does not occur under normal conditions. A chaperoning protein, the heavy chain binding protein, or BIP protein, prevents heavy chains from being secreted. It is only when the light chain displaces the BIP protein in the endoplasmatic reticulum that secretion can occur(13).

The heavy chain dimer found in the serum of human or mice with the so-called "heavy chain disease" lack

the C_H1 domains thought to harbour the BIP site (14). In the absence of this domain the BIP protein can no longer bind and prevent the transport of the heavy chains.

The presence in camels of a IgG1 class composed of heavy and light chains making up between 25% and 50% of the total IgG molecules also raises the problem as to how maturation and class switching occurs and what the role of the light chain is. The camelid light chain appears unusually large and heterogeneous when examined in SDS PAGE.

The largest dimension of an isolated domain is 40 Å and the maximum attainable span between binding sites of a conventional IgG with C_H1 and V_{HH} will be of the order of 160 Å ($2V_{HH} + 2C_H1$) (19). The deletion of C_H1 domain in the two types of heavy chain antibodies devoid of light chains, already sequenced has, as a result, a modification of this maximum span (fig. 6). In the IgG3 the extreme distance between the extremities of the V_{HH} regions will be of the order of 80 Å ($2V_{HH}$). This could be a severe limitation for agglutinating or cross linking. In the IgG2 this is compensated by the extremely long stretch of hinge, composed of a 12-fold repeat of the sequence Pro-X (where X is Gln, Lys or Glu) and located N-terminal to the hinge disulfide bridges. In contrast, in the human IgG3, the very long hinge which also apparently arose as the result of sequence duplication does not contribute to increase the distance spanning the two binding sites as this hinge is inter-spersed with disulfide bridges.

The single V_{HH} domain could also probably allow considerably rotational freedom of the binding site versus the Fc domain.

Unlike myeloma heavy chains which result probably from C_H1 deletion in a single antibody producing cell, or heavy chain antibodies produced by expression

cloning(15); the camelid heavy chain antibodies (devoid of light chains) have emerged in a normal immunological environment and it is expected that they will have undergone the selective refinement in specificity and affinity accompanying B cell maturation.

Expression and purification of the camel V_{HH}21 (DR21 on figure 7) protein from E.coli

The clones can be expressed in several types of expression vectors. As an example using a commercially available vector Immuno PBS (Huse et al : Science (1989) 246, 1275), clones produced in Bluescript ® according to the above described procedure, have been recovered by PCR using the same XhoI containing 5' primer and a new 3' primer, corresponding to residues 113-103 in the framework of the immunoglobulins, in which an Spe site has been constructed : TC TTA ACT AGT GAG GAG ACG GTG ACC TG. This procedure allowed the cloning of the V_{HH} in the Xho/Spe site of the Immuno PBS vector. However, the 3' end of the gene was not in phase with the identification "tag" and the stop codon of the vector. To achieve this, the construct was cut with Spe and the 4 base overhangs were filled in, using the Klenow fragment after which the vector was religated.

- The expression vector plasmid ipBS (immunopBS) (Stratagene) contains a pel B leader sequence which is used for immunoglobulin chain expression in E.coli under the promotor pLAC control, a ribosome binding site, and stop codons. In addition, it contains a sequence for a c-terminal decapeptide tag.

- E.coli JM101 harboring the ipBS-V_{HH}21 plasmid was grown in 1 l of TB medium with 100 µg/ml ampicillin and 0.1 % glucose at 32°C. Expression was induced by the addition of 1 mM IPTG (final concentration) at an OD₅₅₀ of 1.0. After overnight induction at 28°C, the

cells were harvested by centrifugation at 4.000 g for 10 min (4°C) and resuspended in 10 ml TES buffer (0.2 M Tris-HCL pH 8.0, 0.5 mM EDTA, 0.5 M sucrose). The suspension was kept on ice for 2 hours. Periplasmic proteins were removed by osmotic shock by addition of 20 ml TES buffer diluted 1:4 v/v with water, kept on ice for one hour and subsequently centrifugated at 12.000 g for 30 min. at 4°C. The supernatant periplasmic fraction was dialysed against Tris-HCl pH 8.8, NaCl 50mM, applied on a fast Q Sepharose flow (Pharmacia) column, washed with the above buffer prior and eluted with a linear gradient of 50 mM to 1 M NaCl in buffer.

Fractions containing the V_{HH} protein were further purified on a Superdex 75 column (Pharmacia) equilibrated with PBS buffer (0.01 M phosphate pH 7.2, 0.15 M NaCl). The yield of purified V_{HH} protein varies from 2 to 5 mg/1 cell culture.

- Fractions were analyzed by SDS-PAGE(I). Positive identification of the camel V_{HH} antibody fragment was done by Western Blot analysis using antibody raised in rabbits against purified camel IgGH₃ and an anti-rabbit IgG-alkaline phosphatase conjugate (II).

As protein standards (Pharmacia) periplasmic proteins prepared from 1 ml of IPTG-induced JM101/ipBS V_{HH} 21 were used. Figure 8 shows: C,D:fractions from fast S Sepharose column chromatography (C:Eluted at 650 mM NaCl D:Eluted at 700 mM NaCl) E,F:fractions from Superdex 75 column chromatography.

As can be seen, the major impurity is eliminated by ionexchange chromatography and the bulk of the remaining impurities are eliminated by gel filtration.

SUBSTITUTE SHEET

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Table 1
Comparison of the N terminal Camel CH₂ and CH₃ sequences with the translated cDNA sequences of Camel immunoglobulins and with the corresponding human γ sequences. (Numbering according to Kabat et al (1987)(7).

SUBSTITUTE SHEET

Primer Derived	10	20	30	
	G G S V Q T G G S L R L S C E I S G L T F D	# 72/4		
	G G S V Q T G G S L R L S C A V S G F S F S	# 72/3		
	G G S E Q G G G S L R L S C A I S G Y T Y G	# 72/7		
	G G S V Q P G G S L T L S C T V S G A T Y S	# 72/17		
	G G S V Q A G G S L R L S C T G S G F P Y S	# 72/18		
	D V Q L V A S G G G S V G A G G S L R L S C T A S G D S F S	# 72/2		
	E V K L V E S G G G L V E P G G S L R L S C A T S G F T F S	Mouse V _H III _A		
	E V Q L L S G G G L V Q P G G S L R L S C A A S G F T F S	Human V _H III		

Table 2 :
A comparison of N Terminal Fr 1 regions of Camel V_{HH} with a Human V_H III subgroup protein and a mouse V_H IIIA subgroup protein.
The invariable subgroup specific residues are grayed.

	FrameWork 4	J Genes
Human	W G Q G T L V T V S S	J1, J4, J5
	W G R G T L V T V S S	J2
	W G Q G T T V T V S S	J6
	W G Q G T M V T V S S	J3
Murine	W G Q G T T L T V S S	J1
	W G Q G T L V T V S S	J2
	W G Q G T S V T V S A	J3
	W G A G T T V T V S S	J4
		cDNA Clones
Camel	W G Q G T Q V T V S S	Clones
	W G Q G T Q V T V S S	# 72/19 = # 72/3
	W G Q G T L V T V S S	1 Clone
	W G R G T Q V T V S S	# 72/24
	W G Q G T H V T V S S	# 72/21
	W G Q G I Q V T A S S	# 72/16

Table 3

Comparison of some Framework 4 residues found in the Camel V_HH region with the Framework 4 residues corresponding to the consensus region of the Human and Mouse J minigenes.

SUBSTITUTE SHEET

	D	Y	Y	G	S	S	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	D	V	W	G	A	G	T	T	V	T	V	S	S	MOUSE VH3 sequence			
95					100	a	b	c	d	e	f	g	h	i	j	k	-	-	-	-	-	101			105					110								
1	A	L	Q	P	G	G	Y	C	G	Y	G	X	-	-	-	-	-	-	-	-	-	C	L	W	G	Q	G	T	Q	V	T	V	S	S				
2	V	S	L	M	D	R	I	S	Q	H	-	-	-	-	-	-	-	-	-	-	-	G	C	R	G	Q	G	T	Q	V	T	V	S	L				
3	V	P	A	H	L	G	P	G	A	I	L	D	L	K	K	Y	-	-	-	-	-	-	K	Y	W	G	Q	G	T	Q	V	T	V	S	S			
4	F	C	Y	S	T	A	G	D	G	G	S	G	E	-	-	-	-	-	-	-	-	M	Y	W	G	Q	G	T	Q	V	T	V	S	S	Random Sample			
7	E	L	S	G	G	S	C	E	L	P	L	L	F	-	-	-	-	-	-	-	-	D	Y	W	G	Q	G	T	Q	V	T	V	S	S				
9	D	W	K	Y	W	T	C	G	A	Q	T	G	G	Y	F	-	-	-	-	-	-	G	Q	W	G	Q	G	A	Q	V	T	V	S	S				
11	R	L	T	E	M	G	A	C	D	A	R	W	A	T	L	A	T	R	T	F	A	Y	N	Y	W	G	Q	G	T	Q	V	T	V	S	S	18 different camel VHH region		
13	Q	K	K	D	R	T	R	W	A	E	P	R	E	W	-	-	-	-	-	-	-	N	N	W	G	Q	G	T	Q	V	T	A	S	S				
16	G	S	R	F	S	S	P	V	G	S	T	S	R	L	E	S	-	S	D	Y	-	-	N	Y	W	G	Q	G	I	Q	V	T	A	S	S			
17	A	D	P	S	I	Y	Y	S	I	L	X	I	E	Y	-	-	-	-	-	-	-	K	Y	W	G	Q	G	T	Q	V	T	V	S	S				
18	D	S	P	C	Y	M	P	T	M	P	A	P	P	I	R	D	S	F	G	W	-	-	D	D	F	G	Q	G	T	Q	V	T	V	S	S			
19	T	S	S	F	Y	W	Y	C	T	T	A	P	Y	-	-	-	-	-	-	-	-	N	V	W	G	Q	G	T	Q	V	T	V	S	S				
20	T	E	I	E	W	Y	G	C	N	L	R	T	T	F	-	-	-	-	-	-	-	T	R	W	G	Q	G	T	Q	V	T	V	S	S				
21	N	Q	L	A	G	G	W	Y	L	D	P	N	Y	W	L	S	V	G	A	Y	-	-	A	I	W	G	Q	G	T	H	V	T	V	S	S			
24	R	L	T	E	M	G	A	C	D	A	R	W	A	T	L	A	T	R	T	F	A	Y	N	Y	W	G	R	G	T	Q	V	T	V	S	S			
25	D	G	W	T	R	K	E	G	G	I	G	L	P	W	S	V	Q	C	E	D	G	Y	N	Y	W	G	Q	G	T	Q	V	T	V	S	S			
27	D	S	Y	P	C	H	L	L	-	-	-	-	-	-	-	-	-	-	-	-	-	D	V	W	G	Q	G	T	Q	V	T	V	S	S				
29	V	E	Y	P	I	A	D	M	C	S	-	-	-	-	-	-	-	-	-	-	-	R	Y	G	D	P	G	T	Q	V	T	V	S	S				
CDR3																																						

Human and mouse - size range 0-19 aa over 600 entries
Camel 8-24 aa 18 entries

TABLE 4

	human C _H 1	hinge	C _H 2
human gamma 3	KVDKRV	ELKTPLGDTTHTCPRCP <u>EPKCSDTPPPCPRCP</u> <u>EPKSCDTPPPCPRCP</u>	APELLGG PSVFLFP
human gamma 1	KVDRK	— AEPKSCDKTHTCPPCP	APELLGG PSVFLFP
human gamma 2	KVKVTV	— ERKCCVECPPCP	APPVAG - PSVFLFP
human gamma 4	KVDKRV	— ESKYGPPCPSCP	APEFLGG PSVFLFP

TABLE 5(2)

SUBSTITUTE SHEET

SUBSTITUTE SHEET

10	20	40
EVQLVESGGG	LVQPGGSLRL SCAASG	CDR1 WVRQA PGKGLEWVS CDR2
GG SVQGGGSLRL	SCAISG	CDR1 WFREG PGKEREGIA CDR2
GG SVQAGGSLRL	SCASSS	CDR1 WYRQA PGKEREFSVS CDR2

70	80	90	110
RFTIS RDNSKNTLYL	QMNSLRAEDTAVY YCAR	CDR3	WGQGTQVT VSS
RFTIS QDSTLKTHYL	LMNNLKPEDTGT YCAA	CDR3	WGQGTQVT VSS
RFTIS QDSAKNTVYL	QMNSLKPEDTAMY YCKI	CDR3	WGQGTQVT VSS

	camel V_{HH}	hinge	C_H2
camel	WGQGTQVT VSS	GTNEVCKCPKCP	APELPGG PSVFVFP
	WGQGTQVT VSS	EPKIPQPQPKPQPQP	
		QPQPKPQP	
		KPEPECTCPKCP	APELLGG PSVFIFP

TABLE 5(1)

WO 94/04678

55

PCT/EP93/02214

REFERENCES

1. Ward, E.S., Güssow, D., Griffiths, A.D., Jones, P.T. and Winter G. *Nature* 341, 544-546 (1989).
2. Ungar-Waron H., Eliase E., Gluckman A. and Trainin Z. *Isr. J. Vet. Med.*, 43, 198-203 (1987).
3. Bajyana Songa E. and Hamers R., *Ann. Soc. Belge Méd. trop.*, 68, 233-240 (1988).
4. Edelman G.M., Olins D.E., Gally J.A. and Zinder N.D., *Proc. Nat. Acad. Sc.*, 50, 753 (1963).
5. Franek F. and Nezlin R.S., *Biokhimiya*, 28, 193 (1963).
6. Roitt I.M., Brostoff J. and Male D.K., *Immunology*, Gower Med. Pub. London. New-York, p.9.2. (1985).
7. Schiffer M., Girling R.L., Ely K.R. and Edmundson B., *Biochemistry*, 12, 4620-4631 (1973).
8. Fleischman J.B., Pain R.H. and Porter R.R., *Arch. Biochem. Biophys*, Suppl. 1, 174 (1962).
9. Roholt O., Onoue K. and Pressman D., *PNAS* 51, 173-178 (1964).
10. Seligmann M., Mihaesco E., Preud'homme J.L., Danon F. and Brouet J.C., *Immunological Rev.*, 48, 145-167 (1979).
11. Hendershot L., Bole D., Köhler G. and Kearney J.F., *The Journal of Cell Biology*, 104, 761-767 (1987).
12. Hendershot L.M., *The Journal of Cell Biology*, 111, 829-837 (1990).

13. Hamers-Casterman, C., E. Wittouck, W. Van der Loo and R. Hamers, *Journal of Immunogenetics*, 6, 373-381 (1979).
14. Applied Biosystems - Ethanol Precipitation of Electro Eluted Electrodialysed Sample. Issue n° 27.
15. Maniatis, T. E.F. Fritsch and J. Sambrook, *Molecular Cloning. A Laboratory Manual* (1988).
16. Sastry et al., *PNAS*, 86, 5728, (1989).
17. Sanger, F., S. Nicklen and A.R. Coulson, *Proc. Natl. Acad. Sci., U.S.A.*, 74, 5463-5467 (1977).
18. Kabat E.A., Tai Te Wu, M. Reid-Miller, H.M. Perry and K.S. Gottesman, *U.S. Dpt of Health and Human Services, Public Health Service, National Institutes of Health* (1987).
19. Valentine, R.C. and N.M. Geen, *J.M.B.*, 27, 615-617 (1967).

C L A I M S

1. Immunoglobulin characterized in that it comprises two heavy polypeptide chains sufficient for the formation of a complete antigen binding site or several antigen binding sites, this immunoglobulin being further devoid of light polypeptide chains.
2. Immunoglobulin according to claim 1, characterized in that it comprises two heavy polypeptide chains sufficient for the formation of a complete antigen binding site or several antigen binding sites, this immunoglobulin being further devoid of light polypeptide chains and further characterized by the fact that it is the product of the expression in a prokaryotic or in a eukaryotic host cell, of a DNA or of a cDNA having the sequence of an immunoglobulin devoid of light chains as obtainable from lymphocytes or other cells of Camelids.
3. Immunoglobulin according to claim 1 or claim 2, characterized in that the amino acid sequence of its variable region contains in position 45 an amino acid which is different from a leucine, or proline or glutamine residue.
4. Immunoglobulin according to anyone of claims 1 to 3, characterized in that its heavy polypeptide chains are devoid of a so-called first domain in their constant region (C_H1).
5. Immunoglobulin according to anyone of claims 1 to 4, characterized in that it comprises an antigen binding site or several antigen binding sites and especially in that each variable region of each heavy chain contains at least one antigen binding site.
6. Immunoglobulin according to anyone of claims 1 to 5, characterized in that it is a type G immunoglobulin of class 2 (IgG2) or it is a type G immunoglobulin of class 3 (IgG3).

7. Immunoglobulin according to anyone of claims 1 to 6, characterized in that it is a Camelid immunoglobulin.

8. Immunoglobulin according to anyone of claims 1 to 7 as obtainable by purification from the serum of Camelids, characterized in that :

- it is not adsorbed by chromatography on Protein G Sepharose column,
- it is adsorbed by chromatography on Protein A Sepharose column,
- it has a molecular weight of around 100 Kd after elution with a pH 4.5 buffer (0.15 M NaCl, 0.58% acetic acid adjusted to pH 4.5 by NaOH),
- it consists of heavy $\gamma 2$ polypeptide chains of a molecular weight of around 45 Kd preferably 46 Kd after reduction.

9. Immunoglobulin according to anyone of claims 1 to 7, as obtainable by purification from the serum of Camelids is characterized in that the immunoglobulin :

- is adsorbed by chromatography on a Protein A Sepharose column,
- has a molecular weight of around 100 kd after elution with a pH 3.5 buffer (0.15 M NaCl, 0.58% acetic acid),
- is adsorbed by chromatography on a Protein G Sepharose column and eluted with pH 3.5 buffer (0.15 M NaCl, 0.58% acetic acid).
- consists of heavy $\gamma 3$ polypeptide chains of a molecular weight of around 45 Kd in particular between 43 and 47 kd after reduction.

10. Immunoglobulin according to anyone of claims 1 to 9, characterized in that

- it comprises 4 frameworks in its variable region, which frameworks comprise an amino-acid sequence selected from the following sequences :

for the framework 1 domain

60

G G S V Q T G G S L R L S C E I S G L T F D
G G S V Q T G G S L R L S C A V S G F S F S
G G S E Q G G G S L R L S C A I S G Y T Y G
G G S V Q P G G S L T L S C T V S G A T Y S
G G S V Q A G G S L R L S C T G S G F P Y S
G G S V Q A G G S L R L S C V A G F G T S
G G S V Q A G G S L R L S C V S F S P S S

for the framework 4 domain

W G Q G T Q V T V S S
W G Q G T L V T V S S
W G Q G A Q V T V S S
W G Q G T Q V T A S S
R G Q G T Q V T V S L

and/or,

for the CDR3 domain

A L Q P G G Y C G Y G X - - - - - C L
V S L M D R I S Q H - - - - - G C
V P A H L G P G A I L D L K K Y - - - - - K Y
F C Y S T A G D G G S G E - - - - - M Y
E L S G G S C E L P L L F - - - - - D Y
D W K Y W T C G A Q T G G Y F - - - - - G Q
R L T E M G A C D A R W A T L A T R T F A Y N Y
Q K K D R T R W A E P R E W - - - - - N N
G S R F S S P V G S T S R L E S - S D Y - - N Y
A D P S I Y Y S I L X I E Y - - - - - K Y
D S P C Y M P T M P A P P I R D S F G W - - D D
T S S F Y W Y C T T A P Y - - - - - N V
T E I E W Y G C N L R T T F - - - - - T R
N Q L A G G W Y L D P N Y W L S V G A Y - - A I
R L T E M G A C D A R W A T L A T R T F A Y N Y
D G W T R K E G G I G L P W S V Q C E D G Y N Y
D S Y P C H L L - - - - - D V
V E Y P I A D M C S - - - - - R Y

and/or,

61

- in that its constant region comprises C_H2 and C_H3 domains comprising an amino-acid sequence selected from the following sequences :

for the C_H2 domain:

APELLGGPTVFIFPPKPKDVLSITLTP
 APELPGGPSVFVFPTKPKDVLSISGRP
 APELPGGPSVFVFPPKPKDVLSISGRP
 APELLGGPSVFIFPPKPKDVLSISGRP

for the C_H3 domain:

GQTREPQVYTLA
 GQTREPQVYTLAPXRLEL
 GQPREPQVYTLPPSRDEL
 GQPREPQVYTLPPSREEM
 GQPREPQVYTLPPSQEEM

and/or,

- in that its hinge region comprises from 0 to 50 amino-acids, especially in that its hinge region comprises an amino-acid sequence selected from the following sequences :

GTNEVCKCPKCP

or,

EPKIPQPKPQPKPQPKPQPKPQPKPEPECTCPKCP

11. Immunoglobulin according to anyone of claims 1 to 10, characterized in that it is encoded by a sequence selected among those represented on figure 7.

12. Fragment of an immunoglobulin according to anyone of claims 1 to 11, characterized in that it is selected from the following group :

- a fragment corresponding to one heavy polypeptide chain of an immunoglobulin devoid of light chains,
- fragments obtained by enzymatic digestion of the immunoglobulins of the invention, especially those obtained by partial digestion with papain leading to the Fc fragment (constant fragment) and leading to FV_Hh fragment (containing the

antigen binding sites of the heavy chains) or its dimer $F(V_{HH}h)_2$, or a fragment obtained by further digestion with papain of the Fc fragment, leading to the Fc' fragment corresponding to the C-terminal part of the Fc fragment,

- homologous fragments obtained with other proteolytic enzymes,
 - a fragment of at least 10 preferably 20 amino acids of the variable region of the immunoglobulin, or the complete variable region, especially a fragment corresponding to the isolated V_{HH} domains or to the V_{HH} dimers linked to the hinge disulphide,
 - a fragment corresponding to the hinge region of the immunoglobulin, or to at least 6 amino acids of this hinge region,
 - a fragment of the hinge region comprising a repeated sequence of Pro-X,
 - a fragment corresponding to at least 10 preferably 20 amino acids of the constant region or to the complete constant region of the immunoglobulin.
13. Immunoglobulin according to anyone of claims 1 to 12, characterized in that all or a part of its constant region is replaced by all or part of the constant region of a human antibody.
14. Immunoglobulin according to any one of claims 1 to 13, obtainable in prokaryotic cells, especially in E.coli cells by a process comprising the steps of :
- a) cloning in a Bluecript vector of a DNA or cDNA sequence coding for the V_H domain of an immunoglobulin devoid of light chain obtainable for instance from lymphocytes of Camelids,
 - b) recovering the cloned fragment after amplification using a 5' primer containing an Xho site and a 3' primer containing the Spe site having the following sequence

TC TTA ACT AGT GAG GAG ACG GTG ACC TG,

- c) cloning the recovered fragment in phase in the immuno PBS vector after digestion of the vector with Xho and Spe restriction enzymes,
- d) transforming host cells, especially E.coli by transfection with the recombinant immuno PBS vector of step c,
- e) recovering the expression product of the V_{HH} coding sequence, for instance by using antibodies raised against the dromadary V_{HH} domain.

15. Hetero-specific immunoglobulins according to any one of claims 1 to 13 obtainable by a process comprising the steps of:

- obtaining a first DNA or cDNA sequence coding for a V_{HH} domain or part thereof having a determined specificity against a given antigen and comprised between Xho and Spe sites,
- obtaining a second DNA or cDNA sequence coding for a V_{HH} domain or part thereof, having a determined specificity different from the specificity of the first DNA or cDNA sequence and comprised between the Spe and EcoRI sites,
- digesting an immuno PBS vector with EcoRI and XhoI restriction enzymes,
- ligating the obtained DNA or cDNA sequences coding for V_{HH} domains, so that the DNA or cDNA sequences are serially cloned in the vector,
- transforming a host cell, especially E.coli cell by transfection, and recovering the obtained immunoglobulins.

16. Immunoglobulin according to any one of claims 1 to 13, obtainable by a process comprising the steps of:

- obtaining a DNA or cDNA sequence coding for a V_{HH} domain or part thereof, having a determined specific antigen binding site,

- amplifying the obtained DNA or cDNA, using a 5' primer containing an initiation codon and a HindIII site, and a 3' primer containing a termination codon having a XhoI site,
 - recombining the amplified DNA or cDNA into the HindIII (position 2650) and XhoI (position 4067) sites of a plasmid pMM984,
 - transfecting permissive cells especially NB-E cells with the recombinant plasmid,
 - controlling the expression, for instance by an ELISA assay with antibodies directed against a region of a V_{HH} domain and recovering the obtained products.
17. Immunoglobulins according to claim 16, obtainable by a process comprising the further cloning of a second DNA or cDNA sequence having another determined antigen binding site, in the pMM984 plasmid.
18. Immunoglobulin according to any one of claims 11 to 17, characterized in that it is obtainable by a process wherein the vector is Yep 52 and the transformed recombinant cell is a yeast especially S.cerevisiae.
19. Immunoglobulin according to any one of claims 11 to 17, characterized in that it is obtainable by a process wherein the vector is a vector appropriate for expression in plant cells, for example pMon530, and the transformed recombinant cells are plant cells.
20. Immunoglobulin according to any one of claims 16 to 17, characterized in that it has a catalytic activity, especially in that it is directed against an antigen mimicking an activated state of a given substrate, these immunoglobulins having for instance been modified at the level of their catalytic site by random or directed mutagenesis.
21. Nucleotide sequence, characterized in that it codes for all or part of an immunoglobulin according to anyone of claims 1 to 20, which immunoglobulins

65

comprise a peptide sequence selected from the following:

VTVSSGTNEVCKCPKCPAPELPGGPSVVFVFP,
 VTVSSEPKIPQPKPQPKPQPKPQPKPEPECTCPKCPAPELLGGPSVFIFP
 GTNEVCKCPKCP
 APELPGGPSVVFVFP
 EPKIPQPKPQPKPQPKPQPKPEPECTCPKCP
 APELLGGPSVFIFP
 APELLGGPTVFIFPPKPKDVLSITLTP
 APELPGGPSVVFVFTKPKDVLSISGRP
 APELPGGPSVVFVFPKPKDVLSISGRP
 APELLGGPSVFIFPPKPKDVLSISGRP
 GQTREPQVYTLA
 GQTREPQVYTLAPXRLEL
 GQPREPQVYTLPPSRDEL
 GQPREPQVYTLPPSREEM
 GQPREPQVYTLPPSQEEM
 G G S V Q T G G S L R L S C E I S G L T F D
 G G S V Q T G G S L R L S C A V S G F S F S
 G G S E Q G G G S L R L S C A I S G Y T Y G
 G G S V Q P G G S L T L S C T V S G A T Y S
 G G S V Q A G G S L R L S C T G S G F P Y S
 G G S V Q A G G S L R L S C V A G F G T S
 G G S V Q A G G S L R L S C V S F S P S S
 W G Q G T Q V T V S S
 W G Q G T L V T V S S
 W G Q G A Q V T V S S
 W G Q G T Q V T A S S
 R G Q G T Q V T V S L
 and/or,

A L Q P G G Y C G Y G X - - - - - C L
 V S L M D R I S Q H - - - - - G C
 V P A H L G P G A I L D L K K Y - - - - - K Y
 F C Y S T A G D G G S G E - - - - - M Y
 E L S G G S C E L P L L F - - - - - D Y
 D W K Y W T C G A Q T G G Y F - - - - - G Q

66

```

R L T E M G A C D A R W A T L A T R T F A Y N Y
Q K K D R T R W A E P R E W - - - - - N N
G S R F S S P V G S T S R L E S - S D Y - - N Y
A D P S I Y Y S I L X I E Y - - - - - K Y
D S P C Y M P T M P A P P I R D S F G W - - D D
T S S F Y W Y C T T A P Y - - - - - N V
T E I E W Y G C N L R T T F - - - - - T R
N Q L A G G W Y L D P N Y W L S V G A Y - - A I
R L T E M G A C D A R W A T L A T R T F A Y N Y
D G W T R K E G G I G L P W S V Q C E D G Y N Y
D S Y P C H L L - - - - - D V
V E Y P I A D M C S - - - - - R Y

```

22. Nucleotide sequence characterized in that it codes for an immunoglobulin according to anyone of claims 1 to 20, in that it comprises a sequence selected from those represented on figure 7.

23. Process for the preparation of a monoclonal antibody according to anyone of claims 1 to 20, directed against a determined antigen, the antigen binding site of the antibody consisting of heavy polypeptide chains and which antibody is further devoid of light polypeptide chains, which process comprises :

- immortalizing lymphocytes, obtained for example from the peripheral blood of Camelids previously immunized with a determined antigen, with an immortal cell and preferably with myeloma cells, in order to form a hybridoma,
- culturing the immortalized cells formed and recovering the cells producing the antibodies having the desired specificity.

24. Process for the preparation of antibodies directed against determined antigens, comprising the steps of :

- cloning into vectors, especially into phages and more particularly filamentous bacteriophages, DNA

or cDNA sequence obtained from lymphocytes of Camelids previously immunized with determined antigens, capable of producing an immunoglobulin according to anyone of claims 1 to 20,

- transforming prokariotic cells with the above vectors in conditions allowing the production of the antibodies,
- selecting the appropriate antibody by subjecting the transforming cells to antigen-affinity selection,
- recovering the antibodies having the desired specificity.

25. Process according to claim 24, wherein the cloning vector is a plasmid or a eukaryotic virus and the transformed cell is a eukaryotic cell, especially a yeast cell, mammalian cell, plant cell or protozoair cell.

26. Process according to claim 24, wherein the cloning vector is a plasmid capable of expressing the immunoglobulin in the bacterial membrane.

27. Process according to claim 24, wherein the cloning vector is a plasmid capable of expressing the immunoglobulin as a secreted protein.

28. Immunoglobulin according to anyone of claims 1 to 20, characterized in that it is directed against an antigen such as one of a bacteria, a virus, a parasite, or against a protein, hapten, carbohydrate or nucleic acid.

29. Immunoglobulin according to anyone of claims 1 to 20 characterized in that it is directed against an immunoglobulin idiotype.

30. Immunoglobulin according to anyone of claims 1 to 20 characterized in that it is directed against a cellular receptor or membrane protein.

31. Immunoglobulin according to anyone of claims 1 to 20, characterized in that it has a catalytic activity.

32. Immunoglobulin according to anyone of claims 1 to 20, or a fragment according to claim 12, characterized in that it is conjugated with a toxin.

33. Use for a fragment comprising a repeated sequence Pro-X, X being any amino-acid and preferably Gln, Lys or Glu, the sequence containing advantageously at least 3 repeats of Pro-X and especially a fragment composed of a 12-fold repeat of the sequence Pro-X, for coupling protein domains or a protein and a ligand.

34. Use of the hinge region or of a fragment of the hinge region of an immunoglobulin according to anyone of claims 1 to 20, for coupling protein domains or a protein and a ligand.

35. Immunoglobulin according to anyone of claims 1 to 20, characterized in that it is a heterospecific antibody.

36. Recombinant vector characterized in that it comprises a nucleotide sequence according to claim 21 or claim 22, and in that it is a plasmid, a phage especially a bacteriophage, a virus, a YAC, a cosmid.

37. Recombinant cell or organism characterized in that it is modified by a vector according to claim 36.

38. A cDNA library composed of nucleotide sequences coding for a heavy-chain immunoglobulin according to anyone of claims 1 to 20, such as obtained by performing the following steps:

a) treating a sample containing lymphoid cells, especially peripheral lymphocytes, spleen cells, lymph nodes or another lymphoid tissue from a healthy animal, especially selected among the Camelids, in order to separated the B-lymphocytes,

b) separating polyadenylated RNA from the other nucleic acids and components of the cells,

c) reacting the obtained RNA with a reverse transcriptase in order to obtain the corresponding cDNA,

d) contacting the obtained cDNA with 5' primers corresponding to mouse V_H domain of four-chain immunoglobulins, which primer contains a determined restriction site, for example an XhoI site and with 3' primers corresponding to the N-terminal part of a C_H2 domain,

e) amplifying the DNA,

f) cloning the amplified sequence in a vector, especially in a bluescript vector,

g) recovering the clones hybridizing with a probe corresponding to the sequence coding for a constant domain from an isolated heavy-chain immunoglobulin.

39. A modified 4-chain immunoglobulin or a fragment thereof, the V_H regions of which has been partially replaced by specific sequences or amino acids of heavy chain immunoglobulins according to anyone of claims 1 to 20.

40. A modified 4-chain immunoglobulin or a fragment thereof, according to claim 39, wherein the leucine, proline or glutamine in position 45 of the V_H regions has been replaced by other amino acids and preferably by arginine, glutamic acid or cysteine.

41. A modified 4-chain immunoglobulin or a fragment thereof, in which the CDR loops of the region are linked to other parts of the V region by the introduction of paired cysteines, in particular in which the CDR_3 loop is linked to the FW_2 or CDR_1 , and more especially where the cysteine of the CDR_3 of the V_H is linked to a cysteine in position 31 or 33 of FW_2 or in position 45 of CDR_2 .

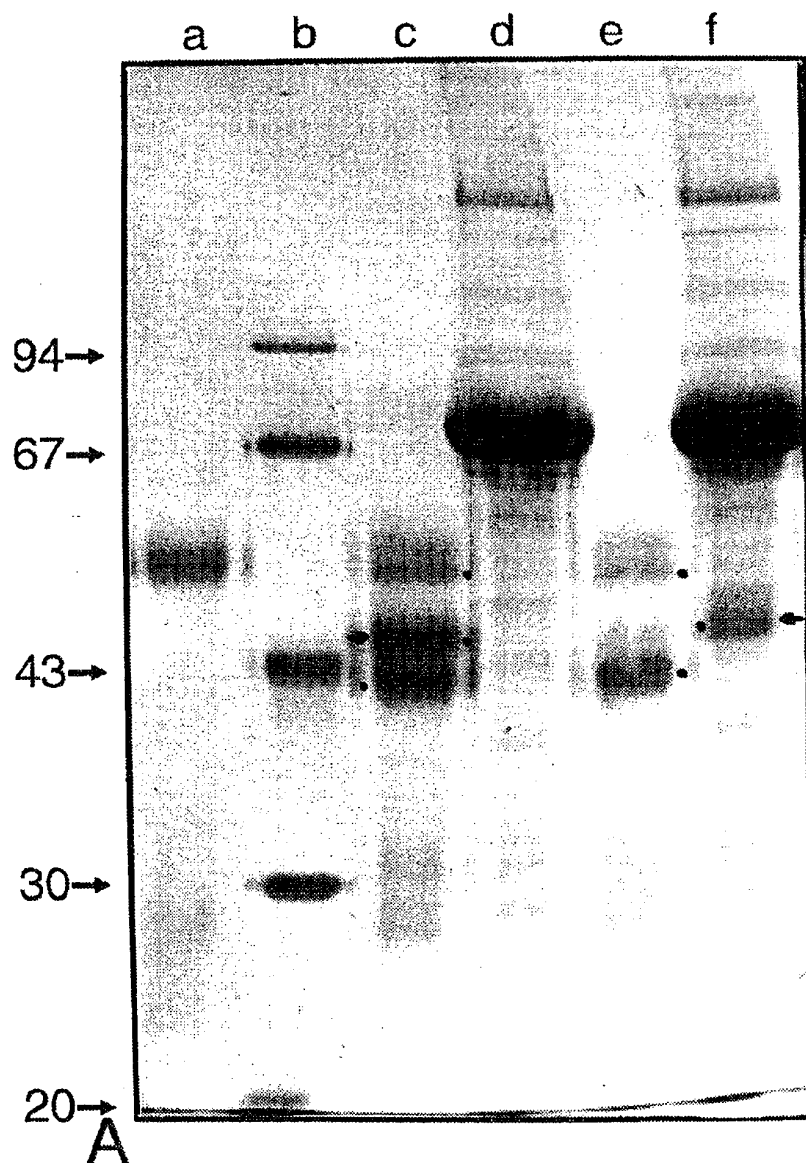


FIGURE 1A

SUBSTITUTE SHEET

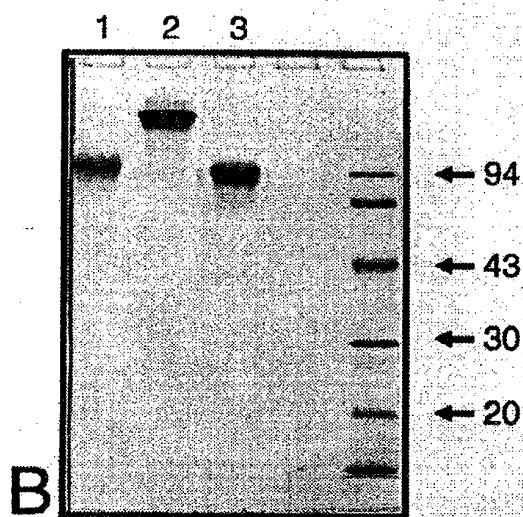


FIGURE 1B

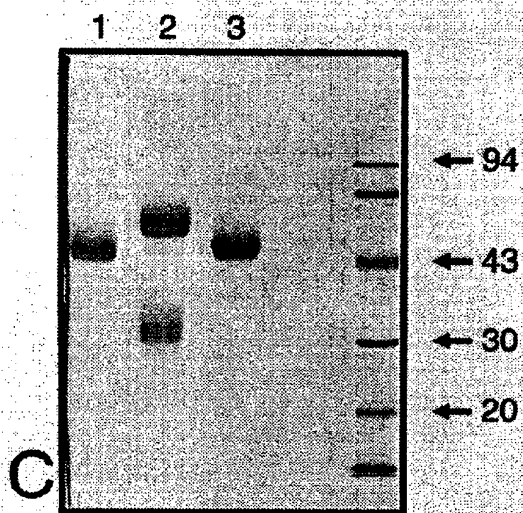


FIGURE 1C

SUBSTITUTE SHEET

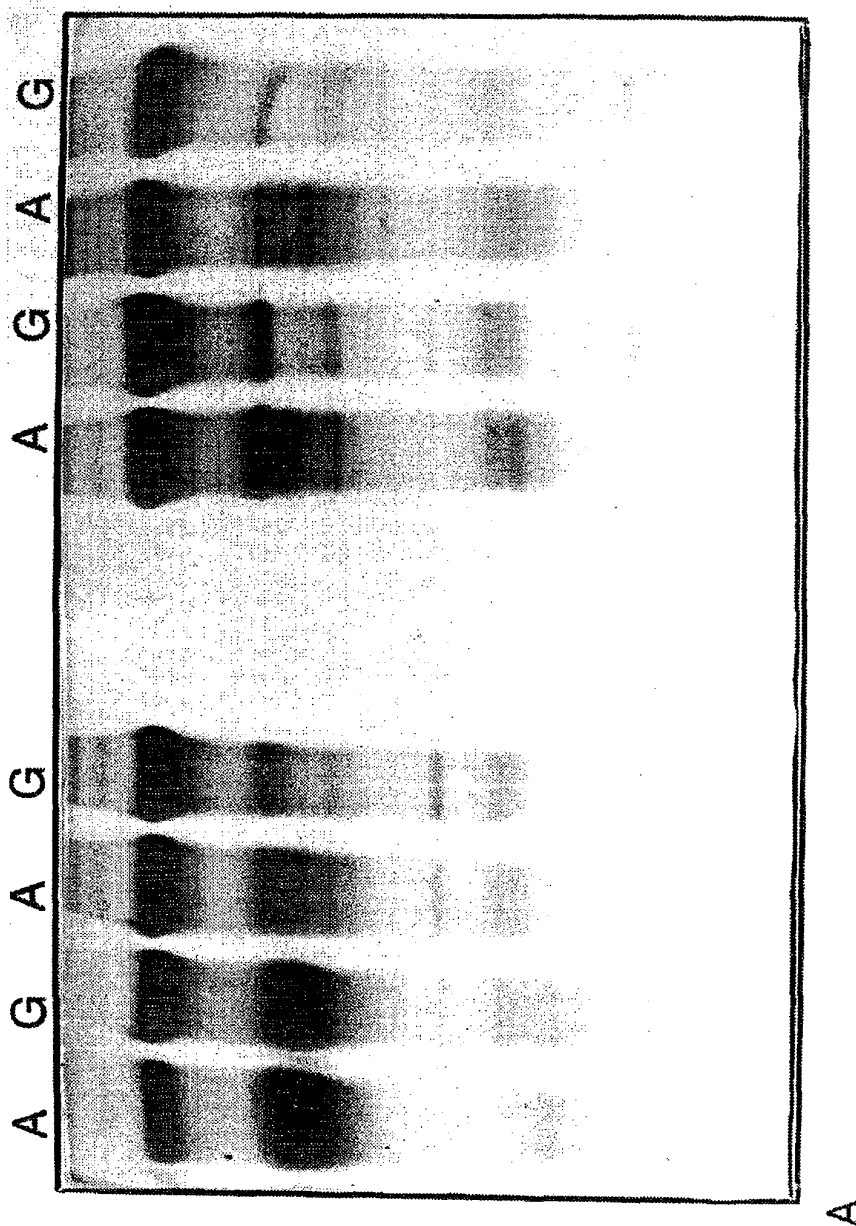


FIGURE 2A

SUBSTITUTE SHEET

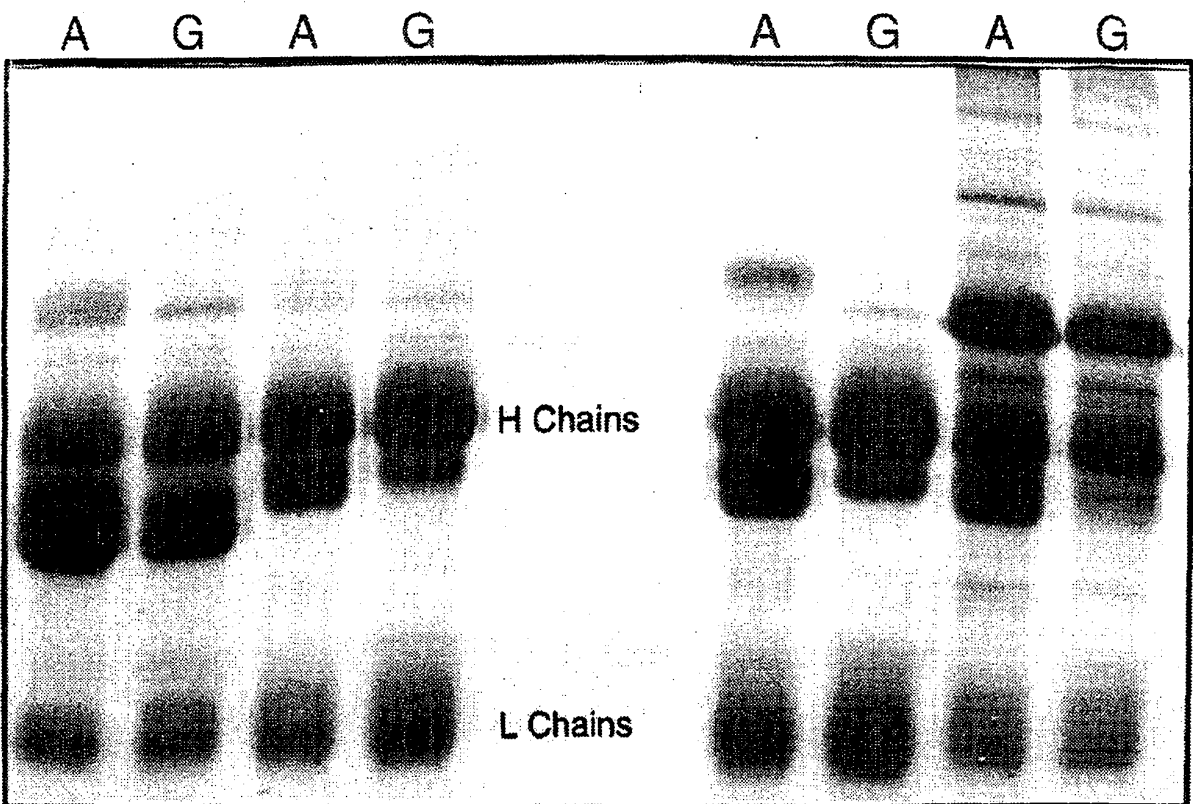
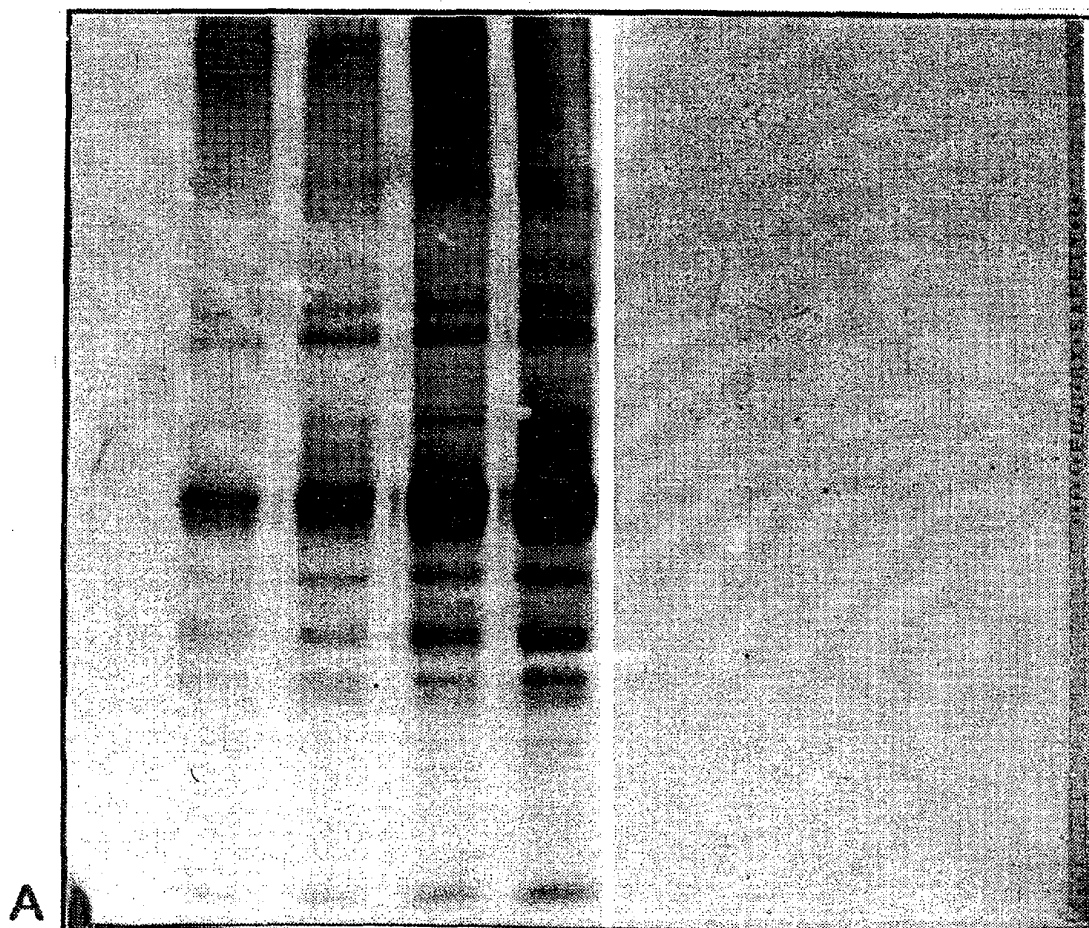


FIGURE 2B

B

SUBSTITUTE SHEET



	Prot. A	Ig1	Ig2	Ig3	Tot.Ser	Ig1	Ig2	Ig3	Tot.Ser
	Control	T. evansi infected				Healthy			
Counts/5ul	65	1258	1214	2700	2978	147	157	160	107

FIGURE 3A

SUBSTITUTE SHEET

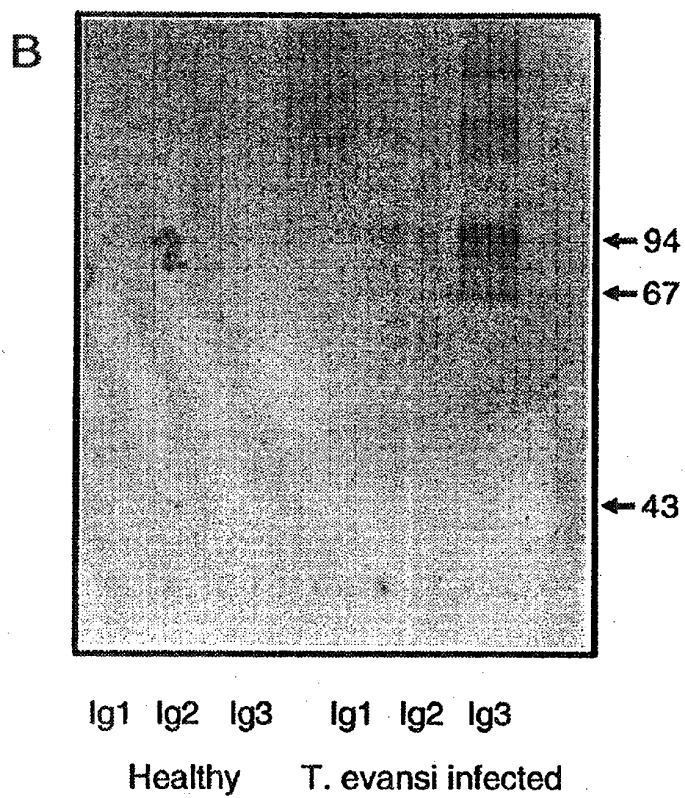


FIGURE 3B

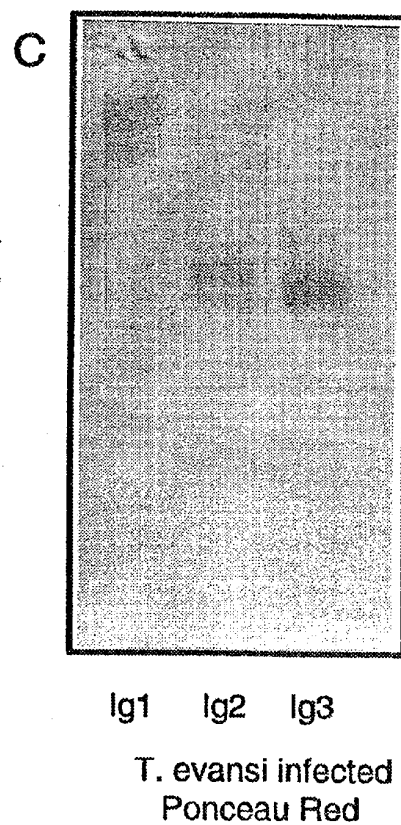


FIGURE 3C

SUBSTITUTE SHEET

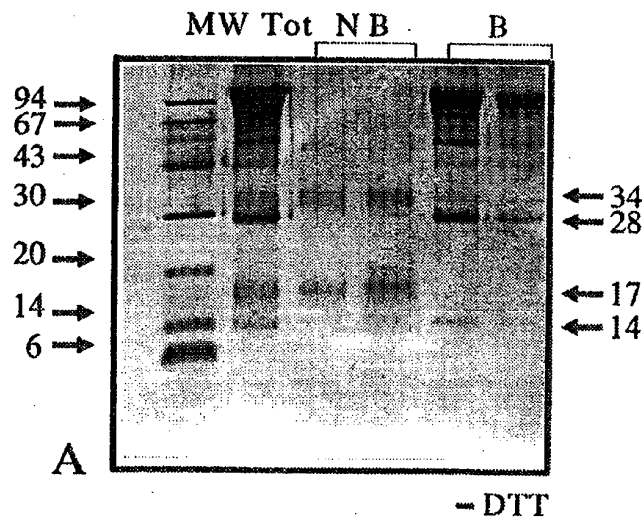


FIGURE 4A

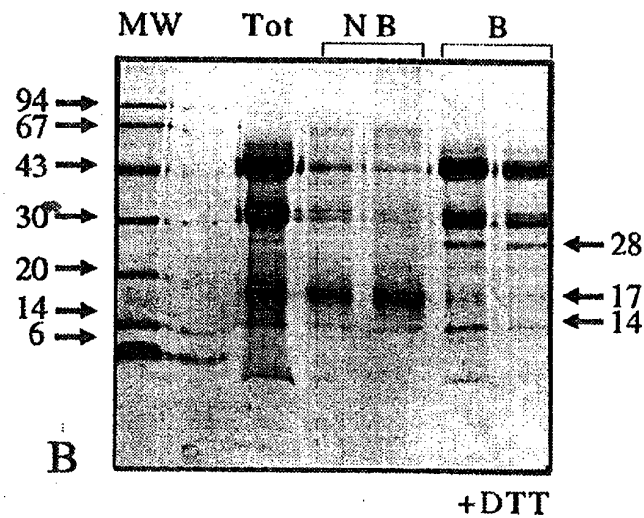


FIGURE 4B

Analysis of IgG₃ Papain Fragments by SDS - PAGE

SUBSTITUTE SHEET

Model for Camel IgG₃

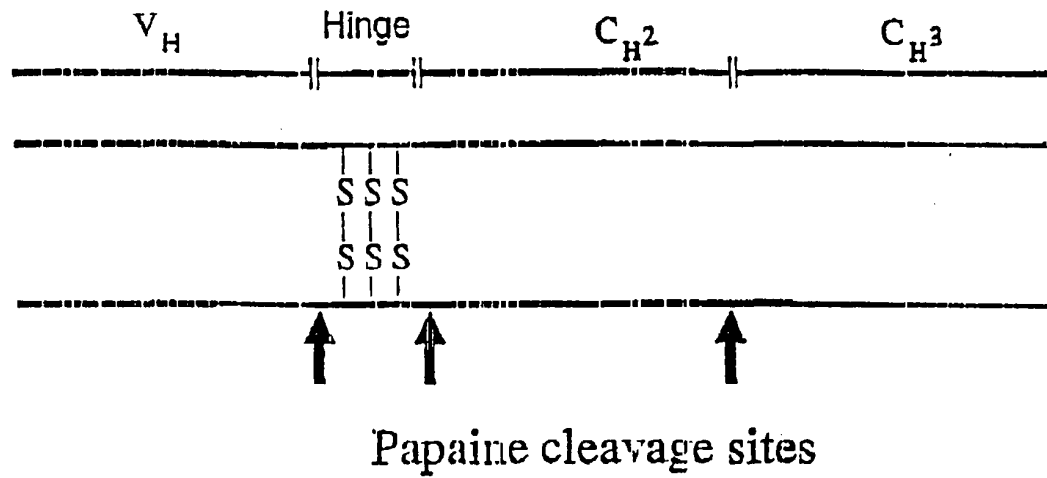


Fig 5: Schematic representation of Camel IgG₃ model.

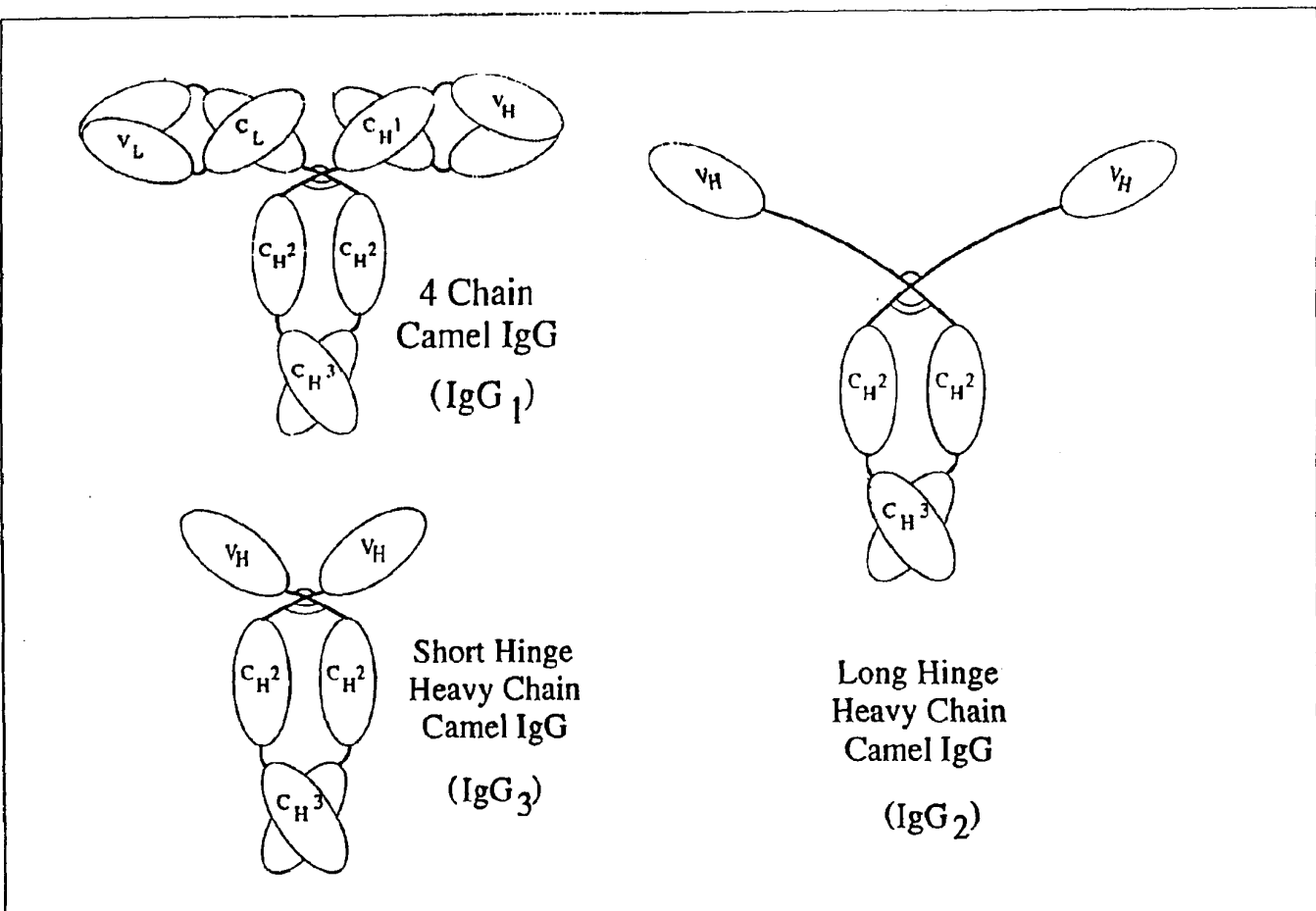


Fig 6: A schematic representation of Camel immunoglobulins IgG₁, the putative IgG₂ and IgG₃. The large (Pro-X)₁₂ of the putative IgG₂ molecule can be modelled into a 6 aa repeat

SUBSTITUTE SHEET

10 / 14

```

DR01006 C-----TCGAG---TCTGGGGGAGG
DR27006 C-----TCGAG---TCTGGGGGAGG
DR03006 C-----AGGTGA-----AACTGCTCGAG---TCTGGAGGAGG
DR11006 C-----TCGAG---TCTGGGGGAGG
DR24006 C-----AGGTGA-----AACTGCTCGAG---TCTGGGGGAGG
DR16006 C-----TCGAG---TCTGGAGGAGG
DR19006 C-----TCGAG---TCTGGAGGAGG
DR07006 C-----TCGAG---TCTGGGGGAGG
DR16006 C-----TCGAG---TCTGGGGGAGG
DR20006 C-----TCGAG---TCAGGGGGAGG
DR25006 C-----TCGAG---TCTGGGGGAGG
DR20006 C-----TCGAG---TCTGGAGGAGG
DR21006 C-----TCGAG---TCTGGGGGAGG
DR09006 C-----AGGTGA-----AACTGCTCGAG---TCTGGGGGAGG
DR17006 C-----TCGAG---TCTGGGGGAGG
DR13006 C-----TCGAG---TCAGGGGGAGG
DR02006 CTCGAGTCAGGTGTCCGGTCTGATGTGCAGCTGGTGGCGTCTGGGGGAGG

```

```

DR01006 ATCGGTGCAGGCTGGAGGGTCTCTGAGACTCTC--GTGCG-CAGCCTCTG
DR27006 CTCGGTGCAGGCTGGAGGGTCTCTGAGACTCTCCTGTGCATCTTCTTA
DR03006 CTCGGTGCAGACTGGAGGATCTCTGAGACTCTCCTGTGCAGT--C-TCTG
DR11006 GTCGGTGCAGGCTGGAGGGTCTCTGAGACTCTCCTGTAAATGT--C-TCTG
DR24006 GTCGGTGCAGGCTGGAGGGTCTCTGAGACTCTCCTGTAAATGT--C-TCTG
DR16006 CTCGGCGCAGGCTGGAGGATCTCTGAGACTCTCCTGTGCAGC--CCACGG
DR19006 CTCGGTTCAGGCTGGAGGGTCCCTTAGACTCTCCTGTGCAGC--C-TCTG
DR07006 CTCGGTGCAGGGTGGAGGGTCTCTGAGACTCTCCTGTGCAA---TCTCTG
DR16006 CTCGGTGCAGGCTGGAGGGTCTCTGAGACTCTCCTGTACAG---GCTCTG
DR20006 CTCGGTACAGGTTGGAGGGTCTCTGAGACTCTCCTGTGTAG---CCTCTA
DR25006 CTCGGTACAACTGGAGGGTCTCTGAGACTCTCTTGCG---AAATCTCTG
DR20006 CTCGGTGCAGGCTGGAGGGTCTCTGAGACTCTCCTGTG---TAGCCTCTG
DR21006 CTCGGTGCAGGTTGGAGGGTCTCTGAAACTCTCCTGTAAAAT---CTCTG
DR09006 CTCGGTGCAGGCTGGGGGGTCTCTGACACTCTCTTGTTG---TATACAC--
DR17006 CTCGGTCCAACCTGGAGGATCTCTGACACTCTCCTGTACAGTT---TCTG
DR13006 CTCGGTGGAGGCTGGAGGGTCTCTGAGACTCTCCTGTACAG---CCTCTG
DR02006 CTCGGTGCAGGCTGGAGGCTCTCTGAGACTCTCCTGTACAG---CCTCTG

```

```

DR01006 GA--TACAGTAATT---GTCCCCTCACTTG-GAGCTGGTATCGCCAGTTT
DR27006 AA--TATATGCCTT---GCACCTACGACAT-GACCTGGTACCGCCAGGCT
DR03006 GA--TTCTCCTTTA---GTACCAGTTGTAT-GGCCTGGTTCCGCCAGGCT
DR11006 GC--TCTCCCAGTA---GTACTTATTGCCT-GGGCTGGTTCCGCCAGGCT
DR24006 GC--TCTCCCAGTA---GTACTTATTGCCT-GGGCTGGTTCCGCCAGGCT
DR16006 GA--TTCCGC-TCA---ATGTTACTACAT-CGCCTGGTTCCGTCAGGCT
DR19006 AC--TACACCATCA---CTGATTATTGCAT-GGCCTGGTTCCGCCAGGCT
DR07006 GA--TACACGTACG---GTAGCTTCTGTAT-GGGCTGGTTCCGCGAGGGT
DR16006 GA--TTCCCCTATA---GTACCTTCTGTCT-GGGGTGGTTCCGCCAGGCT
DR20006 CT--CACACCGACA---GTAGCACCTGTAT-AGGCTGGTTCCGCCAGGCT
DR25006 GA--TTGACTTTTG---ATGATTCTGACGT-GGGGTGGTACCGCCAGGCT
DR20006 GA--TTCAATTTTCG---AACTTCTCGTAT-GGCGTGGTACCGCCAGACT
DR21006 GAGGTACCCCAGATCGTGTTCCTAAATCTTTGGCCTGGTTCCGCCAGGCT

```

FIGURE 7(1)

SUBSTITUTE SHEET

11 / 14

DR09006 -----CAACGATACTGGGACCA-----TGGGATGGTTTCGCCAGGCT
DR17006 --GGGCCACCTACA---GTGACTACAGTATTGGA-TGGATCCGCCAGGCT
DR13006 G-----ATACGTAT-CCT---CTATGGCCTGGTTCCGCCAGGTT
DR02006 GAGA----CAGTTTCAGTAGATT--TGCCATGTCTTGGTTCCGCCAGGCT

DR01006 CCAGGAACGGAGCGCGAGTTTCGTCTCCAGTATGGATCCGGATGGAAATAC
DR27006 CCAGGCAAGGAGCGCGAATTTGTCTCAAGTATAAATATTGATGGTAAGAC
DR03006 TCAGGAAAGCAGCGTGAGGGGGTTCGACGCCATTAATAGTGGCGGTGGTAG
DR11006 CCAGGGAGGGAGCGTGAGGGGGTTCACAGCGATTAA-----CACTGATGG
DR24006 CCAGGGAAGGAGCGTGAGGGGGTTCACAGCGATTAA-----CACTGATGG
DR16006 CCTGGGAAGGGCGTGAGGGGGTTCGCAACAATTAATGGTGGTCG-----
DR19006 CCAGGGAAGGAGCGTGAATTGGTTCGACGCGATTCAAGTTGTCCGTAGTGA
DR07006 CCAGGCAAGGAACGTGAGGGGATCGCAACTATTCTTAATGGTGGTACTAA
DR16006 CCAGGGAAGGAGCGTGAGGGGGTTCGCGGGTATTAATAGTGCAGGAGGTAA
DR20006 CCAGGGAAGGAGCGCGAGGGGGTTCGCAAGTATATATTTTGGTGTGGTGG
DR25006 CCAGGGGATCAGTGCAAAATTGGTCTCAGGTATTCTGAGTGTGGTACT-C
DR20006 CCAGGAAATGTGTGTGAGTTGGTCTCAAGTATTTACAGTGTAGG-----
DR21006 CCAGAGAAGGAGCGCGAGGGGATCGCAGTTCCTTCGACTAAGGATGGTAA
DR09006 CCAGGGAAGAGGTGCGAAAGGGTTCGCGCATATTACGCCTGATGGTATGA-
DR17006 CCAGGGAAGGACCGTGAAGTAGTCGACCGCTAATACTGGTG-----
DR13006 CCAGGGCAGGAGCGCGAGGGGGTTCGCGTTTGTTCAAACGG-----
DR02006 CCAGGGAAGGAGTGC GAATTGGTCTCAAGCATTCAAAGTAATGGAAGGAC

DR01006 CAAGTACA-----CATACTCCGTGAAGGGCCGCTTCACC
DR27006 AACATACG-----CAGACTCCGTGAAGGGCCGATTTCACC
DR03006 GACATACTA-CAACACATATGTCGCCGAGTCCGTGAAGGGCCGATTTCGCC
DR11006 CAGTATCAT-ATACGCA-----GCCGACTCCGTGAAGGGCCGATTTCACC
DR24006 CAGTGTCAT-ATACGCA-----GCCGACTCCGTGAAGGGCCGATTTCACC
DR16006 -----CGA-CGTCACATACTACGCCGACTCCGTGACGGGGCCGATTTCACC
DR19006 TACT--CGC-C-TCACAGACTACGCCGACTCCGTGAAGGGACGATTTCACC
DR07006 -----CACATACTATGCCGACTCCGTGAAGGGCCGATTTCACC
DR16006 -----TACTTACTATGCCGACGCCGTGAAGGGCCGATTTCACC
DR20006 -----TACGAATTATCGCGACTCCGTGAAGGGCCGATTTCACC
DR25006 CATATACAAAGAGTGGAGACTATGCTGAGTCTGTGAGGGGGCCGGTTACC
DR20006 CA-AAACATACTACGTCGACC--GCA-----TGAAGGGCCGATTTCACC
DR21006 GA-----CATTCTATGCCGACTCCGTGAAGGGCCGATTTCACC
DR09006 -----CCTTCATTGATGAACCCGTGAAGGGCCGATTTCAGC
DR17006 -----CGACTAGTAAATTCTACGTCGACTTTGTGAAGGGCCGATTTCACC
DR13006 --CTGACAAT-AGTGCATTATATGGCGACTCCGTGAAGGGCCGATTTCACC
DR02006 AACTGA-----GGCCGATTCCGTGCAAGGGCCGATTTCACC

DR01006 ATGTCCCGAGGCAGCACCGAGTACACAGTATTTCTGCAAAATGGACAATCT
DR27006 ATCTCCCAAGACAGCGCCAAGAACACGGTGTATCTGCAGATGAACAGCCT
DR03006 ATCTCCCAAGACAACGCCAAGACCACGGTATATCTTGATATGAACAACCT
DR11006 ATCTCCCAAGACACCGCCAAGGAAACGGTACATCTCCAGATGAACAACCT
DR24006 ATCTCCCAAGACACCGCCAAGAAAACGGTATATCTCCAGATGAACAACCT
DR16006 ATCTCCCGAGACAGCCCCAAGAATACGGTGTATCTGCAGATGAACAGCCT
DR19006 ATCTCCCAAGGCAACACCAAGAACACAGTGAATCTGCAAAATGAACAGCCT
DR07006 ATCTCCCAAGACAGCACGTTTGAAGACGATGTATCTGCTAATGAACAACCT
DR16006 ATCTCCCAAGGGAATGCCAAGAATACGGTGTCTTGCAAAATGGATAACTT
DR20006 ATCTCCCAACTCAACGCCCAAGAACACAGTGTATCTGCAAAATGAACAGCCT
DR25006 ATCTCCAGAGACAACGCCAAGAACATGATATACCTTCAAAATGAACGACCT
DR20006 ATTTCTAGAGAGAATGCCAAGAATACATTGTATCTACAACCTGAGCGGCCT
DR21006 ATCTTCTTAGATAATGACAAGACCACTTTCTCCTTACAACCTGATCGACT
DR09006 ATCTCCCGAGACAACGCCCAAGAAACGTTGTCTTTGCGAATGAATAGTCT

FIGURE 7(2)

SUBSTITUTE SHEET

12 / 14

DR17006 ATTTCCCAAGACAACGCCAAGAATACGGTATATCTGCAAATGAGCTTCCT
 DR13006 ATCTCCCACGACAACGCCAAGAACACGCTGTATCTGCAAATGCGCAACCT
 DR02006 ATCTCCCGAGACAATTCCAGGAACACAGTGTATCTGCAAATGAACAGCCT

DR01006 GAAACCTGAGGACACGGCGATGTATTACTGTAAAAC-A---GCCCTAC--
 DR27006 GAAACCTGAGGACACGGCGATGTATTACTGTAAAAT-A---GA--TTC--
 DR03006 AACCCTGAAGACACGGCTACGTATTACTGTGCGGCGG---TCCCAGCCC
 DR11006 GCAACCTGAGGATACGGCCACCTATTACTGCGCGGCAA---GACTGACGG
 DR24006 GCAACCTGAGGATACGGCCACCTATTACTGCGCGGCAA---GACTGACGG
 DR16006 GAAACCTGAGGACACGGCCATCTACTTCTGTGCAGCAG---G----CTC
 DR19006 GACACCTGAGGACACGGCCATCTACAGTTGTGCGGCAA---C-----CAG
 DR07006 GAAACCTGAAGACACGGGCACCTATTACTGTGCTG-CA---GAACTAAGT
 DR16006 GAAACCTGAGGACACGGCCATCTATTACTGCGCGG-CG---GATAGTCCA
 DR20006 GAAACCTGAGGACAGCGCCATGTACTACTGTGCAATCA---CTGAAATTG
 DR25006 GAAACCTGAGGACACGGCCATGTATTACTGCGCGGTAGATGGTTGGACCC
 DR20006 CAAACCTGAGGACACGGCCATGTATTACTGTGCG-----CC
 DR21006 GAACCCGGAGGACACTGCCGACTACTACTGCGCTGCAAATCAATTAGC--
 DR09006 GAGGCCGTGAGGACACGGCCGTGTATTACTGTGCGGCAGATTG-----
 DR17006 GAAACCTGAGGACACGGCCATCTATTACTGTGCGGCAG-----CGGACCC
 DR13006 GCAACCTGACGACACTGGCGTGTACTACTGTGCGGCC-----CAA
 DR02006 GAAACCCGAGGACACGGCCGTGTATTACTGTGGGGCAGT-----

DR01006 -----A-AC--CTGGGGGTTATTGTGGGTA-
 DR27006 -----GTAC--CCGTGCCATCTCCTTGATG-
 DR03006 ACTTGGGACCT-----GGCG-CCATT-----CTTGATTTG
 DR11006 AGATGGGGGCTTGTGATGCGAGATGGGCGACCTTAGC--GACAAGGAC-G
 DR24006 AGATGGGGGCTTGTGATGCGAGATGGGCGACCTTAGC--GACAAGGAC-G
 DR16006 GCGTTTTT-CTAGTCCGTGTGGGAGCACTTC-TAGAC---TCGAAAGTAG
 DR19006 TAGTTTTTACTGGTACT-----GCAC-----C---ACG-----G
 DR07006 GGTGGTAGTTGTGAATTGC---CTTTGC-----TATTTGACTA-----
 DR16006 TGTTACATGCCGACTATGC---CCGCTCCCCGATACGAGACAGTTTTGG
 DR20006 AGTGGTATGGGTGCAATTT--AAGGACTACTTTTACT---C-----G
 DR25006 GGAAGGAAG--GGGGAATCGGGTTAC---CCTGGTCGGTCCAATGTGAA
 DR20006 GCTTGAA-----TATC---CTATTGCAGAC--ATGTGTT
 DR21006 ---TGGTGGCTGGTATT-----TGGACCCGAATTACTGG-CTCTCTGTG
 DR09006 ---GAAATACTGGA---CTTGTGGTGC--CCAGA-CTGG-----AG
 DR17006 AAGTATATATTATAGTATC-----CTCCNNAT-----
 DR13006 AAGAAGGATCGTA-----CTAGATGGGC-----CGAGCCT-----
 DR02006 -----CTCCCTAA--TGGACCGAATTTG

DR01006 --TGGGTANTGCCTCTGGGGCCAGGGGACCCAGGTCACCGTCTCCTCACT
 DR27006 --T-----CTGGGGCCAGGGGACCCAGGTCACCGTCTCCTCACT
 DR03006 AAAAAGTATAAGTACTGGGGCCAGGGGACCCAGGTCACCGTCTCCTCACT
 DR11006 TTTGCGTATAACTACTGGGGCCAGGGGACCCAGGTCACCGTCTCCTCACT
 DR24006 TTTGCGTATAACTACTGGGGCCAGGGGACCCAGGTCACCGTCTCCTCACT
 DR16006 CGA-CT-ATAACTATTGGGGCCAGGGGATCCAGGTCACCGTCACCTCACT
 DR19006 CGC-CTTATAACGTCTGGGGTCAGGGGACCCAGGTCACCGTCTCCTCACT
 DR07006 CTGGG-----GCCAGGGGACCCAGGTCACCGTCTCCTCACT
 DR16006 CTGGGATGATTTT-----GGCCAGGGGACCCAGGTCACCGTCTCCTCACT
 DR20006 CTGGG-----GCCAGGGGACCCAGGTCACCGTCTCCTCACT
 DR25006 GATGGTTATAACTATTGGGGCCAGGGGACCCAGGTCACCGTCTCCTCAC-
 DR20006 CGAGAT---ACG---GCGACCCGGGGACCCAGGTCACCGTCTCCTCAC-
 DR21006 GGTGCATATGCCATCTGGGGCCAGGGGACCCAGGTCACCGTCTCCTCAC-
 DR09006 GATACTTCGGACAG-TGGGGTCAGGGGGCCAGGTCACCGTCTCCTCACT
 DR17006 --TGAGTATAAGTACTGGGGCCAGGGGACCCAGGTCACCGTCTCCTCA--

FIGURE 7(3)

SUBSTITUTE SHEET

13 / 14

```
DR13006 CGAGAATGGAACAACACTGGGGCCAGGGGACCCAGGTCACCGTCTCCTCA--
DR02006 CCAACATGGG--TGCCGGGGCCAGGGAACCCAGGTCACCGTCTCCT----

DR01006 AG----TTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAAATTC
DR27006 AG----TTACCCGTACGAGCTTCCGGACTACGGTTCTTAATAGAAATTC
DR03006 AGCTAGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAAATTC
DR11006 AG----TTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAAATTC
DR24006 AGCTAGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAAATTC
DR16006 ----AGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAAATTC
DR19006 ----AGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAAATTC
DR07006 ----AGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAAATTC
DR16006 ----AGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAAATTC
DR20006 ----AGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAAATTC
DR25006 ---TAGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAAATTC
DR20006 ---TAGTTACCCGTACGACGAACCGGACTACGGTTCTTAATAGAAATTC
DR21006 ---TAGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAAATTC
DR09006 AGCTAGTTACCCGTACGACGTTCCGGACTACGGTTCTTAATAGAAATTC
DR17006 -----
DR13006 -----
DR02006 -----TA
```

FIGURE 7(4)

SUBSTITUTE SHEET

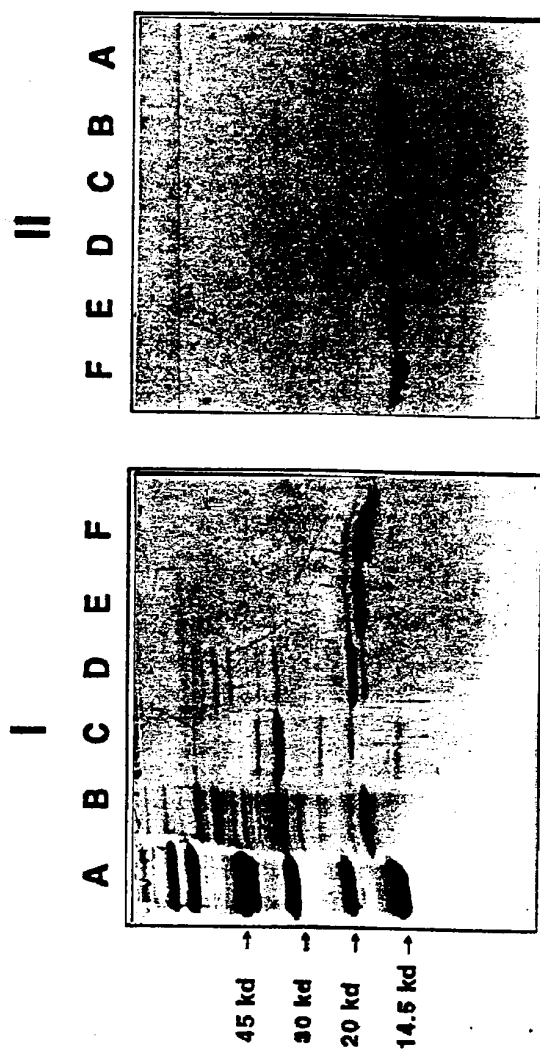


FIGURE 8

SUBSTITUTE SHEET

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 93/02214

A. CLASSIFICATION OF SUBJECT MATTER		
IPC 5	C12N15/13 C12N15/62	C07K15/06 C12P21/08 C12N5/10 A61K39/395
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols)		
IPC 5 C12N C12P C07K A61K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ISRAEL JOURNAL OF VETERINARY MEDECINE vol. 43, no. 3, March 1987 pages 198 - 203 UNGAR -WARON ET AL. 'Dromedary IgG: Purification, characterization and quantitation in sera of dams and newborns' see page 202, column 2, paragraph 3; figure 2 ---	1-5,7-9, 28
X	THE JOURNAL OF CELL BIOLOGY vol. 108, February 1989, NEW YORK US pages 737 - 746 RODITI ET AL. 'Procycline gene expression and loss of the variant surface glycoprotein during differentiation of Trypanosoma brucei' see page 744, paragraph 2 --- -/-	33
☒ Further documents are listed in the continuation of box C. ☐ Patent family members are listed in annex.		
* Special categories of cited documents :		
A document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *Z* document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
9 December 1993		06 -01- 1994
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Cupido, M

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 93/02214

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NATURE. vol. 341, no. 6242 , 12 October 1989 , LONDON GB pages 544 - 546 WARD ET AL. 'Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli' cited in the application see the whole document ---	1-41
A	NATURE. vol. 339 , 4 May 1989 , LONDON GB pages 68 - 70 TRAUNECKER ET AL. 'Highly efficient neutralization of HIV with recombinant CD4-immunoglobulin molecules' see the last line of the abstract and page 70 ---	1-41
P,X	NATURE. vol. 363 , 3 June 1993 , LONDON GB pages 446 - 448 HAMERS-CASTERMAN ET AL. 'Naturally occurring antibodies devoid of light chains' see the whole document -----	1-41

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 0 404 097 B1

(12)

EUROPÄISCHE PATENTSCHRIFT

(45) Veröffentlichungstag und Bekanntmachung des
Hinweises auf die Patenterteilung:
04.09.1996 Patentblatt 1996/36

(51) Int. Cl.⁶: **C07K 16/18**, C12N 15/13,
C12P 21/00

(21) Anmeldenummer: **90111640.0**

(22) Anmeldetag: **20.06.1990**

(54) **Bispezifische und oligospezifische, mono- und oligovalente Rezeptoren, ihre Herstellung und Verwendung**

Bispecific and oligospecific, mono- and oligovalent receptors, production and applications thereof

Récepteurs mono- et oligovalents, bispécifiques et oligospécifiques, ainsi que leur production et application

(84) Benannte Vertragsstaaten:
AT BE CH DE DK ES FR GB GR IT LI LU NL SE

(30) Priorität: **22.06.1989 DE 3920358**

(43) Veröffentlichungstag der Anmeldung:
27.12.1990 Patentblatt 1990/52

(73) Patentinhaber: **BEHRINGWERKE
Aktiengesellschaft
35001 Marburg (DE)**

(72) Erfinder:
• **Bosslet, Klaus, Dr.
D-3550 Marburg (DE)**
• **Hermentin, Peter, Dr.
D-3550 Marburg (DE)**
• **Seemann, Gerhard, Dr.
D-3550 Marburg (DE)**

• **Kuhlmann, Ludwig, Dr.
D-6093 Flörsheim am Main (DE)**
• **Steinsträsser, Axel, Dr. Dr.
D-6237 Liederbach (DE)**

(56) Entgegenhaltungen:
EP-A- 0 141 079 EP-A- 0 369 566
WO-A-86/01407 US-A- 4 652 440

• **JOURNAL OF IMMUNOLOGY**, Band 141, Nr. 11,
11. Dezember 1988; C.B. BEIDLER et al., Seiten
4053-4060
• **TRENDS IN BIOTECHNOLOGY**, Band 6, Nr. 2,
Februar 1988, Cambridge (GB); G. WILLIAMS,
Seiten 36-39, 42
• **BIOTECHNIQUES**, Band 4, Nr. 3, Mai/Juni 1986;
V.T. OI et al., Seiten 214-221

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist. (Art. 99(1) Europäisches Patentübereinkommen).

EP 0 404 097 B1

Beschreibung

Die Erfindung betrifft bispezifische und oligospezifische, mono- und oligovalente Rezeptoren, die gentechnisch durch Fusion von für F(ab) Fragmente von Antikörper zweier oder mehrerer verschiedener Spezifitäten kodierende DNA mittels geeigneter Linker hergestellt werden dadurch gekennzeichnet, daß eine Spezifität über fos-jun Interaktion gegen ein Komplexon gerichtet ist. Vorzugsweise ist eine Spezifität dabei entweder gegen ein auf der Zellmembran oder im Interstitium befindliches Epitop eines Tumor-assoziierten Antigens (TAA) oder gegen ein Epitop im Tumendothel (TE) gerichtet, während die weiteren Spezifitäten hochmolekulare oder niedermolekulare Liganden betreffen und z.B. mit den Komplexonen Äthylendiamintetraacetat bzw. Diäthylentriaminpentaacetat in Y90 komplexierter Form (EDTA-Y90 bzw. DTPA-Y90) reagieren. Die Bindung mit den Komplexonen erfolgt am Komplexon-Rezeptor-Arm über fos-jun Interaktion (oder auch Avidin-Biotin Interaktion). Weitere bevorzugte Spezifitäten haben katalytische Eigenschaften.

Bispezifische Antikörper wurden bisher nach folgenden Methoden hergestellt

- chemische Kopplung von Antikörpern verschiedener Spezifität über heterobifunktionelle Linker (H. Paulus, Behring Inst. Mitt. 78, (1985), 118-132)
- Fusion von bereits vorhandenen Hybriden, die verschiedene monoklonale Antikörper (MAK) ausscheiden, und Isolation des bispezifisch-monovalenten Anteiles (U.S. Staerz und M. J. Bevan, Proc. Natl. Acad. Sci. USA 83, (1986) 1453-1457)
- Transfektion der leichten und schweren Kettengene zweier verschiedener MAK (4 Gene) in murine Myelomzellen oder andere eukaryotische Expressionssysteme und Isolation des bispezifisch-monovalenten Anteiles (U. Zimmermann, Rev. Physio. Biochem. Pharmacol. 105 (1986), 176-260; J. van Dijk et al., Int. J. Cancer 43, (1989), 944-349).

Solche bispezifische Antikörper werden zur Therapie und Diagnostik von malignen Tumoren eingesetzt. Das Prinzip des Verfahrens besteht darin, daß im ersten Schritt durch Injektion des bispezifischen Makromoleküls über längere Zeiträume und mit hohen Dosen eine Absättigung der Epitope, die von einer der beiden Spezifitäten auf den Zielzellen erkannt werden, erreicht wird. Im zweiten Schritt, der aus einer mehrtägigen Unterbrechung der Behandlung besteht, findet die Autoelimination des unspezifisch adsorbierten bispezifischen Antikörpers aus den nicht Zielgeweben statt.

Diese Autoelimination kann durch Injektion eines mit Zuckerresten, vorzugsweise Galactose gekoppelten anti-idiotypischen Antikörpers, der gegen den Anti-Tumor-Arm des bispezifischen Rezeptors gerichtet ist, beschleunigt werden.

Der dritte Schritt des Verfahrens besteht in der i.v. Injektion eines radiomarkierten, hydrophilen, sich nicht in Zellen anreichernden niedermolekularen Liganden mit kurzer Verweilzeit im Organismus, der hohe Komplexkonstanten für beta- und gamma-Strahler wie ^{90}Y , Re^{186} , Re^{188} , Re^{189} , $^{99\text{m}}\text{Tc}$ oder ^{111}In hat und an den die zweite Spezifität des bispezifischen Antikörpers mit hoher Affinität bindet. Durch diesen Schritt wird eine Anreicherung des radioaktiven Liganden, verbunden mit längerem Verbleib am Zielgewebe erreicht, was die selektive Zerstörung des Zielgewebes zur Folge hat bzw. eine Diagnostik beispielsweise von Metastasen ermöglicht.

Die Erfindung stellt nun bispezifische bzw. oligospezifische Rezeptoren bereit, die je nach Erfordernis mono- oder oligovalente Bindungen zu den jeweiligen Epitopen besitzen und gentechnisch mittels geeigneter "Linker" erzeugt werden. Dabei werden die Genfragmente, die für die V_H und C_H1 Abschnitte der Antikörper a und b kodieren, mittels geeigneter synthetischer Oligonukleotide wie in Tab. 1 beispielhaft dargestellt, dergestalt verknüpft, daß der N-Terminus der V_H Domäne des MAK b über einen Polypeptidspacer mit dem C-Terminus der C_H1 Domäne des MAK a kovalent verbunden ist (Fig. 1). Das Genkonstrukt aus V_HaC_H1a -Polypeptidspacer- V_HbC_H1b wird zusammen mit den Genen für die leichten Ketten der Antikörper a und b in eukaryontische Zellen transfiziert (z.B. Mausmyelomzellen). Die C_H1a , C_H1b , C_{ka} und C_{kb} Domänen sind so modifiziert, daß an den Kontaktflächen der konstanten Domänen entgegengesetzte Ladungen aufeinandertreffen ($C_H1a(+)$ $C_{ka}(-)$; $C_H1b(-)$ $C_{kb}(+)$) (+ = positiv, - = negativ) bzw. sich jeweils hydrophobe oder jeweils hydrophile Kontaktflächen gegenüberstehen. Dadurch werden von den Transfektomas bevorzugt Hybridmoleküle ausgeprägt, bei denen die korrekten Paarungen aus schweren und leichten Ketten vorliegen (Fig. 2).

Antikörper a steht hier beispielhaft für einen Antitumor-Antikörper, Antikörper b für einen Antikörper gegen einen niedermolekularen Liganden, vorzugsweise die Komplexe DTPA-Y90 oder EDTA-Y90.

Bi- oder oligospezifischer Rezeptor bedeutet demnach eine gentechnische Konstruktion aus V_H und C_H1 Domänen von Antikörpern verschiedener Spezifität über geeignete Linker, so daß die erforderliche Beweglichkeit zur Assoziation mit den korrespondierenden leichten Ketten gegeben und die Antigenbindung nicht behindert ist.

Valenzen oder Bindungsstellen bezeichnen die Antigenbindungsstellen. Ein bispezifischer monovalenter Rezeptor hat bei zwei Spezifitäten, also je eine Antigenbindungsstelle. Ein bispezifischer trivalenter Rezeptor besitzt folglich eine Antigenbindungsstelle für die eine Spezifität und zwei Antigenbindungsstellen für die andere.

Bispezifische Rezeptoren, die bivalent für das Tumorantigen (MAK a) und monovalent für EDTA-Y90 (MAK b) sind, werden hergestellt, indem man das oben beschriebene schwere Ketten-Genkonstrukt mittels des oben erwähnten Oligonukleotidlinkers mit dem Genabschnitt verknüpft, der für die V_H und C_H1 Domäne des MAK a kodiert (Fig. 3), so daß

das C-terminale Ende der C_H1 Domäne des MAK b mit dem N-terminalen Ende der V_H Domäne des MAK a durch einen Polypeptidspacer verbunden ist. Diese Genkonstrukte werden zusammen mit den Genen für die zu den MAK a und b gehörenden leichten Ketten in eukaryontische Zellen transfiziert (z.B. Myelomzellen). Die C_H1 und C_κ Domänen sind, wie oben beschrieben, mit komplementären Ladungen bzw. jeweils hydrophoben oder hydrophilen Kontaktf lächen versehen. Von den Transfektomas werden bevorzugt Fusionsmoleküle ausgeprägt, die aus zwei F(ab) Fragmenten des MAK a und einem F(ab) Fragment des MAK b bestehen (Fig. 4). Die Beweglichkeit der Peptidlinker ermöglicht die Ausrichtung der beiden F(ab)-Arme des MAK a zur Tumorzelle bei gleichzeitiger Ausrichtung des F(ab)-Armes des MAK b zum interzellulären Raum. Entsprechend können weitere Bindungsstellen identischer oder anderer Spezifität zugefügt werden. Die Reihenfolge der Spezifitäten in den Konstrukten ist dabei frei kombinierbar.

Die Erfindung betrifft folglich bispezifische oder oligospezifische, mono- oder oligovalente Rezeptoren, die sowohl Spezifität für ein auf der Zellmembran oder im Interstitium befindliches Epitop haben, zum Beispiel TAA oder TE, als auch Spezifität für einen nieder- oder hochmolekularen Liganden besitzen, der sich ausschließlich im extrazellulären Raum verteilt. Eine Spezifität wird dabei vorzugsweise von den tumorspezifischen Antikörpern gebildet, wie in der deutschen Patentanmeldung P 39 09 799.4 vorgeschlagen, während die andere Spezifität vorzugsweise gegen DTPA-Y90 oder EDTA-Y90 gerichtet ist und die Bindung mit Komplexonen am Komplexon-Rezeptor-Arm über fos-jun Interaktion (siehe Beispiel 5) erfolgt. Eine bevorzugte Variante der Erfindung besteht im Einbau von katalytisch wirksamen Spezifitäten. Die Reihenfolge der Spezifitäten bzw. Bindungswerten ist dabei frei wählbar, wie beispielhaft in Fig. 4 für drei Valenzen bei zwei Spezifitäten gezeigt.

Besonders bevorzugt bei den erfindungsgemäßen Konstrukten sind solche, die ein V-Gen der Tabellen 2, 3, 4 und/oder 5 enthalten. Antikörper mit diesen Sequenzen und ihre Eigenschaften sind in der deutschen Patentanmeldung P 39 09 799.4 beschrieben. Die "Complementarity Determining Regions" (CDR's) sind dabei nach Kabat und Wu (Sequences of Proteins of Immunological Interest, US Dept. of Health and Human Services, US Government Printing Office (1987)) identifizierbar. Bevorzugt sind ebenfalls Konstrukte, die Spezifitäten gegen die durch vorstehend geschilderte monoklonale Antikörper definierte Epitope enthalten.

Zusätzlich betrifft die Erfindung gentechnische Verfahren zur Herstellung oben beschriebener Konstrukte sowie eine Verwendung o.g. Konstrukte zur Herstellung von Arzneimitteln zur Bekämpfung und Diagnose von Zielzellen. Dabei findet in einem 1. Schritt nach Injektion der Konstrukte eine Absättigung der betroffenen Epitope auf Zielzellen statt, in einem nachfolgenden Intervall werden unspezifisch adsorbierte oder nicht gebundene Konstrukte eliminiert. Der daran anschließende Schritt besteht in der Injektion und nachfolgender spezifischer Bindung eines sich nicht in den Zielzellen anreichernden nieder- oder hochmolekularen Liganden, der per se zytotoxisch ist oder gegebenenfalls in einem weiteren Schritt durch extrakorporale Einwirkungen zur Zytotoxizität "aktiviert" wird. Verfahren dieser Art sind beispielsweise enzymatische Aktivierung, Aktivierung durch Mikrowellenbestrahlung eines "Pro-Drugs" oder Aktivierung durch Laserlicht.

Die Erfindung ist ferner in den Beispielen und den Patentansprüchen enthalten.

Beispiel 1: Herstellung eines anti-DTPA-Y90 bzw. EDTA-Y90-MAK

Als Hapten wurde Isothiocyanato-benzyl-DTPA (Formel 2) auf humanes Serumalbumin (HSA als Carrier) mit einem Derivatisierungsgrad von 19 Benzyl-DTPA Molekülen pro HSA-Molekül entsprechend der in (N.W. Brechbiel et al., Inorganic Chemistry 25, (1986) 2772-2781) beschriebenen Methodik kovalent gekoppelt. 20 µg dieses Hapten-Carrier-Komplexes, in welchen kaltes Y komplexiert wurde, wurden s.c. an Tag 0 mit Freundschem Adjuvans, an Tag 7 und 14 mit inkomplettem Freundschem Adjuvans und an Tag 21 mit PBS in Balb/c Mäuse injiziert. An Tag 24 wurden die Milzen der Mäuse mit den höchsten anti DTPA Antikörpertitern mit der SP2/0-Ag14 Myelomzelllinie (Shulman et al., Nature 276, (1978) 269) fusioniert. Entstehende Hybridome wurden in einem DTPA-spezifischen ELISA auf Produktion von hochaffinen MAK gegen DTPA und EDTA überprüft. Der ELISA bestand aus einer festen Phase, die mit einer HSA-Benzyl-DTPA-Y enthaltenden Lösung beladen wurde. Der zu testende, den MAK enthaltende Überstand wurde mit freiem Komplexon bzw. dessen Metallionkomplexen vorinkubiert und seine Bindung an die spezifische feste Phase gemessen. Hierzu wurde ein Enzymamplifikationssystem benutzt, welches an einen anti Maus-Immunglobulinantikörper gekoppelt ist. Die Details dieser Methodik sind in Anlage 1a und 1b beschrieben.

Mittels dieses Testsystems wurden MAK erhalten, welche die in Anlage 1e beschriebenen Eigenschaften besitzen. Diese MAK binden im Gegensatz zu vielen anderen anti DTPA/EDTA MAK nicht an menschliche Normalgewebe, wie mittels der APAAP-Technik (Cordell et al., J. Histochem. Cytochem. 32: 219, 1984) auf kryopräservierten Geweben ermittelt wurde. Ein in vivo Einsatz dieser MAK im Bereich der Diagnostik und Therapie ist somit möglich.

Als Kompetitoren wurden die Komplexone DTPA, EDTA in nicht komplexierter sowie in komplexierter Form (Anlage 1c) eingesetzt. Zusätzlich wurden als Inhibitoren die strukturverwandten Verbindungen Transaconitsäure und 1,2 Diaminoethan verwendet (siehe Anlage 1e). Besonders geeignet für eine in vivo Anwendung ist der MAK BW 2050/174 der im Gegensatz zu allen anderen MAK eine präferentielle Bindung an EDTA-Y zeigt (siehe Anlage 1e, niedriger Kompetitor Überschuß für EDTA-Y (100fach) höherer Überschuß für andere EDTA-Komplexone. Deshalb wurde das Hybrid 2050/174 stabilisiert und für die Entwicklung des EDTA-Y-Armes im bispezifischen Rezeptor verwendet.

Beispiel 2: Herstellung und Expression eines V_H1a C_H1-Linker-V_H1B C_H1 Genkonstruktes

Die hier verwendeten Techniken wurden, wenn nicht anders angezeigt, aus Molecular Cloning, a Laboratory Manual; Sambrook, Fritsch, Maniatis; Cold Spring Harbor Laboratory, 1982 (S. 11 - 44, 51 - 127, 133 - 134, 141, 146, 150 - 167, 170, 188 - 193, 197 - 199, 248 - 255, 270 - 294, 310 - 328, 364 - 401, 437 - 506) und aus Molecular Cloning, A Laboratory Manual, Second Edition; Sambrook, Fritsch, Maniatis; Cold Spring Harbor Laboratory Press, 1989, (S. 16.2 - 16.22, 16.30 - 16.40, 16.54 - 16.55) entnommen.

Ein humanes IgG₃ C-Gen wurde aus einer humanen Genbank in EMBL3 Phagen isoliert (A.M. Frischaut et al., J. Mol. Biol. 170, 827-842 (1983) und G.H.A. Seemann et al., The EMBO Journal 5 (1986), 547-552).

Aus diesem IgG₃ C-Gen wurden wie in der deutschen Patentanmeldung P 38 25 615.0 beschrieben, Konstruktionen hergestellt, die zum einen nur noch das C_H1 Exon und ein Hinge Exon (Fig. 5) und zum anderen das C_H1 Exon und den 3'NT Bereich eines HLA B27 Gens enthalten (Fig. 6, Fragment M im Plasmid M).

Die V_Ha und V_Hb Gene wurden aus mRNA der Hybridklone a und b amplifiziert, wie von Orlandi et al. (Proc. Natl. Acad. Sci. USA 86, (1989), 3833-3837) beschrieben und in einen M13 Vektor kloniert (V_Ha PCR bzw. V_Hb PCR) (Fig. 7). Das V_Ha Gen wurde als HindIII BamHI Fragment in den eukaryontischen Expressionsvektor pEV_H kloniert (Simon et al., Nucl. Acids. Res. 16, (1988), 354) (Fig. 8). Es entsteht das Plasmid PEV_H a C.

Der humane IgG C Gen Subklon mit dem C_H1 und dem einen Hinge Exon (Fig. 5) enthält zwischen C_H1 Exon und Hinge Exon eine PstI Schnittstelle. Die V_H Gene enthalten am 5' Ende eine PstI Schnittstelle. Das Linkeroligonukleotid wird so entworfen, daß es am 5' Ende mit dem Bereich der PstI Schnittstelle auf dem C_H1 + 1H Subfragment des IgG C Gens und am 3' Ende mit der PstI Schnittstelle des V_Hb Gens überlappt. Das Linkeroligonukleotid wird mittels seiner PstI Schnittstellen in die PstI Schnittstelle eines PUC 18 Plasmids kloniert (Fig. 9). Es entsteht der Plasmidklon L.

Das Plasmid mit dem IgG₃ C Gen Subfragment mit C_H1 Exon und einem Hinge Exon wird mit PstI und BamHI gespalten und mit dem als PstI BamHI Fragment aus V_Hb PCR ausgeschnittenem V_Hb Genfragment ligiert (Fig. 10). Es entsteht das Plasmid X.

Das Plasmid X wird mit PstI gespalten und mit dem ebenfalls mit PstI aus dem Plasmid L ausgeschnittenen Linkerfragment ligiert (Fig. 11). Mit Hilfe der Nukleinsäuresequenzanalyse wird der Klon Z identifiziert, bei dem der Linker in der richtigen Orientierung zwischen C_H1 und V_Hb kloniert ist, ohne den Intron/Exonübergang zwischen Intron 3 und Linkerexon und ohne das Leseraster am Übergang zwischen Linker und V_Hb Gen zu stören.

Das Plasmid pEva C wird mit BamHI gespalten und mit dem mit BamHI aus dem Plasmid M ausgeschnittenen Fragment M ligiert. Durch Restriktionsanalyse wird der Klon Y identifiziert, der das Fragment M in der richtigen Orientierung enthält (Fig. 12).

Das Plasmid Y wird partiell mit BamHI gespalten und mit dem durch HindIII und BamHI Spaltung aus dem Plasmid X ausgeschnittenen Fragment (C_H1-Linker-V_Hb) nach Auffüllen aller Enden ligiert. Durch Nukleotidsequenzanalyse und Restriktionskartierung wird der Plasmidklon PEVT identifiziert, der das Fusionsgen V_Ha C_H1-Linker-V_Hb C_H1 mit der richtigen Orientierung aller Exons enthält (Fig. 13).

Das Plasmid PEVT wird zusammen mit Plasmiden, die die Gene für die leichten Ketten der Antikörper a bzw. b tragen, in geeignete eukaryontische Expressionszellen transfiziert, um das Antikörper a F(ab) Antikörper b F(ab) Fusionsprotein auszuprägen.

Beispiel 3: Transfektion der leichten und schweren Ketten-Gene zweier verschiedener MAK (4 Gene)

Die Isolierung von Immunglobulinen ist in der deutschen Patentanmeldung P 39 09 799.4 beschrieben.

Die in Vektoren klonierten Gene wurden mittels Elektroporation nach Linearisierung der Vektoren in X63Ag8.653 Myelomzellen transfiziert (H. Stopper et al., Biochem. Biophys. Acta 900 (1987), 38-44). Die in Selektivmedien ausgewachsenen Transfektome wurden auf Produktion von bispezifischem-monovalentem MAK in einem spezifischen RIA überprüft. Dieser RIA bestand aus Festphasen-adsorbiertem TAA, auf welchen nach Blockade der unspezifischen "sites" durch Casein die zu testenden Transfektomüberstände gegeben wurden. Nach Zugabe von mit ⁹⁰Y oder ^{99m}Tc komplexiertem DTPA oder EDTA und Wegwaschen des Überschusses konnten diejenigen Transfektome, die bispezifischen-monovalenten anti TAA x anti EDTA MAK ausschieden, an einem erhöhten radioaktiven Signal auf der Festphase detektiert werden.

Transfektom 9 wurde stabilisiert durch "limited dilution" Klonierung und in Zellkultur aufgebaut. Zellkulturüberstände wurden 10fach ankonzentriert, der MAK-Anteil wurde über Protein A chromatographiert (P.L. Ey et al., Immunochimistry 15, (1978), 429) gereinigt und der den bispezifisch-monovalenten MAK enthaltende Anteil mittels Anionenaustauscher-Chromatographie gereinigt. (J. Van Dijk et al., Int. J. Cancer 43, (1989), 344-349).

Beispiel 4: Biologische Wirksamkeit

Gereinigtes, den bispezifischen-monovalenten MAK (BW 431/26 x BW 2050/174) enthaltendes Protein wurde in 500 µg Dosen an den Tagen 0, 3, 5, 8, 10 und 12 i.v. in human Tumorenografts (CoCa 4) tragende Nacktmäuse inji-

ziert. An Tag 27-30 erhielten die Tiere je 50 μCi EDTA-Y90 i.v. injiziert. Eine 2. Gruppe von Tieren erhielt an den gleichen Tagen anstatt des bispezifischen MAK 500 μg des MAK BW 431/26 und, wie oben beschrieben, die EDTA-Y90-Injektionen.

Eine dritte mit Tumor bestückte Gruppe erhielt anstatt des MAK und des EDTA-Y90 Injektionen von PBS (Tumorstudienkontrollen).

Das Tumorstudienwachstum wurde über 6 Wochen verfolgt. Die Injektion von EDTA-Y90 führte in der Gruppe, die den bispezifischen-monovalenten MAK erhielt, zu einer signifikanten Tumorstudienwachstumshemmung, wohingegen die mit MAK BW 431/26 injizierten und mit EDTA-Y90 behandelten Tiere keine Tumorstudienwachstumshemmung zeigten, im Vergleich zu den Tieren, die nur PBS erhielten.

Diese Daten weisen auf die selektive tumorstudientherapeutische Wirksamkeit des bispezifischen-monovalenten MAK in Kombination mit EDTA-Y90 als toxischem Prinzip hin.

Noch günstigere tumorstudientherapeutische Effekte werden durch die oligovalenten/bispezifischen oder oligospezifischen Rezeptoren erhalten, da sie aufgrund der bivalenten Bindung an TAA länger am Tumor verweilen und somit der Ligand ebenfalls länger und in höherer Konzentration am Tumor zurückgehalten wird.

Beispiel 5: Optimierung der biologischen Wirksamkeit bi-oder oligospezifischer Makromoleküle durch Aviditätserhöhung des anti-Komplexonarmes.

Ein entscheidender Faktor, der die effiziente Anheftung des hydrophilen, sich extrazellulär verteilenden Komplexons am anti-Komplexonarm des oligospezifischen Makromoleküls beeinflusst, ist die Avidität dieses Armes zum Komplexon. Die Aviditäten von monoklonalen Antikörpern zu ihren entsprechenden Epitopen liegen im Bereich von 10^5 - 10^{11} l/mol. Da diese Bindungsstärken vielleicht nicht ausreichen, um die für eine effiziente Radioimmuntherapie notwendige Komplexonmasse am Tumor zu lokalisieren, wurde im folgenden Beispiel die extrem starke Interaktion des fos-Leucin-Reißverschlußpeptids (fos-Peptid) mit dem jun-Leucin-Reißverschlußpeptid (jun-Peptid) genutzt (Erin K. O'Shera et al., Science, 245, 1989), um das Komplexon möglichst fest am anti Komplexonarm zu fixieren. Um diese starke fos-jun Interaktion nutzen zu können, muß vorzugsweise das fos-Peptid kovalent mit dem Komplexon (DTPA) verknüpft werden. Hierzu kann im ersten Schritt isothiocyanatobenzyl DTPA mit Hydrazin (oder einem Diaminoalkan) umgesetzt werden. Das so entstandene DTPA-Benzyl-thiocarbamid kann in einem 2. Schritt mit N-(gamma-maleimidobutyryloxy)succinimid oder einem Analogon zu DTPA-Benzyl-(gamma-maleimidobutyl) thiocarbamid umgesetzt werden. In einem 3. Schritt wird dann diese Verbindung mit dem um Glycin-Glycin-Cystein verlängerten fos-Peptid (Bild 1) über die freie SH-Gruppe des aminoterminalen Cysteins verknüpft. Das so entstandene fos-Peptid-DTPA-Konjugat wird in einem 4. Schritt mit Yttriumchlorid komplexiert. Der so entstandene fos-Peptid-DTPA-Y-Konjugatkomplex kann dann zur in vivo Anlagerung an den jun-peptid-Arm des bi- oder oligo-spezifischen Makromoleküls benutzt werden. Nachfolgend ist die Synthese des oben vorskizzierten Beispiels im Detail beschrieben:

A) Herstellung des fos-EDTA-Y-Konjugat-Komplexes

Schritt 1:

Synthese von EDTA-Benzyl-thiocarbamid

Isothiocyanatobenzyl-EDTA (SCN-Bn-ETPA) (30 mg, 54 μmol) wurde 1 h lang in 10 % (v/v) wässrigem Hydrazin gerührt. Dann wurde das Lösungsmittel im Hochvakuum entfernt, der Rückstand über Phosphorpentoxid im Hochvakuum getrocknet und schließlich lyophilisiert. Das Produkt wurde mit DOWEX WX 2 (H⁺-Form) neutralisiert und erneut lyophilisiert (Ausbeute 28 mg).

Schritt 2:

Synthese von EDTA-Benzyl-(gamma-maleimidobutyl)thiocarbamid

Das in Schritt 1 hergestellte ETPA-Benzyl-thiocarbamid (20 mg; 34 μmol) und N-(gamma-Maleimidobutyryloxy)succinimid (8 mg, 29 μmol = 0.9 equiv.) wurden 1 h lang in wasserfreiem Dimethylformamid gerührt. Dann wurde zur Trockne eingedunstet und der Rückstand im Hochvakuum getrocknet.

Schritt 3:

Kopplung des EDTA-Benzyl-(gamma-maleimidobutyl)thiocarbazid an aminoterminales Cystein im fos-Peptid

Eine Lösung des fos-peptids (4.8 mg, 1 µmol) (siehe Schritt 3.1) in Phosphat-gepufferter Kochsalzlösung (2 ml) wurde mit einer Suspension des gemäß Schritt 2 gewonnenen Produktgemisches (4 mg) in Dimethylformamid (400 µl) versetzt und 1 h bei Raumtemperatur inkubiert. Dann wurde das Reaktionsgemisch über eine Sephadex G15-Säule in Phosphat gepufferter Kochsalzlösung Gel-filtriert. Das proteinhaltige Eluat wurde gesammelt und bei -30° C präserviert (Ausbeute 4.2 mg).

Schritt 3.1:

Aminosäuresequenz des fos-Peptids (I) mit N-terminaler GGC Verlängerung.

Ac-CGGyLTDTLQAETDQLEDKKSALQTEIANLLKEKEKLEFILAAy Die Buchstaben stehen für folgende Aminosäuren: A = Alanin, C = Cystein, D = Asparaginsäure, E = Glutaminsäure, G = Glycin, I = Isoleucin, K = Lysin, L = Leucin, M = Methionin, N = Asparagin, Q = Glutamin, R = Arginin, S = Serin, T = Threonin, V = Valin, Y = Tyrosin.

Die Oligopeptidsynthese erfolgte mittels eines automatischen Peptid-Synthesizers (Applied Biosystems Model 430A) entsprechend der Merryfield Festphasen Methode (Stewart und Young, Solid Phase Synthesis, Pierce Chemical Company, 2. Edition, Rockford Ill.) mit der tert-butyloxycarbonyl Schutzgruppe. Die Oligopeptide wurden von dem Phenylacetamidomethylpolystyrene-Träger abgespalten. Nach Abspaltung der Schutzgruppen (Tom et al. 1983, J. Am. Chem. Soc. 105, 6442-6455) wurden die Oligopeptide über reversed phase Chromatographie (PepRPC-Säule, Pharmacia) wie bei Rivier et al. (J. Chromatography 288, 303-328, 1984) beschrieben gereinigt.

Schritt 4:

Herstellung eines fos-Peptid-EDTA-Yttrium-Chelats mit einem nach Schritt 3 hergestellten fos-peptid-EDTA-Konjugates

Das gemäß Schritt 3 hergestellte fos-Peptid-EDTA-Konjugat (4.2 mg) wurde gegen isotonische Kochsalzlösung/0.1 M Natriumcitrat, pH 7.0, in einem Dialyseschlauch mit Ausschlußgrenze, m.w. 1000 (Spectrum) dialysiert und mit 6 mg Yttriumchlorid, die in 3 ml isotonischer Kochsalzlösung/0.1 M Natriumcitrat, pH 7.0, gelöst waren, versetzt. Nach 1 h wurde gegen Phosphat-gepufferte Kochsalzlösung rückdialysiert und die Chelatlösung bei -30° C präserviert. Das im obigen Beispiel beschriebene fos-Peptid-EDTA-Yttrium-Chelat wird dann als Ligand benutzt, um mit hoher Avidität an den jun-Peptid-Arm des bi- oder oligospezifischen Makromoleküles zu binden. Die Konstruktion eines für diese Interaktion besonders gut geeigneten bispezifischen Makromoleküles ist im folgenden Beispiel beschrieben.

B) Konstruktion des Mak-jun Fusionsmoleküles

Die hier verwendeten Techniken wurden, wenn nicht anders angezeigt, aus Maniatis et al. (Laboratory Manual EMBL (1982), Heidelberg, und Sambrook (Molecular Cloning: A Laboratory Manual) entnommen.

Schritt 1:

Ein humanes IgG3 C-Gen wurde aus einer humanen Genbank in EMBL 3 Phagen isoliert (A.M. Frischaut et al., J. Mol. Biol. 170, 827-842, 1983 und G.H.A. Seemann et al., The EMBO Journal 5, 547-552, 1986). Aus diesem IgG3 C-Gen wurde, wie in der deutschen Patentanmeldung P 3825615.0 beschrieben, eine Konstruktion (D) hergestellt, die nur noch das CH1 Exon und das erste Hinge Exon des IgG3 C-Gens enthält (Fig. 14).

Aus der gleichen Genbank wurde, wie ebenfalls in der deutschen Patentanmeldung P 3825615.0 beschrieben, ein humanes HLA B27k Gen isoliert. Aus diesem HLA B27w Gen wurde ein Konstrukt (E) hergestellt, das nur noch das C3 Exon und den 3' NT Bereich des HLA B27k Gens enthält (Fig. 15).

Schritt 2:

Das C1 Exon und der 3' NT Bereich des HLA B27k Gens wurden mit Xba1 aus dem Plasmid E herausgeschnitten, das Fragment isoliert und in die Xba1 Schnittstelle des Konstruktes D kloniert. Durch Restriktionsanalyse und Nukleinsäuresequenzanalyse wurde der Klon F identifiziert, der das C3 Exon und den 3' NT Bereich des HLA B27c Gens in der korrekten 5'-3' Orientierung 3' vom IgG3 C-Gen Fragment enthält (Fig. 16).

Schritt 3:

Das Insert des Klon F wird unter Verwendung der Endonukleasen Hind III und Eco RI aus dem Plasmid ausgeschnitten und zwischen die Hind III und Eco RI Schnittstellen eines M13mp18 Doppelstrang (DS)-Phagen kloniert. Es wird der Phagenklon G isoliert, der das Antikörper/HLA Fusionsgenfragment enthält (Fig. 17).

Schritt 4:

Vom Phagenklon G werden nach der Methode von T.A. Kunkel, 1985, Proc. Natl. Acad. Science, U.S.A., 82, 488-492, Uracileinzelstränge präpariert. Die Einzelstrangphagen wurden mit den mutagenen Oligonukleotiden 1 und 2 (Tab. 6) hybridisiert und die Lücken zwischen den Oligonukleotiden mit Klenon DNA Polymerase und T4 Ligase geschlossen. Nach Transformation in E. coli wurde durch Restriktionsanalyse und Nukleinsäuresequenzanalyse ein Phagenklon identifiziert (G), bei dem die Sst1 Restriktionsschnittstelle am 5' Ende des Hinge Exons deletiert wurde. Gleichzeitig wurden am 3' Ende des Hinge Exons eine Sst1 und eine Sph1 Restriktionsschnittstelle eingeführt (Fig. 18). Zur Deletion der Sst1 Schnittstelle wurde die dritte Base des 2. Codons des Hinge Exons von C in G umgewandelt und für die Einführung der Sst1 und Sph1 Schnittstellen wurden zwischen dem 15. und 16. Codon des Hinge Exons die Basen 5'GAGCTCGGGGCA3' eingeführt (Tab. 7).

Schritt 5:

Doppelstrang DNA des Phagenklons G' wird mit Sph1 vollständig und mit Sst1 partiell gespalten. Die synthetischen Oligonukleotide Jun I und Jun II (Tab. 8) werden zu einem Doppelstrang DNA Fragment zusammengefügt, welches an seinen Enden je eine geschnittene Sph1 und Sst1 Restriktionsschnittstelle enthält und für ein Peptid kodiert, das den Jun Leucin Zipper enthält (O'Shea et al., Science, 245, 646-648, 1989).

Das doppelsträngige DNA Fragment wird in die Sst1 und Sph1 Restriktionsschnittstellen des F' Phagenklons kloniert und der Phagenklon H identifiziert, der ein Genkonstrukt enthält, bei dem im Hinge Exon die Sequenz für das Jun Zipper Peptid inseriert ist (Fig. 19).

Schritt 6:

Das Insert des dS Phagen H wurde mit den Restriktionsendonukleasen Hind III und Eco RI ausgeschnitten, die Enden mit T4 Polymerase aufgefüllt und in einen Sma1 gespaltenen KsF Vektor (Stratagene, 11099 North Torrey Pines Road, La Jolla California 92037) kloniert. Es wurde der Plasmidklon I identifiziert, der das Antikörper/Jun/HLA Fusionsgen in der Orientierung enthält (Fig. 20), in der es auf beiden Seiten von einer BamH1 Schnittstelle flankiert wird.

Schritt 7:

Das Antikörper/Jun/HLA Fusionsgen wurde mit BamH1 aus dem KS Klon H1 ausgeschnitten und in das Expressionsplasmid pABStop (Behringwerke AG) kloniert, welches ein spezifisches funktionelles Immunglobulin V-Gen enthält. Das spezifische V-Gen wurde, wie in der Patentanmeldung P 3909799.4 beschrieben, gewonnen. Es wurde das Expressionsplasmid I identifiziert, das das Antikörper/Jun/HLA Fusionsgenkonstrukt in der korrekten Orientierung hinter dem V_H Gen enthält (Fig. 21).

Die Kointransformation des Plasmids I mit einem Plasmid, der das Gen für die leichte Kette des spezifischen MAK enthält und einen Resistenzgen trägt, führt zur Expression eines spezifischen Antikörper F(ab')₂ Fragmentes, das in der Hinge Region zwei Jun Zipper Peptide enthält, wobei das Jun Zipper Peptid so verändert ist, daß keine Homodimerbildung (Jun/Jun) mehr stattfindet.

Beispiel 6:Optimierung der bi- oder oligospezifischen Rezeptormenge am Tumor und Minimierung derselben im Blut bzw. den Normalgeweben.

Wissenschaftliche Untersuchungen anderer haben gezeigt, daß die Penetration von soliden Tumoren mit Makromolekülen > 50 kDa langsam von statten geht und meist nur die Randregion bzw. einige wenige Areale im Tumor erreicht werden. Diese Untersuchungen basieren auf Experimenten, die eine einmalige Injektion niedriger Mengen von Makromolekülen beinhalten. Im Gegensatz hierzu haben wir gefunden, daß durch repetitive i.v.-Injektion großer Mengen von bi- oder oligospezifischen Rezeptoren (10 x 250 µg Rezeptor/Maus für 10 Tage) eine weitgehende Durchdringung der gesamten Tumormasse in Nacktmaus-Xenografts möglich ist. Des weiteren bleiben die bi- oder oligospezifischen Rezeptoren aufgrund ihrer spezifischen Bindung an TAA für lange Zeiträume (> 20 Tage) in großen

Mengen an der Tumorzellmembran und im Tumorinterstitium hängen. Diese Befunde wurden mittels der indirekten alkalischen Phosphatasetechnik an kryopräservierten Dünnschnitten von humanen Colon- und Pankreas-Tumor-Xenografts erhoben.

Während dieser Zeit (schon nach 10 Tagen) wurden die bi- oder oligospezifischen Rezeptormoleküle aus den TAA-negativen Normalgeweben und dem Blut durch Abbau und Ausscheidung bereits eliminiert. Um diesen Eliminationszeitraum zu verkürzen, wurde ein anti-idiotypischer MAK (anti Id), der nur mit dem anti TAA-Arm nicht gebundener bi- oder oligospezifischer Rezeptormoleküle reagiert, 24 Stunden nach Beendigung der 10 x Injektion von bi- oder oligospezifischen Rezeptoren i.v. injiziert (1 x 50 µg anti Id). Diese einmalige Injektion bewirkte eine beschleunigte Elimination der ungebundenen bi- oder oligospezifischen Rezeptormoleküle aus dem Blut und eine schnellere Metabolisierung in Leber und Milz.

Aufgrund dieser Manipulation kann die Injektion des Komplexons (EDTA-Y90) schon 4 Tage nach Beendigung der Penetrations- und Bindungsphase des bi- oder oligospezifischen Rezeptors erfolgen. Aus diesen Untersuchungen leitet sich folgendes Behandlungsschema (für Nacktmäuse) ab:

- a) Tag 1-10, i.v. Injektion von je 1x250 µg bi- oder oligospezifischem Rezeptor
- b) Tag 11, i.v. Injektion von 1x50 µg anti Id
- c) Tag 14, i.v. Injektion einer therapeutischen Dosis von EDTA-Y90.

Aufgrund vergleichender immunszintigraphischer Daten in Nacktmäusen bzw. Tumorpatienten sollte dieses Schema auch für die Tumorthherapie im Menschen geeignet sein. Allerdings bewegen sich die im Humansystem zu injizierenden Mengen in einer anderen Größenordnung. 10x 5-10 g bispez. Rezeptor, 1x1 g anti Id. Die Injektion des anti Id ist für die Therapie nicht zwingend.

Anlage 1a

Quantitative Inhibition-ELISA für MAK durch DTPA bzw. EDTA Komplexe

Material: Teilbare 96-well Polystyrol Mikrotiterplatten (U-Form) Typ B, Fa. Nunc, Nr. 4-60445

- 1) Pro well werden 50 µl Y Benzyl-DTPA-HSA 19-Konjugat mit einer Konz. von 1 µg Konjugat pro ml PBS, pH 7.2, pipettiert und über Nacht bei Raumtemperatur (RT) inkubiert.
- 2) Der Überstand wird abgesaugt und 3x mit 0.05 M Tris-Citrat- Puffer, pH 7.4, (Waschlsg. 1) gewaschen; (1x waschen = 250 µl Waschlösung pro well einfüllen, 2 min stehen lassen, absaugen)
- 3) Wenn die Mikrotiterplatte nicht direkt benötigt wird, wird sie über Nacht auf Zellstoff bei RT stehengelassen (Öffnung nach unten). Danach wird die Platte in Folien mit Trockenpatronen eingeschweißt (Fa. Gaplast, Postfach 529, 8100 Garmisch-Partenkirchen). Unter diesen Bedingungen sind die Platten bei +4° C mindestens 8 Wochen haltbar.
- 4) 250 µl Blocklösung werden pro well aufgetragen und 30 min bei 37° C inkubiert.
- 5) Während des Blockierens erfolgt die Vorinkubation des verdünnten Hybridomüberstandes mit dem Kompetitor (siehe Anlage 2).
- 6) Von den zu testenden, entsprechend vorverdünnten und vorinkubierten Hybridomaüberständen werden pro well 50 µl aufgetragen und 30 min bei RT inkubiert.
- 7) Anschließend wird 3x mit Waschlösung 2 gewaschen.
- 8) Anschließend werden 50 µl 1:500 in Blocklösung verdünnter, mit alkalischer Phosphatase markierter Ziege anti Maus IgG₁-Antikörper pro well aufgetragen und 30 min bei RT inkubiert.
- 9) Danach wird 3x mit Waschlösung für Enzygnost gewaschen.
- 10) Anschließend werden 50 µl 0,1 mM NADP zugegeben.
- 11) Danach wird 30 min bei RT inkubiert.
- 12) Während der Inkubation mit NADP wird das Verstärkersystem wie folgt angesetzt: pro Platte werden zu 2 Teilen INT und 1 Teil PBS, pH 7.2, zugegeben, des weiteren 1 Teil Diaphorase, des weiteren wird 1 Teil ADH zupipettiert.
- 13) 50 µl dieses Verstärkersystems werden pro well zugegeben.
- 14) Bei deutlichem Farbumschlag von durchsichtig nach rot wird die Reaktion mit 100 µl einer 0,1 N H₂SO₄-Lösung pro well abgestoppt.
- 15) Die Extinktionen werden bei 492 nm im TITERTEK^R MULTISCAN vermessen. Als Blankwert werden 50 µl NADP mit 50 µl Verstärkerlösung und 100 µl 0,1 N H₂SO₄ eingesetzt.

EP 0 404 097 B1

NADP - Fa. Sigma Best.Nr. N-0505
INT - Fa. Sigma Best.Nr. I-8377
ADH - Fa. Sigma Best.Nr. A-3263
DIAPHORASE - Fa. Sigma Best.Nr. D-2381
5 Waschlösung 2 - Fa. Behring, Best.Nr. OSEW96, enthält Tween/PBS

Blocklösung:

10 PBS, pH 7.2, wird durch Zugabe von Casein und 30minütiges Einrühren 3%ig an Casein gemacht und auf pH 7.4 eingestellt. Danach werden Partikel 10' bei 4000 U/min abzentrifugiert.

Verdünnter mit alkalischer Phosphatase markierter Ziege anti Maus IgG₁-Antikörper (Fa. Southern Biotechnology Associates, Cat.Nr. 1030-04)

Herstellung von 0.1 mM NADP:

15 7.65 mg NADP lösen in 100 ml 20 mM Tris, 0.1 mM MgSO₄, pH 9.5; die Lösung kann bei -20° C mehrere Monate gelagert werden.

Herstellung von INT (P-iodo nitrotetrazolium Violet):

20 2.5 mg/ml 30%igem Ethanol im Ultraschallbad lösen; immer frisch ansetzen.

Herstellung von Diaphorase:

25 1 mg Diaphorase/ml PBS, pH 7.2, wird portioniert bei -20° C gelagert.

Herstellung von Alkoholdehydrogenase:

30 0.5 mg ADH/ml PBS, pH 7.2, werden portioniert bei -20° C gelagert.

Anlage 1b

Vorinkubation des Hybridomüberstandes mit dem Kompetitor

35 Die Maus IgG Konzentrationsbestimmung in Hybridomüberständen kann mittels kommerziell erhältlicher quantitativer ELISA Testsysteme bestimmt werden und ist Stand der Technik.

Anhand der Konzentrationsbestimmung im ELISA werden die Hybridomüberstände auf 1.25 µg/ml in PBS ohne Ca⁺⁺ und Mg⁺⁺ verdünnt.

Umrechnung von Gramm in Mol:

40 150 000 g - 1 Mol MAK

1.25 x 10⁻⁶ g- xMol

$$1.25 \mu\text{g} = x = 8.33 \cdot 10^{-12} \text{ Mol}$$

45 Um ein Verhältnis 1+1 von MAK und Inhibitor zu haben, wurden zu 50 µl Hybridomüberstand mit einer Konzentration von 8.33 x 10⁻¹² Mol/ml 10 µl Inhibitor mit einer um den Faktor 5 erhöhten Konzentration von 8.33 x 10⁻¹² Mol/200 µl gegeben.

Der Hybridomüberstand wird mit 100 000fachem, 50 000fachem, 10 000fachem, 5 000 fachem, 1 000 fachem und 100fachem Kompetitorüberschuß für 30' bei RT inkubiert. Hiervon werden 50 µl in den ELISA (siehe Anlage 1a, Punkt 6) pipettiert.

Anlage 1c

Erzeugung der DTPA- bzw. EDTA-Komplexe

55 Die Komplex-Konstante von DTPA bzw. EDTA zu den in Tabelle I dargestellten Metallionen ist extrem hoch, so daß bei äquimolarer Mischung von DTPA bzw. EDTA mit diesen Metallionen eine komplette Sättigung zu erwarten ist. Die entsprechenden Metallionen wurden deswegen in einem 3fachen molaren Überschuß mit den DTPA bzw. EDTA inkubiert. Als Beispiel wurden 170 µl einer 10 mM Cadmiumsulfatlösung in bidest (siehe Anlage 1d) mit 30 µl einer 0.028

EP 0 404 097 B1

molaren DTPA-Stammlösung in bidest für 5' bei RT inkubiert. Zumischung von 10 µl dieser Kompetitorlösung zu dem Hybridomüberstand führt zu einem 100000fachen Überschuß an Kompetitor über den im Hybridomüberstand enthaltenen MAK. Niedrigere Kompetitor zu MAK Verhältnisse wurden dadurch erzielt, daß die Kompetitorlösung entsprechend dem gewünschten molaren Überschuß (siehe Anlage 1b) in der jeweiligen Salzlösung verdünnt wurde.

Anlage 1d

Quelle und relevante physicochemische Parameter der eingesetzten Metallionen

Von folgenden Metallionen wurden 10m molare Lösungen in bidest hergestellt:

Manganchlorid (?)	MG 161,88	
Fa. Merck	Nr. 5934	Ionenradius Mn: 80 pm
Cadmiumsulfate	MG 256,5	
Riedel de Haen	Nr. 31145	Ionenradius Cd: 97 pm
Zinkchlorid	MG 136,28	
Fa. Merck	Nr. 8816	Ionenradius Zn: 74 pm
Kupfersulfat	Mg 159,61	
Fa. Riedel de Haen	Nr. 31294	Ionenradius Cu: 96 pm
Yttriumchlorid	MG 303,36	
Fa. Aldrich	Nr. 20,491-9	Ionenradius Y: 92 pm
Blei(II)-nitrat	MG 331,20	
Fa. Riedel de Haen	Nr. 31137	Ionenradius Pb: 120 pm

Anlage 1e

Quantitativer Inhibitionstest von MAK durch DTPA bzw. EDTA

molarer Überschuß an Kompetitor der zu 50 % Inhibition der Bindung an das Festphasenantigen führt.

MAK-Nr.	DTPA-Y	DTPA	DTPA-Mn	DTPA-Cd	DTPA-Zn	DTPA-Cu
2050/174	10^4	10^3	10^2	10^2	5×10^3	5×10^3
2050/531	5×10^4	10^3	10^2	10^2	5×10^3	5×10^3
2050/532	5×10^4	10^3	10^2	10^2	5×10^3	5×10^3
2050/534	5×10^4	10^3	10^2	10^2	5×10^3	5×10^3
2050/535	10^4	10^2	10^2	10^2	10^3	10^3

MAK-Nr.	DTPA-Pb	1.2.Diaminoethan	Transaconitsäure	EDTA-Y	EDTA	EDTA-Mn
2050/174	10^3	keine Inhibition bis 10^5	keine Inhibition bis 10^5	10^2	10^3	10^3
2050/531	5×10^3	"	"	10^3	10^3	10^3
2050/532	5×10^3	"	"	10^2	10^3	10^3
2050/534	5×10^3	"	"	10^2	10^3	10^2
2050/535	10^3	"	"	10^2	10^2	10^2

MAK-Nr.	EDTA-Cd	EDTA-Zn	EDTA-Cu	EDTA-Pb
2050/174	10^3	10^3	10^3	5×10^3
2050/531	10^3	10^3	10^2	10^5
2050/532	10^3	10^3	5×10^3	10^5
2050/534	10^2	10^3	10^3	5×10^3
2050/535	10^2	10^2	10^2	10^2

Tab. 1

Peptidlinker mit dazugehörigen Nukleotidsequenzen

E P K S C G G E A A P A
 + 5' CTCTCTGCAGAGCCCAAATCTTGTGGCGGCGAGGCAGCTCCCGCAG
 - 3' GAGAGACGTCTCGGGTTTAGAACACCGCCGCTCCGTCGAGGGCGTC
 A A P A A A A G G Q V Q L Q E S
 CTGCACCCGCGAGCAGCCGCGAGGCGGGCAGGTCCAAGTGCAGGAGAGC 3'
 GACGTGGGCGTCGTCGGCGTCGTCGCCCGTCCAGGTTGACGTCCTCTCG 5'

+ = kodierender Strang
 - = komplementärer Strang

EP 0 404 097 B1

Tab. 2 MAK A

VH

5
Q V Q L Q E S G G G L V Q P G G S L R L
CAGGTCCAAC TGCAGGAGTCTGGAGGAGGCTTGGTACAGCCTGGGGGTTCTCTGAGACTC
10 20 30 40 50 60

10
S C A T S G F S D Y Y M N W V R Q P P G
TCCTGCGCAACTTCTGGGTTTCAGTGATTACTACATGAAGTGGTCCGCCAGCCTCCAGGA
70 80 90 100 110 120

15
K A L E W L G F I S N K P N G H T T E Y
AAAGCACTTGAGTGGTGGGTTTTATTTCAAACAAACCTAATGGTCACACAACAGAGTAC
130 140 150 160 170 180

20
S A S V K G R F T I S R D N S Q S I L Y
AGTGCATCTGTGAAGGGTCGGTTCACCATCTCCAGAGATAATTCCCAAAGCATCCTCTAT
190 200 210 220 230 240

25
L Q M N T L R A E D S A T Y Y C A R D K
CTTCAAATGAACACCCTGAGAGCTGAGGACAGTGCCACTTATTATTGTGCAAGAGATAAG
250 260 270 280 290 300

30
G I R W Y F D V W G Q G T T V T V S S
GGAATACGATGGTACTTCGATGTCTGGGGCCAAGGGACCACGGTCACCGTCTCCTCA
310 320 330 340 350

VK

35
A I L S A S P G E K V T M T C R A S S S S
AGCAATCCTGTCTGCATCTCCAGGGGAGAAGGTCACAATGACTTGCAGGGCCAGCTCAAG
10 20 30 40 50 60

40
V S Y M H W Y Q Q K P G S S P K P W I Y
TGTAAGTTACATGCACTGGTACCAGCAGAAGCCAGGATCCTCCCCCAAACCCTGGATTTA
70 80 90 100 110 120

45
A T S N L A S G V P A R F S G S G S G T
TGCCACATCCAACCTGGCTTCTGGAGTCCCTGCTCGCTTCAGTGGCAGTGGGTCTGGGAC
130 140 150 160 170 180

50
S Y S L T I I R V E A E D A A T Y Y C Q
CTCTTACTCTCTCACAATCATCAGAGTGGAGGCTGAAGATGCTGCCACTTATTACTGCCA
190 200 210 220 230 240

55
Q W S S N P L T F G A G T K L E I
GCAGTGGAGTAGTAACCCGCTCACGTTTCGGTGCTGGGACCAAGCTGGAGATC
250 260 270 280 290

EP 0 404 097 B1

Tab. 3 MAK B
VH

5 L Q E S G P D L V K P S Q S L S L T C T
CTGCAGGAGTCAGGACCTGACCTGGTGAAACCTTCTCAGTCACTTTCACTCACCTGCACT
10 10 20 30 40 50 60

10 V T G Y S I T S G Y S W H W I R Q F P G
GTCAGTGGCTACTCCATCACCAGTGGTTATAGCTGGCACTGGATCCGGCAGTTTCCAGGA
70 80 90 100 110 120

15 N K L E W M G Y I Q Y S G I T N Y N P S
AACAACTGGAATGGATGGGCTACATACAGTACAGTGGTATCACTAACTACAACCCCTCT
130 140 150 160 170 180

20 L K S R I S I T R D T S K N Q F F L Q L
CTCAAAAGTCGAATCTCTATCACTCGAGACACATCCAAGAACCAGTTCTTCCTGCAGTTG
190 200 210 220 230 240

25 N S V T T E D T A T Y Y C A R E D Y D Y
AATTCAGTGACTACTGAGGACACAGCCACATATTACTGTGCAAGAGAAGACTATGATTAC
250 260 270 280 290 300

30 H W Y F D V W G A G T T V T V S S
CACTGGTACTTCGATGTCTGGGGCGCAGGGACCACGGTCACCGTCTCCTCA
310 320 330 340 350

VK

35 L T Q S P A I M S A S L G E E I T L T C
CTGACCCAGTCTCCAGCAATCATGTCTGCATCTCTAGGGGAGGAGATCACCCCTAACCTGC
10 20 30 40 50 60

40 S T S S S V S Y M H W Y Q Q K S G T S P
AGTACCAGCTCGAGTGTAAGTTACATGCACTGGTACCAGCAGAAGTCAGGCACTTCTCCC
70 80 90 100 110 120

45 K L L I Y S T S N L A S G V P S R F S G
AAACTCTTGATTTATAGCACATCCAACCTGGCTTCTGGAGTCCCTTCTCGCTTCAGTGGC
130 140 150 160 170 180

50 S G S G T F Y S L T I S S V E A E D A A
AGTGGGTCTGGGACCTTTTATTCTCTCACAATCAGCAGTGTGGAGGCTGAAGATGCTGCC
190 200 210 220 230 240

55 D Y Y C H Q W S S Y P T F G G G T K L E
GATTATTACTGCCATCAGTGGAGTAGTTATCCCACGTTTCGGAGGGGGACCAAGCTGGAG
250 260 270 280 290 300

EP 0 404 097 B1

Tab. 4 MAK C

VH

5 Q V Q L Q Q S G P E L V K P G A S V K M
CAGGTCCAAGTGCAGCAGTCTGGACCTGAGCTGGTAAAGCCTGGGGCTTCAGTGAAGATG
10 20 30 40 50 60

10 S C K A S G Y T F T Y Y V I H W V K Q K
TCCTGCAAGGCTTCTGGATACACATTCACTTACTATGTTATTCACTGGGTGAAACAGAAG
70 80 90 100 110 120

15 P G Q G L E W I G Y I H P Y N A G T E Y
CCTGGGCAGGGCCTTGAGTGGATTGGATACATTATCCTTACAATGCTGGTACTGAGTAC
130 140 150 160 170 180

20 N E K F K G K A T L T S D K S S S T A Y
AATGAGAAGTTCAAAGGCAAGGCCACACTGACTTCAGACAAATCCTCCAGCACAGCCTAC
190 200 210 220 230 240

25 M E L S S L T S E D S A V Y Y C S M G R
ATGGAGCTCAGCAGCCTGACCTCTGAGGACTCTGCGGTCTATTACTGTTCAATGGGACGA
250 260 270 280 290 300

G G D Y W G Q G T T V T V S S
GGGGGTGACTACTGGGGCCAAGGACCACGGTCACCGTCTCCTCA
310 320 330 340

VK

35 L T Q S P A I M S A S P G E K V T M T C
CTGACCCAGTCTCCAGCAATTATGTCTGCATCTCCTGGGGAGAAGGTCACCATGACCTGC
10 20 30 40 50 60

40 S A S S S V S Y M H W Y Q Q K S G T S P
AGTGCCAGCTCAAGTGTAAGTTACATGCACTGGTACCAGCAGAAGTCAGGCACCTCCCC
70 80 90 100 110 120

45 K R W I Y D T S K L A S G V P A R F S G
AAAAGATGGATTTATGACACATCCAACTGGCTTCTGGAGTCCCTGCTCGCTTCAGTGGC
130 140 150 160 170 180

50 S G S G T S Y S L T I S S M E A E D A A
AGTGGGTCTGGGACCTCTTACTCTCTCACAATCAGCAGCATGGAGGCTGAAGATGCTGCC
190 200 210 220 230 240

55 T Y Y C Q Q W S S N P F T F G A G T K L
ACTTATTACTGCCAGCAGTGGAGTAGTAACCCATTACGTTCCGGCGCGGGGACCAAGCTG
250 260 270 280 290 300

EP 0 404 097 B1

Tab. 5 MAK D
VH

5 A E S G P G L V R L T S L S I T C T V S
GCAGAGTCAGGGCCTGGCCTGGTGCGCCTCACGAGCCTGTCCATCACTTGCACTGTCTCT
10 10 20 30 40 50 60

G F S L I S Y G V H W V R Q P P G K G L
GGCTTTTCATTAAATTAGTTATGGTGTACACTGGGTTGCGCAGCCTCCAGGAAAGGGTCTG
10 70 80 90 100 110 120

E W L G V I W A G G S T N Y N S A L M S
GAGTGGCTGGGAGTAATATGGGAGGTGGAAGCACAAATTATAATTCGGCTCTCATGTCC
15 130 140 150 160 170 180

R L S I S K D N S K S Q V F L K M N S L
AGACTGAGCATCAGCAAAGACAACCTCCAAGAGCCAAGTTTCTTAAAAATGAACAGTCTG
20 190 200 210 220 230 240

Q T G D T A I Y Y C A R G G D D Y D G F
CAAACGGTGACACAGCCATATACTACTGTGCCAGAGGGGGGATGATTACGATGGGTTT
25 250 260 270 280 290 300

A Y W G Q G T T V T V S S G E S
GCTTACTGGGGCCAAGGGACCGGTACCGTCTCCTCAGGTGAGTCC
30 310 320 330 340

VK

35 L T Q S P S S L A V S A G E K V T M S C
CTGACCCAGTCTCCATCCTCCCTGGCTGTGTGAGCAGGAGAGAAGGTCACTATGAGCTGC
10 20 30 40 50 60

K S S Q S L L S S T K R K N Y L A W Y Q
AAATCCAGTCAGAGTCTGCTCAGCAGTACAAAGCGAAAGAACTACTTGGCTTGGTACCAG
40 70 80 90 100 110 120

Q K P G Q S P K L L I Y W A S T R E S G
CAGAAACCAGGTCACTCTCCTAACTACTGATCTACTGGGCATCCACTCGGGAATCTGGG
130 140 150 160 170 180

Y P D R F T G S G S G T D F T L T I S S
GTCCCTGATCGCTTCACAGGCAGTGGATCTGGGACAGATTTCACTCTCACCATCAGCAGT
45 190 200 210 220 230 240

V Q A E D L A V Y Y C K Q S Y N L R A F
GTGCAGGCTGAAGACCTGGCAGTTTATTACTGCAAACATCTTATAATCTTCGGGCGTTTC
50 250 260 270 280 290 300

G G G T K L E I K
GGTGGAGGGACCAAGCTGGAGATCAAA
310 320

Tab. 6

Mutagene Oligonukleotide:

1) 5'CTTACCTGGG.CATGCCCCGA.GCTCCCGTGG.GCATGT3'

2) 5'AGTGGGGTTT.TCAGCTCTGCAGAG3'

Tabelle 7

Mutiertes Hinge Exon:

E L K T P L G D T T H T C
AGAGCTCAAA.ACCCCACTTG.GTGACACAAC.TCACACATGC

↓
G

P R C P
CCACGGTGCCCAGGT
GAGCTCGGGGCA
Sst1 *Sph1*

Tab. 8

Jun I Oligonukleotid

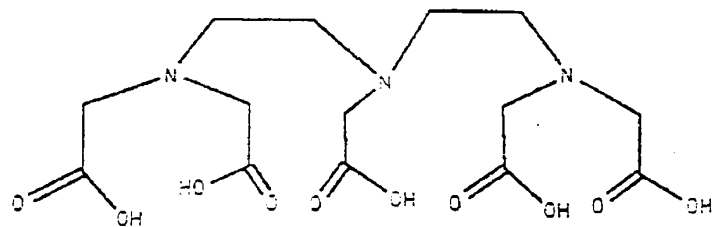
5' CTACGCTCGG. CTAGAGGAAA. AAGTGAAAAC.
CTTGAAAGCG. CAAAACCTCCG. AGCTGGCATC.
CACGGCCAAC. ATGCTCAGGG. AACAGGTGGC.
ACAGCTTAAG. CAGAAAGTCA. TGAACCACCG.
ACCTGCATB3'.

Jun II Oligonukleotid

5' CAGGTCGGTG. GTTCATGACT. TTCTGCTTAA.
GCTGTGCCAC. CTGTTCCCTG. AGCATGTTGG.
CCGTGGATGC. CAGCTCGGAG. TTTTGCGCTT.
TCAAGGTTTT. CACTTTTTCC. TCTAGCCGAG.
CGATGAGCT3'.

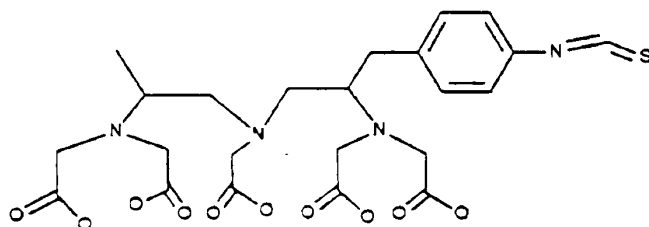
Formel 1

DTPA



Formel 2

Isothiocyanatobenzyl-DTPA

**Patentansprüche**

Patentansprüche für folgende Vertragsstaaten : AT, BE, CH, DE, DK, FR, GB, IT, LI, LU, NL, SE

1. Bispezifische oder oligospezifische mono- oder oligovalente Rezeptoren, gentechnisch erhältlich durch Fusion von für V_H und C_H1 Regionen der Antikörper zweier oder mehrerer verschiedener Spezifitäten kodierende DNA mittels geeigneter Linker und anschließende Expression zusammen mit den dazugehörigen Genabschnitten für die leichten Ketten in Expressionssystemen, dadurch gekennzeichnet, daß eine Spezifität über fos-jun Interaktion gegen ein Komplexon gerichtet ist.
2. Rezeptoren nach Anspruch 1, dadurch gekennzeichnet, daß die andere Spezifität gegen animale oder humane tumorassoziierte Antigene gerichtet ist.
3. Rezeptoren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die andere Spezifität von den monoklonalen Antikörpern mit den variablen Regionen gemäß Tab. 2, 3, 4 oder 5 stammt.
4. Rezeptoren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß bei zwei Spezifitäten drei Bindungsstellen vorhanden sind.
5. Rezeptoren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß bei zwei Spezifitäten 4 Bindungsstellen vorhanden sind.
6. Rezeptoren nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß bei drei Spezifitäten eine gegen Tumore gerichtet ist und die beiden anderen in unterschiedlicher Weise gegen DTPA oder EDTA gerichtet sind.
7. Verfahren zur Herstellung von Rezeptoren nach einem der Ansprüche 1 bis 6, dadurch gekennzeichnet, daß die für die schweren Ketten-Antikörperteile kodierenden DNA-Fragmente mittels geeigneter Linker verbunden und in einem Expressionssystem zusammen mit den Genen für die leichten Ketten exprimiert werden.
8. Arzneimittel enthaltend einen Rezeptor nach einem der Ansprüche 1 bis 6.

Patentansprüche für folgende Vertragsstaaten : ES, GR

1. Verfahren zur Herstellung von bispezifischen oder oligospezifischen mono- oder oligovalenten Rezeptoren wobei eine Spezifität über fos-jun Interaktion gegen ein Komplexon gerichtet ist, dadurch gekennzeichnet, daß für V_H und

C_H1 Regionen der Antikörper zweier oder mehrerer verschiedener Spezifitäten kodierende DNA mittels geeigneter Linker fusioniert und anschließend zusammen mit den dazugehörigen Genabschnitten für die leichten Ketten exprimiert wird.

- 5 2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß die andere Spezifität gegen animale oder humane tumorassoziierte Antigene gerichtet ist.
3. Verfahren nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die andere Spezifität von den monoklonalen Antikörpern mit den variablen Regionen gemäß Tab. 2, 3, 4 oder 5 stammt.
- 10 4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß bei zwei Spezifitäten drei Bindungsstellen vorhanden sind.
5. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß bei zwei Spezifitäten 4 Bindungsstellen vorhanden sind.
- 15 6. Verfahren nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß bei drei Spezifitäten eine gegen Tumore gerichtet ist und die beiden anderen in unterschiedlicher Weise gegen DTPA oder EDTA gerichtet sind.

20 Claims

Claims for the following Contracting States : AT, BE, CH, DE, DK, FR, GB, IT, LI, LU, NL, SE

- 25 1. A bispecific or oligospecific mono- or oligovalent receptor, obtainable by gene manipulation by fusion of DNA coding for V_H and C_H1 regions of the anti-bodies of two or more different specificities, by means of suitable linkers and subsequent expression together with the gene sections belonging thereto for the light chains in expression systems, wherein one specificity is directed against a complexone via fos-jun interaction.
- 30 2. The receptor as claimed in claim 1, wherein the other specificity is directed against animal or human tumor-associated antigens.
3. The receptor as claimed in claim 1 or 2, wherein the other specificity derives from the monoclonal anti-bodies with the variable regions shown in Tab. 2, 3, 4 or 5.
- 35 4. The receptor as claimed in any of claims 1 to 3, wherein three binding sites are present in the case of two specificities.
5. The receptor as claimed in any of claims 1 to 3, wherein 4 binding sites are present in the case of two specificities.
- 40 6. The receptor as claimed in any of claims 1 to 5, wherein, in the case of three specificities, one is directed against tumors and the two others are directed in a different manner against DTPA or EDTA.
7. A process for the preparation of the receptor as claimed in any of claims 1 to 6, which comprises the DNA fragments which code for the heavy chain anti-body portions being connected by means of suitable linkers and expressed in an expression system together with the genes for the light chains.
- 45 8. A pharmaceutical comprising the receptor as claimed in any of claims 1 to 6.

Claims for the following Contracting States : ES, GR

- 50 1. A process for the preparation of a bispecific or oligospecific mono- or oligovalent receptor, one specificity being directed against a complexone via fos-jun interaction, which comprises DNA coding for V_H and C_H1 regions of the antibodies of two or more different specificities being fused by means of suitable linkers and subsequently expressed together with the gene sections belonging thereto for the light chains.
- 55 2. The process as claimed in claim 1, wherein the other specificity is directed against animal or human tumor-associated antigens.

3. The process as claimed in claim 1 or 2, wherein the other specificity derives from the monoclonal anti-bodies with the variable regions shown in Tab. 2, 3, 4 or 5.
- 5 4. The process as claimed in any of claims 1 to 3, wherein three binding sites are present in the case of two specificities.
5. The process as claimed in any of claims 1 to 3, wherein 4 binding sites are present in the case of two specificities.
- 10 6. The process as claimed in any of claims 1 to 5, wherein, in the case of three specificities, one is directed against tumors and the two others are directed in a different manner against DTPA or EDTA.

Revendications

Revendications pour les Etats contractants suivants : AT, BE, CH, DE, DK, FR, GB, IT, LI, LU, NL, SE

- 15 1. Récepteurs mono- ou oligovalents, bispécifiques ou oligospécifiques, pouvant être obtenus, par génie génétique, par fusion d'ADN codant pour les régions V_H et C_H1 des anticorps à deux spécificités différentes ou plus, au moyen de lieurs appropriés, et expression subséquente avec les segments de gènes correspondants codant pour les chaînes légères, dans des systèmes d'expression, caractérisés en ce qu'une spécificité est dirigée contre un complexe par l'intermédiaire d'interaction *fos-jun*.
- 20 2. Récepteurs selon la revendication 1, caractérisés en ce que l'autre spécificité est dirigée contre des anti-gènes humains ou animaux associés à des tumeurs.
- 25 3. Récepteurs selon la revendication 1 ou 2, caractérisés en ce que l'autre spécificité dérive des anticorps monoclonaux comportant les régions variables selon le tableau 2, 3, 4 ou 5.
4. Récepteurs selon l'une des revendications 1 à 3, caractérisés en ce que, dans le cas de deux spécificités, trois sites de liaison sont présents.
- 30 5. Récepteurs selon l'une des revendications 1 à 3, caractérisés en ce que, dans le cas de deux spécificités, quatre sites de liaison sont présents.
6. Récepteurs selon l'une des revendications 1 à 5, caractérisés en ce que, dans le cas de trois spécificités, l'une est dirigé contre des tumeurs et les deux autres sont dirigées, de façon différente, contre le DPTA ou l'EDTA.
- 35 7. Procédé pour la production de récepteurs selon l'une des revendications 1 à 6, caractérisé en ce que les fragments d'ADN codant pour les segments d'anticorps de type chaîne lourde sont assemblés au moyen de lieurs appropriés et exprimés dans un système d'expression, conjointement avec les gènes codant pour les chaînes légères.
- 40 8. Médicament contenant un récepteur selon l'une des revendications 1 à 6.

Revendications pour les Etats contractants suivants : ES, GR

- 45 1. Procédé pour la production de récepteurs mono- ou oligovalents, bispécifiques ou oligospécifiques, une spécificité étant dirigée contre un complexe, par l'intermédiaire d'interaction *fos-jun*, caractérisé en ce que de l'ADN codant pour les régions V_H et C_H1 des anticorps à deux spécificités différentes ou plus sont fusionnés au moyen de lieurs appropriés et ensuite exprimés conjointement avec les segments de gènes correspondants, codant pour les chaînes légères.
- 50 2. Procédé selon la revendication 1, caractérisé en ce que l'autre spécificité est dirigée contre des antigènes humains ou animaux associés à des tumeurs.
3. Procédé selon la revendication 1 ou 2, caractérisé en ce que l'autre spécificité dérive des anticorps monoclonaux comportant les régions variables selon le tableau 2, 3, 4 ou 5.
- 55 4. Procédé selon l'une des revendications 1 à 3, caractérisé en ce que, dans le cas de deux spécificités, trois sites de liaison sont présents.

EP 0 404 097 B1

5. Procédé selon l'une des revendications 1 à 3, caractérisé en ce que, dans le cas de deux spécificités, quatre sites de liaison sont présents.
6. Procédé selon l'une des revendications 1 à 5, caractérisé en ce que, dans le cas de trois spécificités, l'une est dirigé contre des tumeurs et les deux autres sont dirigées, de façon différente, contre le DPTA ou l'EDTA.

5

10

15

20

25

30

35

40

45

50

55

FIG. 1

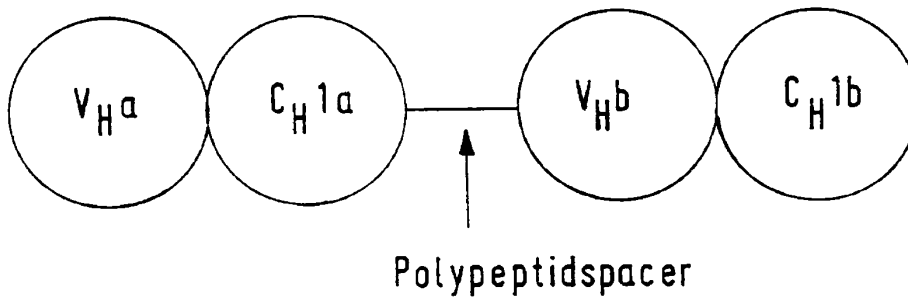
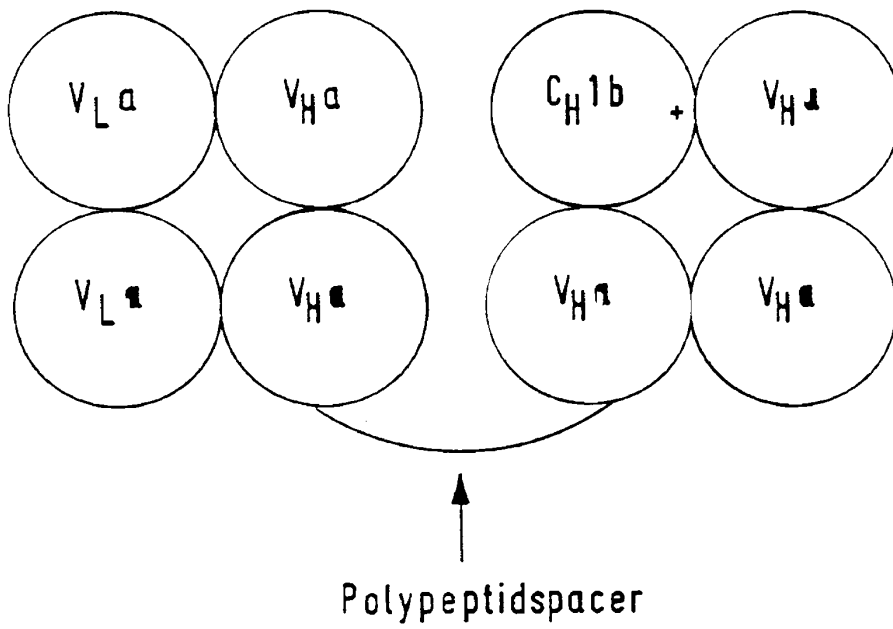
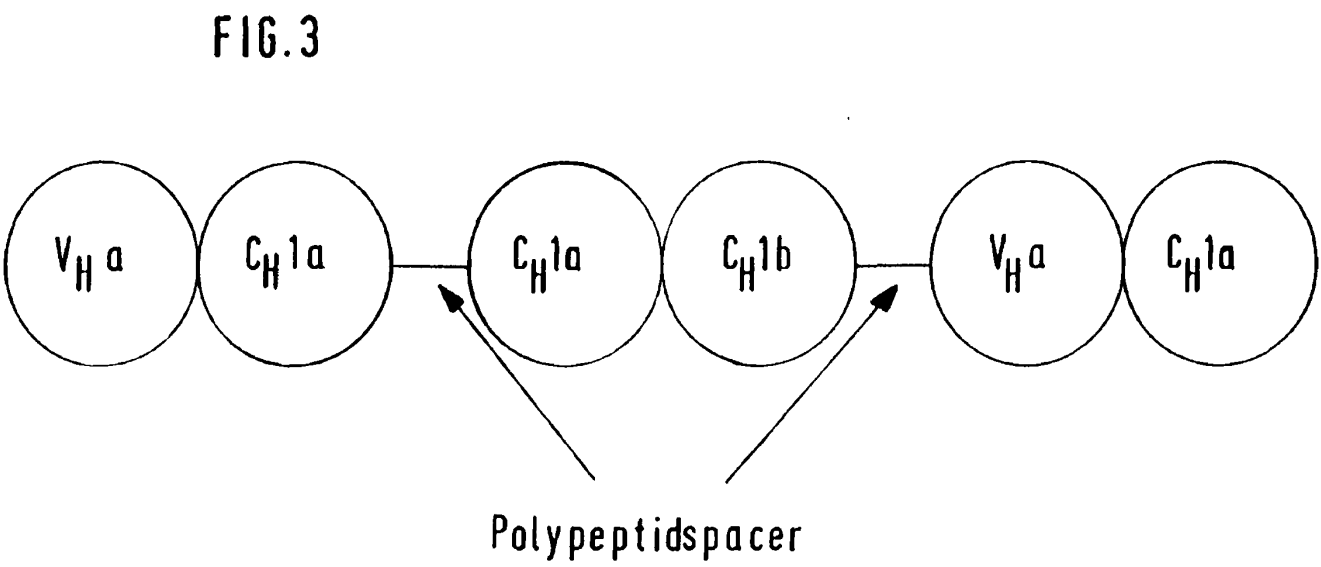
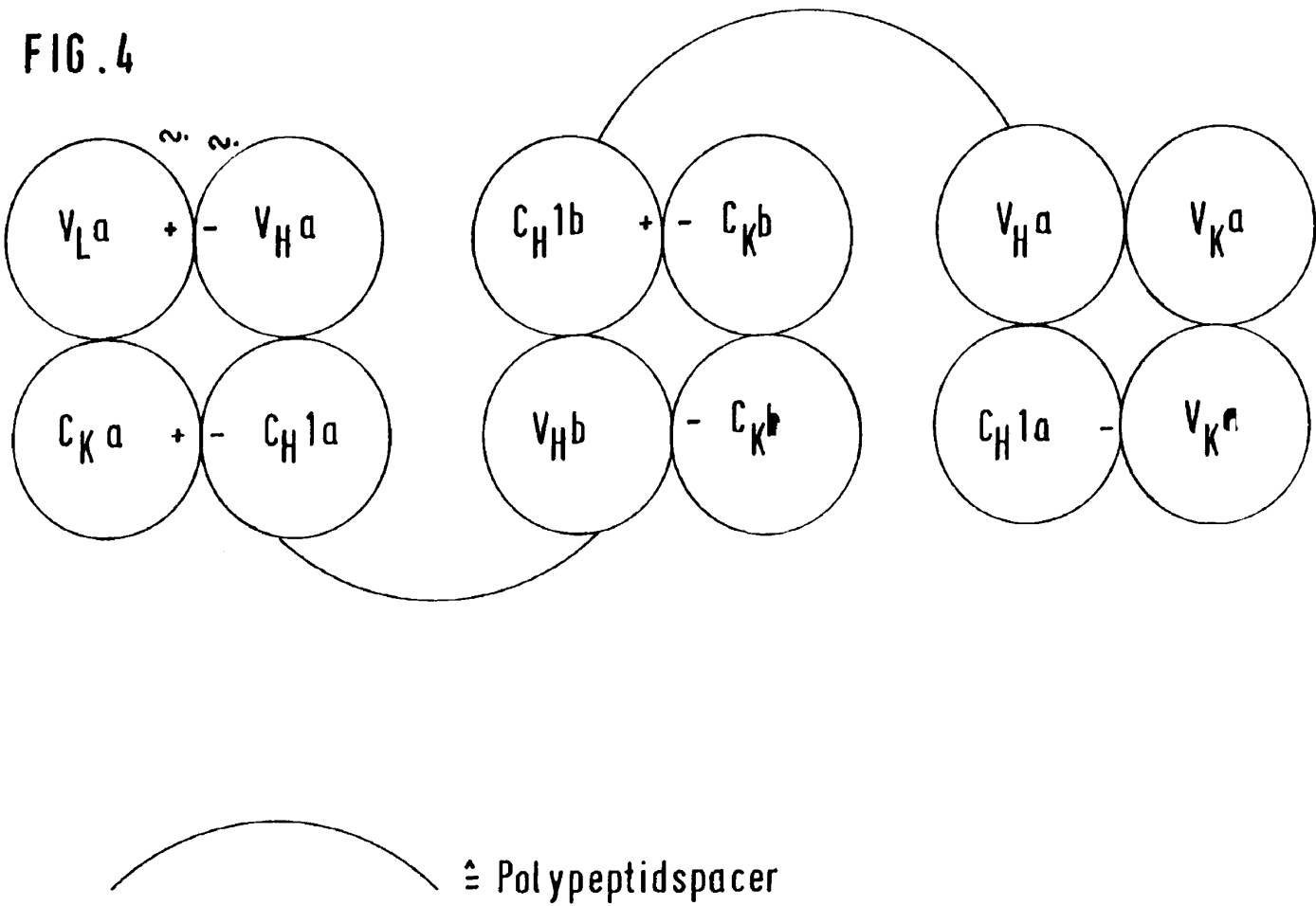
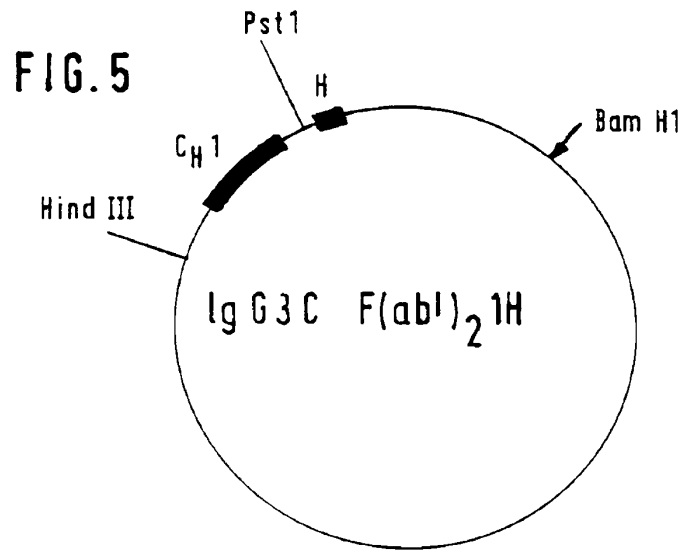
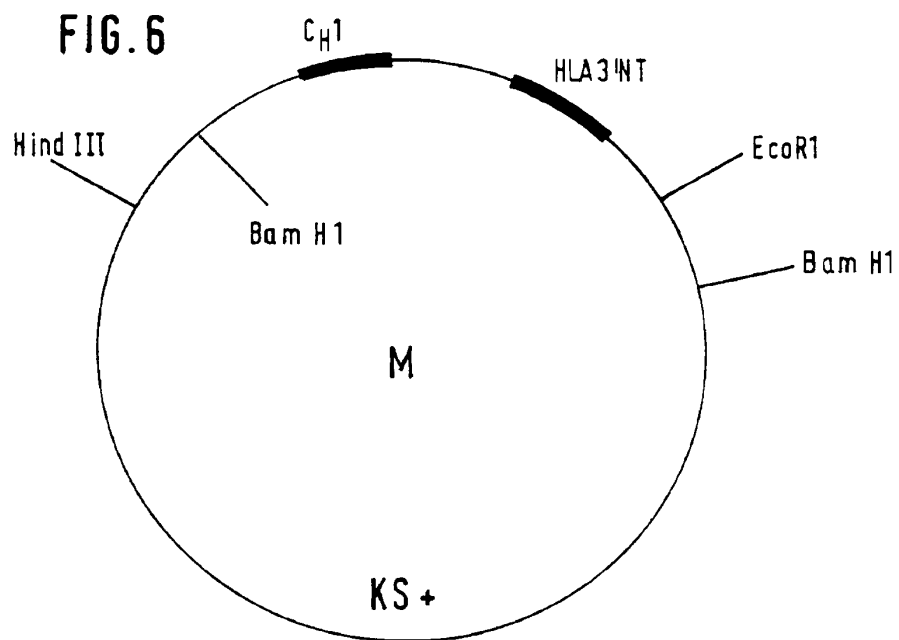


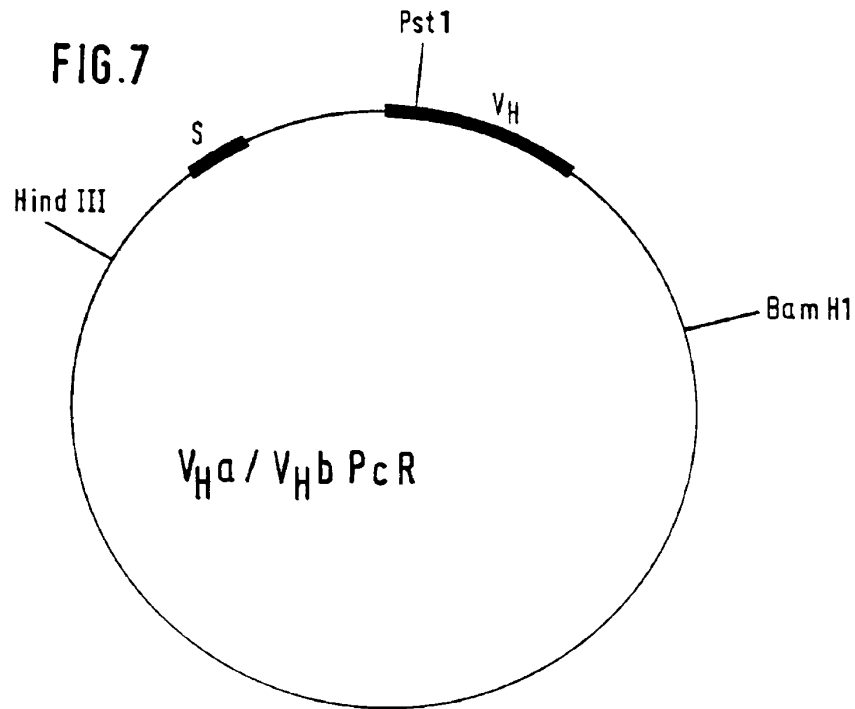
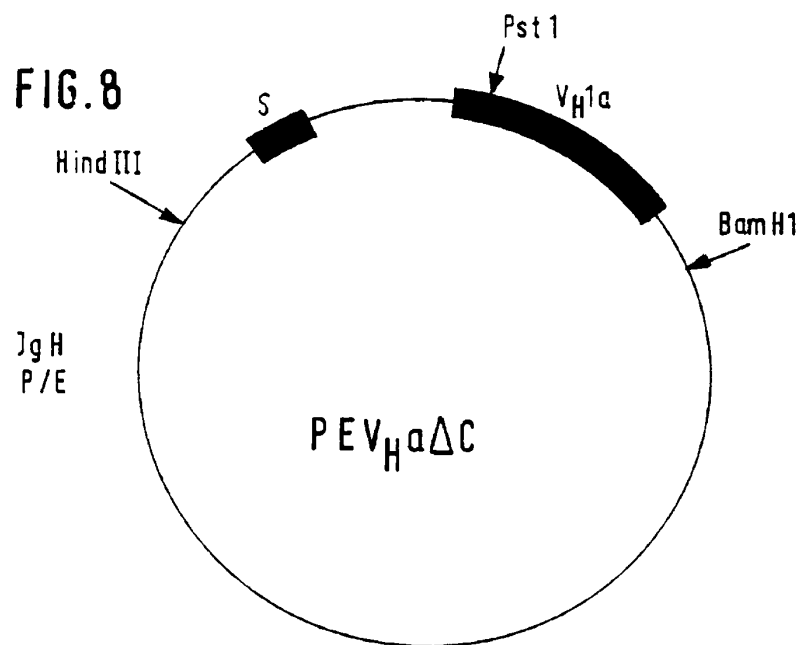
FIG. 2











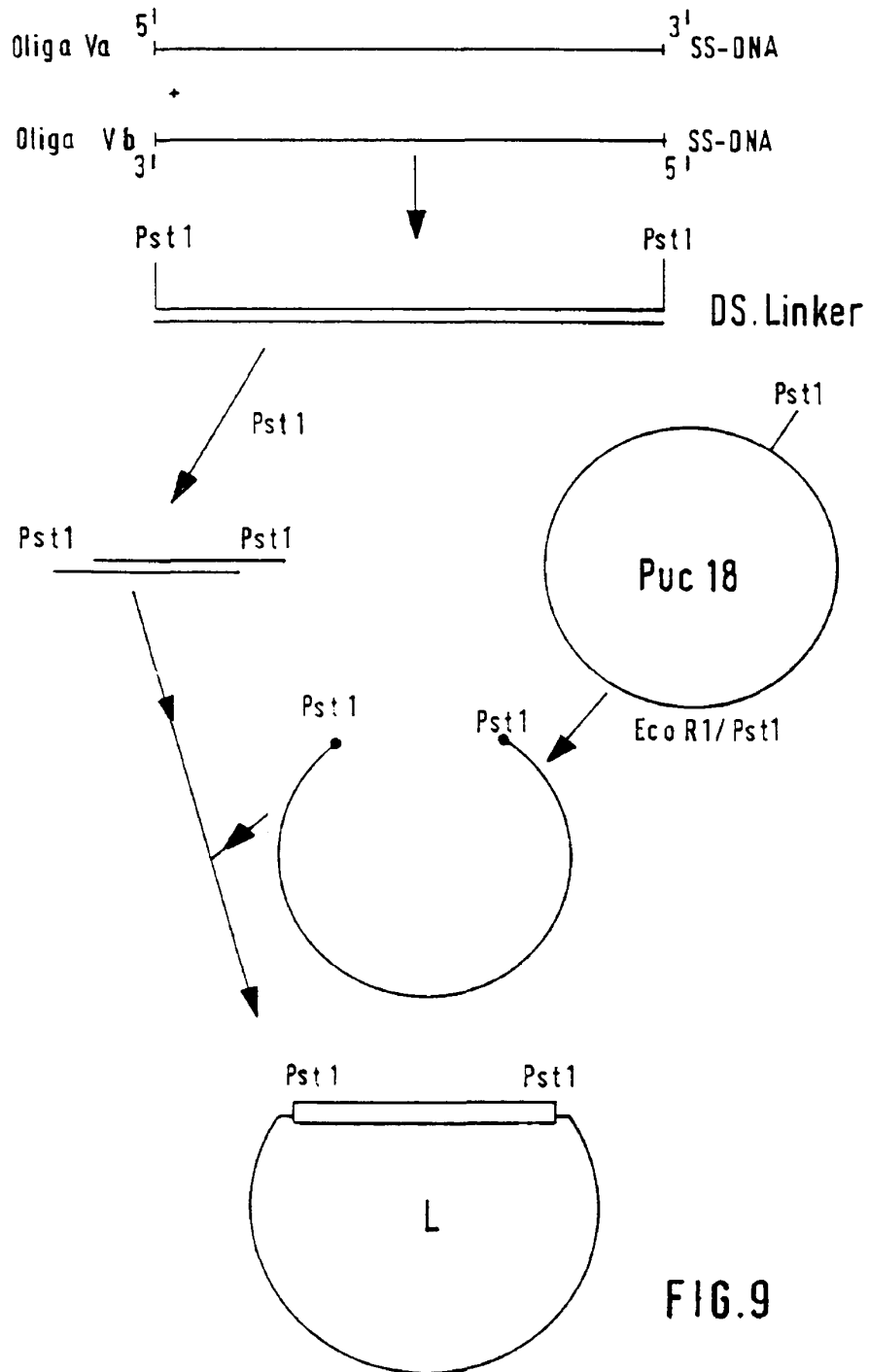


FIG.9

FIG. 10

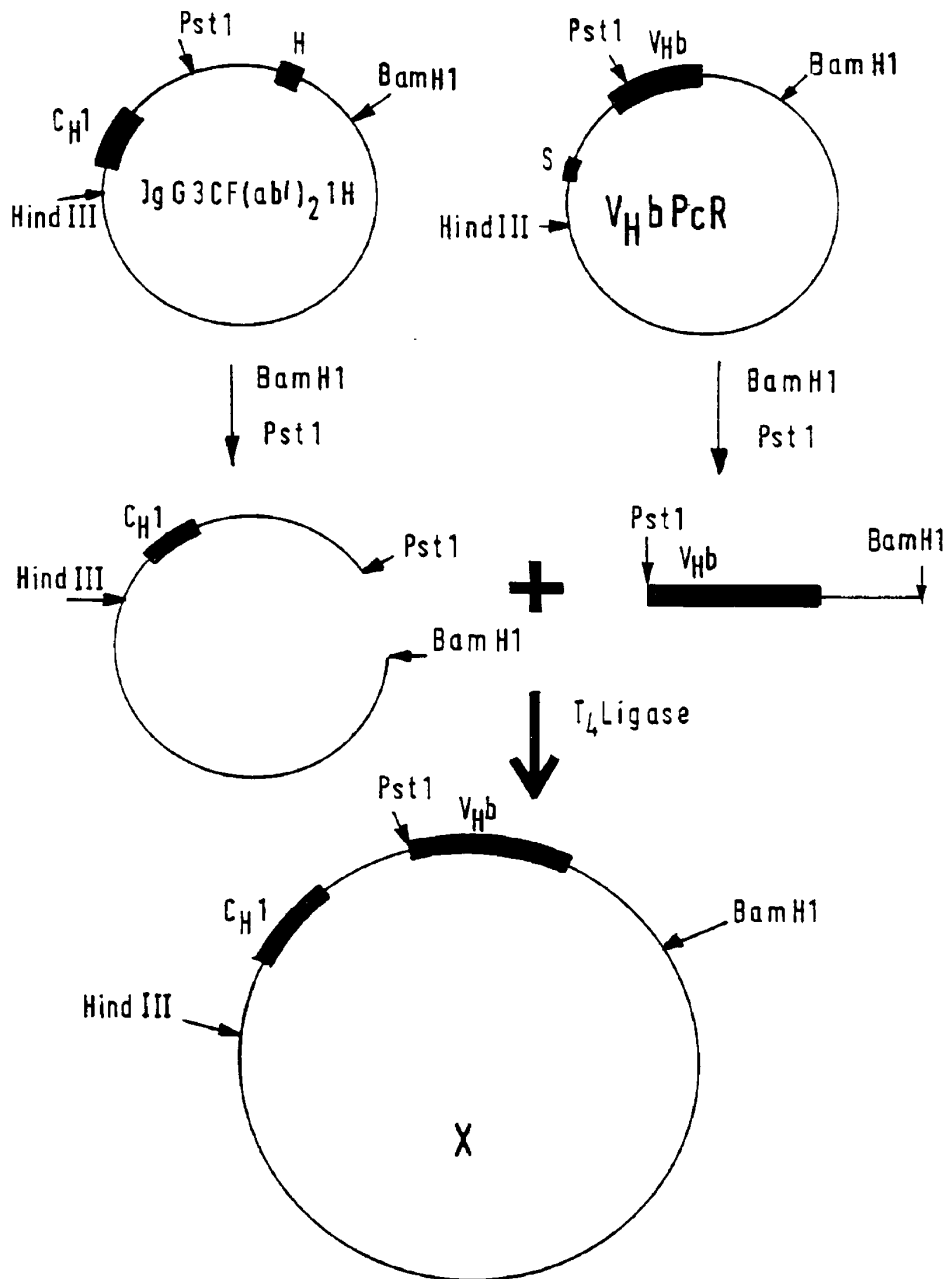


FIG. 11

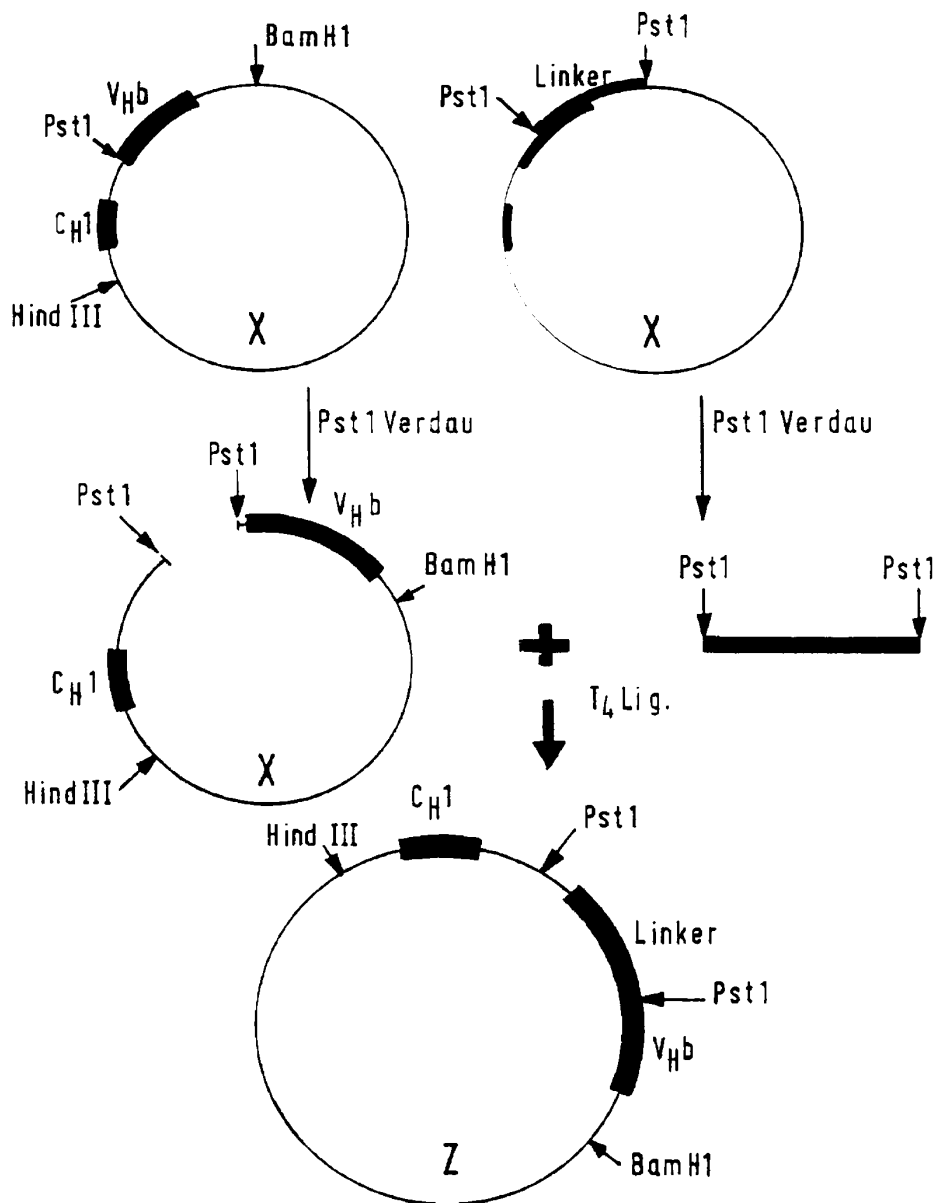


FIG. 12

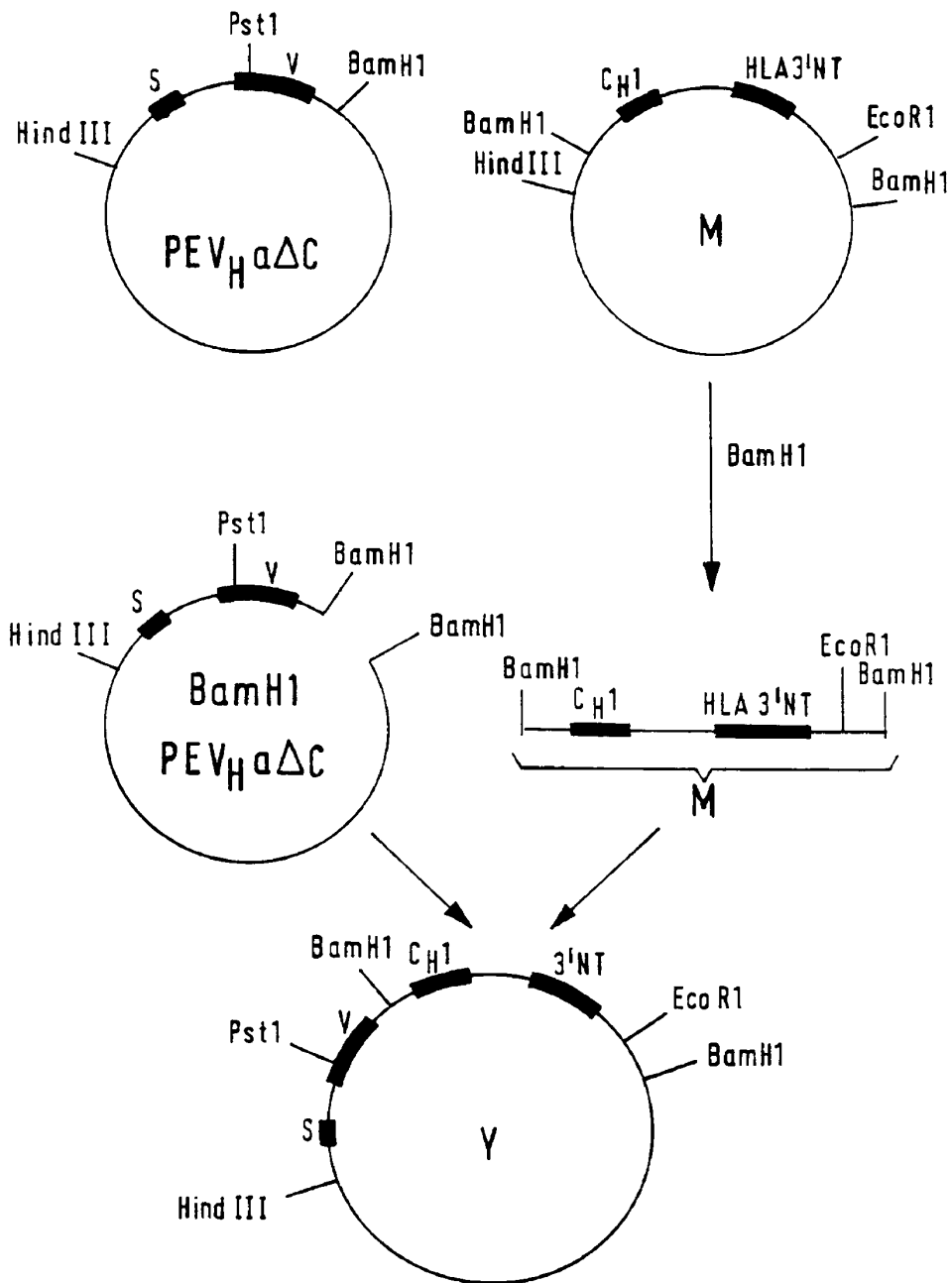


FIG. 13

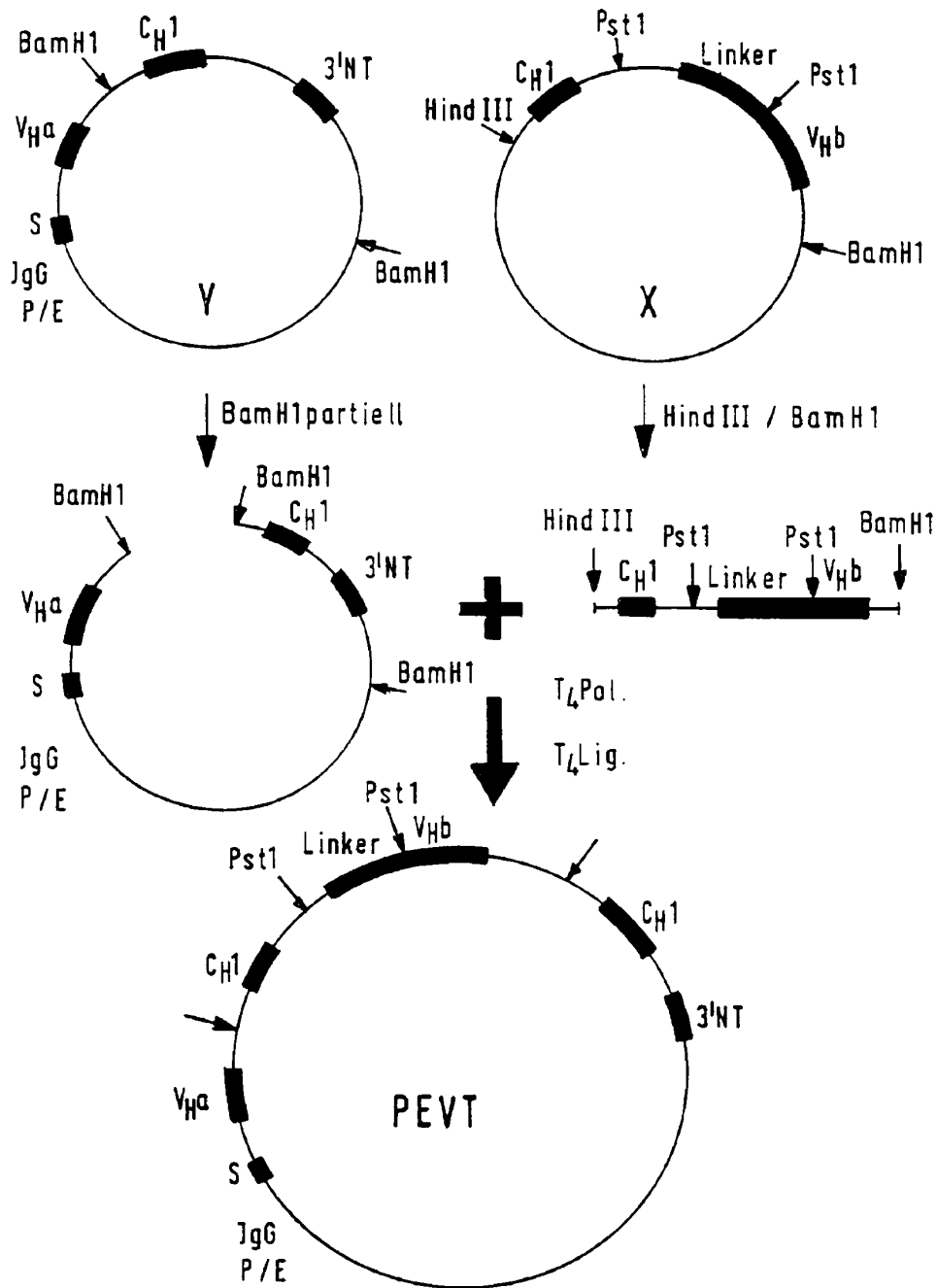


FIG.14

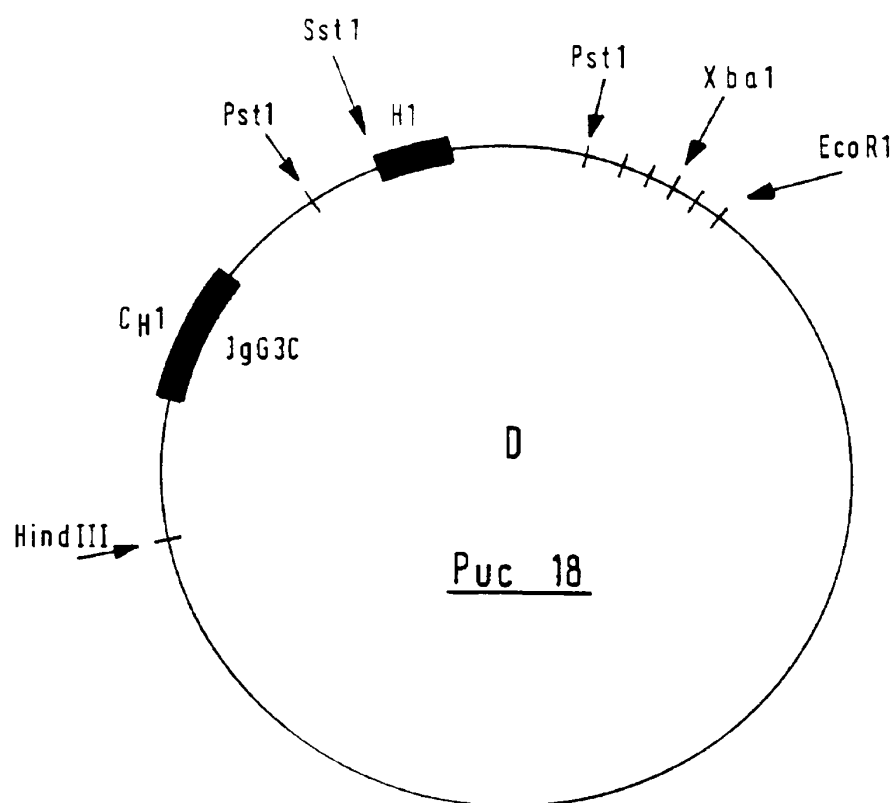


FIG. 15

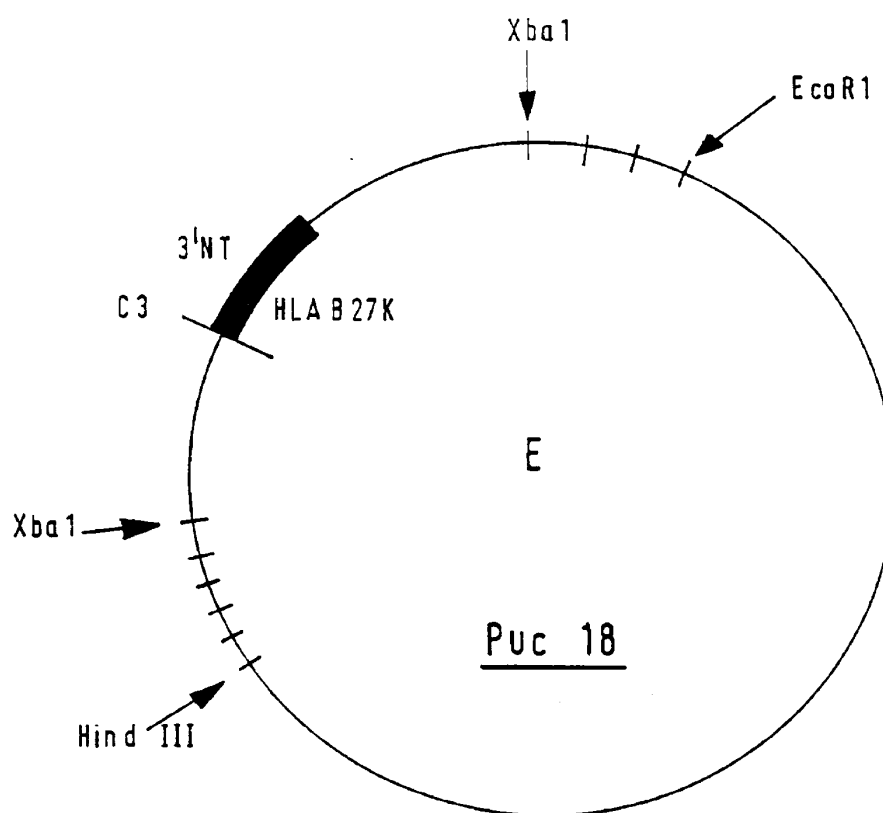


FIG. 16

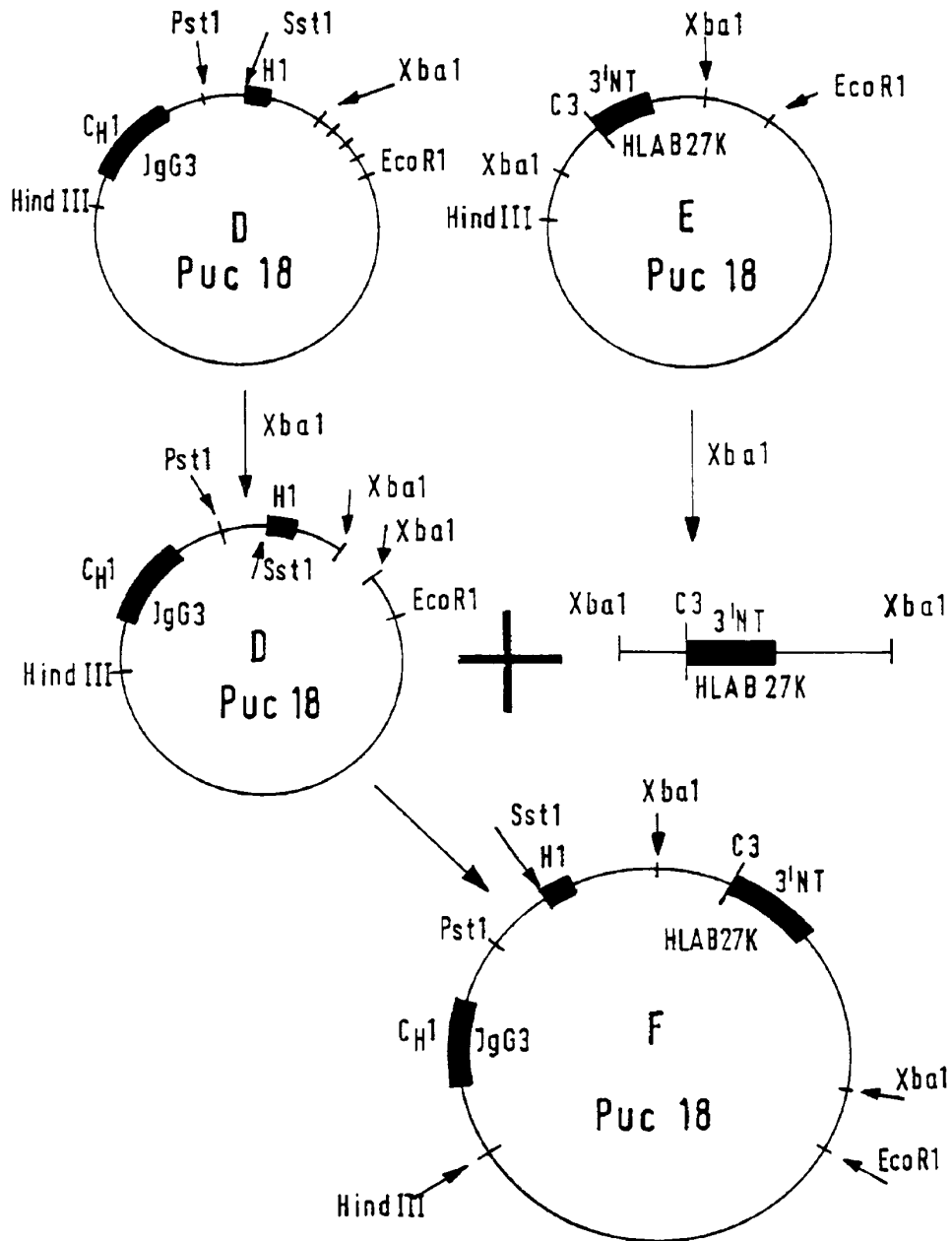


FIG. 17

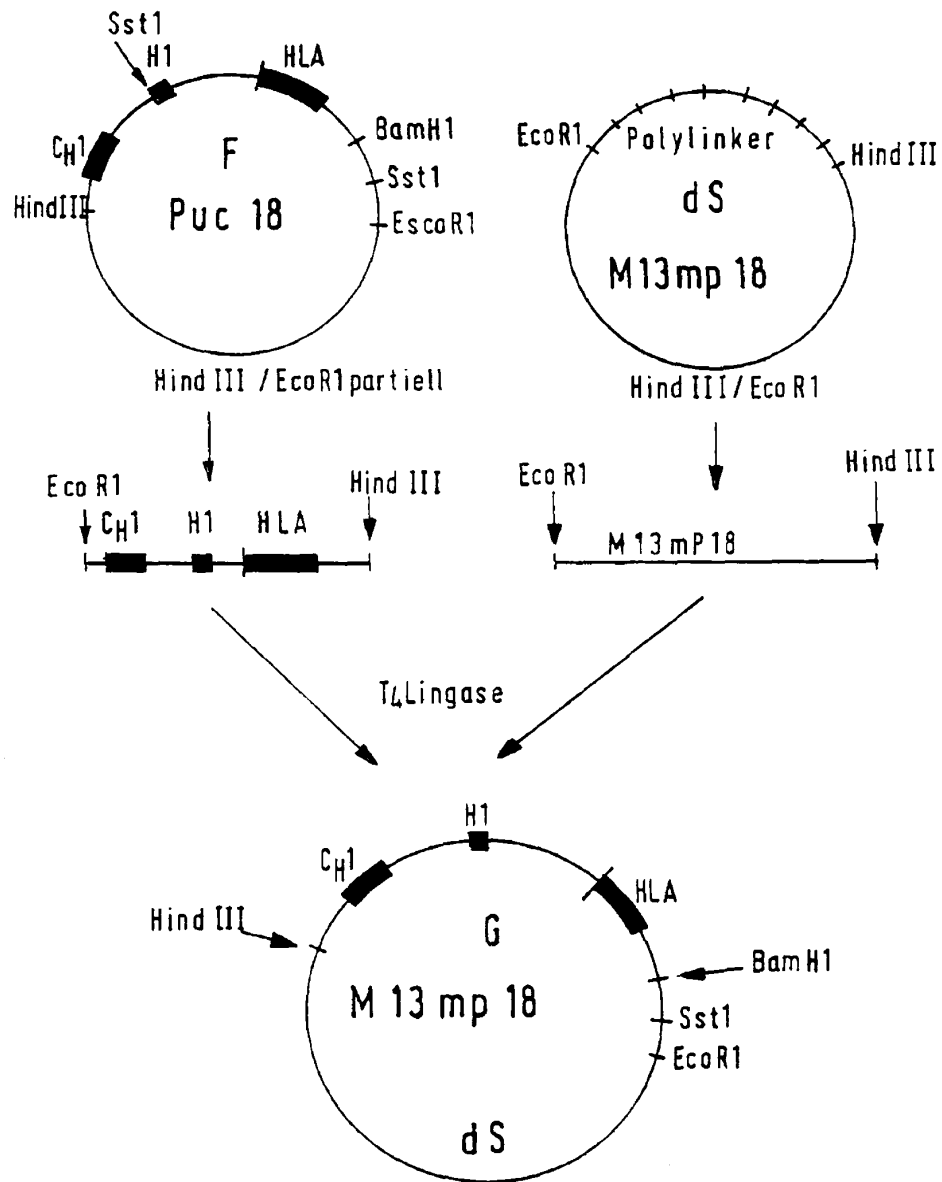


FIG. 18

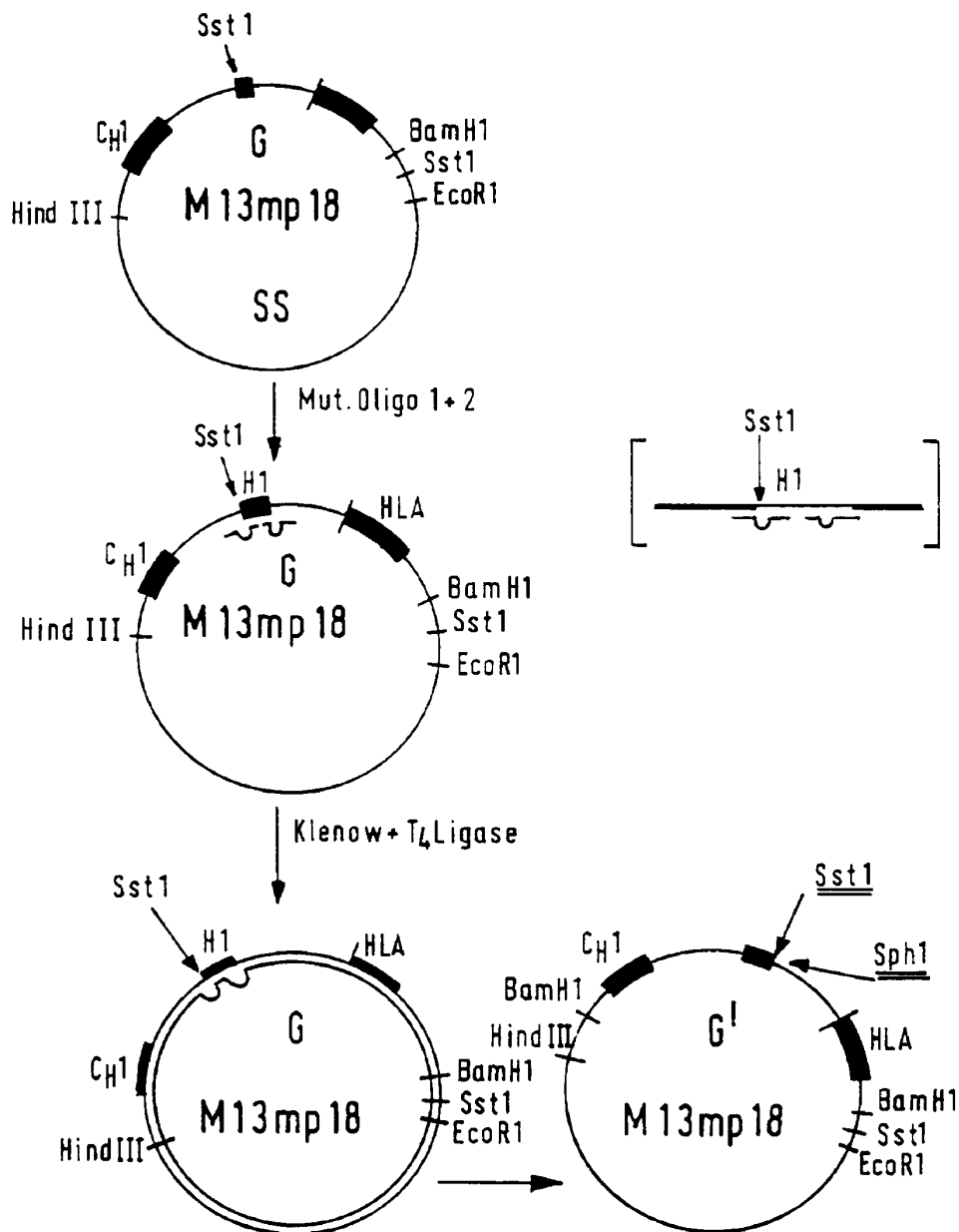


FIG. 19

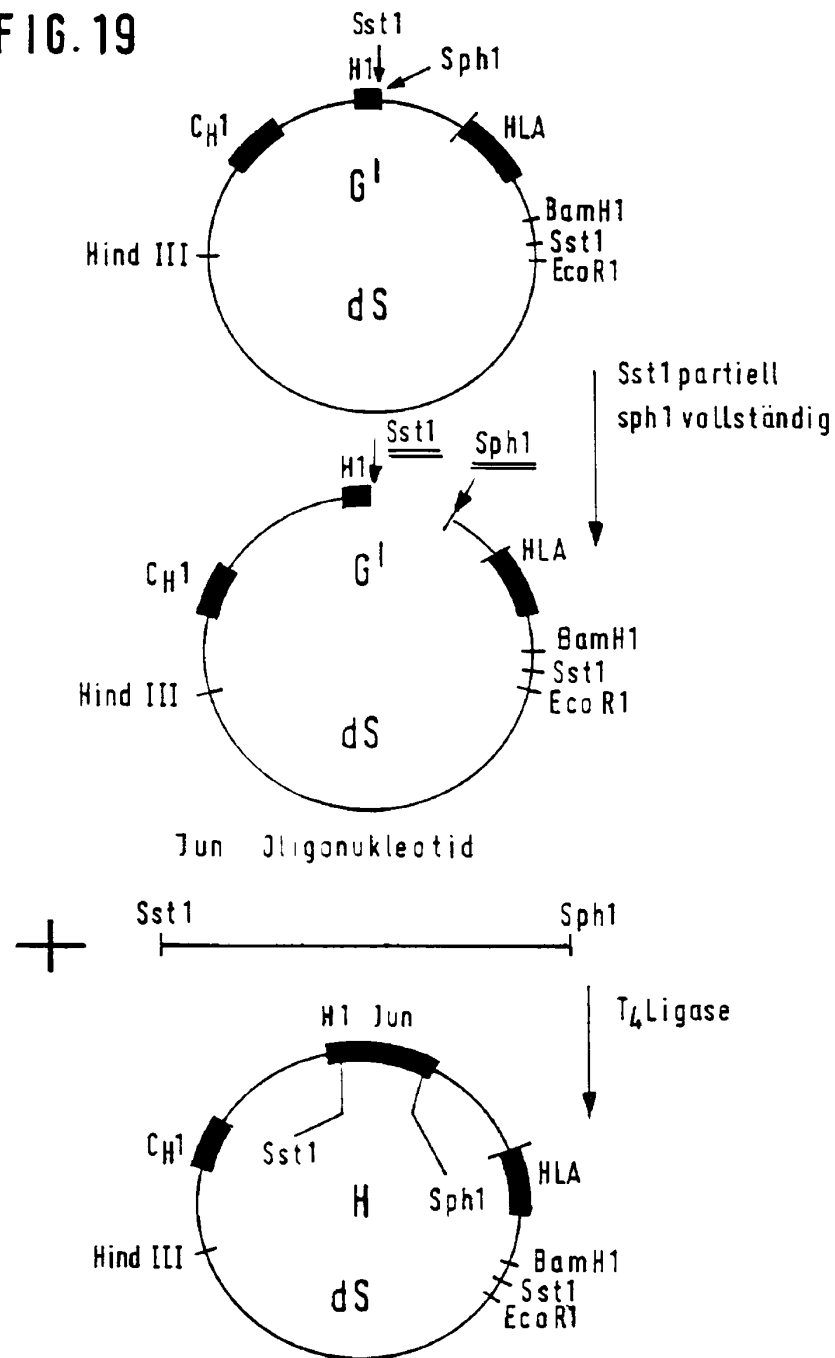


FIG. 20

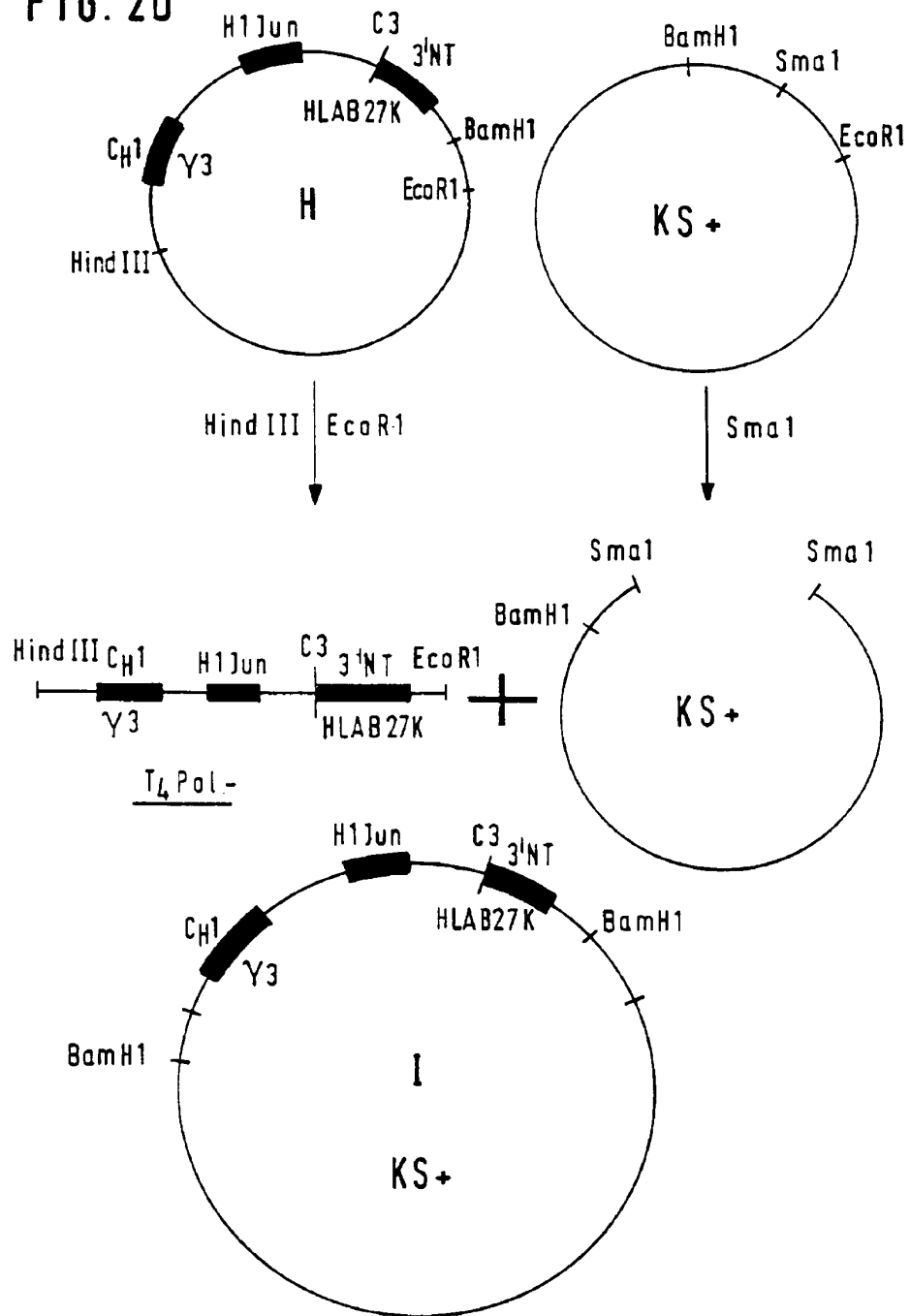
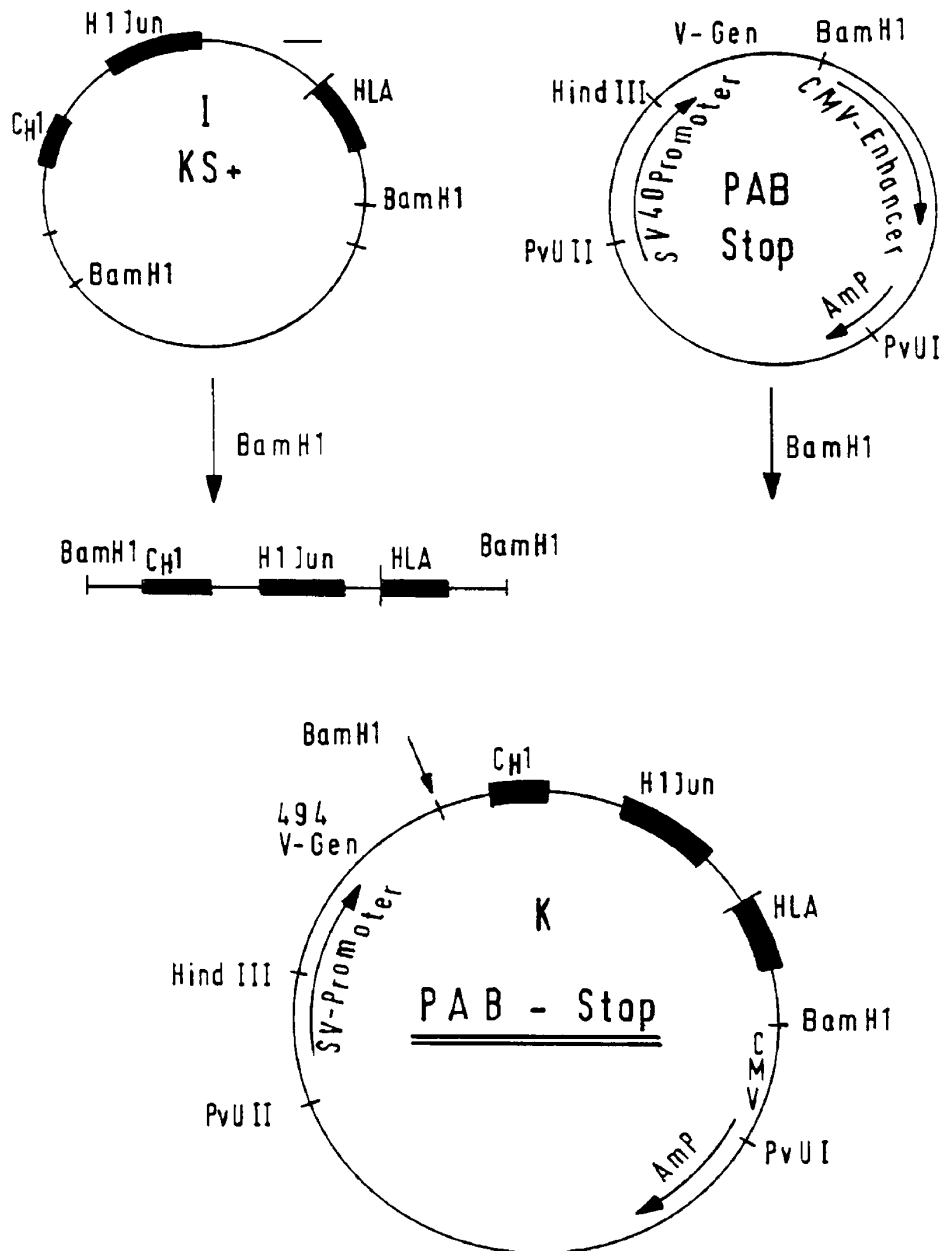


FIG. 21





INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C07K 15/28, 3/20, C07H 21/04 C12P 21/08, C12N 15/00, 15/03	A1	(11) International Publication Number: WO 93/11161 (43) International Publication Date: 10 June 1993 (10.06.93)
(21) International Application Number: PCT/US92/09965 (22) International Filing Date: 20 November 1992 (20.11.92) (30) Priority data: 07/796,936 25 November 1991 (25.11.91) US (71) Applicant: ENZON, INC. [US/US]; 40 Kingsbridge Road, Piscataway, NJ 08854 (US). (72) Inventors: WHITLOW, Marc, D. ; 18727 Cross Country Lane, Gaithersburg, MD 20879 (US). WOOD, James, F. ; 18223 Smoke House Court, Germantown, MD 20874 (US). HARDMAN, Karl ; 1354 Indian Creek Drive, Wynnewood, PA 19096 (US). BIRD, Robert, E. ; 5206 Russett Road, Rockville, MD 20853 (US). FILPULA, David ; 17022 King James Way, Gaithersburg, MD 20877 (US). ROLLENCE, Michele ; 11 Valley Park Court, Damascus, MD 20872 (US).		(74) Agents: GOLDSTEIN, Jorge, A. et al.; Sterne, Kessler, Goldstein & Fox, 1225 Connecticut Avenue, N.W., Suite 300, Washington, DC 20036 (US). (81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE). Published <i>With international search report.</i>
(54) Title: MULTIVALENT ANTIGEN-BINDING PROTEINS (57) Abstract Compositions of, genetic constructions coding for, and methods for producing multivalent antigen-binding proteins are described and claimed. The methods include purification of compositions containing both monomeric and multivalent forms of single polypeptide chain molecules, and production of multivalent proteins from purified monomers. Production of multivalent proteins may occur by a concentration-dependent association of monomeric proteins, or by rearrangement of regions involving dissociation followed by reassociation of different regions. Bivalent proteins, including homobivalent and heterobivalent proteins, are made in the present invention. Genetic sequences coding for bivalent single-chain antigen-binding proteins are disclosed. Uses include all those appropriate for monoclonal and polyclonal antibodies and fragments thereof, including use as a bispecific antigen-binding molecule.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LI	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Monaco	TG	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

Multivalent Antigen-Binding Proteins

This invention was made with Government Support under SBIR Grant 5R44 GM 39662-03 awarded by the National Institutes of Health, National Institute of General Medical Sciences. The Government has certain rights in the invention.

Cross-Reference to Related Applications

This application is a continuation-in-part of U.S. Patent Application Serial Number 07/796,936, filed Nov. 25, 1991, which is a continuation-in-part of U.S. Patent Application Serial No. 07/512,910 filed April 25, 1990, which is a continuation-in-part of Serial No. 07/299,617, filed Jan. 1, 1989, issued as U.S. Patent No. 4,946,778 (Ladner *et al.*), which was a continuation-in-part of Serial No. 092,110, filed Sept. 2, 1987, and Serial No. 902,971, filed Sept. 2, 1986, now abandoned, the contents of all of which are fully incorporated herein by reference.

Background of the Invention

1. Field of the Invention

The present invention relates generally to the production of antigen-binding molecules. More specifically, the invention relates to multivalent forms of antigen-binding proteins. Compositions of, genetic constructions for, methods of use, and methods for producing these multivalent antigen-binding proteins are disclosed.

- 2 -

2. *Description of the Background Art*

Antibodies are proteins generated by the immune system to provide a specific molecule capable of complexing with an invading molecule, termed an antigen. Figure 14 shows the structure of a typical antibody molecule. Natural antibodies have two identical antigen-binding sites, both of which are specific to a particular antigen. The antibody molecule "recognizes" the antigen by complexing its antigen-binding sites with areas of the antigen termed epitopes. The epitopes fit into the conformational architecture of the antigen-binding sites of the antibody, enabling the antibody to bind to the antigen.

The antibody molecule is composed of two identical heavy and two identical light polypeptide chains, held together by interchain disulfide bonds (see Fig. 14). The remainder of this discussion will refer only to one light/heavy pair of chains, as each light/heavy pair is identical. Each individual light and heavy chain folds into regions of approximately 110 amino acids, assuming a conserved three-dimensional conformation. The light chain comprises one variable region (termed V_L) and one constant region (C_L), while the heavy chain comprises one variable region (V_H) and three constant regions (C_{H1} , C_{H2} and C_{H3}). Pairs of regions associate to form discrete structures as shown in Figure 14. In particular, the light and heavy chain variable regions, V_L and V_H , associate to form an " F_V " area which contains the antigen-binding site.

The variable regions of both heavy and light chains show considerable variability in structure and amino acid composition from one antibody molecule to another, whereas the constant regions show little variability. The term "variable" as used in this specification refers to the diverse nature of the amino acid sequences of the antibody heavy and light chain variable regions. Each antibody recognizes and binds antigen through the binding site defined by the association of the heavy and light chain variable regions into an F_V area. The light-chain variable region V_L and the heavy-chain variable region V_H of a particular antibody molecule have specific amino acid sequences that

- 3 -

allow the antigen-binding site to assume a conformation that binds to the antigen epitope recognized by that particular antibody.

Within the variable regions are found regions in which the amino acid sequence is extremely variable from one antibody to another. Three of these so-called "hypervariable" regions or "complementarity-determining regions" (CDR's) are found in each of the light and heavy chains. The three CDR's from a light chain and the three CDR's from a corresponding heavy chain form the antigen-binding site.

Cleavage of the naturally-occurring antibody molecule with the proteolytic enzyme papain generates fragments which retain their antigen-binding site. These fragments, commonly known as Fab's (for Fragment, antigen binding site) are composed of the C_L , V_L , C_H1 and V_H regions of the antibody. In the Fab the light chain and the fragment of the heavy chain are covalently linked by a disulfide linkage.

Recent advances in immunobiology, recombinant DNA technology, and computer science have allowed the creation of single polypeptide chain molecules that bind antigen. These single-chain antigen-binding molecules incorporate a linker polypeptide to bridge the individual variable regions, V_L and V_H , into a single polypeptide chain. A computer-assisted method for linker design is described more particularly in U.S. Patent No. 4,704,692, issued to Ladner *et al.* in November, 1987, and incorporated herein by reference. A description of the theory and production of single-chain antigen-binding proteins is found in U.S. Patent No. 4,946,778 (Ladner *et al.*), issued August 7, 1990, and incorporated herein by reference. The single-chain antigen-binding proteins produced under the process recited in U.S. Patent 4,946,778 have binding specificity and affinity substantially similar to that of the corresponding Fab fragment.

Bifunctional, or bispecific, antibodies have antigen binding sites of different specificities. Bispecific antibodies have been generated to deliver cells, cytotoxins, or drugs to specific sites. An important use has been to deliver host cytotoxic cells, such as natural killer or cytotoxic T cells, to specific cellular targets. (U.D. Staerz, O. Kanagawa, M.J. Bevan, *Nature*

- 4 -

314:628 (1985); S. Songilvilai, P.J. Lachmann, *Clin. Exp. Immunol.* **79**: 315 (1990)). Another important use has been to deliver cytotoxic proteins to specific cellular targets. (V. Raso, T. Griffin, *Cancer Res.* **41**:2073 (1981); S. Honda, Y. Ichimori, S. Iwasa, *Cytotechnology* **4**:59 (1990)). Another
5 important use has been to deliver anti-cancer non-protein drugs to specific cellular targets (J. Corvalan, W. Smith, V. Gore, *Intl. J. Cancer Suppl.* **2**:22 (1988); M. Pimm *et al.*, *British J. of Cancer* **61**:508 (1990)). Such bispecific antibodies have been prepared by chemical cross-linking (M. Brennan *et al.*, *Science* **229**:81 (1985)), disulfide exchange, or the production of hybrid-
10 hybridomas (quadromas). Quadromas are constructed by fusing hybridomas that secrete two different types of antibodies against two different antigens (Kurokawa, T. *et al.*, *Biotechnology* **7**:1163 (1989)).

Summary of the Invention

This invention relates to the discovery that multivalent forms of single-
15 chain antigen-binding proteins have significant utility beyond that of the monovalent single-chain antigen-binding proteins. A multivalent antigen-binding protein has more than one antigen-binding site. Enhanced binding activity, di- and multi-specific binding, and other novel uses of multivalent antigen-binding proteins have been demonstrated or are envisioned here.
20 Accordingly, the invention is directed to multivalent forms of single-chain antigen-binding proteins, compositions of multivalent and single-chain antigen-binding proteins, methods of making and purifying multivalent forms of single-chain antigen-binding proteins, and uses for multivalent forms of single-chain antigen-binding proteins. The invention provides a multivalent antigen-binding
25 protein comprising two or more single-chain protein molecules, each single-chain molecule comprising a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; and a peptide linker linking the first and second
30 polypeptides into a single-chain protein.

- 5 -

Also provided is a composition comprising a multivalent antigen-binding protein substantially free of single-chain molecules.

Also provided is an aqueous composition comprising an excess of multivalent antigen-binding protein over single-chain molecules.

5 A method of producing a multivalent antigen-binding protein is provided, comprising the steps of producing a composition comprising multivalent antigen-binding protein and single-chain molecules, each single-chain molecule comprising a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; a second
10 polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; and a peptide linker linking the first and second polypeptides into a single-chain molecule; separating the multivalent protein from the single-chain molecules; and recovering the multivalent protein.

15 Also provided is a method of producing multivalent antigen-binding protein, comprising the steps of producing a composition comprising single-chain molecules as previously defined; dissociating the single-chain molecules; reassociating the single-chain molecules; separating the resulting multivalent antigen-binding proteins from the single-chain molecules; and recovering the multivalent proteins.

20 Also provided is another method of producing a multivalent antigen-binding protein, comprising the step of chemically cross-linking at least two single-chain antigen-binding molecules.

25 Also provided is another method of producing a multivalent antigen-binding protein, comprising the steps of producing a composition comprising single-chain molecules as previously defined; concentrating said single-chain molecules; separating said multivalent protein from said single-chain molecules; and finally recovering said multivalent protein.

30 Also provided is another method of producing a multivalent antigen-binding protein comprising two or more single-chain molecules, each single-chain molecule as previously defined, said method comprising: providing a genetic sequence coding for said single-chain molecule; transforming a host

- 6 -

cell or cells with said sequence; expressing said sequence in said host or hosts; and recovering said multivalent protein.

Another aspect of the invention includes a method of detecting an antigen in or suspected of being in a sample, which comprises contacting said sample with the multivalent antigen-binding protein of claim 1 and detecting whether said multivalent antigen-binding protein has bound to said antigen.

Another aspect of the invention includes a method of imaging the internal structure of an animal, comprising administering to said animal an effective amount of a labeled form of the multivalent antigen-binding protein of claim 1 and measuring detectable radiation associated with said animal.

Another aspect of the invention includes a composition comprising an association of a multivalent antigen-binding protein with a therapeutically or diagnostically effective agent.

Another aspect of this invention is a single-chain protein comprising: a first polypeptide comprising the binding portion of the variable region of an antibody light chain; a second polypeptide comprising the binding portion of the variable region of an antibody light chain; a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein.

Another aspect of the present invention includes the genetic constructions encoding the combinations of regions V_L-V_L and V_H-V_H for single-chain molecules, and encoding multivalent antigen-binding proteins.

Another part of this invention is a multivalent single-chain antigen-binding protein comprising: a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; a peptide linker linking said first and second polypeptides (a) and (b) into said multivalent protein; a third polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; a fourth polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; a peptide linker linking said third and fourth polypeptides (d) and (e) into said multivalent protein; and a peptide linker linking said second and third polypeptides (b) and (d) into said

- 7 -

multivalent protein. Also included are genetic constructions coding for this multivalent single-chain antigen-binding protein.

Also included are replicable cloning or expression vehicles including plasmids, hosts transformed with the aforementioned genetic sequences, and methods of producing multivalent proteins with the sequences, transformed hosts, and expression vehicles.

Methods of use are provided, such as a method of using the multivalent antigen-binding protein to diagnose a medical condition; a method of using the multivalent protein as a carrier to image the specific bodily organs of an animal; a therapeutic method of using the multivalent protein to treat a medical condition; and an immunotherapeutic method of conjugating a multivalent protein with a therapeutically or diagnostically effective agent. Also included are labelled multivalent proteins, improved immunoassays using them, and improved immunoaffinity purifications.

An advantage of using multivalent antigen-binding proteins instead of single-chain antigen-binding molecules or Fab fragments lies in the enhanced binding ability of the multivalent form. Enhanced binding occurs because the multivalent form has more binding sites per molecule. Another advantage of the present invention is the ability to use multivalent antigen-binding proteins as multi-specific binding molecules.

An advantage of using multivalent antigen-binding proteins instead of whole antibodies, is the enhanced clearing of the multivalent antigen-binding proteins from the serum due to their smaller size as compared to whole antibodies which may afford lower background in imaging applications. Multivalent antigen-binding proteins may penetrate solid tumors better than monoclonals, resulting in better tumor-fighting ability. Also, because they are smaller and lack the Fc component of intact antibodies, the multivalent antigen-binding proteins of the present invention may be less immunogenic than whole antibodies. The Fc component of whole antibodies also contains binding sites for liver, spleen and certain other cells and its absence should thus reduce accumulation in non-target tissues.

- 8 -

Another advantage of multivalent antigen-binding proteins is the ease with which they may be produced and engineered, as compared to the myeloma-fusing technique pioneered by Kohler and Milstein that is used to produce whole antibodies.

5

Brief Description of the Drawings.

The present invention as defined in the claims can be better understood with reference to the text and to the following drawings:

FIG. 1A is a schematic two-dimensional representation of two identical single-chain antigen-binding protein molecules, each comprising a variable light chain region (V_L), a variable heavy chain region (V_H), and a polypeptide linker joining the two regions. The single-chain antigen-binding protein molecules are shown binding antigen in their antigen-binding sites.

FIG. 1B depicts a hypothetical homodivalent antigen-binding protein formed by association of the polypeptide linkers of two monovalent single-chain antigen-binding proteins from Fig. 1A (the Association model). The divalent antigen-binding protein is formed by the concentration-driven association of two identical single-chain antigen-binding protein molecules.

FIG. 1C depicts the hypothetical divalent protein of FIG. 1B with bound antigen molecules occupying both antigen-binding sites.

FIG. 2A depicts the hypothetical homodivalent protein of Figure 1B.

FIG. 2B depicts three single-chain antigen-binding protein molecules associated in a hypothetical trimer.

FIG. 2C depicts a hypothetical tetramer of four single-chain antigen-binding protein molecules.

FIG. 3A depicts two separate and distinct monovalent single-chain antigen-binding proteins, Anti-A single-chain antigen-binding protein and Anti-B single-chain antigen-binding protein, with different antigen specificities, each individually binding either Antigen A or Antigen B.

- 9 -

FIG. 3B depicts a hypothetical bispecific heterodivalent antigen-binding protein formed from the single-chain antigen-binding proteins of Fig. 3A according to the Association model.

FIG. 3C depicts the hypothetical heterodivalent antigen-binding protein of FIG. 3B binding bispecifically, i.e., binding the two different antigens, A and B.

FIG. 4A depicts two identical single-chain antigen-binding protein molecules, each having a variable light chain region (V_L), a variable heavy chain region (V_H), and a polypeptide linker joining the two regions. The single-chain antigen-binding protein molecules are shown binding identical antigen molecules in their antigen-binding sites.

FIG. 4B depicts a hypothetical homodivalent protein formed by the rearrangement of the V_L and V_H regions shown in FIG. 4A (the Rearrangement model). Also shown is bound antigen.

FIG. 5A depicts two single-chain protein molecules, the first having an anti-B V_L and an anti-A V_H , and the second having an anti-A V_L and an anti-B V_H . The figure shows the non-complementary nature of the V_L and V_H regions in each single-chain protein molecule.

FIG. 5B shows a hypothetical bispecific heterodivalent antigen-binding protein formed by rearrangement of the two single-chain proteins of Figure 5A.

FIG. 5C depicts the hypothetical heterodivalent antigen-binding protein of FIG. 5B with different antigens A and B occupying their respective antigen-binding sites.

FIG. 6A is a schematic depiction of a hypothetical trivalent antigen-binding protein according to the Rearrangement model.

FIG. 6B is a schematic depiction of a hypothetical tetravalent antigen-binding protein according to the Rearrangement model.

FIG. 7 is a chromatogram depicting the separation of CC49/212 antigen-binding protein monomer from dimer on a cation exchange high performance liquid chromatographic column. The column is a PolyCAT A

- 10 -

aspartic acid column (Poly WC, Columbia, MD). Monomer is shown as Peak 1, eluting at 27.32 min., and dimer is shown as Peak 2, eluting at 55.52 min.

FIG. 8 is a chromatogram of the purified monomer from Fig. 7. Monomer elutes at 21.94 min., preceded by dimer (20.135 min.) and trimer (18.640 min.). Gel filtration column, Protein-Pak 300SW (Waters Associates, Milford, MA).

FIG. 9 is a similar chromatogram of purified dimer (20.14 min.) from Fig. 7, run on the gel filtration HPLC column of Fig. 8.

FIG. 10A is an amino acid (SEQ ID NO. 11) and nucleotide (SEQ ID NO. 10) sequence of the single-chain protein comprising the 4-4-20 V_L region connected through the 212 linker polypeptide to the CC49 V_H region.

FIG. 10B is an amino acid (SEQ ID NO. 13) and nucleotide (SEQ ID NO. 12) sequence of the single-chain protein comprising the CC49 V_L region connected through the 212 linker polypeptide to the 4-4-20 V_H region.

FIG. 11 is a chromatogram depicting the separation of the monomer (27.83 min.) and dimer (50.47 min.) forms of the CC49/212 antigen-binding protein by cation exchange, on a PolyCAT A cation exchange column (Poly LC, Columbia, MD).

Fig. 12 shows the separation of monomer (17.65 min.), dimer (15.79 min.), trimer (14.19 min.), and higher oligomers (shoulder at about 13.09 min.) of the B6.2/212 antigen-binding protein. This separation depicts the results of a 24-hour treatment of a 1.0 mg/ml B6.2/212 single-chain antigen-binding protein sample. A TSK G2000SW gel filtration HPLC column was used, Toyo Soda, Tokyo, Japan.

Fig. 13 shows the results of a 24-hour treatment of a 4.0 mg/ml CC49/212 antigen-binding protein sample, generating monomer, dimer, and trimer at 16.91, 14.9, and 13.42 min., respectively. The same TSK gel filtration column was used as in Fig. 12.

Fig. 14 shows a schematic view of the four-chain structure of a human IgG molecule.

- 11 -

Fig. 15A is an amino acid (SEQ ID NO. 15) and nucleotide (SEQ ID NO. 14) sequence of the 4-4-20/212 single-chain antigen-binding protein with a single cysteine hinge.

Fig. 15B is an amino acid (SEQ ID NO. 17) and nucleotide (SEQ ID NO. 16) sequence of the 4-4-20/212 single-chain antigen-binding protein with the two-cysteine hinge.

Fig. 16 shows the amino acid (SEQ ID NO. 19) and nucleotide (SEQ ID NO. 18) sequence of a divalent CC49/212 single-chain antigen-binding protein.

Fig. 17 shows the expression of the divalent CC49/212 single-chain antigen-binding protein of Fig. 16 at 42°C, on an SDS-PAGE gel containing total *E. coli* protein. Lane 1 contains the molecular weight standards. Lane 2 is the uninduced *E. coli* production strain grown at 30°C. Lane 3 is divalent CC49/212 single-chain antigen-binding protein induced by growth at 42°C. The arrow shows the band of expressed divalent CC49/212 single-chain antigen-binding protein.

Fig. 18 is a graphical representation of four competition radioimmunoassays (RIA) in which unlabeled CC49 IgG (open circles) CC49/212 single-chain antigen-binding protein (closed circles) and CC49/212 divalent antigen-binding protein (closed squares) and anti-fluorescein 4-4-20/212 single-chain antigen-binding protein (open squares) competed against a CC49 IgG radiolabeled with ¹²⁵I for binding to the TAG-72 antigen on a human breast carcinoma extract.

Figure 19A is an amino acid (SEQ ID NO. 21) and nucleotide (SEQ ID NO. 20) sequence of the single-chain polypeptide comprising the 4-4-20 V_L region connected through the 217 linker polypeptide to the CC49 V_H region.

Figure 19B is an amino acid (SEQ ID NO. 23) and nucleotide (SEQ ID NO. 22) sequence of the single-chain polypeptide comprising the CC49 V_L region connected through the 217 linker polypeptide to the 4-4-20 V_H region.

Figure 20 is a chromatogram depicting the purification of CC49/4-4-20 heterodimer Fv on a cation exchange high performance liquid chromatographic column. The column is a PolyCAT A aspartic acid column (Poly LC,

- 12 -

Columbia, MD). The heterodimer Fv is shown as peak 5, eluting at 30.10 min.

Figure 21 is a coomassie-blue stained 4-20% SDS-PAGE gel showing the proteins separated in Figure 20. Lane 1 contains the molecular weight standards. Lane 3 contains the starting material before separation. Lanes 4-8 contain fractions 2, 3, 5, 6 and 7 respectively. Lane 9 contains purified CC49/212.

Figure 22A is a chromatogram used to determine the molecular size of fraction 2 from Figure 20. A TSK G3000SW gel filtration HPLC column was used (Toyo Soda, Tokyo, Japan).

Figure 22B is a chromatogram used to determine the molecular size of fraction 5 from Figure 20. A TSK G3000SW gel filtration HPLC column was used (Toyo Soda, Tokyo, Japan).

Figure 22C is a chromatogram used to determine the molecular size of fraction 6 from Figure 20. A TSK G3000SW gel filtration HPLC column was used (Toyo Soda, Tokyo, Japan).

Figure 23 shows a Scatchard analysis of the fluorescein binding affinity of the CC49 4-4-20 heterodimer Fv (fraction 5 in Figure 20).

Figure 24 is a graphical representation of three competition enzyme-linked immunosorbent assays (ELISA) in which unlabeled CC49 4-4-20 Fv (closed squares) CC49/212 single-chain Fv (open squares) and MOPC-21 IgG (+) competed against a biotin-labeled CC49 IgG for binding to the TAG-72 antigen on a human breast carcinoma extract. MOPC-21 is a control antibody that does not bind to TAG-72 antigen.

Figure 25 shows a coomassie-blue stained non-reducing 4-20% SDS-PAGE gel. Lanes 1 and 9 contain the molecular weight standards. Lane 3 contains the 4-4-20/212 CPPC single-chain antigen-binding protein after purification. Lane 4, 5 and 6 contain the 4-4-20/212 CPPC single-chain antigen-binding protein after treatment with DTT and air oxidation. Lane 7 contains 4-4-20/212 single-chain antigen-binding protein.

Figure 26 shows a coomassie-blue stained reducing 4-20% SDS-PAGE gel (samples were treated with β -mercaptoethanol prior to being loaded on the

- 13 -

gel). Lanes 1 and 8 contain the molecular weight standards. Lane 3 contains the 4-4-20/212 CPPC single-chain antigen-binding protein after treatment with *bis*-maleimidehexane. Lane 5 contains peak 1 of *bis*-maleimidehexane treated 4-4-20/212 CPCC single-chain antigen-binding protein. Lane 6 contains peak 3 of *bis*-maleimidehexane treated 4-4-20/212 CPPC single-chain antigen-binding protein.

Detailed Description of the Preferred Embodiments

This invention relates to the discovery that multivalent forms of single-chain antigen-binding proteins have significant utility beyond that of the monovalent single-chain antigen-binding proteins. A multivalent antigen-binding protein has more than one antigen-binding site. For the purposes of this application, "valent" refers to the numerosity of antigen binding sites. Thus, a bivalent protein refers to a protein with two binding sites. Enhanced binding activity, bi- and multi-specific binding, and other novel uses of multivalent antigen-binding proteins have been demonstrated or are envisioned here. Accordingly, the invention is directed to multivalent forms of single-chain antigen-binding proteins, compositions of multivalent and single-chain antigen-binding proteins, methods of making and purifying multivalent forms of single-chain antigen-binding proteins, and new and improved uses for multivalent forms of single-chain antigen-binding proteins. The invention provides a multivalent antigen-binding protein comprising two or more single-chain protein molecules, each single-chain molecule comprising a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; and a peptide linker linking the first and second polypeptides into a single-chain protein.

The term "multivalent" means any assemblage, covalently or non-covalently joined, of two or more single-chain proteins, the assemblage having more than one antigen-binding site. The single-chain proteins composing the

- 14 -

assemblage may have antigen-binding activity, or they may lack antigen-binding activity individually but be capable of assembly into active multivalent antigen-binding proteins. The term "multivalent" encompasses bivalent, trivalent, tetravalent, etc. It is envisioned that multivalent forms above bivalent may be useful for certain applications.

A preferred form of the multivalent antigen-binding protein comprises bivalent proteins, including heterobivalent and homobivalent forms. The term "bivalent" means an assemblage of single-chain proteins associated with each other to form two antigen-binding sites. The term "heterobivalent" indicates multivalent antigen-binding proteins that are bispecific molecules capable of binding to two different antigenic determinants. Therefore, heterobivalent proteins have two antigen-binding sites that have different binding specificities. The term "homobivalent" indicates that the two binding sites are for the same antigenic determinant.

The terms "single-chain molecule" or "single-chain protein" are used interchangeably here. They are structurally defined as comprising the binding portion of a first polypeptide from the variable region of an antibody, associated with the binding portion of a second polypeptide from the variable region of an antibody, the two polypeptides being joined by a peptide linker linking the first and second polypeptides into a single polypeptide chain. The single polypeptide chain thus comprises a pair of variable regions connected by a polypeptide linker. The regions may associate to form a functional antigen-binding site, as in the case wherein the regions comprise a light-chain and a heavy-chain variable region pair with appropriately paired complementarity determining regions (CDRs). In this case, the single-chain protein is referred to as a "single-chain antigen-binding protein" or "single-chain antigen-binding molecule."

Alternatively, the variable regions may have unnaturally paired CDRs or may both be derived from the same kind of antibody chain, either heavy or light, in which case the resulting single-chain molecule may not display a functional antigen-binding site. The single-chain antigen-binding protein

- 15 -

molecule is more fully described in U.S. Patent No. 4,946,778 (Ladner *et al.*), and incorporated herein by reference.

Without being bound by any particular theory, the inventors speculate on several models which can equally explain the phenomenon of multivalence. The inventors' models are presented herein for the purpose of illustration only, and are not to be construed as limitations upon the scope of the invention. The invention is useful and operable regardless of the precise mechanism of multivalence.

Figure 1 depicts the first hypothetical model for the creation of a multivalent protein, the "Association" model. Fig. 1A shows two monovalent single-chain antigen-binding proteins, each composed of a V_L , a V_H , and a linker polypeptide covalently bridging the two. Each monovalent single-chain antigen-binding protein is depicted having an identical antigen-binding site containing antigen. Figure 1B shows the simple association of the two single-chain antigen-binding proteins to create the bivalent form of the multivalent protein. It is hypothesized that simple hydrophobic forces between the monovalent proteins are responsible for their association in this manner. The origin of the multivalent proteins may be traceable to their concentration dependence. The monovalent units retain their original association between the V_H and V_L regions. Figure 1C shows the newly-formed homobivalent protein binding two identical antigen molecules simultaneously. Homobivalent antigen-binding proteins are necessarily monospecific for antigen.

Homovalent proteins are depicted in Figs. 2A through 2C formed according to the Association model. Fig. 1A depicts a homobivalent protein, Fig. 2B a trivalent protein, and Fig. 2C a tetravalent protein. Of course, the limitations of two-dimensional images of three-dimensional objects must be taken into account. Thus, the actual spatial arrangement of multivalent proteins can be expected to vary somewhat from these figures.

A heterobivalent antigen-binding protein has two different binding sites, the sites having different binding specificities. Figures 3A through C depict the Association model pathway to the creation of a heterobivalent protein. Figure 3A shows two monovalent single-chain antigen-binding proteins, Anti-

- 16 -

A single-chain antigen-binding protein and Anti-B single-chain antigen-binding protein, with antigen types A and B occupying the respective binding sites. Figure 3B depicts the heterobivalent protein formed by the simple association of the original monovalent proteins. Figure 3C shows the heterobivalent protein having bound antigens A and B into the antigen-binding sites. Figure 3C therefore shows the heterobivalent protein binding in a bispecific manner.

An alternative model for the formation of multivalent antigen-binding proteins is shown in Figures 4 through 6. This "Rearrangement" model hypothesizes the dissociation of the variable region interface by contact with dissociating agents such as guanidine hydrochloride, urea, or alcohols such as ethanol, either alone or in combination. Combinations and relevant concentration ranges of dissociating agents are recited in the discussion concerning dissociating agents, and in Example 2. Subsequent re-association of dissociated regions allows variable region recombination differing from the starting single-chain proteins, as depicted in Fig. 4B. The homobivalent antigen-binding protein of Figure 4B is formed from the parent single-chain antigen-binding proteins shown in Figure 4A, the recombined bivalent protein having V_L and V_H from the parent monovalent single-chain proteins. The homobivalent protein of Figure 4B is a fully functional monospecific bivalent protein, shown actively binding two antigen molecules.

Figures 5A-5C show the formation of heterobivalent antigen-binding proteins via the Rearrangement model. Figure 5A shows a pair of single-chain proteins, each having a V_L with complementarity determining regions (CDRs) that do not match those of the associated V_H . These single-chain proteins have reduced or no ability to bind antigen because of the mixed nature of their antigen-binding sites, and thus are made specifically to be assembled into multivalent proteins through this route. Figure 5B shows the heterobivalent antigen-binding protein formed whereby the V_H and V_L regions of the parent proteins are shared between the separate halves of the heterobivalent protein. Figure 5C shows the binding of two different antigen molecules to the resultant functional bispecific heterobivalent protein. The Rearrangement model also explains the generation of multivalent proteins of

- 17 -

a higher order than bivalent, as it can be appreciated that more than a pair of single-chain proteins can be reassembled in this manner. These are depicted in Figures 6A and 6B.

One of the major utilities of the multivalent antigen-binding protein is in the heterobivalent form, in which one specificity is for one type of hapten or antigen, and the second specificity is for a second type of hapten or antigen. A multivalent molecule having two distinct binding specificities has many potential uses. For instance, one antigen binding site may be specific for a cell-surface epitope of a target cell, such as a tumor cell or other undesirable cell. The other antigen-binding site may be specific for a cell-surface epitope of an effector cell, such as the CD3 protein of a cytotoxic T-cell. In this way, the heterobivalent antigen-binding protein may guide a cytotoxic cell to a particular class of cells that are to be preferentially attacked.

Other uses of heterobivalent antigen-binding proteins are the specific targeting and destruction of blood clots by a bispecific molecule with specificity for tissue plasminogen activator (tPA) and fibrin; the specific targeting of pro-drug activating enzymes to tumor cells by a bispecific molecule with specificity for tumor cells and enzyme; and specific targeting of cytotoxic proteins to tumor cells by a bispecific molecule with specificity for tumor cells and a cytotoxic protein. This list is illustrative only, and any use for which a multivalent specificity is appropriate comes within the scope of this invention.

The invention also extends to uses for the multivalent antigen-binding proteins in purification and biosensors. Affinity purification is made possible by affixing the multivalent antigen-binding protein to a support, with the antigen-binding sites exposed to and in contact with the ligand molecule to be separated, and thus purified. Biosensors generate a detectable signal upon binding of a specific antigen to an antigen-binding molecule, with subsequent processing of the signal. Multivalent antigen-binding proteins, when used as the antigen-binding molecule in biosensors, may change conformation upon binding, thus generating a signal that may be detected.

- 18 -

Essentially all of the uses for which monoclonal or polyclonal antibodies, or fragments thereof, have been envisioned by the prior art, can be addressed by the multivalent proteins of the present invention. These uses include detectably-labelled forms of the multivalent protein. Types of labels are well-known to those of ordinary skill in the art. They include radiolabelling, chemiluminescent labeling, fluorochromic labelling, and chromophoric labeling. Other uses include imaging the internal structure of an animal (including a human) by administering an effective amount of a labelled form of the multivalent protein and measuring detectable radiation associated with the animal. They also include improved immunoassays, including sandwich immunoassay, competitive immunoassay, and other immunoassays wherein the labelled antibody can be replaced by the multivalent antigen-binding protein of this invention.

A first preferred method of producing multivalent antigen-binding proteins involves separating the multivalent proteins from a production composition that comprises both multivalent and single-chain proteins, as represented in Example 1. The method comprises producing a composition of multivalent and single-chain proteins, separating the multivalent proteins from the single-chain proteins, and recovering the multivalent proteins.

A second preferred method of producing multivalent antigen-binding proteins comprises the steps of producing single-chain protein molecules, dissociating said single-chain molecules, reassociating the single-chain molecules such that a significant fraction of the resulting composition includes multivalent forms of the single-chain antigen-binding proteins, separating multivalent antigen-binding proteins from single-chain molecules, and recovering the multivalent proteins. This process is illustrated with more detail in Example 2. For the purposes of this method, the term "producing a composition comprising single-chain molecules" may indicate the actual production of these molecules. The term may also include procuring them from whatever commercial or institutional source makes them available. Use of the term "producing single-chain proteins" means production of single-chain proteins by any process, but preferably according to the process set forth in

- 19 -

U.S. Patent No. 4,946,778 (Ladner *et al.*). Briefly, that patent pertains to a single polypeptide chain antigen-binding molecule which has binding specificity and affinity substantially similar to the binding specificity and affinity of the aggregate light and heavy chain variable regions of an antibody, to genetic sequences coding therefore, and to recombinant DNA methods of producing such molecules, and uses for such molecules. The single-chain protein produced by the Ladner *et al.* methodology comprises two regions linked by a linker polypeptide. The two regions are termed the V_H and V_L regions, each region comprising one half of a functional antigen-binding site.

The term "dissociating said single-chain molecules" means to cause the physical separation of the two variable regions of the single-chain protein without causing denaturation of the variable regions.

"Dissociating agents" are defined herein to include all agents capable of dissociating the variable regions, as defined above. In the context of this invention, the term includes the well-known agents alcohol (including ethanol), guanidine hydrochloride (GuHCl), and urea. Others will be apparent to those of ordinary skill in the art, including detergents and similar agents capable of interrupting the interactions that maintain protein conformation. In the preferred embodiment, a combination of GuHCl and ethanol (EtOH) is used as the dissociating agent. A preferred range for ethanol and GuHCl is from 0 to 50% EtOH, vol/vol, 0 to 2.0 moles per liter (M) GuHCl. A more preferred range is from 10-30% EtOH and 0.5-1.0 M GuHCl, and a most preferred range is 20% EtOH, 0.5 M GuHCl. A preferred dissociation buffer contains 0.5 M guanidine hydrochloride, 20% ethanol, 0.05 M TRIS, and 0.01 M $CaCl_2$, pH 8.0.

Use of the term "re-associating said single-chain molecules" is meant to describe the reassociation of the variable regions by contacting them with a buffer solution that allows reassociation. Such a buffer is preferably used in the present invention and is characterized as being composed of 0.04 M MOPS, 0.10 M calcium acetate, pH 7.5. Other buffers allowing the reassociation of the V_L and V_H regions are well within the expertise of one of ordinary skill in the art.

- 20 -

The separation of the multivalent protein from the single-chain molecules occurs by use of standard techniques known in the art, particularly including cation exchange or gel filtration chromatography.

5 Cation exchange chromatography is the general liquid chromatographic technique of ion-exchange chromatography utilizing anion columns well-known to those of ordinary skill in the art. In this invention, the cations exchanged are the single-chain and multivalent protein molecules. Since multivalent
10 proteins will have some multiple of the net charge of the single-chain molecule, the multivalent proteins are retained more strongly and are thus separated from the single-chain molecules. The preferred cationic exchanger of the present invention is a polyaspartic acid column, as shown in Figure 7. Figure 7 depicts the separation of single-chain protein (Peak 1, 27.32 min.) from bivalent protein (Peak 2, 55.54 min.) Those of ordinary skill in the art will realize that the invention is not limited to any particular type of
15 chromatography column, so long as it is capable of separating the two forms of protein molecules.

Gel filtration chromatography is the use of a gel-like material to separate proteins on the basis of their molecular weight. A "gel" is a matrix of water and a polymer, such as agarose or polymerized acrylamide. The
20 present invention encompasses the use of gel filtration HPLC (high performance liquid chromatography), as will be appreciated by one of ordinary skill in the art. Figure 8 is a chromatogram depicting the use of a Waters Associates' Protein-Pak 300 SW gel filtration column to separate monovalent single-chain protein from multivalent protein, including the monomer (21.940
25 min.), bivalent protein (20.135 min.), and trivalent protein (18.640 min.).

Recovering the multivalent antigen-binding proteins is accomplished by standard collection procedures well known in the chemical and biochemical arts. In the context of the present invention recovering the multivalent protein preferably comprises collection of eluate fractions containing the peak of
30 interest from either the cation exchange column, or the gel filtration HPLC column. Manual and automated fraction collection are well-known to one of

- 21 -

ordinary skill in the art. Subsequent processing may involve lyophilization of the eluate to produce a stable solid, or further purification.

A third preferred method of producing multivalent antigen-binding proteins is to start with purified single-chain proteins at a lower concentration, and then increase the concentration until some significant fraction of multivalent proteins is formed. The multivalent proteins are then separated and recovered. The concentrations conducive to formation of multivalent proteins in this manner are from about 0.5 milligram per milliliter (mg/ml) to the concentration at which precipitates begin to form.

The use of the term "substantially free" when used to describe a composition of multivalent and single-chain antigen-binding protein molecules means the lack of a significant peak corresponding to the single-chain molecule, when the composition is analyzed by cation exchange chromatography, as disclosed in Example 1 or by gel filtration chromatography as disclosed in Example 2.

By use of the term "aqueous composition" is meant any composition of single-chain molecules and multivalent proteins including a portion of water. In the same context, the phrase "an excess of multivalent antigen-binding protein over single-chain molecules" indicates that the composition comprises more than 50% of multivalent antigen-binding protein.

The use of the term "cross-linking" refers to chemical means by which one can produce multivalent antigen-binding proteins from monovalent single-chain protein molecules. For example, the incorporation of a cross-linkable sulfhydryl chemical group as a cysteine residue in the single-chain proteins allows cross-linking by mild reduction of the sulfhydryl group. Both monospecific and multispecific multivalent proteins can be produced from single-chain proteins by cross-linking the free cysteine groups from two or more single-chain proteins, causing a covalent chemical linkage to form between the individual proteins. Free cysteines have been engineered into the C-terminal portion of the 4-4-20/212 single-chain antigen-binding protein, as discussed in Example 5 and Example 8. These free cysteines may then be cross-linked to form multivalent antigen-binding proteins.

- 22 -

The invention also comprises single-chain proteins, comprising: (a) a first polypeptide comprising the binding portion of the variable region of an antibody light chain; (b) a second polypeptide comprising the binding portion of the variable region of an antibody light chain; and (c) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein. A similar single-chain protein comprising the heavy chain variable regions is also a part of this invention. Genetic sequences encoding these molecules are also included in the scope of this invention. Since these proteins are comprised of two similar variable regions, they do not necessarily have any antigen-binding capability.

The invention also includes a DNA sequence encoding a bispecific bivalent antigen-binding protein. Example 4 and Example 7 discusses in detail the sequences that appear in Figs. 10A and 10B that allow one of ordinary skill to construct a heterobivalent antigen-binding molecule. Figure 10A is an amino acid and nucleotide sequence listing of the single-chain protein comprising the 4-4-20 V_L region connected through the 212 linker polypeptide to the CC49 V_H region. Figure 10B is a similar listing of the single-chain protein comprising the CC49 V_L region connected through the 212 linker polypeptide to the 4-4-20 V_H region. Subjecting a composition including these single-chain molecules to dissociating and subsequent re-associating conditions results in the production of a bivalent protein with two different binding specificities.

Synthesis of DNA sequences is well known in the art, and possible through at least two routes. First, it is well-known that DNA sequences may be synthesized through the use of automated DNA synthesizers *de novo*, once the primary sequence information is known. Alternatively, it is possible to obtain a DNA sequence coding for a multivalent single-chain antigen-binding protein by removing the stop codons from the end of a gene encoding a single-chain antigen-binding protein, and then inserting a linker and a gene encoding a second single-chain antigen-binding protein. Example 6 demonstrates the construction of a DNA sequence coding for a bivalent single-chain antigen-binding protein. Other methods of genetically constructing multivalent single-

- 23 -

chain antigen-binding proteins come within the spirit and scope of the present invention.

Having now generally described this invention the same will better be understood by reference to certain specific examples which are included for purposes of illustration and are not intended to limit it unless otherwise specified.

Example 1

Production of Multivalent Antigen-Binding Proteins During Purification

In the production of multivalent antigen-binding proteins, the same recombinant *E. coli* production system that was used for prior single-chain antigen-binding protein production was used. See Bird, *et al.*, *Science* 242:423 (1988). This production system produced between 2 and 20% of the total *E. coli* protein as antigen-binding protein. For protein recovery, the frozen cell paste from three 10-liter fermentations (600-900 g) was thawed overnight at 4°C and gently resuspended at 4°C in 50 mM Tris-HCl, 1.0 mM EDTA, 100 mM KCl, 0.1 mM PMSF, pH 8.0 (lysis buffer), using 10 liters of lysis buffer for every kilogram of wet cell paste. When thoroughly resuspended, the chilled mixture was passed three times through a Manton-Gaulin cell homogenizer to totally lyse the cells. Because the cell homogenizer raised the temperature of the cell lysate to 25 ± 5°C, the cell lysate was cooled to 5 ± 2°C with a Lauda/Brinkman chilling coil after each pass. Complete lysis was verified by visual inspection under a microscope.

The cell lysate was centrifuged at 24,300g for 30 min. at 6°C using a Sorvall RC-5B centrifuge. The pellet containing the insoluble antigen-binding protein was retained, and the supernatant was discarded. The pellet was washed by gently scraping it from the centrifuge bottles and resuspending it in 5 liters of lysis buffer/kg of wet cell paste. The resulting 3.0- to 4.5-liter suspension was again centrifuged at 24,300g for 30 min at 6°C, and the

- 24 -

supernatant was discarded. This washing of the pellet removes soluble *E. coli* proteins and can be repeated as many as five times. At any time during this washing procedure the material can be stored as a frozen pellet at -20°C. A substantial time saving in the washing steps can be accomplished by utilizing a Pellicon tangential flow apparatus equipped with 0.22- μ m microporous filters, in place of centrifugation.

The washed pellet was solubilized at 4°C in freshly prepared 6 M guanidine hydrochloride, 50 mM Tris-HCl, 10 mM CaCl₂, 50 mM KCl, pH 8.0 (dissociating buffer), using 9 ml/g of pellet. If necessary, a few quick pulses from a Heat Systems Ultrasonics tissue homogenizer can be used to complete the solubilization. The resulting suspension was centrifuged at 24,300g for 45 min at 6°C and the pellet was discarded. The optical density of the supernatant was determined at 280 nm and if the OD₂₈₀ was above 30, additional dissociating buffer was added to obtain an OD₂₈₀ of approximately 25.

The supernatant was slowly diluted into cold (4-7°C) refolding buffer (50 mM Tris-HCl, 10 mM CaCl₂, 50 mM KCl, pH 8.0) until a 1:10 dilution was reached (final volume 10-20 liters). Re-folding occurs over approximately eighteen hours under these conditions. The best results are obtained when the GuHCl extract is slowly added to the refolding buffer over a 2-h period, with gentle mixing. The solution was left undisturbed for at least a 20-h period, and 95% ethanol was added to this solution such that the final ethanol concentration was approximately 20%. This solution was left undisturbed until the flocculated material settled to the bottom, usually not less than sixty minutes. The solution was filtered through a 0.2 μ m Millipore Millipak 200. This filtration step may be optionally preceded by a centrifugation step. The filtrate was concentrated to 1 to 2 liters using an Amicon spiral cartridge with a 10,000 MWCO cartridge, again at 4°C.

The concentrated crude antigen-binding protein sample was dialyzed against Buffer A (60 mM MOPS, 0.5 mM Ca acetate, pH 6.0-6.4) until the conductivity was lowered to that of Buffer A. The sample was then loaded on a 21.5 x 250-mm polyaspartic acid PolyCAT A column, manufactured by Poly

- 25 -

LC of Columbia, Maryland. If more than 60 mg of protein is loaded on this column, the resolution begins to deteriorate; thus, the concentrated crude sample often must be divided into several PolyCAT A runs. Most antigen-binding proteins have an extinction coefficient of about $2.0 \text{ ml mg}^{-1} \text{ cm}^{-1}$ at 280 nm and this can be used to determine protein concentration. The antigen-binding protein sample was eluted from the PolyCAT A column with a 50-min linear gradient from Buffer A to Buffer B (see Table 1). Most of the single-chain proteins elute between 20 and 26 minutes when this gradient is used. This corresponds to an eluting solvent composition of approximately 70% Buffer A and 30% Buffer B. Most of the bivalent antigen-binding proteins elute later than 45 minutes, which correspond to over 90% Buffer B.

Figure 7 is a chromatogram depicting the separation of single-chain protein from bivalent CC49/212 protein, using the cation-exchange method just described. Peak 1, 27.32 minutes, represents the monomeric single-chain fraction. Peak 2, 55.52 minutes, represents the bivalent protein fraction.

Figure 8 is a chromatogram of the purified monomeric single-chain antigen-binding protein CC49/212 (Fraction 7 from Fig. 7) run on a Waters Protein-Pak 300SW gel filtration column. Monomer, with minor contaminants of dimer and trimer, is shown. Figure 9 is a chromatogram of the purified bivalent antigen-binding protein CC49/212 (Fraction 15 from Fig. 7) run on the same Waters Protein-Pak 300SW gel filtration column as used in Fig. 8.

- 26 -

TABLE 1				
PolyCAT A Cation-Exchange HPLC Gradients				
Time (min) ^a	Flow (ml/min)	Buffers ^b		
		A	B	C
Initial	15.0	100	0	0
50.0	15.0	0	100	0
55.0	15.0	0	100	0
60.0	15.0	0	0	100
63.0	15.0	0	0	100
64.0	15.0	100	0	0
67.0	15.0	100	0	0
^a Linear gradients are run between each time point. ^b Buffer A, 60 mM MOPS, 0.5 mM Ca acetate, pH 6.0-6.4; Buffer B, 60 mM MOPS, 20mM Ca acetate, pH 7.5-8.0; Buffer C, 40 mM MOPS, 100 mM CaCl ₂ , pH 7.5.				

This purification procedure yielded multivalent antigen-binding proteins that are more than 95% pure as examined by SDS-PAGE and size exclusion HPLC. Modifications of the above procedure may be dictated by the isoelectric point of the particular multivalent antigen-binding protein being purified. Of the monomeric single-chain proteins that have been purified to date, all have had an isoelectric point (pI) between 8.0 and 9.5. However, it is possible that a multivalent antigen-binding protein may be produced with a pI of less than 7.0. In that case, an anion exchange column may be required for purification.

The CC49 monoclonal antibody was developed by Dr. Jeffrey Schlom's group, Laboratory of Tumor Immunology and Biology, National Cancer Institute. It binds specifically to the pan-carcinoma tumor antigen TAG-72. See Muraro, R. *et al.*, *Cancer Research* 48:4588-4596 (1988).

To determine the binding properties of the bivalent and monomeric CC49/212 antigen-binding proteins, a competition radioimmunoassay (RIA)

- 27 -

was set up in which a CC49 IgG (with two antigen binding sites) radiolabeled with ^{125}I was competed against unlabeled CC49 IgG, or monovalent (fraction 7 in Figure 7) or bivalent (fraction 15 in Figure 7) CC49/212 antigen-binding protein for binding to the TAG-72 antigen on a human breast carcinoma extract. (See Figure 18). This competition RIA showed that the bivalent antigen-binding protein competed equally well for the antigen as did IgG, whereas the monovalent single-chain antigen-binding protein needed a ten-fold higher protein concentration to displace the IgG. Thus, the monovalent antigen-binding protein competes with about a ten-fold lower affinity for the antigen than does the bivalent IgG or bivalent antigen-binding protein. Figure 18 also shows the result of the competition RIA of a non-TAG-72 specific single-chain antigen-binding protein, the antiluorescein 4-4-20/212, which does not compete for binding.

Example 2

Process of Making Multivalent Antigen-Binding Proteins Using Dissociating Agents

A. Process Using Guanidine HCl and Ethanol

Multivalent antigen-binding proteins were produced from purified single-chain proteins in the following way. First the purified single-chain protein at a concentration of 0.25-4 mg/ml was dialyzed against 0.5 moles/liter (M) guanidine hydrochloride (GuHCl), 20% ethanol (EtOH), in 0.05 M TRIS, 0.05 M KCl, 0.01 M CaCl_2 buffer pH 8.0. This combination of dissociating agents is thought to disrupt the V_L/V_H interface, allowing the V_H of a first single-chain molecule to come into contact with a V_L from a second single-chain molecule. Other dissociating agents such as urea, and alcohols such as isopropanol or methanol should be substitutable for GuHCl and EtOH. Following the initial dialysis, the protein was dialyzed against the load buffer for the final HPLC purification step. Two separate purification protocols,

- 28 -

cation exchange and gel filtration chromatography, can be used to separate the single-chain protein monomer from the multivalent antigen-binding proteins. In the first method, monomeric and multivalent antigen-binding proteins were separated by using cation exchange HPLC chromatography, using a polyaspartate column (PolyCAT A). This was a similar procedure to that used in the final purification of the antigen-binding proteins as described in Example 1. The load buffer was 0.06 M MOPS, 0.001 M Calcium Acetate pH 6.4. In the second method, the monomeric and multivalent antigen-binding proteins were separated by gel filtration HPLC chromatography using as a load buffer 0.04 M MOPS, 0.10 M Calcium Acetate pH 7.5. Gel filtration chromatography separates proteins based on their molecular size.

Once the antigen-binding protein sample was loaded on the cation exchange HPLC column, a linear gradient was run between the load buffer (0.04 to 0.06 M MOPS, 0.000 to 0.001 M calcium acetate, 0 to 10% glycerol pH 6.0-6.4) and a second buffer (0.04 to 0.06 M MOPS, 0.01 to 0.02 M calcium acetate, 0 to 10% glycerol pH 7.5). It was important to have extensively dialyze the antigen-binding protein sample before loading it on the column. Normally, the conductivity of the sample is monitored against the dialysis buffer. Dialysis is continued until the conductivity drops below 600 μ S. Figure 11 shows the separation of the monomeric (27.83 min.) and bivalent (50.47 min.) forms of the CC49/212 antigen-binding protein by cation exchange. The chromatographic conditions for this separation were as follows: PolyCAT A column, 200 x 4.6mm, operated at 0.62 ml/min.; load buffer and second buffer as in Example 1; gradient program from 100 percent load buffer A to 0 percent load buffer A over 48 mins; sample was CC49/212, 1.66 mg/ml; injection volume 0.2 ml. Fractions were collected from the two peaks from a similar chromatogram and identified as monomeric and bivalent proteins using gel filtration HPLC chromatography as described below.

Gel filtration HPLC chromatography (TSK G2000SW column from Toyo Soda, Tokyo, Japan) was used to identify and separate monomeric single-chain and multivalent antigen-binding proteins. This procedure has been described by Fukano, *et al.*, *J. Chromatography* 166:47 (1978).

- 29 -

Multimerization (creation of multivalent protein from monomeric single-chain protein) was by treatment with 0.5 M GuHCl and 20% EtOH for the times indicated in Table 2A followed by dialysis into the chromatography buffer. Figure 12 shows the separation of monomeric (17.65 min.), bivalent (15.79 min.), trivalent (14.19 min.), and higher oligomers (shoulder at about 13.09 min.) of the B6.2/212 antigen-binding protein. The B6.2/212 single-chain antigen-binding protein is described in Colcher, D., *et al.*, *J. Nat. Cancer Inst.* 82:1191-1197 (1990)). This separation depicts the results of a 24-hour multimerization treatment of a 1.0 mg/ml B6.2/212 antigen-binding protein sample. The HPLC buffer used was 0.04 M MOPS, 0.10 M calcium acetate, 0.04% sodium azide, pH 7.5.

Figure 13 shows the results of a 24-hour treatment of a 4.0 mg/ml CC49/212 antigen-binding protein sample, generating monomeric, bivalent and trivalent proteins at 16.91, 14.9, and 13.42 min., respectively. The HPLC buffer was 40 mM MOPS, 100 mM calcium acetate, pH 7.35. Multimerization treatment was for the times indicated in Table 2.

The results of Example 2A are shown in Table 2A. Table 2A shows the percentage of bivalent and other multivalent forms before and after treatment with 20% ethanol and 0.5M GuHCl. Unless otherwise indicated, percentages were determined using a automatic data integration software package.

- 30 -

Table 2A
Summary of the generation of bivalent and higher
multivalent forms of B6.2/212 and CC49/212
proteins using guanidine hydrochloride and ethanol

protein	Time (hours)	Concentration (mg/ml)	% monomer dimer trimer multimers			
			monomer	dimer	trimer	multimers
CC49/212	0	0.25	86.7	11.6	1.7	0.0
	0	1.0 ²	84.0	10.6	5.5	0.0
	0	4.0	70.0	17.1	12.9 ¹	0.0
	2	0.25 ²	62.9	33.2	4.2	0.0
	2	1.0	24.2	70.6	5.1	0.0
	2	4.0	9.3	81.3	9.5	0.0
	26	0.25	16.0	77.6	6.4	0.0
	26	1.0	9.2	82.8	7.9	0.0
	26	4.0	3.7	78.2	18.1	0.0
B6.2/212	0	0.25	100.0	0.0	0.0	0.0
	0	1.0	100.0	0.0	0.0	0.0
	0	4.0	100.0	0.0	0.0	0.0
	2	0.25 ²	98.1	1.9	0.0	0.0
	2	1.0	100.0	0.0	0.0	0.0
	2	4.0	90.0	5.5	1.0	0.0
	24	0.25	45.6	37.5	10.2	6.7
	24	1.0	50.8	21.4	12.3	15.0
	24	4.0	5.9	37.2	25.7	29.9

¹ Based on cut out peaks that were weighted.² Average of two experiments.

B. Process Using Urea and Ethanol

Multivalent antigen-binding proteins were produced from purified single-chain proteins in the following way. First the purified single-chain protein at a concentration of 0.25-1 mg/ml was dialyzed against 2M urea, 20% ethanol (EtOH), and 50mM Tris buffer pH 8.0, for the times indicated in Table 2B. This combination of dissociating agents is thought to disrupt the V_L/V_H interface, allowing the V_H of a first single-chain molecule to come into contact with a V_L from a second single-chain molecule. Other dissociating agents such as isopropanol or methanol should be substitutable for EtOH.

- 31 -

Following the initial dialysis, the protein was dialyzed against the load buffer for the final HPLC purification step.

Gel filtration HPLC chromatography (TSK G2000SW column from Toyo Soda, Tokyo, Japan) was used to identify and separate monomeric single-chain and multivalent antigen-binding proteins. This procedure has been described by Fukano, *et al.*, *J. Chromatography* **166**:47 (1978).

The results of Example 2B are shown in Table 2B. Table 2B shows the percentage of bivalent and other multivalent forms before and after treatment with 20% ethanol and urea. Percentages were determined using an automatic data integration software package.

Table 2B

Summary of the generation of bivalent and higher multivalent forms of B6.2/212 and CC49/212 proteins using urea and ethanol

protein	Time (hours)	Concentration (mg/ml)	monomer	% dimer	trimer	multimers
B6.2	0	0.25	44.1	37.6	15.9	2.4
	0	1.0	37.7	33.7	19.4	9.4
	3	0.25	22.2	66.5	11.3	0.0
	3	1.0	13.7	69.9	16.4	0.0

Example 3

Determination of Binding Constants

Three anti-fluorescein single-chain antigen-binding proteins have been constructed based on the anti-fluorescein monoclonal antibody 4-4-20. The three 4-4-20 single-chain antigen-binding proteins differ in the polypeptide linker connecting the V_H and V_L regions of the protein. The three linkers used were 202', 212 and 216 (see Table 3). Bivalent and higher forms of the 4-4-20 antigen-binding protein were produced by concentrating the purified monomeric single-chain antigen-binding protein in the cation exchange load buffer (0.06 M MOPS, 0.001 M calcium acetate pH 6.4) to 5 mg/ml. The

- 32 -

bivalent and monomeric forms of the 4-4-20 antigen-binding proteins were separated by cation exchange HPLC (polyaspartate column) using a 50 min. linear gradient between the load buffer (0.06 M MOPS, 0.001 M calcium acetate pH 6.4) and a second buffer (0.06 M MOPS, 0.02 M calcium acetate pH 7.5). Two 0.02 ml samples were separated, and fractions of the bivalent and monomeric protein peaks were collected on each run. The amount of protein contained in each fraction was determined from the absorbance at 278 nm from the first separation. Before collecting the fractions from the second separation run, each fraction tube had a sufficient quantity of 1.03×10^5 M fluorescein added to it, such that after the fractions were collected a 1-to-1 molar ratio of protein-to-fluorescein existed. Addition of fluorescein stabilized the bivalent form of the 4-4-20 antigen-binding proteins. These samples were kept at 2°C (on ice).

The fluorescein dissociation rates were determined for each of these samples following the procedures described by Herron, J.N., in *Fluorescence Hapten: An Immunological Probe*, E.W. Voss, Ed., CRC Press, Boca Raton, FL (1984). A sample was first diluted with 20 mM HEPES buffer pH 8.0 to 5.0×10^{-8} M 4-4-20 antigen-binding protein. 560 μ l of the 5.0×10^{-8} M 4-4-20 antigen-binding protein sample was added to a cuvette in a fluorescence spectrophotometer equilibrated at 2°C and the fluorescence was read. 140 μ l of 1.02×10^{-5} M fluoresceinamine was added to the cuvette, and the fluorescence was read every 1 minute for up to 25 minutes (see Table 4).

The binding constants (K_d) for the 4-4-20 single-chain antigen-binding protein monomers diluted in 20 mM HEPES buffer pH 8.0 in the absence of fluorescein were also determined (see Table 4).

The three polypeptide linkers in these experiments differ in length. The 202', 212 and 216 linkers are 12, 14 and 18 residues long, respectively. These experiments show that there are two effects of linker length on the 4-4-20 antigen-binding proteins: first, the shorter the linker length the higher the fraction of bivalent protein formed; second, the fluorescein dissociation rates of the monomeric single-chain antigen-binding proteins are effected more by the linker length than are the dissociation rates of the bivalent antigen-binding

- 33 -

proteins. With the shorter linkers 202' and 212, the bivalent antigen-binding proteins have slower dissociation rates than the monomers. Thus, the linkers providing optimum production and binding affinities for monomeric and bivalent antigen-binding proteins may be different. Longer linkers may be more suitable for monomeric single-chain antigen-binding proteins, and shorter linkers may be more suitable for multivalent antigen-binding proteins.

Table 3				
Linker Designs				
V _L	Linker	V _H	Linker Name	Reference
-KLEIE	GKSSGSGSES ¹	TQKLD-	202'	Bird <i>et al.</i>
-KLEIK	GSTSGSGKSSEGKG ²	EVKLD-	212	Bedzyk <i>et al.</i>
-KLEIK	GSTSGSGKSSEGSGSTKG ³	EVKLD-	216	This application
-KLVLK	GSTSGKPSEGKG ⁴	EVKLD-	217	This application

(1) SEQ ID NO. 1

(2) SEQ ID NO. 2

(3) SEQ ID NO. 3

(4) SEQ ID NO. 4

Table 4			
Effects of Linkers on the SCA Protein Monomers and Dimers			
	Linker		
	202'	212	216
Monomer Fraction	0.47	0.66	0.90
K _a	$0.5 \times 10^9 \text{ M}^{-1}$	$1.0 \times 10^9 \text{ M}^{-1}$	$1.3 \times 10^9 \text{ M}^{-1}$
Dissociation rate	$8.2 \times 10^{-3} \text{ s}^{-1}$	$4.9 \times 10^{-3} \text{ s}^{-1}$	$3.3 \times 10^{-3} \text{ s}^{-1}$
Dimer Fraction	0.53	0.34	0.10
Dissociation rate	$4.6 \times 10^{-3} \text{ s}^{-1}$	$3.5 \times 10^{-3} \text{ s}^{-1}$	$3.5 \times 10^{-3} \text{ s}^{-1}$
Monomer/Dimer Dissociation rate ratio	1.8	1.4	0.9

Example 4

Genetic Construction of a Mixed-Fragment Bivalent Antigen-Binding Protein

The genetic constructions for one particular heterobivalent antigen-binding protein according to the Rearrangement model are shown in Figures 10A and 10B. Figure 10A is an amino acid and nucleotide sequence listing of the 4-4-20 V_L/212/CC49 V_H construct, coding for a single-chain protein with a 4-4-20 V_L, linked via a 212 polypeptide linker to a CC49 V_H. Figure 10B is a similar listing showing the CC49 V_L/212/4-4-20 V_H construct, coding for a single-chain protein with a CC49 V_L, linked via a 212 linker to a 4-4-20 V_H. These single-chain proteins may recombine according to the Rearrangement model to generate a heterobivalent protein comprising a CC49 antigen-binding site linked to a 4-4-20 antigen-binding site, as shown in Figure 5B.

"4-4-20 V_L" means the variable region of the light chain of the 4-4-20 mouse monoclonal antibody (Bird, R.E. *et al.*, *Science* **242**:423 (1988)). The number "212" refers to a specific 14-residue polypeptide linker that links the 4-4-20 V_L and the CC49 V_H. *See* Bedzyk, W.D. *et al.*, *J. Biol. Chem.* **265**:18615-18620 (1990). "CC49 V_H" is the variable region of the heavy chain of the CC49 antibody, which binds to the TAG-72 antigen. The CC49 antibody was developed at The National Institutes of Health by Schlom, *et al.* *Generation and Characterization of B72.3 Second Generation Monoclonal Antibodies Reactive With The Tumor-associated Glycoprotein 72 Antigen*, *Cancer Research* **48**:4588-4596 (1988).

Insertion of the sequences shown in FIGS. 10A and 10B, by standard recombinant DNA methodology, into a suitable plasmid vector will enable one of ordinary skill in the art to transform a suitable host for subsequent expression of the single-chain proteins. *See* Maniatis *et al.*, *Molecular Cloning, A Laboratory Manual*, p. 104, Cold Spring Harbor Laboratory (1982), for general recombinant techniques for accomplishing the aforesaid goals; *see also* U.S. Patent 4,946,778 (Ladner *et al.*) for a complete

- 35 -

description of methods of producing single-chain protein molecules by recombinant DNA technology.

To produce multivalent antigen-binding proteins from the two single-chain proteins, 4-4-20V_L-212/CC49V_H and CC49V_L/212/4-4-20V_H, the two single-chain proteins are dialyzed into 0.5 M GuHCl/20% EtOH being combined in a single solution either before or after dialysis. The multivalent proteins are then produced and separated as described in Example 2.

Example 5

Preparation of Multivalent Antigen-Binding Proteins by Chemical Cross-Linking

Free cysteines were engineered into the C-terminal of the 4-4-20/212 single-chain antigen-binding protein, in order to chemically crosslink the protein. The design was based on the hinge region found in antibodies between the C_H1 and C_H2 regions. In order to try to reduce antigenicity in humans, the hinge sequence of the most common IgG class, IgG1, was chosen. The 4-4-20 Fab structure was examined and it was determined that the C-terminal sequence GluH216-ProH217-ArgH218, was part of the C_H1 region and that the hinge between C_H1 and C_H2 starts with ArgH218 or GlyH219 in the mouse 4-4-20 IgG2A antibody. Figure 14 shows the structure of a human IgG. The hinge region is indicated generally. Thus the hinge from human IgG1 would start with LysH218 or SerH219. (See Table 5).

The C-terminal residue in most of the single-chain antigen-binding proteins described to date is the amino acid serine. In the design for the hinge region, the C-terminal serine in the 4-4-20/212 single-chain antigen-binding protein was made the first serine of the hinge and the second residue of the hinge was changed from a cysteine to a serine. This hinge cysteine normally forms a disulfide bridge to the C-terminal cysteine in the light chain.

- 36 -

TABLE 5

		218	
5	IgG2A mouse ¹	E P R G P T I K P	C P P C L C -
	IgG1 human ²	A E P K S C D K T H T C P P C -	
	SCA* ³	- - V T V S	
	SCA* Hinge design 1 ⁴	- - V T V S S D K T H T C	
	SCA* Hinge design 2 ⁵	- - V T V S S D K T H T C P P C	

* - single-chain antigen-binding protein

- 10 (1) SEQ ID NO. 5
 (2) SEQ ID NO. 6
 (3) SEQ ID NO. 7
 (4) SEQ ID NO. 8
 (5) SEQ ID NO. 9

15 There are possible advantages to having two C-terminal cysteines, for they might form an intramolecular disulfide bond, making the protein recovery easier by protecting the sulfurs from oxidation. The hinge regions were added by introduction of a BstE II restriction site in the 3'-terminus of the gene encoding the 4-4-20/212 single-chain antigen-binding protein (see Figures 15A-
 20 15B).

The monomeric single-chain antigen-binding protein containing the C-terminal cysteine can be purified using the normal methods of purifying a single-chain antigen-binding proteins, with minor modifications to protect the free sulfhydryls. The cross-linking could be accomplished in one of two
 25 ways. First, the purified single-chain antigen-binding protein could be treated with a mild reducing agent, such as dithiothreitol, then allowed to air oxidize to form a disulfide-bond between the individual single-chain antigen-binding proteins. This type of chemistry has been successful in producing heterodimers from whole antibodies (Nisonoff *et al.*, Quantitative Estimation
 30 of the Hybridization of Rabbit Antibodies, *Nature* 4826:355-359 (1962); Brennan *et al.*, Preparation of Bispecific Antibodies by Chemical Recombination of Monoclonal Immunoglobulin G₁ Fragments, *Science* 229:81-83 (1985)). Second, chemical crosslinking agents such as *bismaleimidehexane* could be used to cross-link two single-chain antigen-binding proteins by their
 35 C-terminal cysteines. See Partis *et al.*, *J. Prot. Chem.* 2:263-277 (1983).

- 37 -

Example 6

Genetic Construction of Bivalent Antigen-Binding Proteins

Bivalent antigen-binding proteins can be constructed genetically and subsequently expressed in *E. coli* or other known expression systems. This can be accomplished by genetically removing the stop codons at the end of a gene encoding a monomeric single-chain antigen-binding protein and inserting a linker and a gene encoding a second single-chain antigen-binding protein. We have constructed a gene for a bivalent CC49/212 antigen-binding protein in this manner (see Figure 16). The CC49/212 gene in the starting expression plasmid is in an Aat II to Bam H1 restriction fragment (see Bird *et al.*, Single-Chain Antigen-Binding Proteins, *Science* 242:423-426 (1988); and Whitlow *et al.*, Single-Chain F_v Proteins and Their Fusion Proteins, *Methods* 2:97-105 (1991)). The two stop codons and the Bam H1 site at the C-terminal end of the CC49/212 antigen-binding protein gene were replaced by a single residue linker (Ser) and an Aat II restriction site. The resulting plasmid was cut with Aat II and the purified Aat II to Aat II restriction fragment was ligated into Aat II cut CC49/212 single-chain antigen-binding protein expression plasmid. The resulting bivalent CC49/212 single-chain antigen-binding protein expression plasmid was transfected into an *E. coli* expression host that contained the gene for the cI857 temperature-sensitive repressor. Expression of single-chain antigen-binding protein in this system is induced by raising the temperature from 30°C to 42°C. Fig. 17 shows the expression of the divalent CC49/212 single-chain antigen-binding protein of Fig. 16 at 42°C, on an SDS-PAGE gel containing total *E. coli* protein. Lane 1 contains the molecular weight standards. Lane 2 is the uninduced *E. coli* production strain grown at 30°C. Lane 3 is divalent CC49/212 single-chain antigen-binding protein induced by growth at 42°C. The arrow shows the band of expressed divalent CC49/212 single-chain antigen-binding protein.

- 38 -

*Example 7**Construction, Purification, and Testing of 4-4-20/CC49
Heterodimer F_v With 217 Linkers.*

5 The goals of this experiment were to produce, purify and analyze for
activity a new heterodimer F_v that would bind to both fluorescein and the pan-
carcinoma antigen TAG-72. The design consisted of two polypeptide chains,
which associated to form the active heterodimer F_v. Each polypeptide chain
can be described as a mixed single-chain F_v (mixed sFv). The first mixed sFv
(GX 8952) comprised a 4-4-20 variable light chain (V_L) and a CC-49 variable
10 heavy chain (V_H) connected by a 217 polypeptide linker (Figure 19A). The
second mixed sFv (GX 8953) comprised a CC-49 V_L and a 4-4-20 V_H
connected by a 217 polypeptide linker (Figure 19B). The sequence of the 217
polypeptide linker is shown in Table 3. Construction of analogous CC49/4-4-
20 heterodimers connected by a 212 polypeptide linker as described in
15 Example 4.

*Results**A. Purification*

One 10-liter fermentation of each mixed sFv was grown on casein
digest-glucose-salts medium at 32°C to an optical density at 600 nm of 15 to
20. The mixed sFv expression was induced by raising the temperature of the
fermentation to 42°C for one hour. 277gm (wet cell weight) of *E. coli* strain
GX 8952 and 233gm (wet cell weight) of *E. coli* strain GX 8953 were
harvested in a centrifuge at 7000g for 10 minutes. The cell pellets were kept
and the supernate discarded. The cell pellets were frozen at -20°C for
25 storage.

- 39 -

2.55 liters of "lysis/wash buffer" (50mM Tris/ 200mM NaCl/ 1 mM EDTA, pH 8.0) was added to both of the mixed sFv's cell pellets, which were previously thawed and combined to give 510gm of total wet cell weight. After complete suspension of the cells they were then passed through a Gaulin homogenizer at 9000psi and 4°C. After this first pass the temperature increased to 23°C. The temperature was immediately brought down to 0°C using dry ice and methanol. The cell suspension was passed through the Gaulin homogenizer a second time and centrifuged at 8000 rpm with a Dupont GS-3 rotor for 60 minutes. The supernatant was discarded after centrifugation and the pellets resuspended in 2.5 liters of "lysis/wash buffer" at 4°C. This suspension was centrifuged for 45 minutes at 8000 rpm with the Dupont GS-3 rotor. The supernatant was again discarded and the pellet weighed. The pellet weight was 136.1 gm.

1300ml of 6M Guanidine Hydrochloride/50mM Tris/50mM KCl/10mM CaCl₂ pH 8.0 at 4°C was added to the washed pellet. An overhead mixer was used to speed solubilization. After one hour of mixing, the heterodimer GuHCl extract was centrifuged for 45 minutes at 8000 rpm and the pellet was discarded. The 1425ml of heterodimer Fv 6M GuHCl extract was slowly added (16 ml/min) to 14.1 liters of "Refold Buffer" (50mM Tris/50mM KCl/10mM CaCl₂, pH 8.0) under constant mixing at 4°C to give an approximate dilution of 1:10. Refolding took place overnight at 4°C.

After 17 hours of refolding the anti-fluorescein activity was checked by a 40% quenching assay, and the amount of active protein calculated. 150mg total active heterodimer Fv was found by the 40% quench assay, assuming a 54,000 molecular weight.

4 liters of prechilled (4°C) 190 proof ethanol was added to the 15 liters of refolded heterodimer with mixing for 3 hours. The mixture sat overnight at 4°C. A flocculent precipitate had settled to the bottom after this overnight treatment. The nearly clear solution was filtered through a Millipak-200 (0.22μ) filter so as to not disturb the precipitate. A 40% quench assay showed that 10% of the anti-fluorescein activity was recovered in the filtrate.

- 40 -

The filtered sample of heterodimer was dialyzed, using a Pellicon system containing 10,000 dalton MWCO membranes, with "dialysis buffer" 40mM MOPS/0.5mM Calcium Acetate (CaAc), pH 6.4 at 4°C. 20 liters of dialysis buffer was required before the conductivity of the retentate was equal to that of the dialysis buffer (~500 μ S). After dialysis the heterodimer sample was filtered through a Millipak-20 filter, 0.22 μ . After this step a 40% quench assay showed there was 8.8 mg of active protein.

The crude heterodimer sample was loaded on a Poly CAT A cation exchange column at 20ml/min. The column was previously equilibrated with 60mM MOPS, 1 mM CaAc pH 6.4, at 4°C, (Buffer A). After loading, the column was washed with 150ml of "Buffer A" at 15ml/min. A 50min linear gradient was performed at 15ml/min using "Buffer A" and "Buffer B" (60mM MOPS, 20mM CaAc pH 7.5 at 4°C). The gradient conditions are presented in Table 6. "Buffer C" comprises 60mM MOPS, 100mM CaCl₂, pH 7.5.

Table 6				
Time	%A	%B	%C	Flow
0:00	100.0	0.0	0.0	15ml/min
50:00	0.0	100.0	0.0	15ml/min
52:00	0.0	100.0	0.0	15ml/min
54:00	0.0	0.0	100.0	15ml/min
58:00	0.0	0.0	100.0	15ml/min
60:00	100.0	0.0	0.0	15ml/min

Approximately 50ml fractions were collected and analyzed for activity, purity, and molecular weight by size-exclusion chromatography. The fractions were not collected by peaks, so contamination between peaks is likely. Fractions 3 through 7 were pooled (total volume - 218ml), concentrated to 50ml and dialyzed against 4 liters of 60mM MOPS, 0.5mM CaAc pH 6.4 at 4°C overnight. The dialyzed pool was filtered through a 0.22 μ filter and

- 41 -

checked for absorbance at 280nm. The filtrate was loaded onto the PolyCAT A column, equilibrated with 60mM MOPS, 1 mM CaAc pH 6.4 at 4°C, at a flow rate of 10ml/min. Buffer B was changed to 60mM MOPS, 10mM CaAc pH 7.5 at 4°C. The gradient was run as in Table 6. The fractions were collected by peak and analyzed for activity, purity, and molecular weight. The chromatogram is shown in Figure 20. Fraction identification and analysis is presented in Table 7.

Table 7			
Fraction Analysis of the Heterodimer Fv protein			
Fraction No.	A ₂₈₀ reading	Total Volume (ml)	HPLC-SE Elution Time (min)
2	0.161	36	20.525
3	0.067	40	
4	0.033	40	
3	0.178	45	19.133
6	0.234	50	19.163
7	0.069	50	
8	0.055	40	

Fractions 2 to 7 and the starting material were analyzed by SDS gel electrophoresis, 4-20%. A picture and description of the gel is presented in Figure 21.

B. HPLC Size Exclusion Results

Fractions 2, 5, and 6 correspond to the three main peaks in Figure 20 and therefore were chosen to be analyzed by HPLC size exclusion. Fraction 2 corresponds to the peak that runs at 21.775 minutes in the preparative purification (Figure 20), and runs on the HPLC sizing column at 20.525 minutes, which is in the monomeric position (Figure 22A). Fractions 5 and 6 (30.1 and 33.455 minutes, respectively, in Figure 20) run on the HPLC sizing column (Figures 22B and 22C) at 19.133 and 19.163 minutes,

- 42 -

respectively (see Table 7). Therefore, both of these peaks could be considered dimers. 40% Quenching assays were performed on all fractions of this purification. Only fraction 5 gave significant activity. 2.4 mg of active CC49 4-4-20 heterodimer Fv was recovered in fraction 5, based on the Scatchard analysis described below.

C. N-terminal sequencing of the fractions

The active heterodimer Fv fraction should contain both polypeptide chains. N-terminal sequence analysis showed that fractions 5 and 6 displayed N-terminal sequences consistent with the presence of both CC49 and 4-4-20 polypeptides and fraction 2 displayed a single sequence corresponding to the CC49/212/4-4-20 polypeptide only. We believe that fraction 6 was contaminated by fraction 5 (see Figure 20), since only fraction 5 had significant activity.

D. Anti-fluorescein activity by Scatchard analysis

The fluorescein association constants (K_a) were determined for fractions 5 and 6 using the fluorescence quenching assay described by Herron, J.N., in *Fluorescence Hapten: An Immunological Probe*, E.W. Voss, ed., CRC Press, Boca Raton, FL (1984). Each sample was diluted to approximately 5.0×10^{-8} M with 20 mM HEPES buffer pH 8.0. 590 μ l of the 5.0×10^{-8} M sample was added to a cuvette in a fluorescence spectrophotometer equilibrated at room temperature. In a second cuvette 590 μ l of 20 mM HEPES buffer pH 8.0 was added. To each cuvette was added 10 μ l of 3.0×10^{-7} M fluorescein in 20 mM HEPES buffer pH 8.0, and the fluorescence recorded. This is repeated until 140 μ l of fluorescein had been added. The resulting Scatchard analysis for fraction 5 shows a binding constant of $1.16 \times 10^9 \text{ M}^{-1}$ for fraction #5 (see Figure 23). This is very close to the 4-4-20/212 sFv constant of $1.1 \times 10^9 \text{ M}^{-1}$ (see Pantoliano *et al.*, *Biochemistry* 30:10117-10125 (1991)). The R intercept on the Scatchard analysis represents the fraction of active material. For fraction 5, 61% of the

- 43 -

material was active. The graph of the Scatchard analysis on fraction 6 shows a binding constant of $3.3 \times 10^8 \text{ M}^{-1}$ and 14% active. The activity that is present in fraction 6 is most likely contaminants from fraction 5.

E. Anti-TAG-72 activity by competition ELISA

5 The CC49 monoclonal antibody was developed by Dr. Jeffrey Schlom's group, Laboratory of Tumor Immunology and Biology, National Cancer Institute. It binds specifically to the pan-carcinoma tumor antigen TAG-72. See Muraro, R., *et al.*, *Cancer Research* 48:4588-4596 (1988).

10 To determine the binding properties of the bivalent CC49/4-4-20 Fv (fraction 5) and the CC49/212 sFv, a competition enzyme-linked immunosorbent assay (ELISA) was set up in which a CC49 IgG labeled with biotin was competed against unlabeled CC49/4-4-20 Fv and the CC49/212 sFv for binding to TAG-72 on a human breast carcinoma extract (see Figure 24). The amount of biotin-labeled CC49 IgG was determined using a preformed
15 complex with avidin and biotin coupled to horse radish peroxidase and O-phenylenediamine dihydrochloride (OPD). The reaction was stopped with 4N H_2SO_4 (sulfuric acid), after 10 min. and the optical density read at 490nm. This competition ELISA showed that the bivalent CC49/4-4-20 Fv binds to the
20 TAG-72 antigen. The CC49/4-4-20 Fv needed a two hundred-fold higher protein concentration to displace the IgG than the single-chain Fv.

Example 8

Cross-Linking Antigen-Binding Dimers

25 We have chemically crosslinked dimers of 4-4-20/212 antigen-binding protein with the two cysteine C-terminal extension (4-4-20/212 CPPC single-chain antigen-binding protein) in two ways. In Example 5 we describe the design and genetic construction of the 4-4-20/212 CPPC single-chain antigen-binding protein (hinge design 2 in Table 5). Figure 15B shows the nucleic

- 44 -

acid and protein sequences of this protein. After purifying the 4-4-20/212 CPPC single-chain antigen-binding protein, using the methods described in Whitlow and Filpula, *Meth. Enzymol.* 2:97 (1991), dimers were formed by two methods. First, the free cysteines were mildly reduced with dithiothreitol (DTT) and then the disulfide-bonds between the two molecules were allowed to form by air oxidation. Second, the chemical crosslinker *bis*-maleimidehexane was used to produce dimers by crosslinking the free cysteines from two 4-4-20/212 CPPC single-chain antigen-binding proteins.

A 0.1 mg/ml solution of the 4-4-20/212 CPPC single-chain antigen-binding protein was mildly reduced using 1 mM DTT, 50 mM HEPES, 50mM NaCl, 1 mM EDTA buffer pH 8.0 at 4°C. The samples were dialyzed against 50mM HEPES, 50 mM NaCl, 1 mM EDTA buffer pH 8.0 at 4°C overnight, to allow the oxidation of free sulfhydryls to intermolecular disulfide-bonds. Figure 25 shows a non-reducing SDS-PAGE gel after the air oxidation; it shows that approximately 10% of the 4-4-20/212 CPPC protein formed dimers with molecular weights around 55,000 Daltons.

A 0.1 mg/ml solution of the 4-4-20/212 CPPC single-chain antigen-binding protein was treated with 2 mM *bis*-maleimidehexane. Unlike forming a disulfide-bond between two free cysteines in the previous example, the *bis*-maleimidehexane crosslinker material should be stable to reducing agents such as β -mercaptoethanol. Figure 26 shows that approximately 5% of the treated material produced dimer with a molecular weight of 55,000 Daltons on a reducing SDS-PAGE gel (samples were treated with β -mercaptoethanol prior to being loaded on the gel). We further purified the *bis*-maleimidehexane treated 4-4-20/212 CPPC protein on PolyCAT A cation exchange column after the protein had been extensively dialyzed against buffer A. Figure 26 shows that we were able to enhance the fraction containing the dimer to approximately 15%.

- 45 -

Conclusions

We have produced a heterodimer Fv from two complementary mixed sFv's which has been shown to have the size of a dimer of the sFv's. The N-terminal analysis has shown that the active heterodimer Fv contains two polypeptide chains. The heterodimer Fv has been shown to be active for both fluorescein and TAG-72 binding.

All publications cited herein are incorporated fully into this disclosure by reference.

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention and the following claims. As examples, the steps of the preferred embodiment constitute only one form of carrying out the process in which the invention may be embodied.

(1) GENERAL INFORMATION:

-47-

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 18 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Ser Gly Ser Thr
 1 5 10 15
 Lys Gly

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 12 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Gly Ser Thr Ser Gly Lys Pro Ser Glu Gly Lys Gly
 1 5 10

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Glu Pro Arg Gly Pro Thr Ile Lys Pro Cys Pro Pro Cys Leu Cys
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 15 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Ala Glu Pro Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys
 1 5 10 15

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Val Thr Val Ser
 1

- 48 -

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 11 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Val Thr Val Ser Ser Asp Lys Thr His Thr Cys
 1 5 10

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 14 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: both

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Val Thr Val Ser Ser Asp Lys Thr His Thr Cys Pro Pro Cys
 1 5 10

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 731 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: both
 (D) TOPOLOGY: both

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 1..729

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GAC GTC GTT ATG ACT CAG ACA CCA CTA TCA CTT CCT GTT AGT CTA GGT	48
Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly	
1 5 10 15	
GAT CAA GCC TCC ATC TCT TGC AGA TCT AGT CAG AGC CTT GTA CAC AGT	96
Asp Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser	
20 25 30	
AAT GGA AAC ACC TAT TTA CGT TGG TAC CTG CAG AAG CCA GGC CAG TCT	144
Asn Gly Asn Thr Tyr Leu Arg Trp Tyr Leu Gln Lys Pro Gly Gln Ser	
35 40 45	
CCA AAG GTC CTG ATC TAC AAA GTT TCC AAC CGA TTT TCT GGG GTC CCA	192
Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro	
50 55 60	
GAC AGG TTC AGT GGC AGT GGA TCA GGG ACA GAT TTC ACA CTC AAG ATC	240
Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile	
65 70 75 80	
AGC AGA GTG GAG GCT GAG GAT CTG GGA GTT TAT TTC TGC TCT CAA AGT	288
Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser	
85 90 95	
ACA CAT GTT CCG TGG ACG TTC GGT GGA GGC ACC AAG CTT GAA ATC AAA	336
Thr His Val Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys	
100 105 110	
GGT TCT ACC TCT GGT TCT GGT AAA TCC TCT GAA GGC AAA GGT CAG GTT	384
Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Lys Gly Gln Val	
115 120 125	

SUBSTITUTE SHEET

- 49 -

CAG CTG CAG CAG TCT GAC GCT GAG TTG GTG AAA CCT GGG GCT TCA GTG Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala Ser Val 130 135 140	432
AAG ATT TCC TGC AAG GCT TCT GGC TAC ACC TTC ACT GAC CAT GCA ATT Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His Ala Ile 145 150 155 160	480
CAC TGG GTG AAA CAG AAC CCT GAA CAG GGC CTG GAA TGG ATT GGA TAT His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr 165 170 175	528
TTT TCT CCC GGA AAT GAT GAT TTT AAA TAC AAT GAG AGG TTC AAG GGC Phe Ser Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly 180 185 190	576
AAG GCC ACA CTG ACT GCA GAC AAA TCC TCC AGC ACT GCC TAC GTG CAG Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr Val Gln 195 200 205	624
CTC AAC AGC CTG ACA TCT GAG GAT TCT GCA GTG TAT TTC TGT ACA AGA Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys Thr Arg 210 215 220	672
TCC CTG AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA GTC ACC GTC TCC Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr Val Ser 225 230 235 240	720
TAA TAG GAT CC * * Asp	731

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 243 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly 1 5 10 15
Asp Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser 20 25 30
Asn Gly Asn Thr Tyr Leu Arg Trp Tyr Leu Gln Lys Pro Gly Gln Ser 35 40 45
Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro 50 55 60
Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile 65 70 75 80
Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser 85 90 95
Thr His Val Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys 100 105 110
Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Lys Gly Gln Val 115 120 125
Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala Ser Val 130 135 140
Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His Ala Ile 145 150 155 160
His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr 165 170 175

SUBSTITUTE SHEET

-50-

Phe Ser Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly
 180 185 190

Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr Val Gln
 195 200 205

Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys Thr Arg
 210 215 220

Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr Val Ser
 225 230 235 240

* * Asp

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 744 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: both
 (D) TOPOLOGY: both

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 1..744

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAC GTC GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA GTT GGC Asp Val Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly 1 5 10 15	48
GAG AAG GTT ACT TTG AGC TGC AAG TCC AGT CAG AGC CTT TTA TAT AGT Glu Lys Val Thr Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser 20 25 30	96
GGT AAT CAA AAG AAC TAC TTG GCC TGG TAC CAG CAG AAA CCA GGG CAG Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln 35 40 45	144
TCT CCT AAA CTG CTG ATT TAC TGG GCA TCC GCT AGG GAA TCT GGG GTC Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg Glu Ser Gly Val 50 55 60	192
CCT GAT CGC TTC ACA GGC AGT GGA TCT GGG ACA GAT TTC ACT CTC TCC Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser 65 70 75 80	240
ATC AGC AGT GTG AAG ACT GAA GAC CTG GCA GTT TAT TAC TGT CAG CAG Ile Ser Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln 85 90 95	288
TAT TAT AGC TAT CCC CTC ACG TTC GGT GCT GGG ACC AAG CTT GTG CTG Tyr Tyr Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu 100 105 110	336
AAA GGC TCT ACT TCC GGT AGC GGC AAA TCT TCT GAA GGT AAA GGT GAA Lys Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Lys Gly Glu 115 120 125	384
GTT AAA CTG GAT GAG ACT GGA GGA GGC TTG GTG CAA CCT GGG AGG CCC Val Lys Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro Gly Arg Pro 130 135 140	432
ATG AAA CTC TCC TGT GTT GCC TCT GGA TTC ACT TTT AGT GAC TAC TGG Met Lys Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp 145 150 155 160	480
ATG AAC TGG GTC CGC CAG TCT CCA GAG AAA GGA CTG GAG TGG GTA GCA Met Asn Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala 165 170 175	528
CAA ATT AGA AAC AAA CCT TAT AAT TAT GAA ACA TAT TAT TCA GAT TCT Gln Ile Arg Asn Lys Pro Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser 180 185 190	576

-51-

GTG AAA GGC AGA TTC ACC ATC TCA AGA GAT GAT TCC AAA AGT AGT GTC	624
Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Ser Ser Val	
195 - 200 205	
TAC CTG CAA ATG AAC AAC TTA AGA GTT GAA GAC ATG GGT ATC TAT TAC	672
Tyr Leu Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly Ile Tyr Tyr	
210 - 215 220	
TGT ACG GGT TCT TAC TAT GGT ATG GAC TAC TGG GGT CAA GGA ACC TCA	720
Cys Thr Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser	
225 230 235 240	
GTC ACC GTC TCC TAA TAA GGA TCC	744
Val Thr Val Ser * * Gly Ser	
245	

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 248 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Asp Val Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly	15
1 5 10	
Glu Lys Val Thr Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser	30
20 25 30	
Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln	45
35 40 45	
Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg Glu Ser Gly Val	60
50 55 60	
Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser	80
65 70 75 80	
Ile Ser Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln	95
85 90 95	
Tyr Tyr Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu	110
100 105 110	
Lys Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Lys Gly Glu	125
115 120 125	
Val Lys Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro Gly Arg Pro	140
130 135 140	
Met Lys Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp	160
145 150 155 160	
Met Asn Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala	175
165 170 175	
Gln Ile Arg Asn Lys Pro Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser	190
180 185 190	
Val Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Ser Ser Val	205
195 200 205	
Tyr Leu Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly Ile Tyr Tyr	220
210 215 220	
Cys Thr Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser	240
225 230 235 240	
Val Thr Val Ser * * Gly Ser	
245	

-52-

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 761 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: both
 (D) TOPOLOGY: both

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 1..756

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAC GTC GTT ATG ACT CAG ACA CCA CTA TCA CTT CCT GTT AGT CTA GGT Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly 1 5 10 15	48
GAT CAA GCC TCC ATC TCT TGC AGA TCT AGT CAG AGC CTT GTA CAC AGT Asp Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser 20 25 30	96
AAT GGA AAC ACC TAT TTA CGT TGG TAC CTG CAG AAG CCA GGC CAG TCT Asn Gly Asn Thr Tyr Leu Arg Trp Tyr Leu Gln Lys Pro Gly Gln Ser 35 40 45	144
CCA AAG GTC CTG ATC TAC AAA GTT TCC AAC CGA TTT TCT GGG GTC CCA Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro 50 55 60	192
GAC AGG TTC AGT GGC AGT GGA TCA GGG ACA GAT TTC ACA CTC AAG ATC Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile 65 70 75 80	240
AGC AGA GTG GAG GCT GAG GAT CTG GGA GTT TAT TTC TGC TCT CAA AGT Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser 85 90 95	288
ACA CAT GTT CCG TGG ACG TTC GGT GGA GGC ACC AAG CTT GAA ATC AAA Thr His Val Pro Trp Thr Phe Gly Gly Thr Lys Leu Glu Ile Lys 100 105 110	336
GGT TCT ACC TCT GGT TCT GGT AAA TCT TCT GAA GGT AAA GGT GAA GTT Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Lys Gly Glu Val 115 120 125	384
AAA CTG GAT GAG ACT GGA GGA GGC TTG GTG CAA CCT GGG AGG CCC ATG Lys Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro Gly Arg Pro Met 130 135 140	432
AAA CTC TCC TGT GTT GCC TCT GGA TTC ACT TTT AGT GAC TAC TGG ATG Lys Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp Met 145 150 155 160	480
AAC TGG GTC CGC CAG TCT CCA GAG AAA GGA CTG GAG TGG GTA GCA CAA Asn Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala Gln 165 170 175	528
ATT AGA AAC AAA CCT TAT AAT TAT GAA ACA TAT TAT TCA GAT TCT GTG Ile Arg Asn Lys Pro Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser Val 180 185 190	576
AAA GGC AGA TTC ACC ATC TCA AGA GAT GAT TCC AAA AGT AGT GTC TAC Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Ser Ser Val Tyr 195 200 205	624
CTG CAA ATG AAC AAC TTA AGA GTT GAA GAC ATG GGT ATC TAT TAC TGT Leu Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly Ile Tyr Tyr Cys 210 215 220	672
ACG GGT TCT TAC TAT GGT ATG GAC TAC TGG GGT CAA GGA ACC TCG GTC Thr Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser Val 225 230 235 240	720

- 53 -

ACC GTC TCC AGT GAT AAG ACC CAT ACA TGC TAA TAGGATCC
 Thr Val Ser Ser Asp Lys Thr His Thr Cys *
 245 250

761

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 251 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly
 1 5 10 15
 Asp Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser
 20 25 30
 Asn Gly Asn Thr Tyr Leu Arg Trp Tyr Leu Gln Lys Pro Gly Gln Ser
 35 40 45
 Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro
 50 55 60
 Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
 65 70 75 80
 Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser
 85 90 95
 Thr His Val Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys
 100 105 110
 Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Lys Gly Glu Val
 115 120 125
 Lys Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro Gly Arg Pro Met
 130 135 140
 Lys Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp Met
 145 150 155 160
 Asn Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala Gln
 165 170 175
 Ile Arg Asn Lys Pro Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser Val
 180 185 190
 Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Ser Ser Val Tyr
 195 200 205
 Leu Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly Ile Tyr Tyr Cys
 210 215 220
 Thr Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser Val
 225 230 235 240
 Thr Val Ser Ser Asp Lys Thr His Thr Cys *
 245 250

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 770 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: both
 (D) TOPOLOGY: both

- 54 -

(ix) FEATURE:

(A)-NAME/KEY: CDS

(B) LOCATION: 1..765

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GAC GTC GTT ATG ACT CAG ACA CCA CTA TCA CTT CCT GTT AGT CTA GGT Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly 1 5 10 15	48
GAT CAA GCC TCC ATC TCT TGC AGA TCT AGT CAG AGC CTT GTA CAC AGT Asp Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser 20 25 30	96
AAT GGA AAC ACC TAT TTA CGT TGG TAC CTG CAG AAG CCA GGC CAG TCT Asn Gly Asn Thr Tyr Leu Arg Trp Tyr Leu Gln Lys Pro Gly Gln Ser 35 40 45	144
CCA AAG GTC CTG ATC TAC AAA GTT TCC AAC CGA TTT TCT GGG GTC CCA Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro 50 55 60	192
GAC AGG TTC AGT GGC AGT GGA TCA GGG ACA GAT TTC ACA CTC AAG ATC Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile 65 70 75 80	240
AGC AGA GTG GAG GCT GAG GAT CTG GGA GTT TAT TTC TGC TCT CAA AGT Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser 85 90 95	288
ACA CAT GTT CCG TGG ACG TTC GGT GGA GGC ACC AAG CTT GAA ATC AAA Thr His Val Pro Trp Thr Phe Gly Gly Thr Lys Leu Glu Ile Lys 100 105 110	336
GGT TCT ACC TCT GGT TCT GGT AAA TCT TCT GAA GGT AAA GGT GAA GTT Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Lys Gly Glu Val 115 120 125	384
AAA CTG GAT GAG ACT GGA GGA GGC TTG GTG CAA CCT GGG AGG CCC ATG Lys Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro Gly Arg Pro Met 130 135 140	432
AAA CTC TCC TGT GTT GCC TCT GGA TTC ACT TTT AGT GAC TAC TGG ATG Lys Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp Met 145 150 155 160	480
AAC TGG GTC CGC CAG TCT CCA GAG AAA GGA CTG GAG TGG GTA GCA CAA Asn Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala Gln 165 170 175	528
ATT AGA AAC AAA CCT TAT AAT TAT GAA ACA TAT TAT TCA GAT TCT GTG Ile Arg Asn Lys Pro Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser Val 180 185 190	576
AAA GGC AGA TTC ACC ATC TCA AGA GAT GAT TCC AAA AGT AGT GTC TAC Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Ser Ser Val Tyr 195 200 205	624
CTG CAA ATG AAC AAC TTA AGA GTT GAA GAC ATG GGT ATC TAT TAC TGT Leu Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly Ile Tyr Tyr Cys 210 215 220	672
ACG GGT TCT TAC TAT GGT ATG GAC TAC TGG GGT CAA GGA ACC TCG GTC Thr Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser Val 225 230 235 240	720
ACC GTC TCC AGT GAT AAG ACC CAT ACA TGC CCT CCA TGC TAA TAGGATCC Thr Val Ser Ser Asp Lys Thr His Thr Cys Pro Pro Cys * 245 250 255	770

- 55 -

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 254 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```

Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly
 1      5      10      15
Asp Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser
 20      25      30
Asn Gly Asn Thr Tyr Leu Arg Trp Tyr Leu Gln Lys Pro Gly Gln Ser
 35      40      45
Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro
 50      55      60
Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
 65      70      75      80
Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser
 85      90      95
Thr His Val Pro Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys
100      105      110
Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Lys Gly Glu Val
115      120      125
Lys Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro Gly Arg Pro Met
130      135      140
Lys Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp Met
145      150      155      160
Asn Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala Gln
165      170      175
Ile Arg Asn Lys Pro Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser Val
180      185      190
Lys Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Ser Ser Val Tyr
195      200      205
Leu Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly Ile Tyr Tyr Cys
210      215      220
Thr Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser Val
225      230      235      240
Thr Val Ser Ser Asp Lys Thr His Thr Cys Pro Pro Cys
245      250

```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1460 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: both
- (D) TOPOLOGY: both

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1398

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

- 56 -

GAC GTC GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA GTT GGC Asp Val Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly 1 5 10 15	48
GAG AAG GTT ACT TTG AGC TGC AAG TCC AGT CAG AGC CTT TTA TAT AGT Glu Lys Val Thr Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser 20 25 30	96
GGT AAT CAA AAG AAC TAC TTG GCC TGG TAC CAG CAG AAA CCA GGG CAG Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln 35 40 45	144
TCT CCT AAA CTG CTG ATT TAC TGG GCA TCC GCT AGG GAA TCT GGG GTC Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg Glu Ser Gly Val 50 55 60	192
CCT GAT CGC TTC ACA GGC AGT GGA TCT GGG ACA GAT TTC ACT CTC TCC Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser 65 70 75 80	240
ATC AGC AGT GTG AAG ACT GAA GAC CTG GCA GTT TAT TAC TGT CAG CAG Ile Ser Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln 85 90 95	288
TAT TAT AGC TAT CCC CTC ACG TTC GGT GCT GGG ACC AAG CTT GTG CTG Tyr Tyr Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu 100 105 110	336
AAA GGC TCT ACT TCC GGT AGC GGC AAA TCC TCT GAA GGC AAA GGT CAG Lys Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Lys Gly Gln 115 120 125	384
GTT CAG CTG CAG CAG TCT GAC GCT GAG TTG GTG AAA CCT GGG GCT TCA Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala Ser 130 135 140	432
GTG AAG ATT TCC TGC AAG GCT TCT GGC TAC ACC TTC ACT GAC CAT GCA Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His Ala 145 150 155 160	480
ATT CAC TGG GTG AAA CAG AAC CCT GAA CAG GGC CTG GAA TGG ATT GGA Ile His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly 165 170 175	528
TAT TTT TCT CCC GGA AAT GAT GAT TTT AAA TAC AAT GAG AGG TTC AAG Tyr Phe Ser Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe Lys 180 185 190	576
GGC AAG GCC ACA CTG ACT GCA GAC AAA TCC TCC AGC ACT GCC TAC GTG Gly Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr Val 195 200 205	624
CAG CTC AAC AGC CTG ACA TCT GAG GAT TCT GCA GTG TAT TTC TGT ACA Gln Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys Thr 210 215 220	672
AGA TCC CTG AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA GTC ACC GTC Arg Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr Val 225 230 235 240	720
TCC TCA GAC GTC GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA Ser Ser Asp Val Val Met Ser Gln Ser Pro Ser Ser Ser Leu Pro Val Ser 245 250 255	768
GTT GGC CAG AAG GTT ACT TTG AGC TGC AAG TCC AGT CAG AGC CTT TTA Val Gly Glu Lys Val Thr Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu 260 265 270	816
TAT AGT GGT AAT CAA AAG AAC TAC TTG GCC TGG TAC CAG CAG AAA CCA Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro 275 280 285	864
GGG CAG TCT CCT AAA CTG CTG ATT TAC TGG GCA TCC GCT AGG GAA TCT Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg Glu Ser 290 295 300	912

- 57 -

GGG GTC CCT GAT CGC TTC ACA GGC AGT GGA TCT GGG ACA GAT TTC ACT Gly Val Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr 305 - 310 315 320	960
CTC TCC ATC AGC AGT GTG AAG ACT GAA GAC CTG GCA GTT TAT TAC TGT Leu Ser Ile Ser Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys 325 330 335	1008
CAG CAG TAT TAT AGC TAT CCC CTC ACG TTC GGT GCT GGG ACC AAG CTT Gln Gln Tyr Tyr Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu 340 345 350	1056
GTG CTG AAA GGC TCT ACT TCC GGT AGC GGC AAA TCC TCT GAA GGC AAA Val Leu Lys Gly Ser Thr Ser Gly Ser Gly Lys Ser Ser Glu Gly Lys 355 360 365	1104
GGT CAG GTT CAG CTG CAG CAG TCT GAC GCT GAG TTG GTG AAA CCT GGG Gly Gln Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly 370 375 380	1152
GCT TCA GTG AAG ATT TCC TGC AAG GCT TCT GGC TAC ACC TTC ACT GAC Ala Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp 385 390 395 400	1200
CAT GCA ATT CAC TGG GTG AAA CAG AAC CCT GAA CAG GGC CTG GAA TGG His Ala Ile His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp 405 410 415	1248
ATT GGA TAT TTT TCT CCC GGA AAT GAT GAT TTT AAA TAC AAT GAG AGG Ile Gly Tyr Phe Ser Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg 420 425 430	1296
TTC AAG GGC AAG GCC ACA CTG ACT GCA GAC AAA TCC TCC AGC ACT GCC Phe Lys Gly Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala 435 440 445	1344
TAC GTG CAG CTC AAC AGC CTG ACA TCT GAG GAT TCT GCA GTG TAT TTC Tyr Val Gln Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe 450 455 460	1392
TGT ACA AGA TCC CTG AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA GTC Cys Thr Arg Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val 465 470 475 480	1440
ACC GTC TCC TAA TAG GAT CC Thr Val Ser * * Asp 485	1460

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 485 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Asp Val Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly 1 5 10 15
Glu Lys Val Thr Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser 20 25 30
Gly Asn Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln 35 40 45
Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg Glu Ser Gly Val 50 55 60
Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser 65 70 75 80

Ile	Ser	Ser	Val	Lys	Thr	Glu	Asp	Leu	Ala	Val	Tyr	Ser	Cys	Gln	Gln
				85						90				95	
Tyr	Tyr	Ser	Tyr	Pro	Leu	Thr	Phe	Gly	Ala	Gly	Thr	Lys	Leu	Val	Leu
			100					105					110		
Lys	Gly	Ser	Thr	Ser	Gly	Ser	Gly	Lys	Ser	Ser	Glu	Gly	Lys	Gly	Gln
		115					120					125			
Val	Gln	Leu	Gln	Gln	Ser	Asp	Ala	Glu	Leu	Val	Lys	Pro	Gly	Ala	Ser
		130				135					140				
Val	Lys	Ile	Ser	Cys	Lys	Ala	Ser	Gly	Tyr	Thr	Phe	Thr	Asp	His	Ala
		145			150					155					160
Ile	His	Trp	Val	Lys	Gln	Asn	Pro	Glu	Gln	Gly	Leu	Glu	Trp	Ile	Gly
			165						170					175	
Tyr	Phe	Ser	Pro	Gly	Asn	Asp	Asp	Phe	Lys	Tyr	Asn	Glu	Arg	Phe	Lys
			180					185					190		
Gly	Lys	Ala	Thr	Leu	Thr	Ala	Asp	Lys	Ser	Ser	Ser	Thr	Ala	Tyr	Val
		195					200					205			
Gln	Leu	Asn	Ser	Leu	Thr	Ser	Glu	Asp	Ser	Ala	Val	Tyr	Phe	Cys	Thr
		210					215				220				
Arg	Ser	Leu	Asn	Met	Ala	Tyr	Trp	Gly	Gln	Gly	Thr	Ser	Val	Thr	Val
		225			230					235					240
Ser	Ser	Asp	Val	Val	Met	Ser	Gln	Ser	Pro	Ser	Ser	Leu	Pro	Val	Ser
			245						250					255	
Val	Gly	Glu	Lys	Val	Thr	Leu	Ser	Cys	Lys	Ser	Ser	Gln	Ser	Leu	Leu
			260					265					270		
Tyr	Ser	Gly	Asn	Gln	Lys	Asn	Tyr	Leu	Ala	Trp	Tyr	Gln	Gln	Lys	Pro
		275					280					285			
Gly	Gln	Ser	Pro	Lys	Leu	Leu	Ile	Tyr	Trp	Ala	Ser	Ala	Arg	Glu	Ser
		290				295					300				
Gly	Val	Pro	Asp	Arg	Phe	Thr	Gly	Ser	Gly	Ser	Gly	Thr	Asp	Phe	Thr
		305			310					315					320
Leu	Ser	Ile	Ser	Ser	Val	Lys	Thr	Glu	Asp	Leu	Ala	Val	Tyr	Tyr	Cys
			325						330					335	
Gln	Gln	Tyr	Tyr	Ser	Tyr	Pro	Leu	Thr	Phe	Gly	Ala	Gly	Thr	Lys	Leu
			340					345					350		
Val	Leu	Lys	Gly	Ser	Thr	Ser	Gly	Ser	Gly	Lys	Ser	Ser	Glu	Gly	Lys
		355					360					365			
Gly	Gln	Val	Gln	Leu	Gln	Gln	Ser	Asp	Ala	Glu	Leu	Val	Lys	Pro	Gly
		370				375					380				
Ala	Ser	Val	Lys	Ile	Ser	Cys	Lys	Ala	Ser	Gly	Tyr	Thr	Phe	Thr	Asp
					390					395					400
His	Ala	Ile	His	Trp	Val	Lys	Gln	Asn	Pro	Glu	Gln	Gly	Leu	Glu	Trp
			405						410					415	
Ile	Gly	Tyr	Phe	Ser	Pro	Gly	Asn	Asp	Asp	Phe	Lys	Tyr	Asn	Glu	Arg
			420					425					430		

-59-

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 725 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: both
 (D) TOPOLOGY: both

- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 1..723

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GAC GTC GTT ATG ACT CAG ACA CCA CTA TCA CTT CCT GTT AGT CTA GGT Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly 1 5 10 15	48
GAT CAA GCC TCC ATC TCT TGC AGA TCT AGT CAG AGC CTT GTA CAC AGT Asp Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser 20 25 30	96
AAT GGA AAC ACC TAT TTA CGT TGG TAC CTG CAG AAG CCA GGC CAG TCT Asn Gly Asn Thr Tyr Leu Arg Trp Tyr Leu Gln Lys Pro Gly Gln Ser 35 40 45	144
CCA AAG GTC CTG ATC TAC AAA GTT TCC AAC CGA TTT TCT GGG GTC CCA Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro 50 55 60	192
GAC AGG TTC AGT GGC AGT GGA TCA GGG ACA GAT TTC ACA CTC AAG ATC Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile 65 70 75 80	240
AGC AGA GTG GAG GCT GAG GAT CTG GGA GTT TAT TTC TGC TCT CAA AGT Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser 85 90 95	288
ACA CAT GTT CCG TGG ACG TTC GGT GGA GGC ACC AAG CTT GAA ATC AAA Thr His Val Pro Trp Thr Phe Gly Gly Thr Lys Leu Glu Ile Lys 100 105 110	336
GGT TCT ACC TCT GGT AAA CCA TCT GAA GGC AAA GGT CAG GTT CAG CTG Gly Ser Thr Ser Gly Lys Pro Ser Glu Gly Lys Gly Gln Val Gln Leu 115 120 125	384
CAG CAG TCT GAC GCT GAG TTG GTG AAA CCT GGG GCT TCA GTG AAG ATT Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala Ser Val Lys Ile 130 135 140	432
TCC TGC AAG GCT TCT GGC TAC ACC TTC ACT GAC CAT GCA ATT CAC TGG Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His Ala Ile His Trp 145 150 155 160	480
GTG AAA CAG AAC CCT GAA CAG GGC CTG GAA TGG ATT GGA TAT TTT TCT Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser 165 170 175	528
CCC GGA AAT GAT GAT TTT AAA TAC AAT GAG AGG TTC AAG GGC AAG GCC Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala 180 185 190	576
ACA CTG ACT GCA GAC AAA TCC TCC AGC ACT GCC TAC GTG CAG CTC AAC Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr Val Gln Leu Asn 195 200 205	624
AGC CTG ACA TCT GAG GAT TCT GCA GTG TAT TTC TGT ACA AGA TCC CTG Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys Thr Arg Ser Leu 210 215 220	672

- 60 -

```

AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA GTC ACC GTC TCC TAA TAG      720
Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr Val Ser * *
225                230                235                240

GAT CC      725
Asp

```

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 241 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```

Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly
 1          5          10          15
Asp Gln Ala Ser Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser
 20          25          30
Asn Gly Asn Thr Tyr Leu Arg Trp Tyr Leu Gln Lys Pro Gly Gln Ser
 35          40          45
Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe Ser Gly Val Pro
 50          55          60
Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
 65          70          75          80
Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser
 85          90          95
Thr His Val Pro Trp Thr Phe Gly Gly Thr Lys Leu Glu Ile Lys
100          105          110
Gly Ser Thr Ser Gly Lys Pro Ser Glu Gly Lys Gly Gln Val Gln Leu
115          120          125
Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala Ser Val Lys Ile
130          135          140
Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His Ala Ile His Trp
145          150          155          160
Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser
165          170          175
Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala
180          185          190
Thr Leu Thr Ala Asp Lys Ser Ser Ser Thr Ala Tyr Val Gln Leu Asn
195          200          205
Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys Thr Arg Ser Leu
210          215          220
Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr Val Ser * *
225                230                235                240

Asp

```

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 738 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: both
 (D) TOPOLOGY: both

-61-

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 1..738

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GAC	GTC	GTG	ATG	TCA	CAG	TCT	CCA	TCC	TCC	CTA	CCT	GTG	TCA	GTT	GGC	48
Asp	Val	Val	Met	Ser	Gln	Ser	Pro	Ser	Ser	Leu	Pro	Val	Ser	Val	Gly	
1				5					10					15		
GAG	AAG	GTT	ACT	TTG	AGC	TGC	AAG	TCC	AGT	CAG	AGC	CTT	TTA	TAT	AGT	96
Glu	Lys	Val	Thr	Leu	Ser	Cys	Lys	Ser	Ser	Gln	Ser	Leu	Leu	Tyr	Ser	
		20						25					30			
GGT	AAT	CAA	AAG	AAC	TAC	TTG	GCC	TGG	TAC	CAG	CAG	AAA	CCA	GGG	CAG	144
Gly	Asn	Gln	Lys	Asn	Tyr	Leu	Ala	Trp	Tyr	Gln	Gln	Lys	Pro	Gly	Gln	
		35					40					45				
TCT	CCT	AAA	CTG	CTG	ATT	TAC	TGG	GCA	TCC	GCT	AGG	GAA	TCT	GGG	GTC	192
Ser	Pro	Lys	Leu	Leu	Ile	Tyr	Trp	Ala	Ser	Ala	Arg	Glu	Ser	Gly	Val	
		50				55					60					
CCT	GAT	CGC	TTC	ACA	GGC	AGT	GGA	TCT	GGG	ACA	GAT	TTC	ACT	CTC	TCC	240
Pro	Asp	Arg	Phe	Thr	Gly	Ser	Gly	Ser	Gly	Thr	Asp	Phe	Thr	Leu	Ser	
	65			70					75					80		
ATC	AGC	AGT	GTG	AAG	ACT	GAA	GAC	CTG	GCA	GTT	TAT	TAC	TGT	CAG	CAG	288
Ile	Ser	Ser	Val	Lys	Thr	Glu	Asp	Leu	Ala	Val	Tyr	Tyr	Cys	Gln	Gln	
			85					90					95			
TAT	TAT	AGC	TAT	CCC	CTC	ACG	TTC	GGT	GCT	GGG	ACC	AAG	CTT	GTG	CTG	336
Tyr	Tyr	Ser	Tyr	Pro	Leu	Thr	Phe	Gly	Ala	Gly	Thr	Lys	Leu	Val	Leu	
			100				105						110			
AAA	GGC	TCT	ACT	TCC	GGT	AAA	CCA	TCT	GAA	GGT	AAA	GGT	GAA	GTT	AAA	384
Lys	Gly	Ser	Thr	Ser	Gly	Lys	Pro	Ser	Glu	Gly	Lys	Gly	Glu	Val	Lys	
		115				120					125					
CTG	GAT	GAG	ACT	GGA	GGA	GGC	TTG	GTG	CAA	CCT	GGG	AGG	CCC	ATG	AAA	432
Leu	Asp	Glu	Thr	Gly	Gly	Gly	Leu	Val	Gln	Pro	Gly	Arg	Pro	Met	Lys	
	130					135					140					
CTC	TCC	TGT	GTT	GCC	TCT	GGA	TTC	ACT	TTT	AGT	GAC	TAC	TGG	ATG	AAC	480
Leu	Ser	Cys	Val	Ala	Ser	Gly	Phe	Thr	Phe	Ser	Asp	Tyr	Trp	Met	Asn	
	145				150				155					160		
TGG	GTC	CGC	CAG	TCT	CCA	GAG	AAA	GGA	CTG	GAG	TGG	GTA	GCA	CAA	ATT	528
Trp	Val	Arg	Gln	Ser	Pro	Glu	Lys	Gly	Leu	Glu	Trp	Val	Ala	Gln	Ile	
			165					170					175			
AGA	AAC	AAA	CCT	TAT	AAT	TAT	GAA	ACA	TAT	TAT	TCA	GAT	TCT	GTG	AAA	576
Arg	Asn	Lys	Pro	Tyr	Asn	Tyr	Glu	Thr	Tyr	Tyr	Ser	Asp	Ser	Val	Lys	
			180				185					190				
GGC	AGA	TTC	ACC	ATC	TCA	AGA	GAT	GAT	TCC	AAA	AGT	AGT	GTC	TAC	CTG	624
Gly	Arg	Phe	Thr	Ile	Ser	Arg	Asp	Asp	Ser	Lys	Ser	Ser	Val	Tyr	Leu	
		195				200					205					
CAA	ATG	AAC	AAC	TTA	AGA	GTT	GAA	GAC	ATG	GGT	ATC	TAT	TAC	TGT	ACG	672
Gln	Met	Asn	Asn	Leu	Arg	Val	Glu	Asp	Met	Gly	Ile	Tyr	Tyr	Cys	Thr	
	210					215					220					
GGT	TCT	TAC	TAT	GGT	ATG	GAC	TAC	TGG	GGT	CAA	GGA	ACC	TCA	GTC	ACC	720
Gly	Ser	Tyr	Tyr	Gly	Met	Asp	Tyr	Trp	Gly	Gln	Gly	Thr	Ser	Val	Thr	
	225				230				235					240		
GTC	TCC	TAA	TAA	GGA	TCC											738
Val	Ser	*	*	Gly	Ser											
				245												

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 246 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

- 62 -

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```

Asp Val Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly
 1           5           10           15
Glu Lys Val Thr Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser
 20           25           30
Gly Ash Gln Lys Asn Tyr Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln
 35           40           45
Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg Glu Ser Gly Val
 50           55           60
Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser
 65           70           75           80
Ile Ser Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln
 85           90           95
Tyr Tyr Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu
100           105           110
Lys Gly Ser Thr Ser Gly Lys Pro Ser Glu Gly Lys Gly Glu Val Lys
115           120           125
Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro Gly Arg Pro Met Lys
130           135           140
Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp Met Asn
145           150           155           160
Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala Gln Ile
165           170           175
Arg Asn Lys Pro Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser Val Lys
180           185           190
Gly Arg Phe Thr Ile Ser Arg Asp Asp Ser Lys Ser Ser Val Tyr Leu
195           200           205
Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly Ile Tyr Tyr Cys Thr
210           215           220
Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser Val Thr
225           230           235           240
Val Ser * * Gly Ser
245

```

- 63 -

What Is Claimed Is:

1. A multivalent antigen-binding protein comprising two or more single-chain molecules, each single-chain molecule comprising:
 - (a) a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;
 - 5 (b) a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; and
 - (c) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain molecule.
- 10 2. The multivalent protein of claim 1 wherein said first polypeptide comprises the binding portion of the variable region of an antibody light chain, and said second polypeptide comprises the binding portion of the variable region of an antibody heavy chain.
- 15 3. The multivalent protein of claim 1 wherein said first polypeptide comprises the binding portion of the variable region of an antibody light chain, and said second polypeptide comprises the binding portion of the variable region of an antibody light chain.
- 20 4. The multivalent protein of claim 1 wherein said first polypeptide comprises the binding portion of the variable region of an antibody heavy chain, and said second polypeptide comprises the binding portion of the variable region of an antibody heavy chain.
5. The multivalent protein of claims 1, 2, 3, or 4 comprising a bivalent antigen-binding protein.
6. The multivalent protein of claim 5 comprising a heterobivalent antigen-binding protein.

- 64 -

7. The multivalent protein of claim 5 comprising a homobivalent antigen-binding protein.

8. A composition comprising a multivalent antigen-binding protein substantially free of single-chain molecules, wherein said multivalent protein comprises two or more single-chain molecules, each single-chain molecule comprising:

(a) a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

(b) a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; and

(c) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain molecule.

9. The composition of claim 8 wherein said first polypeptide comprises the binding portion of the variable region of an antibody light chain, and said second polypeptide comprises the binding portion of the variable region of an antibody heavy chain.

10. The composition of claim 8 wherein said first polypeptide comprises the binding portion of the variable region of an antibody light chain, and said second polypeptide comprises the binding portion of the variable region of an antibody light chain.

11. The composition of claim 8 wherein said first polypeptide comprises the binding portion of the variable region of an antibody heavy chain, and said second polypeptide comprises the binding portion of the variable region of an antibody heavy chain.

12. The composition of claims 8, 9, 10, or 11, comprising a bivalent antigen-binding protein substantially free of single-chain molecules.

- 65 -

13. The composition of claim 12 wherein said bivalent protein is heterobivalent.

14. The composition of claim 12 wherein said bivalent protein is homobivalent.

5 15. An aqueous composition comprising an excess of multivalent antigen-binding protein over single-chain molecules, said multivalent protein comprising two or more single-chain molecules, each single-chain molecule comprising:

- 10 (a) a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;
- (b) a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; and
- (c) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein.

15 16. The aqueous composition of claim 15 wherein at least one of said single-chain molecules comprises:

- (a) a first polypeptide comprising the binding portion of the variable region of an antibody light chain;
- 20 (b) a second polypeptide comprising the binding portion of the variable region of an antibody heavy chain; and
- (c) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein.

17. The aqueous composition of claim 15 wherein at least one of said single-chain molecules comprises:

- 25 (a) a first polypeptide comprising the binding portion of the variable region of an antibody light chain;
- (b) a second polypeptide comprising the binding portion of the variable region of an antibody light chain; and

- 66 -

(c) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein.

18. The composition of claim 15 wherein at least one of said single-chain molecules comprises:

5 (a) a first polypeptide comprising the binding portion of the variable region of an antibody heavy chain;

(b) a second polypeptide comprising the binding portion of the variable region of an antibody heavy chain; and

10 (c) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein.

19. A method of producing a multivalent antigen-binding protein, comprising the steps of:

15 (a) producing a composition comprising multivalent antigen-binding protein and single-chain molecules, each single-chain molecule comprising:

(i) a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

(ii) a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; and

20 (iii) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain molecule;

(b) separating said multivalent protein from said single-chain molecules; and

(c) recovering said multivalent protein.

25 20. The method of claim 19 wherein separating said multivalent protein from said single-chain molecules comprises utilizing cation exchange chromatography.

- 67 -

21. The method of claim 19 wherein separating said multivalent protein from said single-chain molecules comprises utilizing gel filtration chromatography.

22. A method of producing a multivalent antigen-binding protein comprising the steps of:

(a) producing a composition comprising single-chain molecules, each single-chain molecule comprising:

(i) a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

(ii) a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; and

(iii) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain molecule;

(b) dissociating said single-chain molecules;

(c) re-associating said single-chain molecules;

(d) separating multivalent antigen-binding proteins from said single-chain molecules; and

(e) recovering said multivalent proteins.

23. The method of claim 22 wherein said dissociation is caused by dialysis against a dissociating solution.

24. The method of claim 22 wherein said reassociation is caused by dialysis against a refolding solution or a refolding agent.

25. A method of producing a multivalent antigen-binding protein, comprising the step of cross-linking at least two single-chain molecules to each other, each single-chain molecule comprising:

(a) a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

- 68 -

(b) a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; and

(c) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain molecule.

5 26. The method of claim 25 wherein said cross-linking is effected by chemical means.

27. A method of producing a multivalent antigen-binding protein, comprising the steps of:

10 (a) producing a composition comprising single-chain molecules, each single-chain molecule comprising:

(i) a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

(ii) a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; and

15 (iii) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain molecule;

(b) concentrating said single-chain molecules;

(c) separating said multivalent protein from said single-chain molecules; and

20 (d) recovering said multivalent protein.

28. The method of claim 27 wherein said concentrating step occurs from approximately 0.5 mg/ml single-chain molecule to the concentration at which precipitation starts.

25 29. A method of detecting an antigen in or suspected of being in a sample, which comprises:

(a) contacting said sample with the multivalent antigen-binding protein of claim 1; and

- 69 -

- (b) detecting whether said multivalent antigen-binding protein has bound to said antigen.

5 30. A method of imaging the internal structure of an animal, comprising administering to said animal an effective amount of a labeled form of the multivalent antigen-binding protein of claim 1 and measuring detectable radiation associated with said animal.

31. A composition comprising an association of a multivalent antigen-binding protein as claimed in any one of claims 1-4, 8-11, or 15-18 with a therapeutically or diagnostically effective agent.

10 32. A single-chain protein comprising:
 (a) a first polypeptide comprising the binding portion of the variable region of an antibody light chain;
 (b) a second polypeptide comprising the binding portion of the variable region of an antibody light chain;
15 (c) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein.

33. A single-chain protein comprising:
 (a) a first polypeptide comprising the binding portion of the variable region of an antibody heavy chain;
20 (b) a second polypeptide comprising the binding portion of the variable region of an antibody heavy chain;
 (c) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein.

25 34. A single-chain protein comprising:
 (a) a first polypeptide comprising the V_L or V_H of a CC49 monoclonal antibody;

- 70 -

(b) a second polypeptide comprising the V_L or V_H of a CC49 monoclonal antibody; and

(c) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein.

5 35. The single-chain protein of claim 34 wherein said linker is selected from the group consisting of the 202', 212, 216, and 217 linkers.

36. A single-chain protein comprising:

(a) a first polypeptide comprising the V_L or V_H of a CC49 monoclonal antibody;

10 (b) a second polypeptide comprising the V_L or V_H of a 4-20 monoclonal antibody; and

(c) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein.

15 37. The single-chain protein of claim 36 wherein said linker is selected from the group consisting of the 202', 212, 216, and 217 linkers.

38. A genetic sequence which codes for the single-chain protein of claim 32, comprising:

20 (a) a DNA sequence coding for a first polypeptide comprising the binding portion of the variable region of an antibody light chain;

(b) a DNA sequence coding for a second polypeptide comprising the binding portion of the variable region of an antibody light chain;

25 (c) a DNA sequence coding for a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein.

39. A genetic sequence which codes for the single-chain protein of claim 33, comprising:

- 71 -

(a) a DNA sequence coding for a first polypeptide comprising the binding portion of the variable region of an antibody heavy chain;

(b) a DNA sequence coding for a second polypeptide comprising the binding portion of the variable region of an antibody heavy chain;

(c) a DNA sequence coding for a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein.

40. A genetic sequence which codes for the single-chain protein of claim 34, comprising:

(a) a DNA sequence coding for the V_L or V_H of a CC49 monoclonal antibody;

(b) a DNA sequence coding for the V_L or V_H of a CC49 monoclonal antibody;

(c) a DNA sequence coding for a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein.

41. The genetic sequence of claim 40 wherein said DNA sequence codes for a peptide linker selected from the group consisting of the 202', 212, 216, and 217 linkers.

42. A genetic sequence which codes for the single-chain protein of claim 36, comprising:

(a) a DNA sequence coding for the V_L or V_H of a CC49 monoclonal antibody;

(b) a DNA sequence coding for the V_L or V_H of a 4-4-20 monoclonal antibody;

(c) a DNA sequence coding for a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain protein.

- 72 -

43. The genetic sequence of claim 42 wherein said DNA sequence codes for a peptide linker selected from the group consisting of the 202', 212, 216, and 217 linkers.

44. A multivalent single-chain antigen-binding protein comprising:

5 (a) a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

(b) a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

10 (c) a peptide linker linking said first and second polypeptides (a) and (b) into said multivalent protein;

(d) a third polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

(e) a fourth polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

15 (f) a peptide linker linking said third and fourth polypeptides (d) and (e) into said multivalent protein; and

(g) a peptide linker linking said second and third polypeptides (b) and (d) into said multivalent protein.

45. A multivalent single-chain antigen-binding protein comprising:

20 (a) a first polypeptide comprising the binding portion of the variable region of an antibody light chain;

(b) a second polypeptide comprising the binding portion of the variable region of an antibody heavy chain;

25 (c) a peptide linker linking said first and second polypeptides (a) and (b) into said multivalent protein;

(d) a third polypeptide comprising the binding portion of the variable region of an antibody light chain;

(e) a fourth polypeptide comprising the binding portion of the variable region of an antibody heavy chain;

- 73 -

(f) a peptide linker linking said third and fourth polypeptides (d) and (e) into said multivalent protein; and

(g) a peptide linker linking said second and third polypeptides (b) and (d) into said multivalent protein.

5 46. A genetic sequence which codes for the multivalent antigen-binding protein of claim 44 or 45, comprising:

(a) a DNA sequence coding for a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

10 (b) a DNA sequence coding for a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

(c) a DNA sequence coding for a peptide linker linking said first and second polypeptides (a) and (b) into said multivalent protein

15 (d) a DNA sequence coding for a third polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

(e) a DNA sequence coding for a fourth polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

20 (f) a DNA sequence coding for a peptide linker linking said third and fourth polypeptides (d) and (e) into said multivalent protein; and

(g) a DNA sequence coding for a peptide linker linking said second and third polypeptides (b) and (d) into said multivalent protein.

25 47. A replicable cloning or expression vehicle comprising the DNA sequence of any one of claims 38-43.

48. The vehicle of claim 47 which is a plasmid.

49. A host cell transformed with the vehicle of claim 47.

- 74 -

50. The host cell of claim 49 which is a bacterial cell, a yeast cell or other fungal cell, or a mammalian cell line.

51. A method of producing a multivalent antigen-binding protein comprising two or more single-chain molecules, each single-chain molecule comprising:

5

(a) a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

(b) a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain; and

10

(c) a peptide linker linking said first and second polypeptides (a) and (b) into said single-chain molecule, said method comprising:

(i) providing a genetic sequence coding for said single-chain molecule;

15

(ii) transforming one or more host cells with said sequence;

(iii) expressing said sequence in said host or hosts; and

(iv) recovering a multivalent protein from said host or hosts.

20

52. A method of producing a multivalent single-chain antigen-binding protein comprising two or more single-chain molecules, each single-chain molecule comprising:

(a) a first polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

25

(b) a second polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

(c) a peptide linker linking said first and second polypeptides (a) and (b) into said multivalent protein;

30

(d) a third polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

- 75 -

(e) a fourth polypeptide comprising the binding portion of the variable region of an antibody heavy or light chain;

(f) a peptide linker linking said third and fourth polypeptides (d) and (e) into said multivalent protein; and

5 (g) a peptide linker linking said second and third polypeptides (b) and (d) into said multivalent protein, said method comprising:

(i) providing a genetic sequence coding for said single-chain molecule;

10 (ii) transforming one or more host cells with said sequence;

(iii) expressing said sequence in said host or hosts; and

(iv) recovering a multivalent protein from said host or hosts.

15 53. The method of claim 51 or 52 wherein recovering said multivalent protein comprises separating said multivalent protein from said single-chain molecules.

54. The method of claim 51 or 52 wherein recovering said multivalent protein comprises:

- 20 (a) dissociating said single-chain molecules;
(b) re-associating said single-chain molecules;
(c) separating multivalent antigen-binding proteins from said single-chain molecules; and
(d) recovering said multivalent proteins.

25 55. The method of claim 51 or 52 which further comprises purifying said recovered multivalent protein.

56. The method of claim 51 or 52 wherein said host cell is a bacterial cell, a yeast cell or other fungal cell, or a mammalian cell line.

- 76 -

57. The method of claim 56 wherein said host cell is *E. coli* or *Bacillus subtilis*.

58. The multivalent antigen-binding protein of claim 1 in detectably-labelled form.

5 59. In an immunoassay method which utilizes an antibody in detectably-labelled form, the improvement comprising using the multivalent protein of claim 58 instead of said antibody.

60. The immunoassay of claim 59 wherein said immunoassay is a competitive immunoassay.

10 61. The immunoassay of claim 59 wherein said immunoassay is a sandwich immunoassay.

62. In an immunotherapeutic method which utilizes an antibody conjugated to a therapeutic agent, the improvement comprising using the multivalent protein of claim 1 instead of said antibody.

15 63. In a method of immunoaffinity purification which utilizes an antibody therefor, the improvement which comprises using the molecule of claim 1 instead of said antibody.

1/39

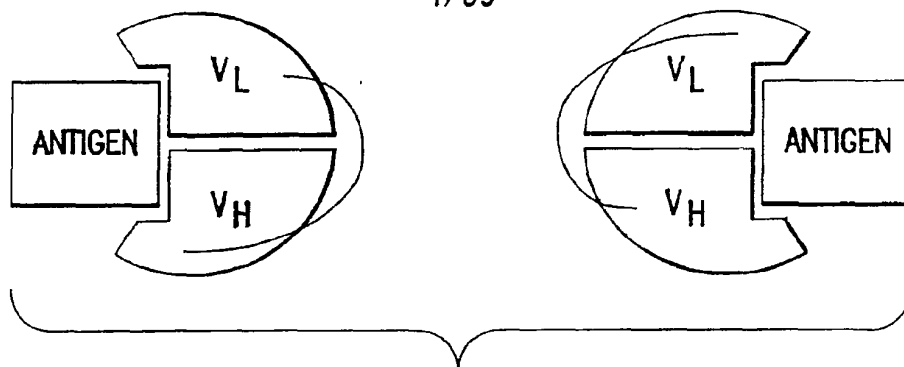


FIG.1A

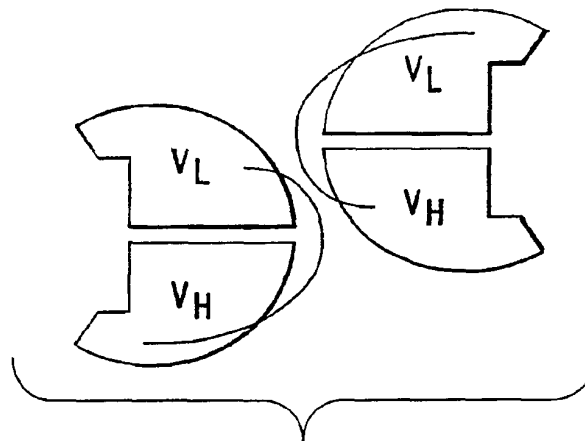


FIG.1B

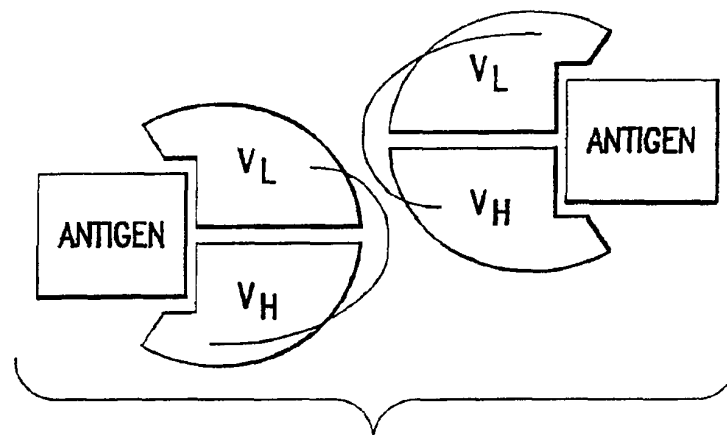


FIG.1C

SUBSTITUTE SHEET

2/39

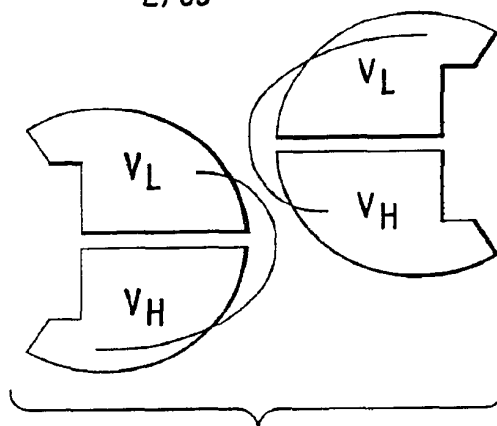


FIG.2A

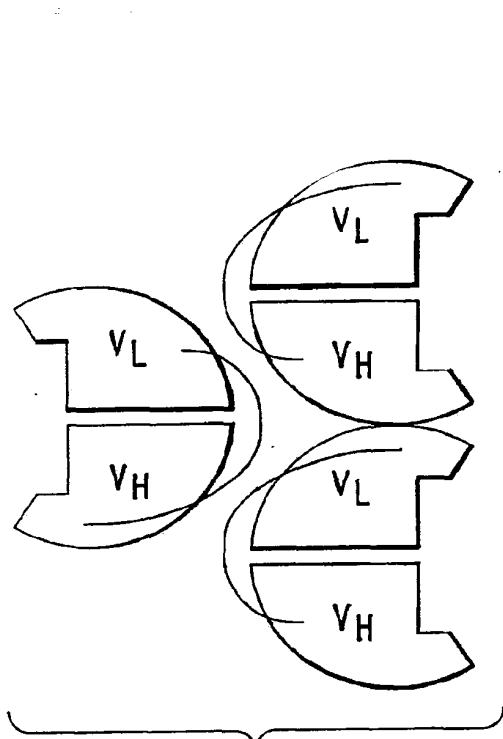


FIG.2B

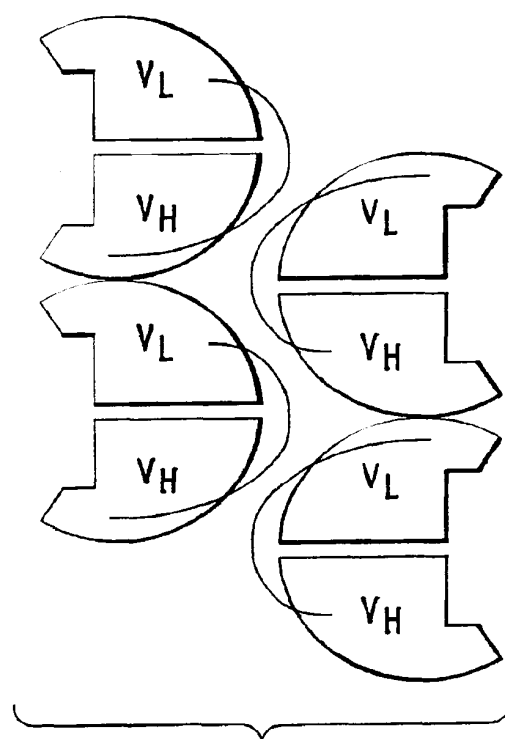


FIG.2C

SUBSTITUTE SHEET

3/39

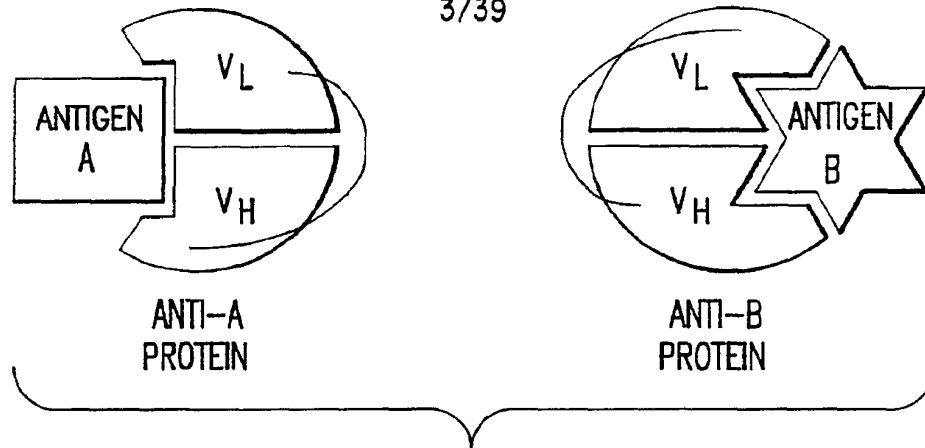


FIG.3A

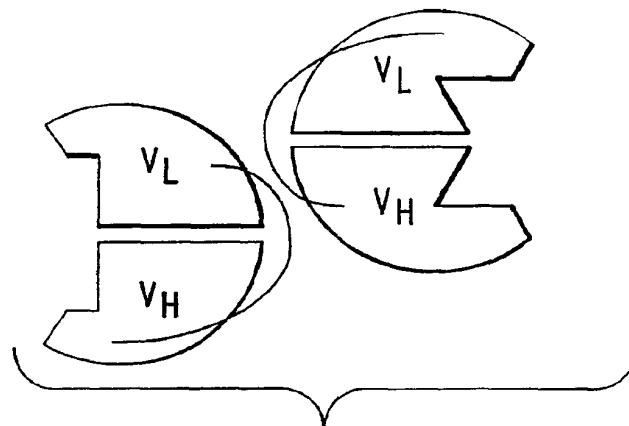


FIG.3B

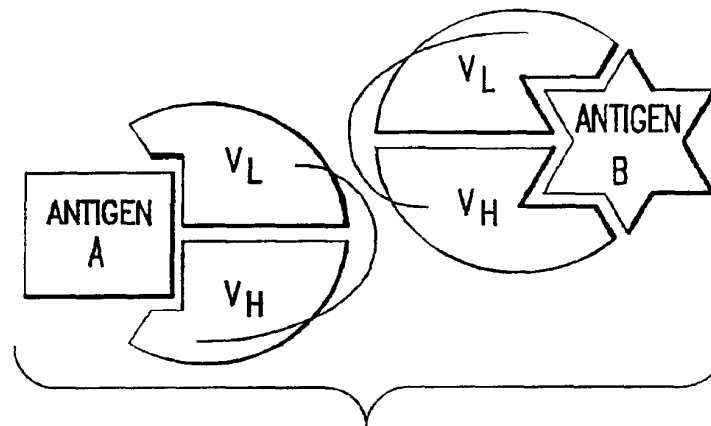


FIG.3C

· SUBSTITUTE SHEET

4/39

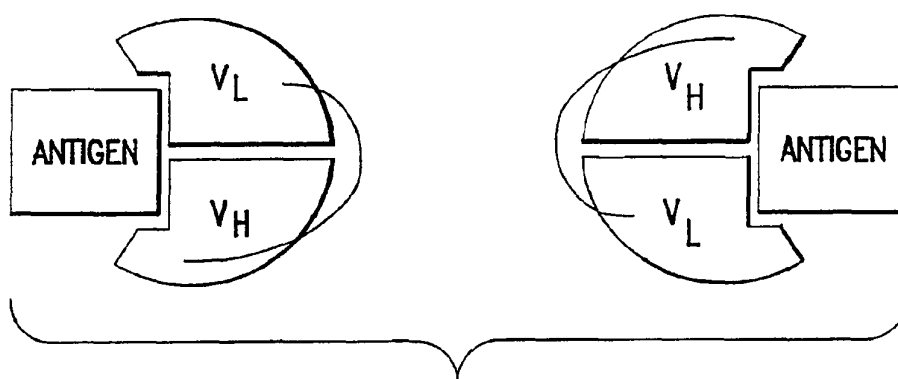


FIG.4A

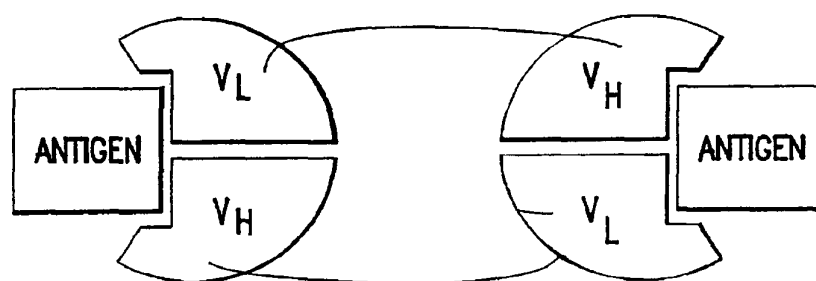


FIG.4B

SUBSTITUTE SHEET

5/39

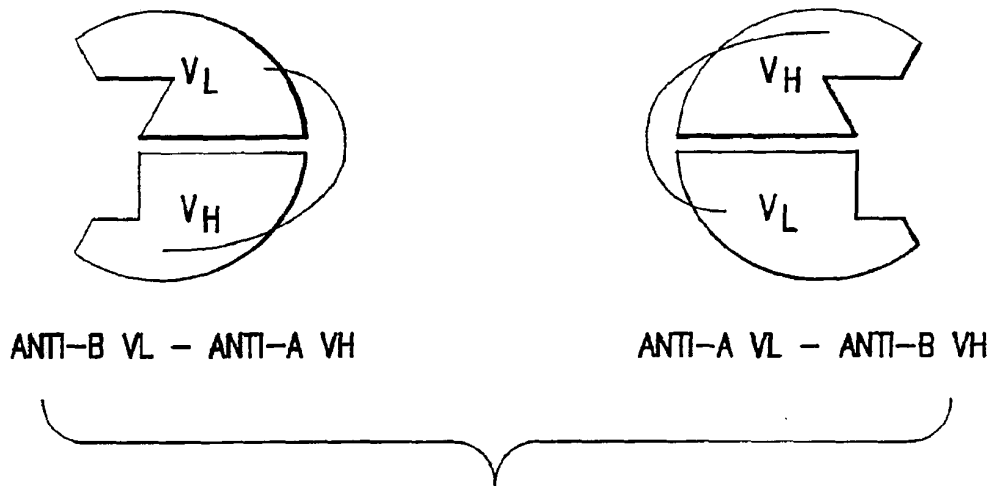


FIG.5A

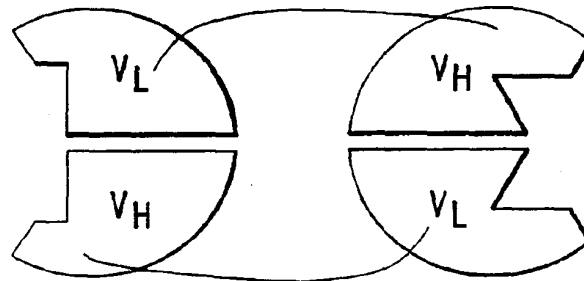


FIG.5B

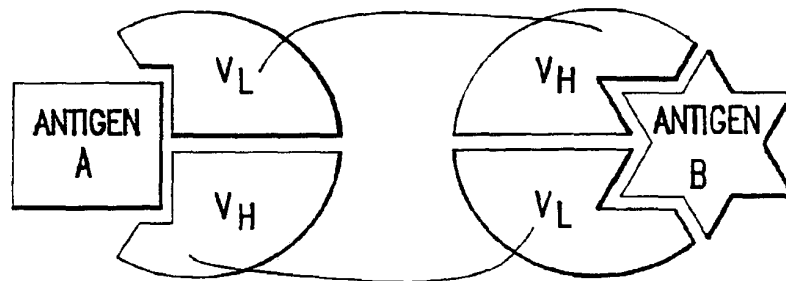


FIG.5C

SUBSTITUTE SHEET

6/39

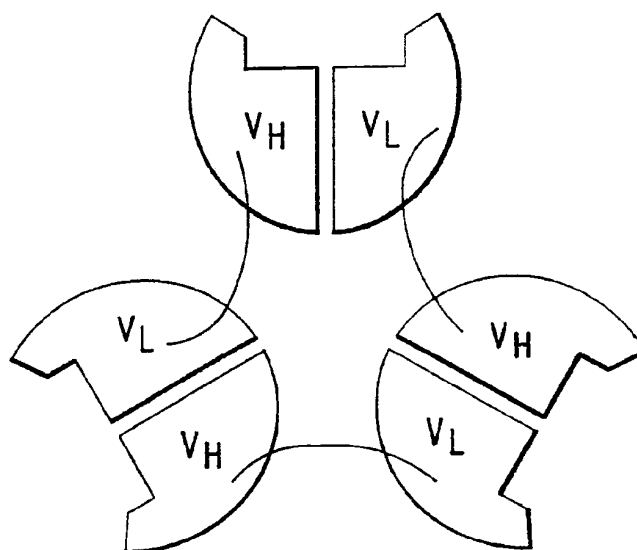


FIG.6A

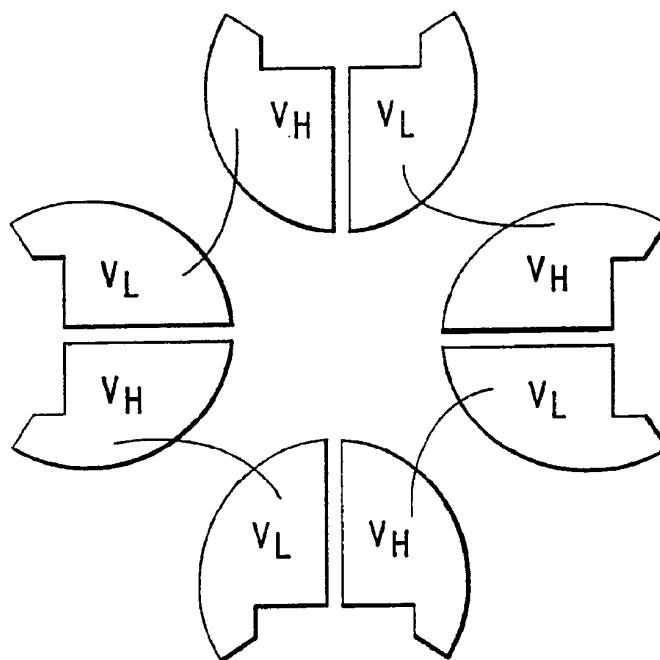
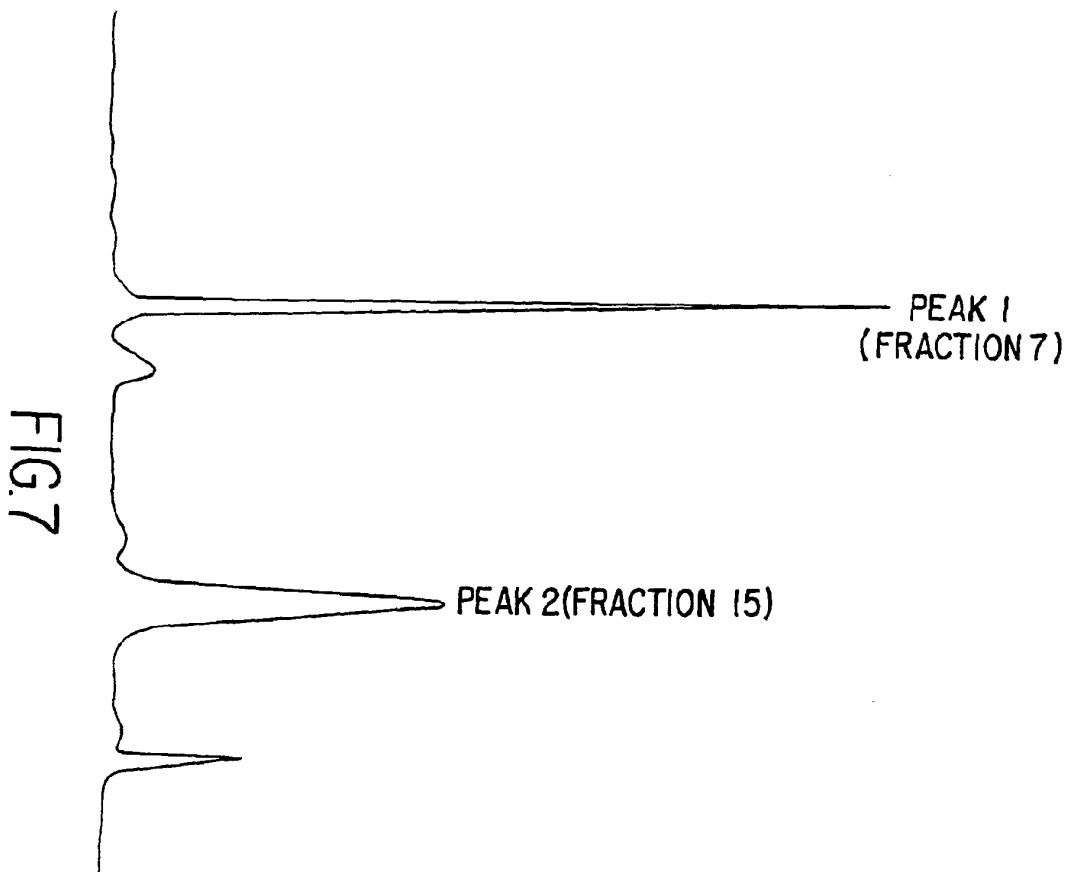


FIG.6B

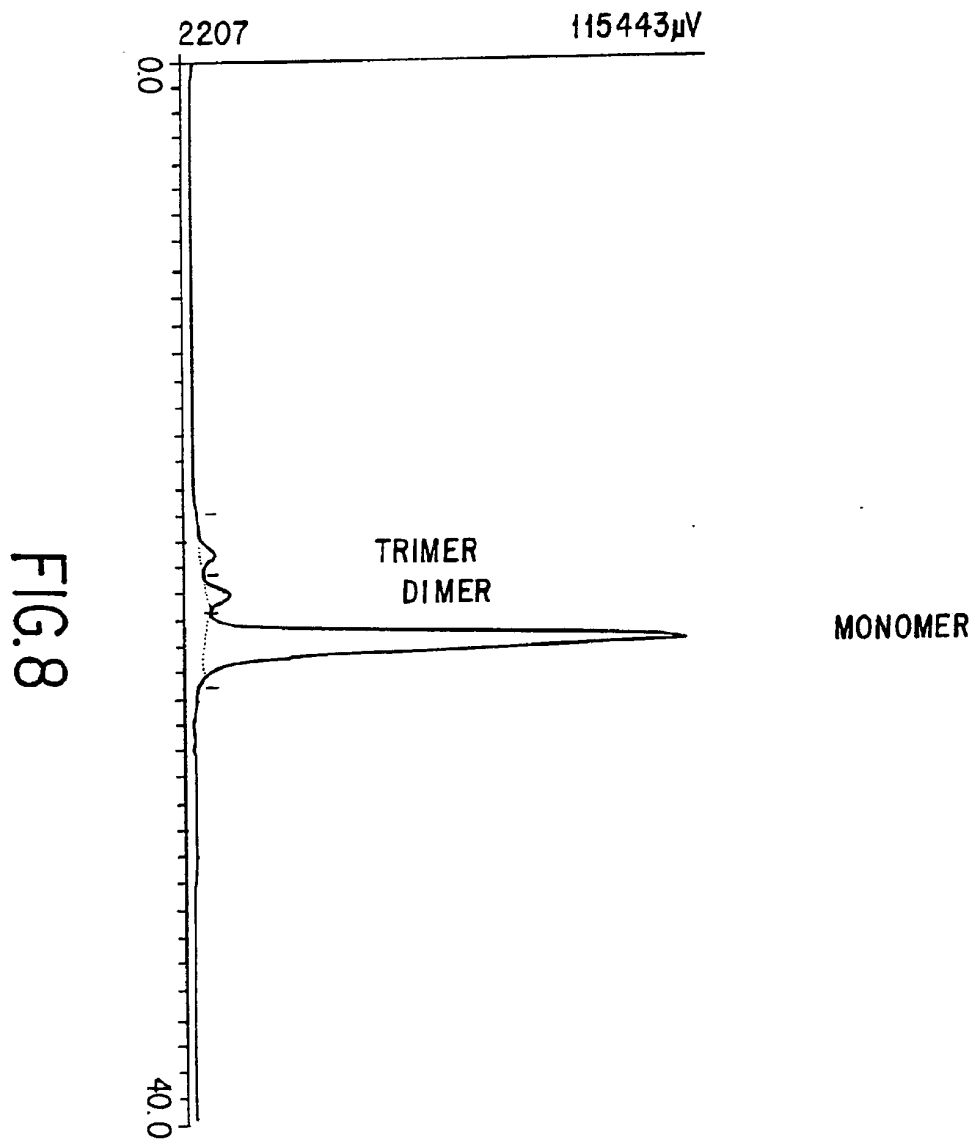
SUBSTITUTE SHEET

SUBSTITUTE SHEET



7/39

SUBSTITUTE SHEET



8/39

9/39

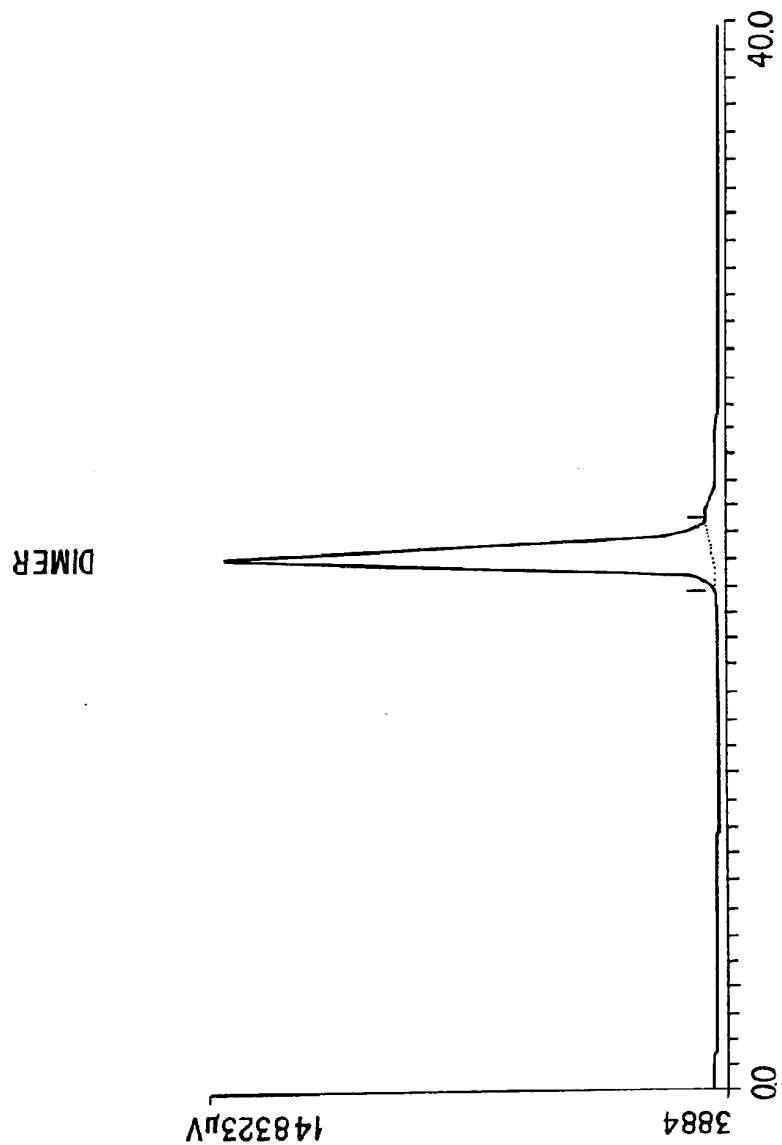


FIG. 9

SUBSTITUTE SHEET

10/39

4-4-20 V_L/212/CC49 V_H gene

4-4-20 V_L 10 20
 Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly Asp Gln Ala Ser
 GAC GTC GTT ATG ACT CAG ACA CCA CTA TCA CTT CCT GTT AGT CTA GGT GAT CAA GCC TCC
 Aat II

30 40
 Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser Asn Gly Asn Thr Tyr Leu Arg Trp
 ATC TCT TGC AGA TCT AGT CAG AGC CTT GTA CAC AGT AAT GGA AAC ACC TAT TTA CGT TGG

50 60
 Tyr Leu Gln Lys Pro Gly Gln Ser Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe
 TAC CTG CAG AAG CCA GGC CAG TCT CCA AAG GTC CTG ATC TAC AAA GTT TCC AAC CGA TTT

70 80
 Ser Gly Val Pro Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
 TCT GGG GTC CCA GAC AGG TTC AGT GGC AGT GGA TCA GGG ACA GAT TTC ACA CTC AAG ATC

90 100
 Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser Thr His Val Pro
 AGC AGA GTG GAG GCT GAG GAT CTG GGA GTT TAT TTC TGC TCT CAA AGT ACA CAT GTT CCG

110 212 Linker 120
 Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys Gly Ser Thr Ser Gly Ser Gly Lys
 TGG ACG TTC GGT GGA GGC ACC AAG CTT GAA ATC AAA GGT TCT ACC TCT GGT TCT GGT AAA
 Hind III

CC49 V_H 130 140
 Ser Ser Glu Gly Lys Gly Gln Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro
 TCC TCT GAA GGC AAA GGT CAG GTT CAG CTG CAG CAG TCT GAC GCT GAG TTG GTG AAA CCT
 PvuII PstI

150 160
 Gly Ala Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His Ala Ile
 GGG GCT TCA GTG AAG ATT TCC TGC AAG GCT TCT GGC TAC ACC TTC ACT GAC CAT GCA ATT

170 180
 His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser Pro Gly
 CAC TGG GTG AAA CAG AAC CCT GAA CAG GGC CTG GAA TGG ATT GGA TAT TTT TCT CCC GGA

FIG.10A

SUBSTITUTE SHEET

11/39

4-4-20 V_L/212/CC49 V_H gene

190 200
 Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala Thr Leu Thr Ala Asp Lys
 AAT GAT GAT TTT AAA TAC AAT GAG AGG TTC AAG GGC AAG GCC ACA CTG ACT GCA GAC AAA

210 220
 Ser Ser Ser Thr Ala Tyr Val Gln Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr
 TCC TCC AGC ACT GCC TAC GTG CAG CTC AAC AGC CTG ACA TCT GAG GAT TCT GCA GTG TAT

230 240
 TTC TGT ACA AGA TCC CTG AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA GTC ACC GTC TCC
 Phe Cys Thr Arg Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr Val Ser

*** ** Asp
TAA TAG GAT CC
 Bam H1

FIG.10A(CONT.)

SUBSTITUTE SHEET

12/39

CC49 VL/212/4-4-20 VH gene

CC49 VL 10 20
 Asp Val Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly Glu Lys Val Thr
 GAC GTC GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA GTT GGC GAG AAG GTT ACT
 Aat II

30 40
 Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala
 TTG AGC TGC AAG TCC AGT CAG AGC CTT TTA TAT AGT GGT AAT CAA AAG AAC TAC TTG GCC

50 60
 Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg
 TGG TAC CAG CAG AAA CCA GGG CAG TCT CCT AAA CTG CTG ATT TAC TGG GCA TCC GCT AGG

70 80
 Glu Ser Gly Val Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser
 GAA TCT GGG GTC CCT GAT CGC TTC ACA GGC AGT GGA TCT GGG ACA GAT TTC ACT CTC TCC

90 100
 Ile Ser Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr Ser Tyr
 ATC AGC AGT GTG AAG ACT GAA GAC CTG GCA GTT TAT TAC TGT CAG CAG TAT TAT AGC TAT

110 212 Linker 120
 Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu Lys Gly Ser Thr Ser Gly Ser Gly
 CCC CTC ACG TTC GGT GCT GGG ACC AAG CTT GTG CTG AAA GGC TCT ACT TCC GGT AGC GGC
 Hind III

4-4-20 VH 140
 Lys Ser Ser Glu Gly Lys Gly Glu Val Lys Leu Asp Glu Thr Gly Gly Gly Leu Val Gln
 AAA TCT TCT GAA GGT AAA GGT GAA GTT AAA CTG GAT GAG ACT GGA GGA GGC TTG GTG CAA

150 160
 Pro Gly Arg Pro Met Lys Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp
 CCT GGG AGG CCC ATG AAA CTC TCC TGT GTT GCC TCT GGA TTC ACT TTT AGT GAC TAC TGG

170 180
 Met Asn Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala Gln Ile Arg Asn
 ATG AAC TGG GTC CGC CAG TCT CCA GAG AAA GGA CTG GAG TGG GTA GCA CAA ATT AGA AAC

FIG.10B

13/39

CC49 V_L/212/4-4-20 V_H gene

190200

Lys Pro Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser Val Lys Gly Arg Phe Thr Ile Ser
 AAA CCT TAT AAT TAT GAA ACA TAT TAT TCA GAT TCT GTG AAA GGC AGA TTC ACC ATC TCA

210220

Arg Asp Asp Ser Lys Ser Ser Val Tyr Leu Gln Met Asn Asn Leu Arg Val Glu Asp Met
 AGA GAT GAT TCC AAA AGT AGT GTC TAC CTG CAA ATG AAC AAC TTA AGA GTT GAA GAC ATG

230240

Gly Ile Tyr Tyr Cys Thr Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser
 GGT ATC TAT TAC TGT ACG GGT TCT TAC TAT GGT ATG GAC TAC TGG GGT CAA GGA ACC TCA

Val Thr Val Ser * * Gly Ser
 GTC ACC GTC TCC TAA TAA GGA TCC
Bam HI

FIG.10B(CONT.)

SUBSTITUTE SHEET

14/39

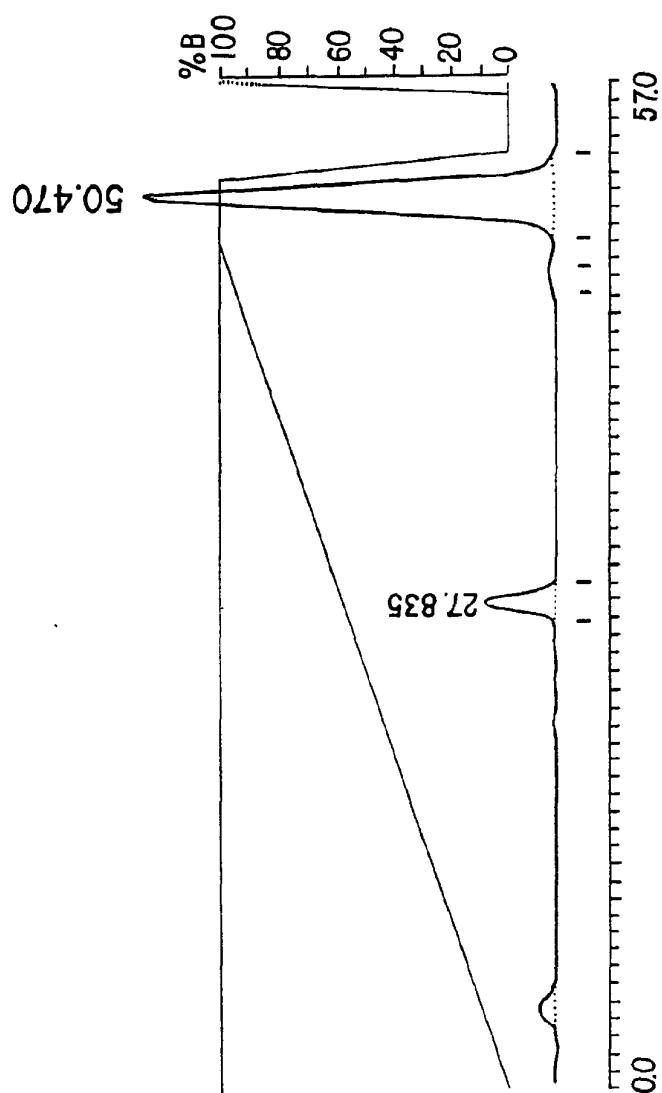


FIG.11

SUBSTITUTE SHEET

15/39

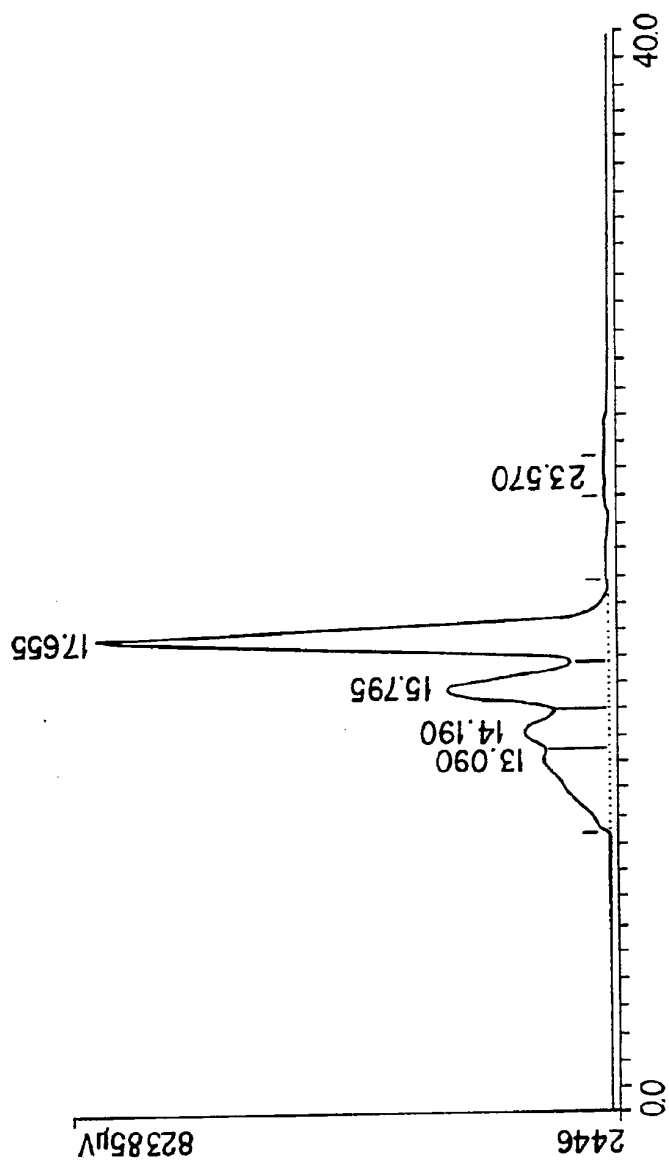


FIG. 12

SUBSTITUTE SHEET

16/39

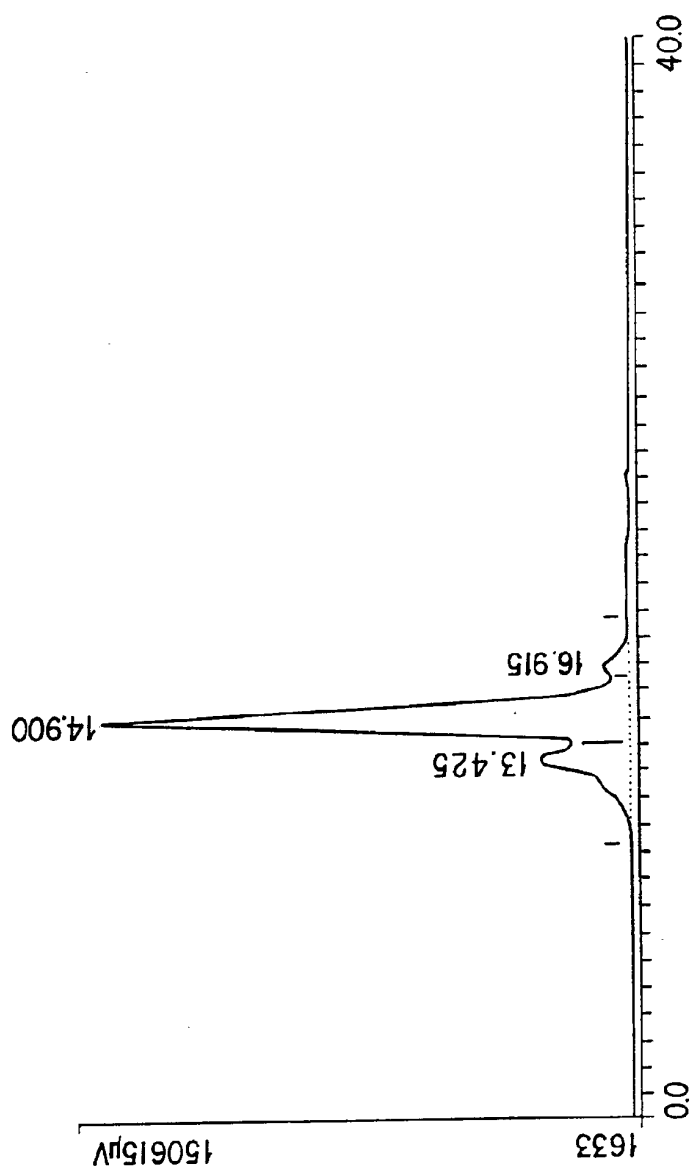


FIG.13

SUBSTITUTE SHEET

17/39

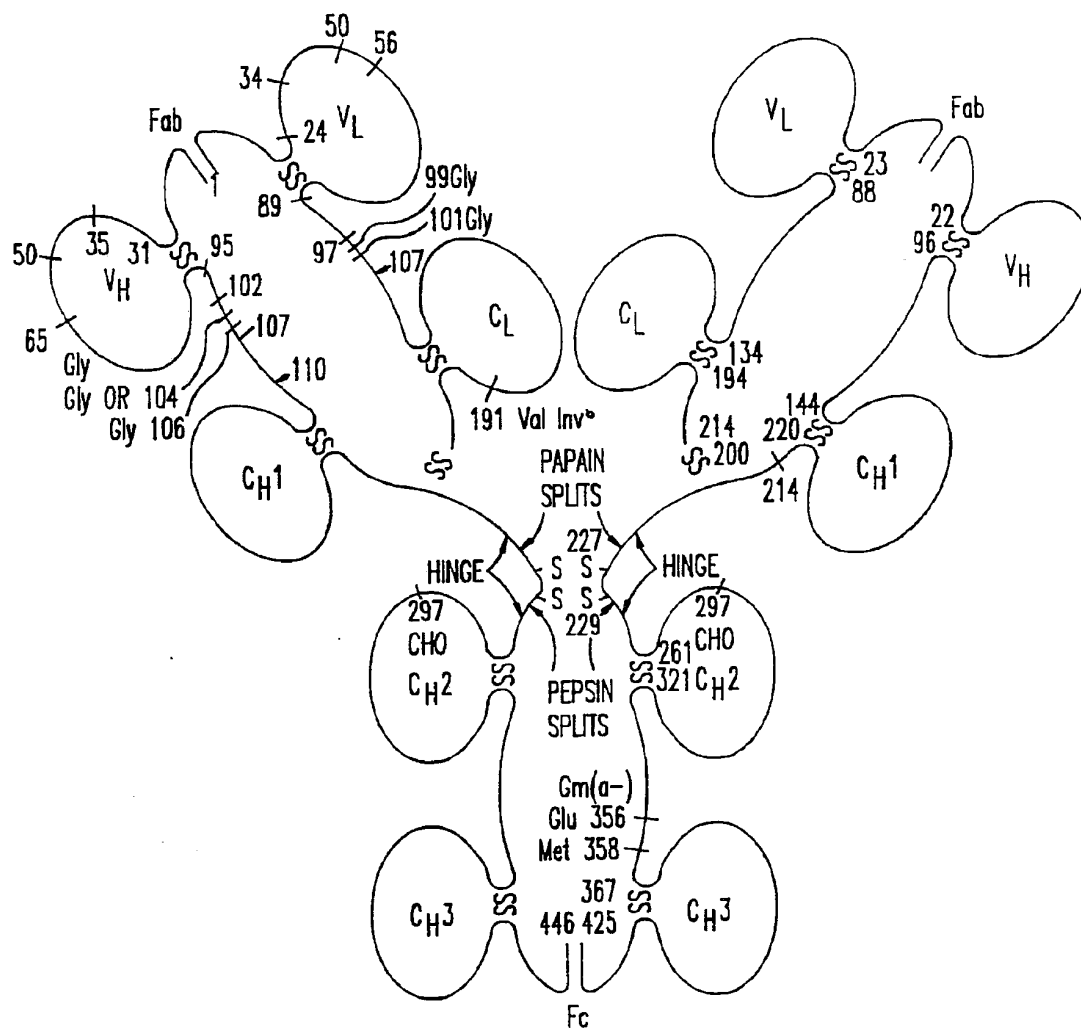


FIG.14

SUBSTITUTE SHEET

18/39

4-4-20/212 protein with single cysteine hinge

4-4-20 VL 10 20
 Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly Asp Gln Ala Ser
 GAC GTC GTT ATG ACT CAG ACA CCA CTA TCA CTT CCT GTT AGT CTA GGT GAT CAA GCC TCC
 Aat II
 30 40
 Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser Asn Gly Asn Thr Tyr Leu Arg Trp
 ATC TCT TGC AGA TCT AGT CAG AGC CTT GTA CAC AGT AAT GGA AAC ACC TAT TTA CGT TGG
 50 60
 Tyr Leu Gln Lys Pro Gly Gln Ser Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe
 TAC CTG CAG AAG CCA GGC CAG TCT CCA AAG GTC CTG ATC TAC AAA GTT TCC AAC CGA TTT
 70 80
 Ser Gly Val Pro Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
 TCT GGG GTC CCA GAC AGG TTC AGT GGC AGT GGA TCA GGG ACA GAT TTC ACA CTC AAG ATC
 90 100
 Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser Thr His Val Pro
 AGC AGA GTG GAG GCT GAG GAT CTG GGA GTT TAT TTC TGC TCT CAA AGT ACA CAT GTT CCG
 110 212 Linker 120
 Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys Gly Ser Thr Ser Gly Ser Gly Lys
 TGG ACG TTC GGT GGA GGC ACC AAG CTT GAA ATC AAA GGT TCT ACC TCT GGT TCT GGT AAA
 Hind III
 4-4-20 VH 140
 Ser Ser Glu Gly Lys Gly Glu Val Lys Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro
 TCT TCT GAA GGT AAA GGT GAA GTT AAA CTG GAT GAG ACT GGA GGA GGC TTG GTG CAA CCT
 150 160
 Gly Arg Pro Met Lys Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp Met
 GGG AGG CCC ATG AAA CTC TCC TGT GTT GCC TCT GGA TTC ACT TTT AGT GAC TAC TGG ATG
 170 180
 Asn Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala Gln Ile Arg Asn Lys
 AAC TGG GTC CGC CAG TCT CCA GAG AAA GGA CTG GAG TGG GTA GCA CAA ATT AGA AAC AAA

FIG.15A

SUBSTITUTE SHEET

19/39

4-4-20/212 protein with single cysteine hinge

	190		200
Pro Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg			
CCT TAT AAT TAT GAA ACA TAT TAT TCA GAT TCT GTG AAA GGC AGA TTC ACC ATC TCA AGA			
	210		220
Asp Asp Ser Lys Ser Ser Val Tyr Leu Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly			
GAT GAT TCC AAA AGT AGT GTC TAC CTG CAA ATG AAC AAC TTA AGA GTT GAA GAC ATG GGT			
	230		240
Ile Tyr Tyr Cys Thr Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser Val			
ATC TAT TAC TGT ACG GGT TCT TAC TAT GGT ATG GAC TAC TGG GGT CAA GGA ACC TCG GTC			
			Bst E11
	Hinge	250	
Thr Val Ser <u>Ser Asp Lys Thr His Thr Cys</u> *** **			
ACC GTC TCC AGT GAT AAG ACC CAT ACA TGC TAA TAG GAT CC			
			Bam H1

pGx 5532, Gx 8932

FIG.15A(CONT.)

SUBSTITUTE SHEET

20/39

4-4-20/212 protein with two cysteine hinge

4-4-20 V_L 10 20
 Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly Asp Gln Ala Ser
 GAC GTC GTT ATG ACT CAG ACA CCA CTA TCA CTT CCT GTT AGT CTA GGT GAT CAA GCC TCC
 Aat II

30 40
 Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser Asn Gly Asn Thr Tyr Leu Arg Trp
 ATC TCT TGC AGA TCT AGT CAG AGC CTT GTA CAC AGT AAT GGA AAC ACC TAT TTA CGT TGG

50 60
 Tyr Leu Gln Lys Pro Gly Gln Ser Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe
 TAC CTG CAG AAG CCA GGC CAG TCT CCA AAG GTC CTG ATC TAC AAA GTT TCC AAC CGA TTT

70 80
 Ser Gly Val Pro Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
 TCT GGG GTC CCA GAC AGG TTC AGT GGC AGT GGA TCA GGG ACA GAT TTC ACA CTC AAG ATC

90 100
 Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser Thr His Val Pro
 AGC AGA GTG GAG GCT GAG GAT CTG GGA GTT TAT TTC TGC TCT CAA AGT ACA CAT GTT CCG

110 212 Linker 120
 Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys Gly Ser Thr Ser Gly Ser Gly Lys
 TGG ACG TTC GGT GGA GGC ACC AAG CTT GAA ATC AAA GGT TCT ACC TCT GGT TCT GGT AAA
 Hind III

4-4-20 V_H 130 140
 Ser Ser Glu Gly Lys Gly Glu Val Lys Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro
 TCT TCT GAA GGT AAA GGT GAA GTT AAA CTG GAT GAG ACT GGA GGA GGC TTG GTG CAA CCT

150 160
 Gly Arg Pro Met Lys Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp Met
 GGG AGG CCC ATG AAA CTC TCC TGT GTT GCC TCT GGA TTC ACT TTT AGT GAC TAC TGG ATG

170 180
 Asn Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala Gln Ile Arg Asn Lys
 AAC TGG GTC CGC CAG TCT CCA GAG AAA GGA CTG GAG TGG GTA GCA CAA ATT AGA AAC AAA

FIG.15B

SUBSTITUTE SHEET

21/39

4-4-20/212 protein with two cysteine hinge

190 200
 Pro Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg
 CCT TAT AAT TAT GAA ACA TAT TAT TCA GAT TCT GTG AAA GGC AGA TTC ACC ATC TCA AGA

210 220
 Asp Asp Ser Lys Ser Ser Val Tyr Leu Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly
 GAT GAT TCC AAA AGT AGT GTC TAC CTG CAA ATG AAC AAC TTA AGA GTT GAA GAC ATG GGT

230 240
 Ile Tyr Tyr Cys Thr Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser Val
 ATC TAT TAC TGT ACG GGT TCT TAC TAT GGT ATG GAC TAC TGG GGT CAA GGA ACC TCG GTC
 Bst EII

Hinge 250
 Thr Val Ser Ser Asp Lys Thr His Thr Cys Pro Pro Cys *** ***
 ACC GTC TCC AGT GAT AAG ACC CAT ACA TGC CCT CCA TGC TAA TAG GAT CC
 Ban HI

pGx 5533, Gx 8933

FIG.15B(CONT.)

SUBSTITUTE SHEET

22/39

CC49/212 SCATM protein genetic dimer

CC49 V _L	10	20
Asp Val Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly Glu Lys Val Thr		
<u>GAC GTC</u> GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA GTT GGC GAG AAG GTT ACT		
Aat II		
	30	40
Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala		
TTG AGC TGC AAG TCC AGT CAG AGC CTT TTA TAT AGT GGT AAT CAA AAG AAC TAC TTG GCC		
	50	60
Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg		
TGG TAC CAG CAG AAA CCA GGG CAG TCT CCT AAA CTG CTG ATT TAC TGG GCA TCC GCT AGG		
	70	80
Glu Ser Gly Val Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser		
GAA TCT GGG GTC CCT GAT CGC TTC ACA GGC AGT GGA TCT GGG ACA GAT TTC ACT CTC TCC		
	90	100
Ile Ser Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr Ser Tyr		
ATC AGC AGT GTG AAG ACT GAA GAC CTG GCA GTT TAT TAC TGT CAG CAG TAT TAT AGC TAT		
	110	212 Linker
Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu Lys <u>Gly Ser Thr Ser Gly Ser Gly</u>		120
CCC CTC ACG TTC GGT GCT GGG ACC <u>AAG CTT</u> GTG CTG AAA GGC TCT ACT TCC GGT AGC GGC		
Hind III		
CC49 V _H		140
<u>Lys Ser Ser Glu Gly Lys Gly</u> Gln Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys		
<u>AAA TCC TCT GAA GGC AAA GGT CAG GTT</u> <u>CAG CTG CAG</u> CAG TCT GAC GCT GAG TTG GTG AAA		
PvuII PstI		
	150	160
Pro Gly Ala Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His Ala		
CCT GGG GCT TCA GTG AAG ATT TCC TGC AAG GCT TCT GGC TAC ACC TTC ACT GAC CAT GCA		
	170	180
Ile His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser Pro		
ATT CAC TGG GTG AAA CAG AAC CCT GAA CAG GGC CTG GAA TGG ATT GGA TAT TTT TCT CCC		

FIG.16A

SUBSTITUTE SHEET

23/39

CC49/212 SCATM protein genetic dimer

	190	200
Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala Thr Leu Thr Ala Asp		
GGA AAT GAT GAT TTT AAA TAC AAT GAG AGG TTC AAG GGC AAG GCC ACA CTG ACT GCA GAC		
	210	220
Lys Ser Ser Ser Thr Ala Tyr Val Gln Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val		
AAA TCC TCC AGC ACT GCC TAC GTG CAG CTC AAC AGC CTG ACA TCT GAG GAT TCT GCA GTG		
	230	240
Tyr Phe Cys Thr Arg Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr Val		
TAT TTC TGT ACA AGA TCC CTG AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA GTC ACC GTC		
CC49 VL	250	260
Ser Ser Asp Val Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly Glu Lys		
TCC TCA <u>GAC GTC</u> GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA GTT GGC GAG AAG		
Aat II		
	270	280
Val Thr Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr		
GTT ACT TTG AGC TGC AAG TCC AGT CAG AGC CTT TTA TAT AGT GGT AAT CAA AAG AAC TAC		
	290	300
Leu Ala Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser		
TTG GCC TGG TAC CAG CAG AAA CCA GGG CAG TCT CCT AAA CTG CTG ATT TAC TGG GCA TCC		
	310	320
Ala Arg Glu Ser Gly Val Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr		
GCT AGG GAA TCT GGG GTC CCT GAT CGC TTC ACA GGC AGT GGA TCT GGG ACA GAT TTC ACT		
	330	340
Leu Ser Ile Ser Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr		
CTC TCC ATC AGC AGT GTG AAG ACT GAA GAC CTG GCA GTT TAT TAC TGT CAG CAG TAT TAT		
	350	212 Linker 360
Ser Tyr Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu Lys <u>Gly Ser Thr Ser Gly</u>		
AGC TAT CCC CTC ACG TTC GGT GCT GGG ACC AAG CTT GTG CTG AAA <u>GGC TCT ACT TCC GGT</u>		
	Hind III	
CC49 VH		380
Ser Gly Lys Ser Ser Glu Gly Lys Gly Gln Val Gln Leu Gln Gln Ser Asp Ala Glu Leu		
AGC GGC AAA TCC TCT GAA GGC AAA GGT CAG GTT CAG CTG CAG CAG TCT GAC GCT GAG TTG		
	PvuII PstI	

FIG.16B

SUBSTITUTE SHEET

24/39

CC49/212 SCATM protein genetic dimer

390	400
Val Lys Pro Gly Ala Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp	
GTG AAA CCT GGG GCT TCA GTG AAG ATT TCC TGC AAG GCT TCT GGC TAC ACC TTC ACT GAC	
410	420
His Ala Ile His Trp Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr Phe	
CAT GCA ATT CAC TGG GTG AAA CAG AAC CCT GAA CAG GGC CTG GAA TGG ATT GGA TAT TTT	
430	440
Ser Pro Gly Asn Asp Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala Thr Leu Thr	
TCT CCC GGA AAT GAT GAT TTT AAA TAC AAT GAG AGG TTC AAG GGC AAG GCC ACA CTG ACT	
450	460
Ala Asp Lys Ser Ser Ser Thr Ala Tyr Val Gln Leu Asn Ser Leu Thr Ser Glu Asp Ser	
GCA GAC AAA TCC TCC AGC ACT GCC TAC GTG CAG CTC AAC AGC CTG ACA TCT GAG GAT TCT	
470	480
Ala Val Tyr Phe Cys Thr Arg Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val	
GCA GTG TAT TTC TGT ACA AGA TCC CTG AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA GTC	

Thr Val Ser *** *** Asp
 ACC GTC TCC TAA TAG GAT CC
 Bam HI

FIG.16C

SUBSTITUTE SHEET

25/39

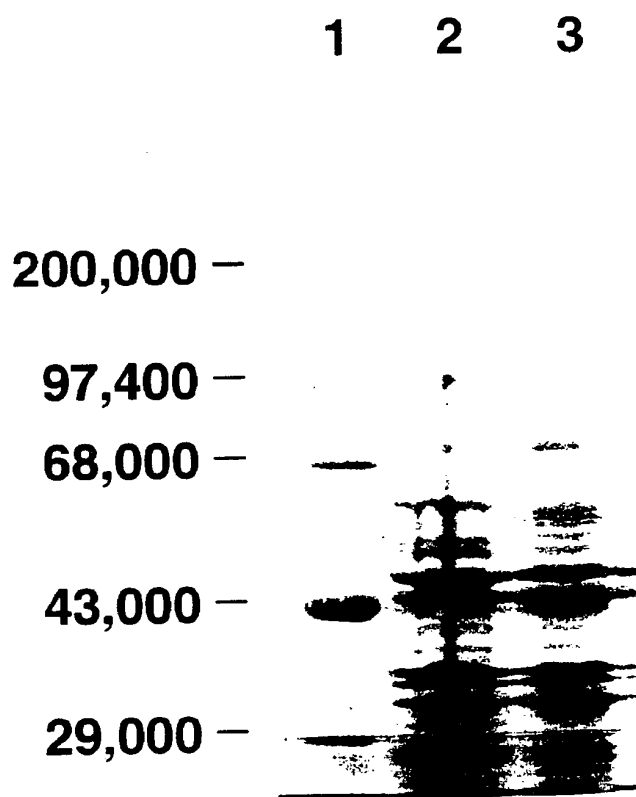


FIG. 17

SUBSTITUTE SHEET

26/39

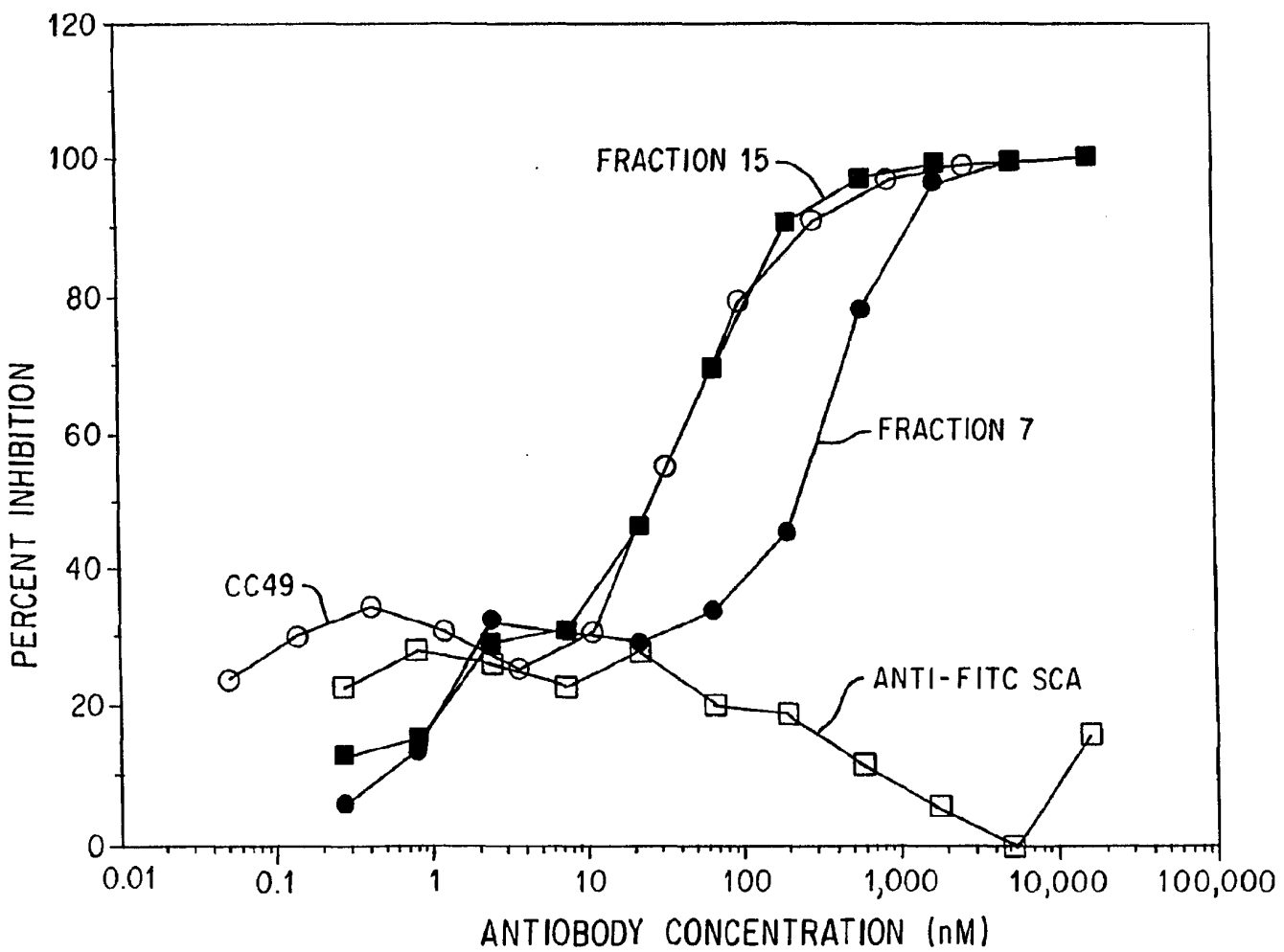


FIG. 18

27/39

4-4-20 V_L/217/CC49 V_H gene

4-4-20 V_L 10 20
 Asp Val Val Met Thr Gln Thr Pro Leu Ser Leu Pro Val Ser Leu Gly Asp Gln Ala Ser
GAC GTC GTT ATG ACT CAG ACA CCA CTA TCA CTT CCT GTT AGT CTA GGT GAT CAA GCC TCC
 Aat II

30 40
 Ile Ser Cys Arg Ser Ser Gln Ser Leu Val His Ser Asn Gly Asn Thr Tyr Leu Arg Trp
 ATC TCT TGC AGA TCT AGT CAG AGC CTT GTA CAC AGT AAT GGA AAC ACC TAT TTA CGT TGG

50 60
 Tyr Leu Gln Lys Pro Gly Gln Ser Pro Lys Val Leu Ile Tyr Lys Val Ser Asn Arg Phe
 TAC CTG CAG AAG CCA GGC CAG TCT CCA AAG GTC CTG ATC TAC AAA GTT TCC AAC CGA TTT

70 80
 Ser Gly Val Pro Asp Arg Phe Ser Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Lys Ile
 TCT GGG GTC CCA GAC AGG TTC AGT GGC AGT GGA TCA GGG ACA GAT TTC ACA CTC AAG ATC

90 100
 Ser Arg Val Glu Ala Glu Asp Leu Gly Val Tyr Phe Cys Ser Gln Ser Thr His Val Pro
 AGC AGA GTG GAG GCT GAG GAT CTG GGA GTT TAT TTC TGC TCT CAA AGT ACA CAT GTT CCG

110 217 Linker 120
 Trp Thr Phe Gly Gly Gly Thr Lys Leu Glu Ile Lys Gly Ser Thr Ser Gly Lys Pro Ser
 TGG ACG TTC GGT GGA GGC ACC AAG CTT GAA ATC AAA GGT TCT ACC TCT GGT AAA CCA TCT
 Hind III

CC49 V_H 130 140
 Glu Gly Lys Gly Gln Val Gln Leu Gln Gln Ser Asp Ala Glu Leu Val Lys Pro Gly Ala
 GAA GGC AAA GGT CAG GTT CAG CTG CAG CAG TCT GAC GCT GAG TTG GTG AAA CCT GGG GCT
 PvuII PstI

150 160
 Ser Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr Asp His Ala Ile His Trp
 TCA GTG AAG ATT TCC TGC AAG GCT TCT GGC TAC ACC TTC ACT GAC CAT GCA ATT CAC TGG

170 180
 Val Lys Gln Asn Pro Glu Gln Gly Leu Glu Trp Ile Gly Tyr Phe Ser Pro Gly Asn Asp
 GTG AAA CAG AAC CCT GAA CAG GGC CTG GAA TGG ATT GGA TAT TTT TCT CCC GGA AAT GAT

FIG.19A

SUBSTITUTE SHEET

28/39

4-4-20 VL/217/CC49 VH gene

	190		200
Asp Phe Lys Tyr Asn Glu Arg Phe Lys Gly Lys Ala Thr Leu Thr Ala Asp Lys Ser Ser			
GAT TTT AAA TAC AAT GAG AGG TTC AAG GGC AAG GCC ACA CTG ACT GCA GAC AAA TCC TCC			

	210		220
Ser Thr Ala Tyr Val Gln Leu Asn Ser Leu Thr Ser Glu Asp Ser Ala Val Tyr Phe Cys			
AGC ACT GCC TAC GTG CAG CTC AAC AGC CTG ACA TCT GAG GAT TCT GCA GTG TAT TTC TGT			

	230		240
Thr Arg Ser Leu Asn Met Ala Tyr Trp Gly Gln Gly Thr Ser Val Thr Val Ser *** **			
ACA AGA TCC CTG AAT ATG GCC TAC TGG GGT CAA GGA ACC TCA GTC ACC GTC TCC TAA TAG			

Asp
 GAT CC
 Bam HI

FIG.19A(CONT.)

SUBSTITUTE SHEET

29/39

CC49 V_L/217/4-4-20 V_H gene

CC49 V_L 10 20
 Asp Val Val Met Ser Gln Ser Pro Ser Ser Leu Pro Val Ser Val Gly Glu Lys Val Thr
GAC GTC GTG ATG TCA CAG TCT CCA TCC TCC CTA CCT GTG TCA GTT GGC GAG AAG GTT ACT
 Aat II
 30 40
 Leu Ser Cys Lys Ser Ser Gln Ser Leu Leu Tyr Ser Gly Asn Gln Lys Asn Tyr Leu Ala
 TTG AGC TGC AAG TCC AGT CAG AGC CTT TTA TAT AGT GGT AAT CAA AAG AAC TAC TTG GCC
 50 60
 Trp Tyr Gln Gln Lys Pro Gly Gln Ser Pro Lys Leu Leu Ile Tyr Trp Ala Ser Ala Arg
 TGG TAC CAG CAG AAA CCA GGG CAG TCT CCT AAA CTG CTG ATT TAC TGG GCA TCC GCT AGG
 70 80
 Glu Ser Gly Val Pro Asp Arg Phe Thr Gly Ser Gly Ser Gly Thr Asp Phe Thr Leu Ser
 GAA TCT GGG GTC CCT GAT CGC TTC ACA GGC AGT GGA TCT GGG ACA GAT TTC ACT CTC TCC
 90 100
 Ile Ser Ser Val Lys Thr Glu Asp Leu Ala Val Tyr Tyr Cys Gln Gln Tyr Tyr Ser Tyr
 ATC AGC AGT GTG AAG ACT GAA GAC CTG GCA GTT TAT TAC TGT CAG CAG TAT TAT AGC TAT
 110 217 Linker 120
 Pro Leu Thr Phe Gly Ala Gly Thr Lys Leu Val Leu Lys Gly Ser Thr Ser Gly Lys Pro
 CCC CTC ACG TTC GGT GCT GGG ACC AAG CTT GTG CTG AAA GGC TCT ACT TCC GGT AAA CCA
 Hind III
 4-4-20 V_H 130 140
 Ser Glu Gly Lys Gly Glu Val Lys Leu Asp Glu Thr Gly Gly Gly Leu Val Gln Pro Gly
 TCT GAA GGT AAA GGT GAA GTT AAA CTG GAT GAG ACT GGA GGA GGC TTG GTG CAA CCT GGG
 150 160
 Arg Pro Met Lys Leu Ser Cys Val Ala Ser Gly Phe Thr Phe Ser Asp Tyr Trp Met Asn
 AGG CCC ATG AAA CTC TCC TGT GTT GCC TCT GGA TTC ACT TTT AGT GAC TAC TGG ATG AAC
 170 180
 Trp Val Arg Gln Ser Pro Glu Lys Gly Leu Glu Trp Val Ala Gln Ile Arg Asn Lys Pro
 TGG GTC CGC CAG TCT CCA GAG AAA GGA CTG GAG TGG GTA GCA CAA ATT AGA AAC AAA CCT

FIG.19B

SUBSTITUTE SHEET

30/39

CC49 V_L/217/4-4-20 V_H gene

	190		200
Tyr Asn Tyr Glu Thr Tyr Tyr Ser Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Arg Asp			
TAT AAT TAT GAA ACA TAT TAT TCA GAT TCT GTG AAA GGC AGA TTC ACC ATC TCA AGA GAT			
	210		220
Asp Ser Lys Ser Ser Val Tyr Leu Gln Met Asn Asn Leu Arg Val Glu Asp Met Gly Ile			
GAT TCC AAA AGT AGT GTC TAC CTG CAA ATG AAC AAC TTA AGA GTT GAA GAC ATG GGT ATC			
	230		240
Tyr Tyr Cys Thr Gly Ser Tyr Tyr Gly Met Asp Tyr Trp Gly Gln Gly Thr Ser Val Thr			
TAT TAC TGT ACG GGT TCT TAC TAT GGT ATG GAC TAC TGG GGT CAA GGA ACC TCA GTC ACC			

Val Ser *** *** Gly Ser
 GTC TCC TAA TAA GGA TCC
 Bam HI

FIG.19B(CONT.)

SUBSTITUTE SHEET

31/39

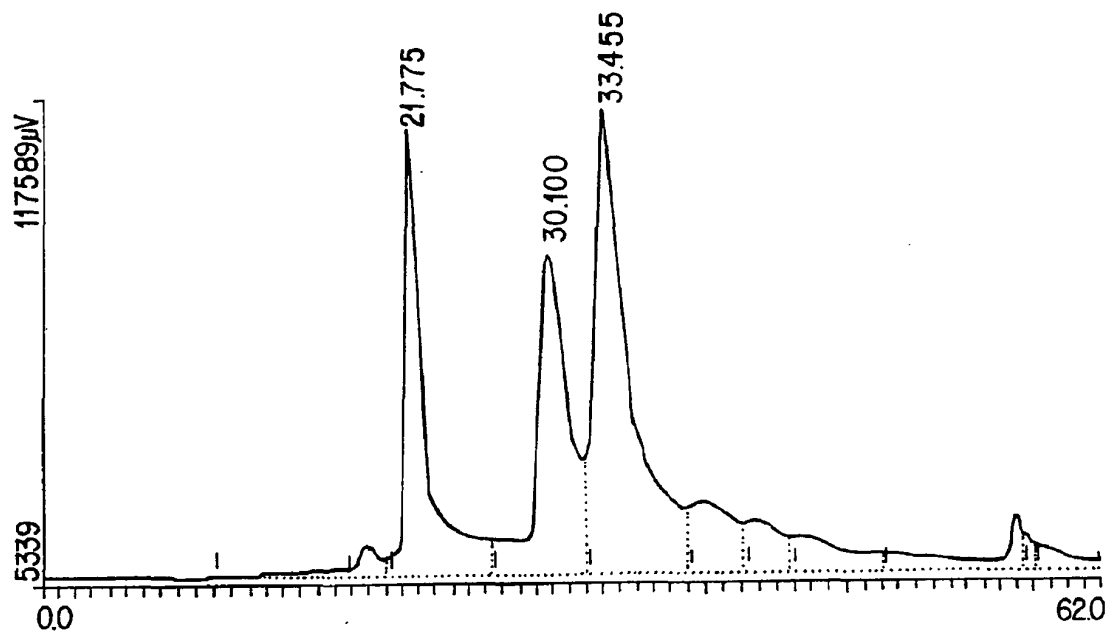
PROCESSING FILE: PolyCatA/Proc.CC-49Prep

METHOD: PREP POLY CAT A#2

INJECT VOL: 44

SAMPLING INT: 0.3 SECONDS

CHROMATOGRAM:



ANALYSIS: CHANNEL A

PEAK NO.	TIME	TYPE	HEIGHT(μV)	AREA(μV-SEC)	AREA%
1	17.090	N1	1651	348239	0.778
2	18.940	N2	8014	669441	1.496
3	21.775	N3	104401	8617252	19.263
4	30.100	N4	74925	9753616	21.804
5	33.455	N5	106864	15749605	35.208
6	38.940	N6	17296	2833701	6.334
7	42.010	N7	12645	1637917	3.661
8	44.640	N8	9287	1968584	4.400
9	57.055	N9	13767	2012338	4.498
10	57.610	N10	9323	210914	0.471
11	58.240	X11	6824	930855	2.080
TOTAL AREA				44732462	99.993

FIG.20

SUBSTITUTE SHEET

32/39

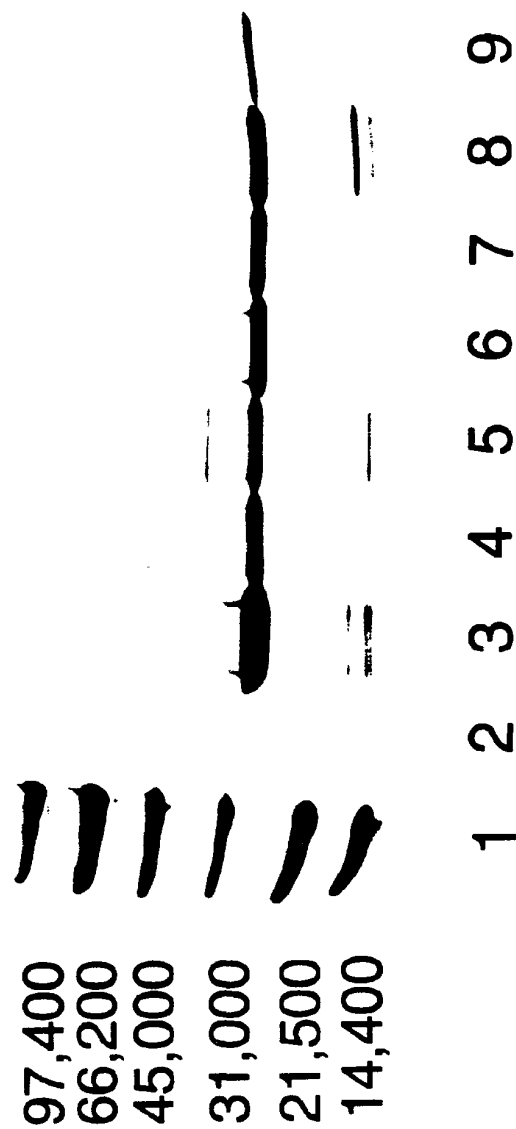


FIG. 21

SUBSTITUTE SHEET

33/39

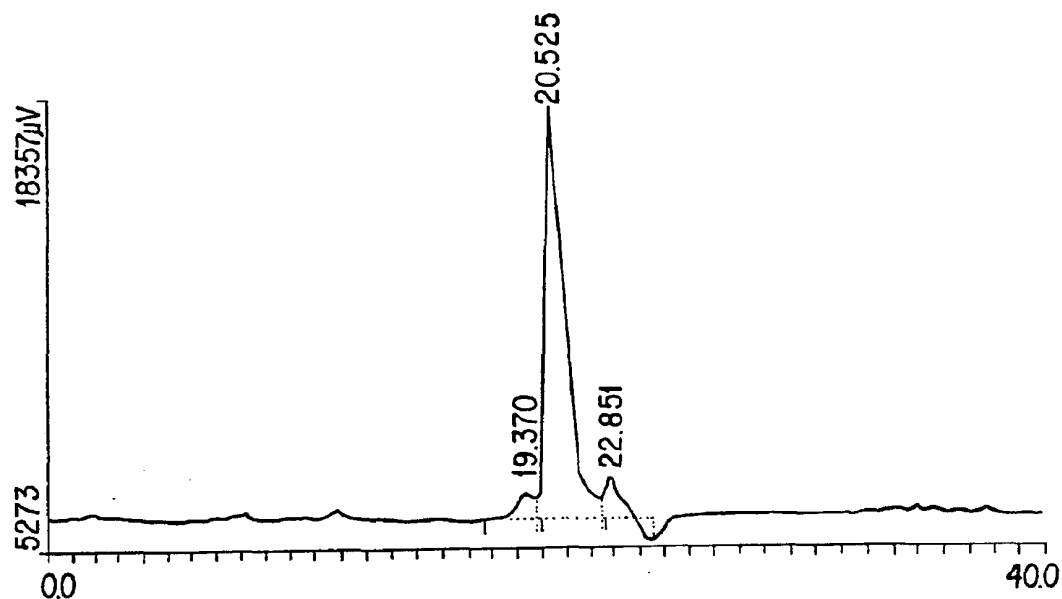
PROCESSING FILE: PolyCatA/Proc.CC-49Prep

METHOD: CC-49 QC SIZE-EXCLUSION

INJECT VOL: .05

SAMPLING INT: 0.1 SECONDS

CHROMATOGRAM:



ANALYSIS:		CHANNEL A			
PEAK NO.	TIME	TYPE	HEIGHT(μV)	AREA(μV-SEC)	AREA%
1	19.370	N1	797	41706	5.694
2	20.525	N2	11789	657280	89.737
3	22.851	N3	1227	33466	4.569
TOTAL AREA				732452	100.000

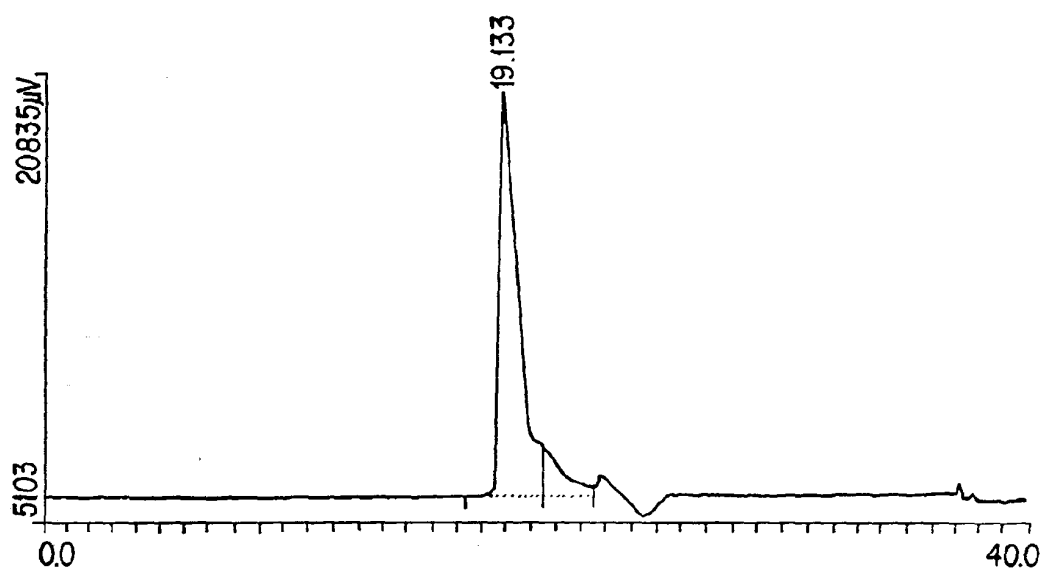
FIG.22A

SUBSTITUTE SHEET

34/39

PROCESSING FILE: PolyCatA/Proc.CC-49Prep
METHOD: CC-49 QC SIZE-EXCLUSION
INJECT VOL: .05
SAMPLING INT: 0.1 SECONDS

CHROMATOGRAM:



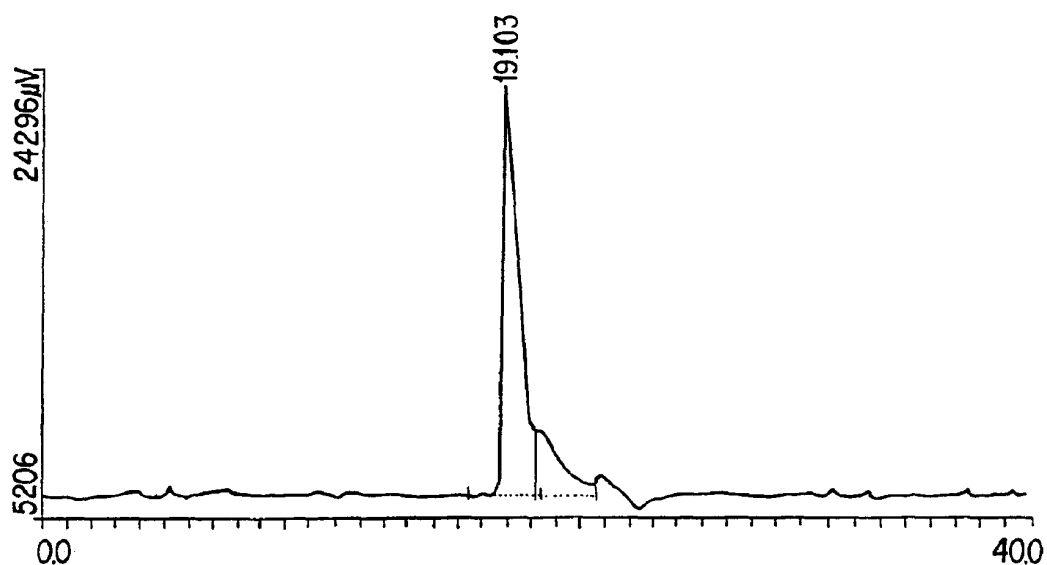
ANALYSIS:		CHANNEL A			
PEAK NO.	TIME	TYPE	HEIGHT(μV)	AREA(μV-SEC)	AREA%
1	19.133	N1	14211	749671	88.214
2	20.538	N2	1863	100154	11.785
TOTAL AREA				849825	99.999

FIG.22B

SUBSTITUTE SHEET

35/39

PROCESSING FILE: PolyCatA/Proc.CC-49Prep
METHOD: CC-49 QC SIZE-EXCLUSION
INJECT VOL: .05
SAMPLING INT: 0.1 SECONDS
CHROMATOGRAM:



ANALYSIS:		CHANNEL A			
PEAK NO.	TIME	TYPE	HEIGHT(μV)	AREA(μV-SEC)	AREA%
1	19.163	N1	17550	876502	83.039
2	20.435	N2	2981	179029	16.961
TOTAL AREA				1055531	100.000

FIG.22C

SUBSTITUTE SHEET

36/39

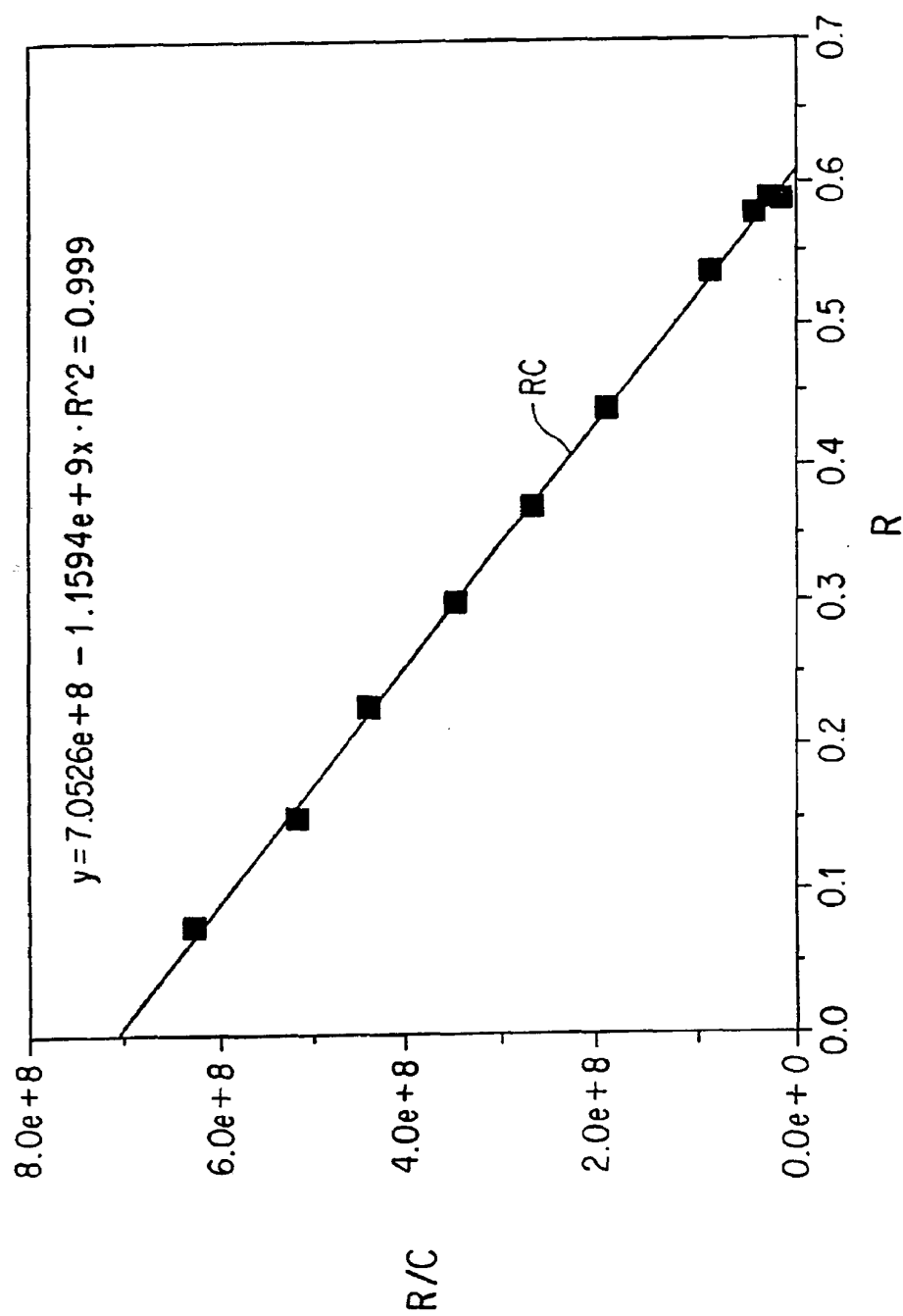


FIG. 23

SUBSTITUTE SHEET

37/39

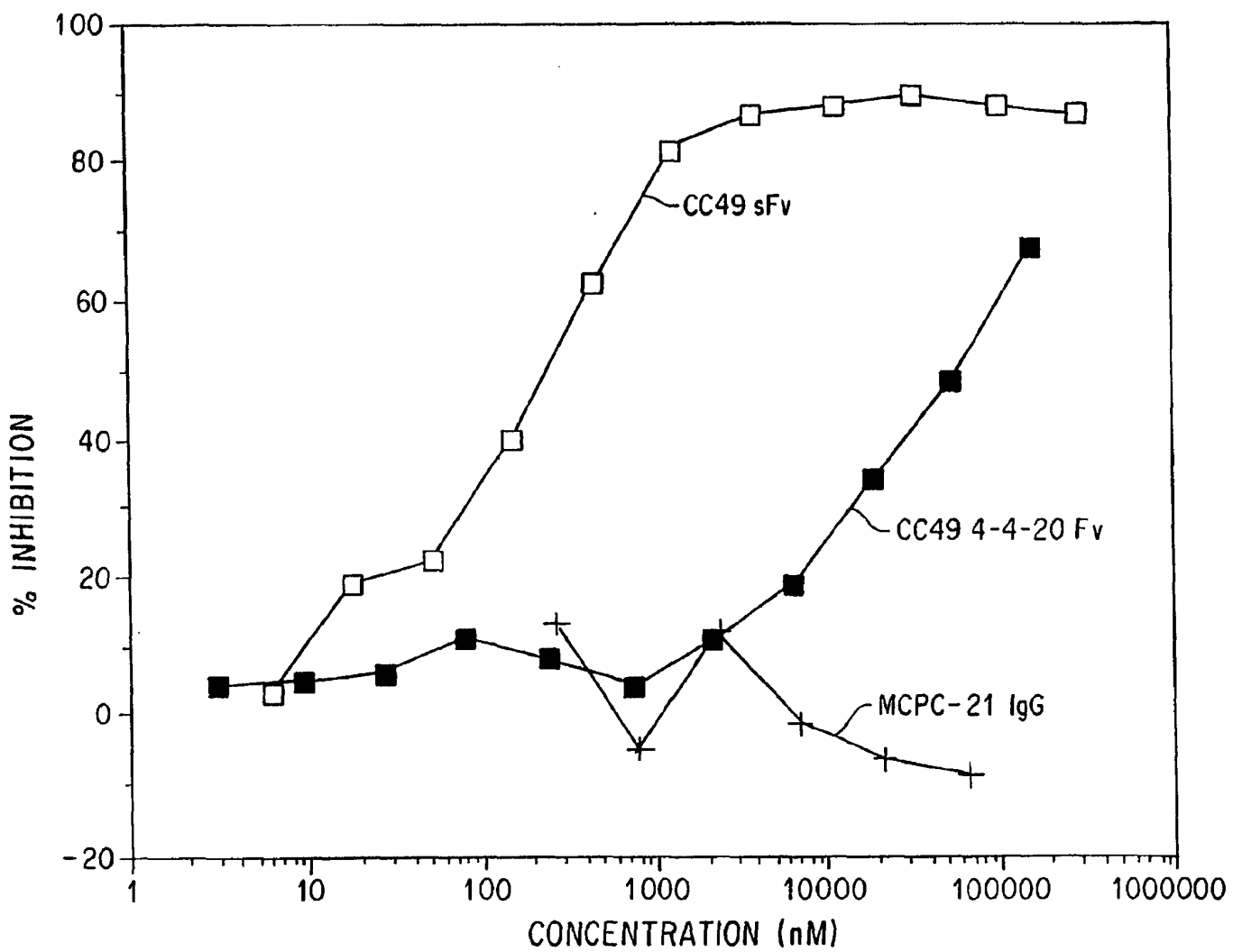


FIG. 24

38/39

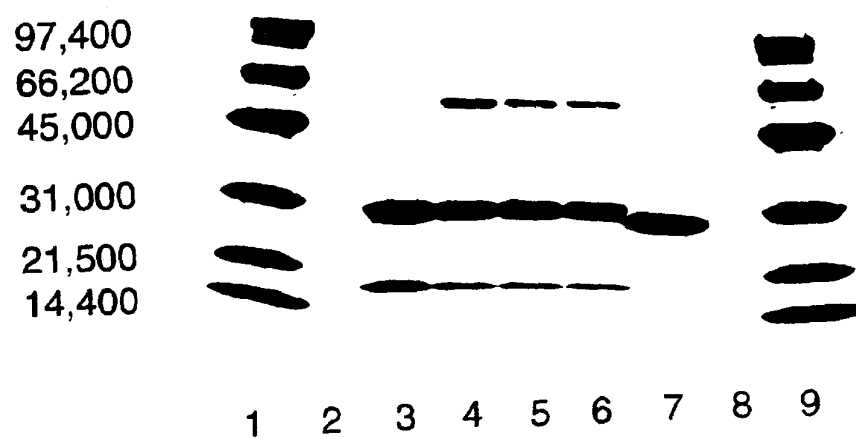


FIG. 25

SUBSTITUTE SHEET

39/39

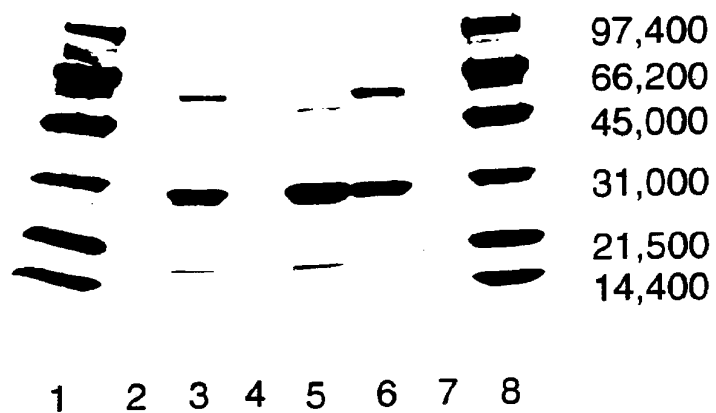


FIG. 26

SUBSTITUTE SHEET

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/09965**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(5) : C07K 15/28, 3/20; C07H 21/04; C12P 21/08; C12N 15/00, 15/03;

US CL : Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 530/387.3, 413; 435/7.92, 7.93, 7.94, 69.6, 69.7, 70.21, 172.2, 172.3, 240.27, 252.3, 320.1; 424/1.1, 85.8; 536/23.4, 23.53

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

DIALOG (FILES 5, 73, 155, 351); U.S. AUTOMATED PATENT SYSTEM (FILE USPAT, 1971-PRESENT).

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	WO 88/09344 (HUSTON et al.) 01 DECEMBER 1988, see entire document.	<u>1-33,38-39,44-63</u> 34-37,40-43
X Y	US, A, 4,946,778 (LADNER et al.) 07 AUGUST 1990, see entire document.	<u>32-33,38-39</u> 1-31,34-37,40-63
Y	CANCER RESEARCH, Vol. 48, issued 15 AUGUST 1988, Muraro et al., "Generation and Characterization of B72.3 Second Generation Monoclonal Antibodies Reactive with the Tumor-associated Glycoprotein 72 Antigen," pages 4588-4596, see entire document.	34-37,40-43
Y	SCIENCE, Vol. 242, issued 21 OCTOBER 1988, Bird et al., "Single-Chain Antigen-Binding Proteins," pages 423-426. See entire document.	32-42
Y	JOURNAL OF BIOLOGICAL CHEMISTRY, Vol. 265, No. 30, issued 25 OCTOBER 1990, Bedzyk et al., "Immunological and Structural Characterization of a High Affinity Anti-fluorescein Single-chain Antibody," pages 18615-18620. See entire document.	32-42

☐ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be part of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	
Date of the actual completion of the international search 23 FEBRUARY 1993	Date of mailing of the international search report 05 MAR 1993
Name and mailing address of the ISA/ Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	Authorized officer ROBERT D. BUDENS <i>Robert Budens</i>
Facsimile No. NOT APPLICABLE	Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/09965

A. CLASSIFICATION OF SUBJECT MATTER: US CL :

530/387.3, 413; 435/7.92, 7.93, 7.94, 69.6, 69.7, 172.3, 252.3, 320.1; 424/1.1, 85.8; 536/23.53

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

- I. Claims 1-14, 19-29, 31, 44-46, 51-58, a first product, method of making and method of using, drawn to multivalent antigen-binding proteins, compositions, methods of making multivalent proteins and method of using multivalent proteins to detect antigens, classified in Class 530, subclass 387.3 and Class 435, subclasses 7.1, 69.7, 172.3, 320.1, 252.3.
- II. Claims 15-18, a second product, drawn to compositions containing multivalent proteins and single chain proteins, classified in Class 530, subclass 387.3.
- III. Claim 30, a second method of use, directed to a method of imaging, classified in Class 424, subclass 85.8.
- IV. Claims 32-37, a third product, drawn to single chain proteins classified in Class 530, subclass 387.3.
- V. Claims 38-43 and 47-50, a fourth product, drawn to genetic sequences, vectors and hosts, classified in Class 536, subclass 23.53, Class 435, subclasses 320.1 and 252.3.
- VI. Claims 59-61, a third method of use, drawn to immunoassay methods, classified in Class 435, subclasses 7.92, 7.93 and 7.94.
- VII. Claim 62, a fourth method of use, drawn to a method of immunotherapy using multivalent proteins, classified in Class 424, subclass 85.8.
- VIII. Claim 63, a fifth method of use, drawn to a method of immunoaffinity purification using multivalent proteins, classified in Class 530, subclass 413.

The inventions of Groups I-II and IV-V are directed to multiple products which differ in their physical properties such as primary sequence, molecular weights and chemical composition and are not so linked as to have a common special technical feature.

Further, the methods of Groups I, III and VI-VIII differ in their utilities, reagents and method steps and are not so linked as to have a common special technical feature.



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁴ : C12P 21/00, C07H 17/00 C07K 15/04, A61K 39/395	A1	(11) International Publication Number: WO 88/ 01649 (43) International Publication Date: 10 March 1988 (10.03.88)
(21) International Application Number: PCT/US87/02208 (22) International Filing Date: 2 September 1987 (02.09.87) (31) Priority Application Number: 902,971 (32) Priority Date: 2 September 1986 (02.09.86) (33) Priority Country: US (71) Applicant: GENEX CORPORATION [US/US]; 16020 Industrial Drive, Gaithersburg, MD 20877 (US). (72) Inventors: LADNER, Robert, Charles ; 3827 Green Valley, Iamsville, MD 21754 (US). BIRD, Robert, Earl ; 3903 Morrell Court, Kensington, MD 20895 (US). (74) Agents: GOLDSTEIN, Jorge, A. et al.; Saidman, Sterne, Kessler & Goldstein, 1225 Connecticut Avenue, N.W., Washington, DC 20036 (US).	(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), DK, FI, FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), NO, SE (European patent). Published <i>With international search report.</i>	
(54) Title: SINGLE POLYPEPTIDE CHAIN BINDING MOLECULES (57) Abstract A single polypeptide chain binding molecule which has binding specificity substantially similar to the binding specificity of the light and heavy chain aggregate variable region of an antibody, genetic sequences coding therefor, recombinant DNA methods of producing such molecule and uses for such molecule.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	ML	Mali
AU	Australia	GA	Gabon	MR	Mauritania
BB	Barbados	GB	United Kingdom	MW	Malawi
BE	Belgium	HU	Hungary	NL	Netherlands
BG	Bulgaria	IT	Italy	NO	Norway
BJ	Benin	JP	Japan	RO	Romania
BR	Brazil	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	LI	Liechtenstein	SN	Senegal
CH	Switzerland	LK	Sri Lanka	SU	Soviet Union
CM	Cameroon	LU	Luxembourg	TD	Chad
DE	Germany, Federal Republic of	MC	Monaco	TG	Togo
DK	Denmark	MG	Madagascar	US	United States of America
FI	Finland				

SINGLE POLYPEPTIDE CHAIN BINDING MOLECULES

BACKGROUND OF THE INVENTION

This application is a continuation-in-part of Application Serial No. 902,971, filed September 2, 1986, the contents of which are herein fully incorporated by reference.

Field of the Invention

The present invention relates to single polypeptide chain binding molecules having the three dimensional folding, and thus the binding ability and specificity, of the variable region of an antibody. Methods of producing these molecules by genetic engineering are also disclosed.

Description of the Background Art

The advent of modern molecular biology and immunology has brought about the possibility of producing large quantities of biologically active materials in highly reproduceable form and with low cost. Briefly, the gene sequence coding for a desired natural protein is isolated, replicated (cloned) and introduced into a foreign host such as a bacterium, a yeast (or other fungi) or a mammalian cell line in culture, with appropriate regulatory control signals. When the signals are activated, the gene is transcribed and translated, and expresses the desired protein. In this manner, such useful biologically active materials as hormones, enzymes or antibodies have been cloned and expressed in foreign hosts.

-2-

One of the problems with this approach is that it is limited by the "one gene, one polypeptide chain" principle of molecular biology. In other words, a genetic sequence codes for a single polypeptide chain. Many biologically active polypeptides, however, are aggregates of two or more chains. For example, antibodies are three-dimensional aggregates of two heavy and two light chains. In the same manner, large enzymes such as aspartate transcarbamylase, for example, are aggregates of six catalytic and six regulatory chains, these chains being different. In order to produce such complex materials by recombinant DNA technology in foreign hosts, it becomes necessary to clone and express a gene coding for each one of the different kinds of polypeptide chains. These genes can be expressed in separate hosts. The resulting polypeptide chains from each host would then have to be reaggregated and allowed to refold together in solution. Alternatively, the two or more genes coding for the two or more polypeptide chains of the aggregate could be expressed in the same host simultaneously, so that refolding and reassociation into the native structure with biological activity will occur after expression. The approach, however, necessitates expression of multiple genes, and as indicated, in some cases, in multiple and different hosts. These approaches have proved to be inefficient.

Even if the two or more genes are expressed in the same organism it is quite difficult to get them all expressed in the required amounts.

A classical example of multigene expression to form multimeric polypeptides is the expression by recombinant DNA technology of antibodies. Genes for heavy and light chains have been introduced into ap-

-3-

propriate hosts and expressed, followed by reaggregation of these individual chains into functional antibody molecules (see for example Munro, Nature, 312:597 (1984); Morrison, S.L. Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986)); Wood et al., Nature, 314: 446-449 (1985)).

Antibody molecules have two generally recognized regions, in each of the heavy and light chains. These regions are the so-called "variable" region which is responsible for binding to the specific antigen in question, and the so-called "constant" region which is responsible for biological effector responses such as complement binding, etc. The constant regions are not necessary for antigen binding. The constant regions have been separated from the antibody molecule, and biologically active (i.e. binding) variable regions have been obtained.

The variable regions of an antibody are composed of a light chain and a heavy chain. Light and heavy chain variable regions have been cloned and expressed in foreign hosts, and maintain their binding ability (Moore et al., European Patent Publication 0088994 (published September 21, 1983)).

Further, it is by now well established that all antibodies of a certain class and their Fab fragments whose structures have been determined by X-ray crystallography, even when from different species, show closely similar variable regions despite large differences in the hypervariable segments. The immunoglobulin variable region seems to be tolerant toward mutations in the combining loops. Therefore, other than in the hypervariable regions, most of the so

-4-

called "variable" regions of antibodies, which are defined by both heavy and light chains, are in fact quite constant in their three dimensional arrangement. See, for example, Huber, R., "Structural Basis for Antigen-Antibody Recognition," Science, 233:702-703 (1986).

It would be very efficient if one could produce single polypeptide-chain molecules which have the same biological activity as the multiple chain aggregates such as, for example, multiple chain antibody aggregates or enzyme aggregates. Given the "one gene-one-polypeptide chain" principle, such single chain molecules would be more readily produceable, and would not necessitate multiple hosts or multiple genes in the cloning and expression. In order to accomplish this, it is first necessary to devise a method for generating single chain structures from two-chain aggregate structures, wherein the single chain will retain the three-dimensional folding of the separate natural aggregate of two polypeptide chains.

While the art has discussed the study of proteins in three dimensions, and has suggested modifying their architecture (see, for example, the article "Protein Architecture: Designing from the Ground Up," by Van Brunt, J., BioTechnology, 4: 277-283 (April, 1986)), the problem of generating single chain structures from multiple chain structures, wherein the single chain structure will retain the three-dimensional architecture of the multiple chain aggregate, has not been satisfactorily addressed.

Given that methods for the preparation of genetic sequences, their replication, their linking to expres-

-5-

sion control regions, formation of vectors therewith and transformation of appropriate hosts are well understood techniques, it would indeed be greatly advantageous to be able to produce, by genetic engineering, single polypeptide chain binding proteins having the characteristics and binding ability of multi chain variable regions of antibody molecules.

SUMMARY OF THE INVENTION

The present invention starts with a computer based system and method to determine chemical structures for converting two naturally aggregated but chemically separated light and heavy polypeptide chains from an antibody variable region into a single polypeptide chain which will fold into a three dimensional structure very similar to the original structure made of the two polypeptide chains.

The single polypeptide chain obtained from this method can then be used to prepare a genetic sequence coding therefor. The genetic sequence can then be replicated in appropriate hosts, further linked to control regions, and transformed into expression hosts, wherein it can be expressed. The resulting single polypeptide chain binding protein, upon refolding, has the binding characteristics of the aggregate of the original two (heavy and light) polypeptide chains of the variable region of the antibody.

The invention therefore comprises:

A single polypeptide chain binding molecule which has binding specificity substantially similar to the binding specificity of the light and heavy chain aggregate variable region of an antibody.

-6-

The invention also comprises genetic sequences coding for the above mentioned single polypeptide chain, cloning and expression vectors containing such genetic sequences, hosts transformed with such vectors, and methods of production of such polypeptides by expression of the underlying genetic sequences in such hosts.

The invention also extends to uses for the binding proteins, including uses in diagnostics, therapy, in vivo and in vitro imaging, purifications, and biosensors. The invention also extends to the single chain binding molecules in immobilized form, or in detectably labelled forms for utilization in the above mentioned diagnostic, imaging, purification or biosensor applications. It also extends to conjugates of the single polypeptide chain binding molecules with therapeutic agents such as drugs or specific toxins, for delivery to a specific site in an animal, such as a human patient.

Essentially all of the uses that the prior art has envisioned for monoclonal or polyclonal antibodies, or for variable region fragments thereof, can be considered for the molecules of the present invention.

The advantages of single chain over conventional antibodies are smaller size, greater stability and significantly reduced cost. The smaller size of single chain antibodies may reduce the body's immunologic reaction and thus increase the safety and efficacy of therapeutic applications. Conversely, the single chain antibodies could be engineered to be highly antigenic. The increased stability and lower cost permits greater use in biosensors and protein purification.

-7-

tion systems. Because it is a smaller and simpler protein, the single chain antibody is easier to further modify by protein engineering so as to improve both its binding affinity and its specificity. Improved affinity will increase the sensitivity of diagnosis and detection and detection systems while improved specificity will reduce the number of false positives observed.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention as defined in the claims can be better understood with reference to the text and to the following drawings, as follows:

Figure 1 is a block diagram of the hardware aspects of the serial processor mode of the present invention.

Figure 2 is a block diagram of an alternate embodiment of the hardware aspects of the present invention.

Figure 3 is a block diagram of the three general steps of the present invention.

Figure 4 is a block diagram of the steps in the site selection step in the single linker embodiment.

Figure 5A is a schematic two dimensional simplified representation of the light chain L and heavy chain H of two naturally aggregated antibody variable region F_v polypeptide chains used to illustrate the site selection process.

Figure 5B is a two dimensional representation of the three dimensional relationship of the two aggregated polypeptide chains showing the light chain L (----) and the heavy chain H (-) of the variable region of one antibody.

-8-

Figure 6A is a simplified two dimensional schematic diagram of the two polypeptide chains showing the location of the residue Tau 1 and the residue Sigma 1.

Figure 6B is a two dimensional representation of the actual relationship of the two polypeptide chains showing the residue Tau 1 and the residue Sigma 1.

Figure 7 shows in very simplified schematic way the concept of the direction linkers that are possible between the various possible sites on the light chain L and the heavy chain H in the residue Tau 1 and residue Sigma 1 respectively.

Figure 8A is a two dimensional simplified schematic diagram of a single chain antibody linking together two separate chains ((Heavy) and (light)) by linker 1 (----) to produce a single chain antibody.

Figure 8B is a two dimensional representation showing a single chain antibody produced by linking two aggregated polypeptide chains using linker 1.

Figure 9 shows a block diagram of candidate selection for correct span.

Figure 10 shows a block diagram of candidate selection for correct direction from N terminal to C terminal.

Figure 11 shows a comparison of direction of a gap to direction of a candidate.

Figure 12 shows a block diagram of candidate selection for correct orientation at both ends.

Figure 13 shows a block diagram of selection of sites for the two-linker embodiment.

Figure 14 shows examples of rules by which candidates may be ranked.

-9-

Figure 15A shows a two-dimensional simplified representation of the variable domain of an Fv light chain, L, and the variable domain of an Fv heavy chain, H, showing the first two sites to be linked.

Figure 15B shows a two-dimensional representation of the three-dimensional relationships between the variable domain of an Fv light chain, L, and the variable domain of an Fv heavy chain, H, showing the regions in which the second sites to be linked can be found and the linker between the first pair of sites.

Figure 16A shows the two-dimensional simplified representation of the variable domain of an Fv light chain, L, and the variable domain of an Fv heavy chain, H, showing the regions in which the second sites to be linked can be found and the linker between the first pair of sites.

Figure 16B shows the two-dimensional representation of the three-dimensional relationships between the variable domain of an Fv light chain, L, and the variable domain of an Fv heavy chain, H, showing the regions in which the second sites to be linked can be found and the linker between the first pair of sites.

Figure 17A shows the two-dimensional simplified representation of the variable domain of an Fv light chain, L, and the variable domain of an Fv heavy chain, H, showing the second linker and the portions of the native protein which are lost.

Figure 17B shows the two-dimensional representation of the three-dimensional relationships between the variable domain of an Fv light chain, L, and the variable domain of an Fv heavy chain, H, showing the second linker and the portions of native protein which are lost.

-10-

Figure 18 shows the two-dimensional simplified representation of the variable domain of an Fv light chain, L, and the variable domain of an Fv heavy chain, H, showing the complete construction.

Figure 19 shows a block diagram of the parallel processing mode of the present invention.

Figure 20A shows five pieces of molecular structure. The uppermost segment consists of two peptides joined by a long line. The separation between the peptides is 12.7 Å. The first C of each peptide lies on the X-axis. The two dots indicate the standard reference point in each peptide.

Below the gap are four linker candidates (labeled 1, 2, 3 & 4), represented by a line joining the alpha carbons. In all cases, the first and penultimate alpha carbons are on lines parallel to the X-axis, spaced 8.0 Å apart. Note that the space between dots in linker 1 is much shorter than in the gap.

Figure 20B shows the initial peptides of linkers 2, 3, and 4 which have been aligned with the first peptide of the gap. For clarity, the linkers have been translated vertically to their original positions.

The vector from the first peptide in the gap to the second peptide in the gap lies along the X-axis, a corresponding vector for linkers 3 and 4 also lies along the X-axis. Linker 2, however, has this vector pointing up and to the right, thus linker 2 is rejected.

Figure 20C shows the ten atoms which compose the initial and final peptides of linkers 3 and 4, which have been least-squares fit to the corresponding atoms from the gap. These peptides have been drawn in.

-11-

Note that in the gap and in linker 4 the final peptide points down and lies more-or-less in the plane of the paper. In linker 3, however, this final peptide points down and to the left and is twisted about 90 degrees so that the carbonyl oxygen points toward the viewer. Thus linker 3 is rejected.

Sections B and C are stereo diagrams which may be viewed with the standard stereo viewer provided.

Figure 21 shows the nucleotide sequence and translation of the sequence for the heavy chain of a mouse anti bovine growth hormone (BGH) monoclonal antibody.

Figure 22 shows the nucleotide sequence and translation of the sequence for the light chain of the same monoclonal antibody as that shown in Figure 21.

Figure 23 is a plasmid restriction map containing the variable heavy chain sequence (pGX3772) and that containing the variable light sequence (pGX3773) shown in figures 21 and 22.

Figure 24 shows construction TRY40 comprising the nucleotide sequence and its translation sequence of a single polypeptide chain binding protein prepared according to the methods of the invention.

Figure 25 shows a restriction map of the expression vector pGX3776 carrying a single chain binding protein, the sequence of which is shown in Figure 24. In this and subsequent plasmid maps (Figures 27 and 29) the hashed bar represents the promoter O_L/P_R sequence and the solid bar represents heavy chain variable region sequences.

Figure 26 shows the sequences of TRY61, another single chain binding protein of the invention.

-12-

Figure 27 shows expression plasmid pGX4904 carrying the genetic sequence shown in Figure 26.

Figure 28 shows the sequences of TRY59, another single chain binding protein of the invention.

Figure 29 shows the expression plasmid pGX 4908 carrying the genetic sequence shown in Figure 28.

Figures 30A, 30B, 30C, and 30D (stereo) are explained in detail in Example 1. They show the design and construction of double linked single chain antibody TRY40.

Figures 31A and 31B (stereo) are explained in detail in Example 2. They show the design and construction of single linked single chain antibody TRY61.

Figures 32A and 32B (stereo) are explained in detail in Example 3. They show the design and construction of single linked single chain antibody TRY59.

Figure 33 is explained in Example 4 and shows the sequence of TRY104b.

Figure 34 shows a restriction map of the expression vector pGX4910 carrying a single linker construction, the sequence of which is shown in Figure 33.

Figure 35 shows the assay results for BGH binding activity wherein strip one represents TRY61 and strip two represents TRY40.

Figure 36 is explained in Example 4 and shows the results of competing the F_{ab} portion of 3C2 monoclonal with TRY59 protein.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

TABLE OF CONTENTS

I. General Overview

-13-

II. Hardware and Software Environment**III. Single Linker Embodiment****A. Plausible Site Selection****B. Selection of Candidates**

1. Selecting Candidates with Proper Distance Between the N Terminal and the C Terminal.
2. Selecting Candidates with Proper Direction From the N Terminal and the C Terminal.
3. Selecting Candidates With Proper Orientation between the Termini.

C. Ranking and Eliminating Candidates**IV. Double and Multiple Linker Embodiments****A. Plausible Site Selection****B. Candidate Selection and Candidate Rejection Steps****V. Parallel Processing Embodiment****VI. Preparation and Expression of Genetic Sequences and Uses****I. General Overview**

The present invention starts with a computer based system and method for determining and displaying possible chemical structures (linkers) for converting two naturally aggregated but chemically separate heavy and light (H and L) polypeptide chains from the variable region of a given antibody into a single polypeptide chain which will fold into a three dimensional struc-

-14-

ture very similar to the original structure made of two polypeptide chains. The original structure is referred to hereafter as "native protein."

The first general step of the three general design steps of the present invention involves selection of plausible sites to be linked. In the case of a single linker, criteria are utilized to select a plausible site on each of the two polypeptide chains (H and L in the variable region) which will result in 1) a minimum loss of residues from the native protein chains and 2) a linker of minimum number of amino acids consistent with the need for stability. A pair of sites defines a gap to be bridged or linked.

A two-or-more-linker approach is adopted when a single linker can not achieve the two stated goals. In both the single-linker case and the two-or-more-linker case, more than one gap may be selected for use in the second general step.

The second general step of the present invention involves examining a data base to determine possible linkers to fill the plausible gaps selected in the first general step, so that candidates can be enrolled for the third general step. Specifically, a data base contains a large number of amino acid sequences for which the three-dimensional structure is known. In the second general step, this data base is examined to find which amino acid sequences can bridge the gap or gaps to create a plausible one-polypeptide structure which retains most of the three dimensional features of the native (i.e. original aggregate) variable region molecule. The testing of each possible linker proceeds in three general substeps. The first general substep utilizes the length of the possible candidate.

-15-

Specifically, the span or length (a scalar quantity) of the candidate is compared to the span of each of the gaps. If the difference between the length of the candidate and the span of any one of the gaps is less than a selected quantity, then the present invention proceeds to the second general substep with respect to this candidate. Figure 20A shows one gap and four possible linkers. The first linker fails the first general substep because its span is quite different from the span of the gap.

In the second general substep, called the direction substep, the initial peptide of the candidate is aligned with the initial peptide of each gap. Specifically, a selected number of atoms in the initial peptide of the candidate are rotated and translated as a rigid body to best fit the corresponding atoms in the initial peptide of each gap. The three dimensional vector (called the direction of the linker) from the initial peptide of the candidate linker to the final peptide of the candidate linker is compared to the three dimensional vector (call the direction of the gap) from the initial peptide of each gap to the final peptide of the same gap. If the ends of these two vectors come within a preselected distance of each other, the present invention proceeds to the third general substep of the second general step with respect to this candidate linker.

Figure 20B shows one gap and three linkers. All the linkers have the correct span and the initial peptides have been aligned. The second linker fails the second general substep because its direction is quite different from that of the gap; the other two linkers are carried forward to the third general substep of the second general step.

-16-

In the third general substep of the second design of the step of the present invention, the orientations of the terminal peptides of each linker are compared to the orientations of the terminal peptides of each gap. Specifically, a selected number of atoms (3, 4, or 5, 5 in the preferred embodiment) from the initial peptide of the candidate plus the same selected number of atoms (3, 4, or 5; 5 in the preferred embodiment) from the final peptide of the candidate are taken as a rigid body. The corresponding atoms from one of the gaps (viz 5 from the initial peptide and 5 from the final peptide) are taken as a second rigid body. These two rigid bodies are superimposed by a least-squares fit. If the error for this fit is below some preselected value, then the candidate passes the third general substep of the second general step and is enrolled for the third general step of the present invention. If the error is greater than or equal to the preselected value, the next gap is tested. When all gaps have been tested without finding a sufficiently good fit, the candidate is abandoned.

The third general step of the present invention results in the ranking of the linker candidates from most plausible to least plausible. The most plausible candidate is the fragment that can bridge the two plausible sites of one of the gaps to form a single polypeptide chain, where the bridge will least distort the resulting three dimensional folding of the single polypeptide chain from the natural folding of the aggregate of the two originally chemically separate chains.

In this third general step of the present invention, an expert operator uses an interactive computer-graphics approach to rank the linker candidates from

-17-

most plausible to least plausible. This ranking is done by observing the interactions between the linker candidate with all retained portions of the native protein. A set of rules are used for the ranking. These expert system rules can be built into the system so that the linkers are displayed only after they have satisfied the expert system rules that are utilized.

The present invention can be programmed so that certain expert rules are utilized as a first general substep in the third general step to rank candidates and even eliminate unsuitable candidates before visual inspection by an expert operator, which would be the second general substep of the third general step. These expert rules assist the expert operator in ranking the candidates from most plausible to least plausible. These expert rules can be modified based on experimental data on linkers produced by the system and methods of the present invention.

The most plausible candidate is a genetically producible single polypeptide chain binding molecule which has a very significantly higher probability (a million or more as compared to a random selection) of folding into a three dimensional structure very similar to the original structure made of the heavy and light chains of the antibody variable region than would be produced if random selection of the linker was done. In this way, the computer based system and method of the present invention can be utilized to engineer single polypeptide chains by using one or more linkers which convert naturally aggregated but chemically separated polypeptide chains into the desired single chain.

The elected candidate offers to the user a linked chain structure having a very significantly increased

-18-

probability of proper folding than would be obtained using a random selection process. This means that the genetic engineering aspect of creating the desired single polypeptide chain is significantly reduced, since the number of candidates that have to be genetically engineered in practice is reduced by a corresponding amount. The most plausible candidate can be used to genetically engineer an actual molecule.

The parameters of the various candidates can be stored for later use. They can also be provided to the user either visually or recorded on a suitable media (paper, magnetic tape, color slides, etc.). The results of the various steps utilized in the design process can also be stored for later use or examination.

The design steps of the present invention operate on a conventional minicomputer system having storage devices capable of storing the amino acid sequence-structure data base, the various application programs utilized and the parameters of the possible linker candidates that are being evaluated.

The minicomputer CPU is connected by a suitable serial processor structure to an interactive computer-graphics display system. Typically, the interactive computer-graphics display system comprises a display terminal with resident three-dimensional application software and associated input and output devices, such as X/Y plotters, position control devices (potentiometers, an x-y tablet, or a mouse), and keyboard.

The interactive computer-graphics display system allows the expert operator to view the chemical structures being evaluated in the design process of the

-19-

present invention. Graphics and programs are used to select the gaps (Gen. Step 1), and to rank candidates (Gen. Step 3). Essentially, it operates in the same fashion for the single linker embodiment and for the two or more linker embodiments.

For example, during the first general step of the present invention, the computer-graphics interactive display system allows the expert operator to visually display the two naturally aggregated but chemically separate polypeptide chains. Using three dimensional software resident in the computer-graphics display system, the visual representation of the two separate polypeptide chains can be manipulated as desired. For example, the portion of the chain(s) being viewed can be magnified electronically, and such magnification can be performed in a zoom mode. Conversely, the image can be reduced in size, and this reduction can also be done in a reverse zoom mode. The position of the portion of the molecule can be translated, and the displayed molecule can be rotated about any one of the three axes (x, y and z). Specific atoms in the chain can be selected with an electronic pointer. Selected atoms can be labeled with appropriate text. Specific portions of native protein or linker can be identified with color or text or brightness. Unwanted portions of the chain can be erased from the image being displayed so as to provide the expert operator with a visual image that represents only a selected aspect of the chain(s). Atoms selected by pointing or by name can be placed at the center of the three dimensional display; subsequent rotation uses the selected atom as the origin. These and other display aspects provide

-20-

the expert operator with the ability to visually represent portions of the chains which increase the ability to perform the structural design process.

One of the modes of the present invention utilizes a serial computational architecture. This architecture using the present equipment requires approximately four to six hours of machine and operator time in order to go through the various operations required for the three general steps for a particular selection of gaps. Obviously, it would be desirable to significantly reduce the time since a considerable portion thereof is the time it takes for the computer system to perform the necessary computational steps.

An alternate embodiment of the present invention utilizes a parallel processing architecture. This parallel processing architecture significantly reduces the time required to perform the necessary computational steps. A hypercube of a large number of nodes can be utilized so that the various linkers that are possible for the selected sites can be rapidly presented to the expert system operator for evaluation.

Since there are between 200 and 300 known protein structures, the parallel processing approach can be utilized. There currently are computers commercially available that have as many as 1,024 computing nodes.

Using a parallel processing approach, the data base of observed peptide structures can be divided into as many parts as there are computing nodes. For example, if there are structures for 195 proteins with 219 amino acids each, one would have structures for 195x218 dipeptides, 195x217 tripeptides, 195x216 tetrapeptides, etc. One can extract all peptides up to

-21-

some length n . For example, if n were 30, one would have $195 \times 30 \times 204$ peptides. Of course, proteins vary in length, but with 100 to 400 proteins of average length 200 (for example), and for peptide linkers up to length 30 amino acids (or any other reasonable number), one will have between 1,000,000 and 4,000,000 peptide structures. Once the peptides have been extracted and labeled with the protein from which they came, one is free to divide all the peptides as evenly as possible among the available computing nodes.

The parallel processing mode operates as follows. The data base of known peptides is divided among the available nodes. Each gap is sent to all the nodes. Each node takes the gap and tests it against those peptides which have been assigned to it and returns information about any peptides which fit the gap and therefore are candidate linkers. As the testing for matches between peptides and gaps proceeds independently in each node, the searching will go faster by a factor equal to the number of nodes.

A first embodiment of the present invention utilizes a single linker to convert the naturally aggregated but chemically separate heavy and light chains into a single polypeptide chain which will fold into a three dimensional structure very similar to the original structure made of two polypeptide chains.

A second embodiment utilizes two or more linkers to convert the two heavy and light chains into the desired single polypeptide chain. The steps involved in each of these embodiments utilizing the present invention are illustrated in the explanation below.

-22-

Once the correct amino acid sequence for a single chain binding protein has been defined by the computer assisted methodology, it is possible, by methods well known to those with skill in the art, to prepare an underlying genetic sequence coding therefor.

In preparing this genetic sequence, it is possible to utilize synthetic DNA by synthesizing the entire sequence de novo. Alternatively, it is possible to obtain cDNA sequences coding for certain preserved portions of the light and heavy chains of the desired antibody, and splice them together by means of the necessary sequence coding for the peptide linker, as described.

Also by methods known in the art, the resulting sequence can be amplified by utilizing well known cloning vectors and well known hosts. Furthermore, the amplified sequence, after checking for correctness, can be linked to promoter and terminator signals, inserted into appropriate expression vectors, and transformed into hosts such as procaryotic or eucaryotic hosts. Bacteria, yeasts (or other fungi) or mammalian cells can be utilized. Upon expression, either by itself or as part of fusion polypeptides, as will otherwise be known to those of skill in the art, the single chain binding protein is allowed to refold in physiological solution, at appropriate conditions of pH, ionic strength, temperature, and redox potential, and purified by standard separation procedures. These would include chromatography in its various different types, known to those with skill in the art.

The thus obtained purified single chain binding protein can be utilized by itself, in detectably la-

-23-

belled form, in immobilized form, or conjugated to drugs or other appropriate therapeutic agents, in diagnostic, imaging, biosensors, purifications, and therapeutic uses and compositions. Essentially all uses envisioned for antibodies or for variable region fragments thereof can be considered for the molecules of the present invention.

II. Hardware and Software Environment

A block diagram of the hardware aspects of the present invention is found in Figure 1. A central processing unit (CPU) 102 is connected to a first bus (designated massbus 104) and to a second bus (designated Unibus 106). A suitable form for CPU 102 is a model Vax 11/780 made by Digital Equipment Corporation of Maynard, Massachusetts. Any suitable type of CPU, however, can be used.

Bus 104 connects CPU 102 to a plurality of storage devices. In the best mode, these storage devices include a tape drive unit 106. The tape drive unit 106 can be used, for example, to load into the system the data base of the amino acid sequences whose three dimensional structures are known. A suitable form for tape drive 106 is a Digital Equipment Corporation model TU 78 drive, which operates at 125 inches per second, and has a 1600-6250 bit per inch (BPI) dual capability. Any suitable type of tape drive can be used, however.

Another storage device is a pair of hard disk units labeled generally by reference numeral 108. A suitable form for disk drive 108 comprises two Digital Equipment Corporation Rm05 disk drives having, for

-24-

example, 256 Mbytes of storage per disk. Another disk drive system is also provided in the serial processor mode and is labeled by reference numeral 110. This disk drive system is also connected to CPU 102 by bus 104. A suitable form for the disk system 110 comprises three Digital Equipment Corporation model Ra 81 hard disk drives having, for example, 450 Mbytes of storage per disk.

Dynamic random access memory is also provided by a memory stage 112 also connected to CPU 102 by bus 104. Any suitable type of dynamic memory storage device can be used. In the serial processor mode, the memory is made up of a plurality of semi-conductor storage devices found in a DEC model Ecc memory unit. Any suitable type of dynamic memory can be employed.

The disk drives 108 and 110 store several different blocks of information. For example, they store the data base containing the amino acid sequences and structures that are read in by the tape drive 106. They also store the application software package required to search the data base in accordance with the procedures of the present invention. They also store the documentation and executables of the software. The hypothetical molecules that are produced and structurally examined by the present invention are represented in the same format used to represent the protein structures in the data base. Using this format, these hypothetical molecules are also stored by the disk drives 108 and 110 for use during the structural design process and for subsequent use after the process has been completed.

-25-

A Digital Equipment Corporation VAX/VMS DEC operating system allows for multiple users and assures file system integrity. It provides virtual memory, which relieves the programmer of having to worry about the amount of memory that is used. Initial software was developed under versions 3.0 to 3.2 of the VAX/VMS operating system. The serial processor mode currently is running on version 4.4. DEC editors and FORTRAN compiler were utilized.

The CPU 102 is connected by Bus 106 to a multiplexer 114. The multiplexer allows a plurality of devices to be connected to the CPU 102 via Bus 106. A suitable form for multiplexer 114 is a Digital Equipment Corporation model Dz 16 terminal multiplexer. In the preferred embodiment, two of these multiplexers are used. The multiplexer 114 supports terminals (not shown in Figure 1) and the serial communications (at 19.2 Kbaud, for example) to the computer-graphics display system indicated by the dash lined box 116.

The computer-graphics display system 116 includes an electronics stage 118. The electronic stage 118 is used for receiving the visual image prepared by CPU 102 and for displaying it to the user on a display (typically one involving color) 120. The electronic stage 118 in connection with the associated subsystems of the computer-graphics display system 116 provide for local control of specific functions, as described below. A suitable form of the electronics system 118 is a model PS 320 made by Evans & Sutherland Corp. of Salt Lake, Utah. A suitable form for the display 120 is either a 25 inch color monitor or a 19 inch color monitor from Evans & Sutherland.

-26-

Dynamic random access memory 122 is connected to the electronic stage 118. Memory 122 allows the electronic system 118 to provide the local control of the image discussed below. In addition, a keyboard 124 of conventional design is connected to the electronic stage 118, as is an x/y tablet 126 and a plurality of dials 128. The keyboard 124, x/y tablet 126, and dials 128 in the serial processor mode are also obtained from Evans & Sutherland.

The computer generated graphics system 116, as discussed above, receives from CPU 102 the image to be displayed. It provides local control over the displayed image so that specific desired user initiated functions can be performed, such as:

- (1) zoom (so as to increase or decrease the size of the image being displayed;
- (2) clipping (where the sides, front or back of the image being displayed are removed);
- (3) intensity depth queing (where objects further away from the viewer are made dimmer so as to provide a desired depth effect in the image being displayed);
- (4) translation of the image in any of the three axes of the coordinate system utilized to plot the molecules being displayed;
- (5) rotation in any of the three directions of the image being displayed;
- (6) on/off control of the logical segments of the picture. For example, a line connecting the alpha carbons of the native protein might be one logical segment; labels on some or all of the residues of the native protein might be a second logical segment; a trace of the alpha carbons of the linker(s) might be a

-27-

third segment; and a stick figure connecting Carbon, Nitrogen, Oxygen, and Sulphur atoms of the linker(s) and adjacent residue of the native protein might be a fourth logical segment. The user seldom wants to see all of these at once; rather the operator first becomes oriented by viewing the first two segments at low magnification. Then the labels are switched off and the linker carbon trace is turned on. Once the general features of the linker are seen, the operator zooms to higher magnification and turns on the segments which hold more detail;

(7) selection of atoms in the most detailed logical segment. Despite the power of modern graphics, the operator can be overwhelmed by too much detail at once. Thus the operator will pick one atom and ask to see all amino acids within some radius of that atom, typically 6 Angstroms, but other radii can be used. The user may also specify that certain amino acids will be included in addition to those that fall within the specified radius of the selected atom;

(8) changing of the colors of various portions of the image being displayed so as to indicate to the viewer particular information using visual queing.

As stated above, the serial processor mode of the present invention currently is running the application software on version 4.4 of the Vax/Vms operating system used in conjunction with CPU 102. The application programs were programmed using the FLECS (FORTRAN Language with Extended Control Sections) programming language written in 1974 by Terry Beyer of the University of Oregon, Eugene, Oregon. FLECS is a FORTRAN preprocessor, which allows more logical programming.

-28-

All of the code used in the serial processor mode was developed in FLECS. It can be appreciated, however, that the present invention encompasses other operating systems and programming languages.

The macromolecules displayed on color display 120 of the computer-graphics display system 116 utilize an extensively modified version of version 5.6 of FRODO. FRODO is a program for displaying and manipulating macromolecules. FRODO was written by T.A. Jones at Max Planck Institute for Biochemistry, Munich, West Germany, for building or modeling in protein crystallography. FRODO version 5.6 was modified so as to be driven by command files; programs were then written to create the command files. It is utilized by the electronic stage 118 to display and manipulate images on the color display 120. Again, any suitable type of program can be used for displaying and manipulating the macromolecules, the coordinates of which are provided to the computer-graphics display system 116 by the CPU 102.

Design documentation and memos were written using PDL (Program Design Language) from Caine, Farber & Gordon of Pasadena, California. Again, any suitable type of program can be used for the design documents and memos.

Figure 2 shows a block diagram for an improved version of the hardware system of the present invention. Like numbers refer to like items of Figure 1. Only the differences between the serial processor mode system of Figure 1 and the improved system of Figure 2 are discussed below.

-29-

The CPU 102' is the latest version of the Vax 11/780 from Digital Equipment Corporation. The latest processor from DEC in the VAX product family is approximately ten times faster than the version shown in the serial processor mode of Figure 1.

Instead of the two Rm05 disk drives 108 of Figure 1, the embodiment of Figure 2 utilizes five RA 81 disk drive units 110'. This is to upgrade the present system to more state of the art disk drive units, which provide greater storage capability and faster access.

Serial processor 106 is connected directly to the electronic stage 118' of the computer-graphics display system 116. The parallel interface in the embodiment of Figure 2 replaces the serial interface approach of the serial processor mode of Figure 1. This allows for faster interaction between CPU 102' and electronic stage 118' so as to provide faster data display to the expert operator.

Disposed in front of color display 120 is a stereo viewer 202. A suitable form for stereo viewer 202 is made by Terabit, Salt Lake City, Utah. Stereo viewer 202 would provide better 3-D perception to the expert operator than can be obtained presently through rotation of the molecule.

In addition, this embodiment replaces the FRODO macromolecule display programs with a program designed to show a series of related hypothetical molecules. This newer program performs the operations more quickly so that the related hypothetical molecules can be presented to the expert operator in a short enough time that makes examination less burdensome on the operator.

-30-

The programs can be modified so as to cause the present invention to eliminate candidates in the second general step where obvious rules have been violated by the structures that are produced. For example, one rule could be that if an atom in a linker comes closer than one Angstrom to an atom in the native structure the candidate would be automatically eliminated.

In addition, the surface accessibility of molecules could be determined and a score based on the hydrophobic residues in contact with the solvent could be determined. After the hydrophobic residues have been calculated, the candidates could be ranked so that undesired candidates could automatically be eliminated. The protein is modeled in the present invention without any surrounding matter. Proteins almost always exist in aqueous solution; indeed, protein crystals contain between 20% and 90% water and dissolved salts which fill the space between the protein molecules. Certain kinds of amino acids have side-chains which make favorable interactions with aqueous solutions (serine, threonine, arginine, lysine, histidine, aspartic acid, glutamic acid, proline, asparagine, and glutamine) and are termed hydrophilic. Other amino acids have side chains which are apolar and make unfavorable interactions with water (phenylalanine, tryptophan, leucine, isoleucine, valine, methionine, and tyrosine) and are termed hydrophobic. In natural proteins, hydrophilic amino acids are almost always found on the surface, in contact with solvent; hydrophobic amino acids are almost always inside the protein in contact with other hydrophobic amino acids.

-31-

The remaining amino acids (alanine, glycine, and cysteine) are found both inside proteins and on their surfaces. The designs of the present invention should resemble natural proteins as much as possible, so hydrophobic residues are placed inside and hydrophilic residues are placed outside as much as possible.

Programs could be utilized to calculate an energy for each hypothetical structure. In addition, programs could make local adjustments to the hypothetical molecules to minimize the energy. Finally, molecular dynamics could be used to identify particularly unstable parts of the hypothetical molecule. Although existing programs could calculate a nominal energy for each hypothetical structure, it has not yet been demonstrated that such calculations can differentiate between sequences which will fold and those that will not. Energy minimization could also be accomplished with extant programs, but energy minimization also can not differentiate between sequences which will fold and those that will not. Molecular dynamics simulations currently cannot be continued long enough to simulate the actual folding or unfolding of a protein and so cannot distinguish between stable and unstable molecules.

Two megabytes of storage 128' in the computer generated display system 116 is added so that several different molecules can be stored at the display level. These molecules then can be switched back and forth on the color display 120 so that the expert operator can sequentially view them while making expert decisions. The parallel interface that is shown in Figure 2 would allow the coordinates to be trans-

-32-

ferred faster from the CPU 102' to the electronics stage 118' of the computer generated display system 116.

The parallel processing architecture embodiment of the present invention is described below in Section V. This parallel architecture embodiment provides even faster analysis and display.

III. Single Linker Embodiment

This first embodiment of the present invention determines and displays possible chemical structures for using a single linker to convert the naturally aggregated but chemically separate heavy and light polypeptide chains into a single polypeptide chain which will fold into a three dimensional structure very similar to the original structure made of two polypeptide chains.

A. Plausible Site Selection

There are two main goals of the plausible site selection step 302 of the present invention shown in very generalized block diagram form in Figure 3. The first goal is to select a first plausible site on the first chain that is the minimum distance from the second plausible site on the second chain. The first point on the first chain and the second point on the second chain comprise the plausible site.

The second goal of the site selection is to select plausible sites that will result in the least loss of native protein. Native protein is the original protein composed of the two aggregated polypeptide chains of the variable region. It is not chemically possible to convert two chains to one without altering some of

-33-

the amino acids. Even if only one amino acid was added between the carboxy terminal of the first domain and the amino terminal of the second domain, the charges normally present at these termini would be lost. In the variable regions of antibodies, the termini of the H and L chains are not very close together. Hypothetical linkers which join the carboxy terminus of one chain to the amino terminus of the other do not resemble the natural variable region structures. Although such structures are not impossible, it is more reasonable to cut away small parts of the native protein so that compact linkers which resemble the native protein will span the gap. Many natural proteins are known to retain their structure when one or more residues are removed from either end.

In the present embodiment, only a single linker (amino acid sequence or bridge for bridging or linking the two plausible sites to form a single polypeptide chain) is used. Figure 4 shows in block diagram form the steps used to select plausible sites in the single linker. The steps of Figure 4 are a preferred embodiment of step 302 of Figure 3.

A domain 1 is picked in a step 402 (see Figure 4). A schematic diagram of two naturally aggregated but chemically separate polypeptide chains is shown in Figure 5A. For purposes of illustration, assume that L is the light chain of the antibody variable region (the first polypeptide chain) and is domain 1. As shown in Figure 5A, light chain L is on the left side, and heavy chain H is on the right side.

The next step 404 is to pick the domain 2, which, as indicated, is the heavy chain H of the antibody variable region on the right side of Figure 5A.

-34-

The linker that will be selected will go from domain 1 (the light chain L) towards domain 2 (heavy chain, H). As the linker will become part of the single polypeptide chain, it must have the same directionality as the polypeptides it is linking; *i.e.* the amino end of the linker must join the carboxy terminal of some amino acid in domain 1, and the carboxy terminal of the linker must join the amino terminal of some residue in domain 2. A starting point (first site) on domain 1 is selected, as represented by step in 406 in Figure 4. The starting point is chosen to be close to the C (C for carboxy) terminal of domain 1, call this amino acid tau 1. It is important to pick tau 1 close to the C terminal to minimize loss of native protein structure. Residue tau 1 is shown schematically in two dimensions in figure 6A; it is also shown in figure 6B where it is presented in a two-dimensional representation of the naturally aggregated but chemically separate H and L polypeptide chains.

Next, the final point (second site) close the N (N for amino) terminal of domain 2 is selected, as indicated by step 408 of Figure 4. The final site is an amino acid of domain 2 which will be called sigma 1. It is important that amino acid sigma 1 be close to the N terminal of domain 2 to minimize loss of native protein structure. Amino acid sigma 1 is shown schematically in figure 6A and in the more realistic representation of figure 6B.

Figure 7 shows in simplified form the concept that the linker goes from a first site at amino acid tau 1 in domain 1 to a second site at amino acid sigma 1 in

-35-

domain 2. There are a plurality of possible first sites and a plurality of second sites, as is shown in figure 7. A computer program prepares a table which contains for each amino acid in domain 1 the identity of the closest amino acid in domain 2 and the distance. This program uses the position of the alpha carbon as the position of the entire amino acid. The expert operator prepares a list of plausible amino acids in domain 1 to be the first site, tau 1, and a list of plausible amino acids in domain 2 to be the second site, sigma 1. Linkers are sought from all plausible sites tau 1 to all plausible sites sigma 1. The expert operator must exercise reasonable judgement in selecting the sites tau 1 and sigma 1 in deciding that certain amino acids are more important to the stability of the native protein than are other amino acids. Thus the operator may select sites which are not actually the closest.

The complete designed protein molecule in accordance with the present invention consists of the domain 1 (of the light chain L) up to the amino acid tau 1, the linker, as shown by the directional-line in Figure 8A and in Figure 8B, and the domain 2 from amino acid sigma 1 to the C terminus of the heavy chain, H. As shown in Figures 8A and 8B, in the representative example, this results in the following loss of native protein.

The first loss in native protein is from the residue after residue tau 1 to the C terminus of domain 1 (light chain L). The second loss of native protein is from the N terminus of domain 2 (heavy chain, H) to the amino acid before sigma 1.

-36-

As is best understood from Figure 8A, the introduction of linker 1 produces a single polypeptide chain from the two naturally aggregated chains. The polypeptide chain begins with the N terminal of domain 1. Referring now to Figure 8B, the chain proceeds through almost the entire course of the native light chain, L, until it reaches amino acid tau 1. The linker then connects the carboxy terminal of a very slightly truncated domain 1 to residue sigma 1 in the very slightly truncated domain 2. Since a minimum amount of native protein is eliminated, and the linker is selected to fit structurally as well as possible (as described below in connection with general steps 2 and 3 of the present invention), the resulting single polypeptide chain has a very high probability (several orders of magnitude greater than if the linker was selected randomly) to fold into a three-dimensional structure very similar to the original structure made of two polypeptide chains.

The single polypeptide chain results in a much more stable protein which contains a binding site very similar to the binding site of the original antibody. In this way a single polypeptide chain can be engineered from the naturally occurring two-polypeptide chain variable region, so as to create a polypeptide of only one chain, but maintaining the binding site of the antibody.

In the current mode of the present invention, the expert operator selects the sites with minimal help from the computer. The computer prepares the table of closest-residue-in-other-domain. The computer can provide more help in the following ways.

-37-

(1) Prepare a list of conserved and variable residues for variable regions of antibodies (Fv region). Residues which vary from Fv to Fv would be much better starting or ending sites for linkage than are residues which are conserved over many different Fv sequences.

(2) Prepare a list of solvent accessibilities. Amino acids exposed to solvent can be substituted with less likelihood of destabilizing the native structure than amino acids buried within the native structure. Exposed amino acids are better choices to start or end linkage.

With respect to each of the plurality of possible first sites (on domain 1 or light chain L) there are available a plurality of second sites (on domain 2 or heavy chain H) (See Figures 7 and 8A). As the second site is selected closer to the N terminus of domain 2, the distance to any of the plausible first sites increases. Also, as the first site is selected closer to the C terminus of domain 1 the distance to any of the plausible second sites increases. It is this tension between shortness of linker and retention of native protein which the expert operator resolves in choosing gaps to be linked. The penalty for including extra sites in the list of gaps are:

(1) searching in general step 2 will be slower; and

(2) more candidates will pass from step 2 many of which must be rejected in step 3. As step 3 is currently a manual step, this is the more serious penalty.

Figure 8B shows diagrammatically by a directional arrow the possible links that can occur between the various

-38-

sites near the C terminal of domain 1 and the various sites near the N terminal of domain 2.

B. Selection of Candidates

In the second of the three general steps of the present invention as used in the single linker embodiment, plausible candidates for linking the site 1 on domain 1 with site 2 on domain 2 are selected from a much larger group of candidates. This process of winnowing out candidates results in the expert operator and/or expert system having a relatively small group of candidates to rank from most plausible to least plausible in the third general step of the present invention, as described in subsection C below.

Currently, there are approximately 250 protein structures, determined at 2.0 A or higher resolution, in the public domain. The structures of these very complicated molecules are determined using sophisticated scientific techniques such as X-ray crystallography, neutron diffraction, and nuclear magnetic resonance. Structure determination produces a file of data for each protein. The Brookhaven Protein Data Bank (BPDB) exemplifies a repository of protein structural information. Each file in BPDB contains many records of different types. These records carry the following information:

- (1) Name of the protein and standard classification number,
- (2) Organism from which protein was obtained,
- (3) Name and address of contributor,
- (4) Amino-acid sequence of each polypeptide chain, if known,
- (5) Connectivity of disulfides, if any,

-39-

(6) Names and connectivities of any prosthetic groups, if any,

(7) References to literature,

(8) Transformation from reported coordinates to crystallographic coordinates,

(9) Coordinates of each atom determined.

There is at least one record for each atom for which a coordinate was determined. Some parts of some proteins are disordered and do not diffract X-rays, so no sensible coordinates can be given. Thus there may be amino acids in the sequence for which only some or none of the atoms have coordinates. Coordinates are given in Angstrom units ($100,000,000 \text{ \AA} = 1 \text{ cm}$) on a rectangular Cartesian grid. As some parts of a protein may adopt more than one spatial configuration, there may be two or more coordinates for some atoms. In such cases, fractional occupancies are given for each alternative position. Atoms move about, some more freely than others. X-ray data can give an estimate of atomic motion which is reported as a temperature (a.k.a. Debye-Waller) factor.

Any other data base which included, implicitly or explicitly, the following data would be equally useful:

(1) Amino acid sequence of each polypeptide chain.

(2) Connectivity of disulfides, if any,

(3) Names and connectivities of any prosthetic groups, if any,

(4) Coordinates (x, y, z) of each atom in each observed configuration.

(5) Fractional occupancy of each atom,

(6) Temperature factor of each atom.

-40-

Proteins usually exist in aqueous solution. Although protein coordinates are almost always determined for proteins in crystals, direct contacts between proteins are quite rare. protein crystals contain from 20% to 90% water by volume. Thus one usually assumes that the structure of the protein in solution will be the same as that in the crystal. It is now generally accepted that the solution structure of a protein will differ from the crystal structure only in minor details. Thus, given the coordinates of the atoms, one can calculate quite easily the solvent accessibility of each atom.

In addition, the coordinates implicitly give the charge distribution throughout the protein. This is of use in estimating whether a hypothetical molecule (made of native protein and one or more linkers) will fold as designed. The typical protein whose structure is known comprises a chain of amino acids (there are 21 types of amino acids) in the range of 100 to 300 amino acids.

Each of these amino acids alone or in combination with the other amino acids as found in the known protein molecule can be used as a fragment to bridge the two sites. The reason that known protein molecules are used is to be able to use known protein fragments for the linker or bridge.

Even with only 250 proteins of known structure, the number of possible known fragments is very large. A linker can be from one to twenty or thirty amino acids long. Let "Lmax" be the maximum number of amino acids allowed in a linker, for example, Lmax might be

-41-

25. Consider a protein of "Naa" amino acids. Proteins have Naa in the range 100 to 800, 250 is typical. From this protein one can select Naa-1 distinct two-amino-acid linkers, Naa-2 distinct three-amino-acid linkers,...and (Naa+1-Lmax) distinct linkers containing exactly Lmax amino acids. The total number of linkers containing Lmax or fewer linkers is "Nlink,"

$$\begin{aligned}
 \text{Nlink} &= \sum_{j=1, L_{\max}} (Naa+1-j) \\
 &= Naa \times (L_{\max}) - (L_{\max} \times L_{\max})/2 + L_{\max} / 2
 \end{aligned}$$

If Naa is 250 and Lmax is 25, Nlink will be 5975. If the number of known proteins is "Nprot," then the total number of linkers, "Nlink_total" will be

$$\text{Nlink_total} = \sum_{k=1, N_{\text{prot}}} \sum_{j=1, L_{\max}} (Naa(k)+1-j)$$

$$= \sum_{k=1, N_{\text{prot}}} [Naa(k) \times (L_{\max}) - (L_{\max} \times L_{\max})/2 + L_{\max}/2]$$

$$\begin{aligned}
 &= N_{\text{prot}} \times (L_{\max}/2 - L_{\max} \times L_{\max})/2 + L_{\max} \times \sum_{K=1, N_{\text{prot}}} Naa(k)
 \end{aligned}$$

-42-

Where $N_{aa}(k)$ is the number of amino acids in the k th protein. With 250 proteins, each containing 250 amino acids (on average), and L_{max} set to 25, N_{link_total} is 1,425,000.

This is the number of linkers of known structure. If one considers the number of possible amino acid sequences up to length L_{max} (call it " $N_{link_possible}$ "), it is much larger.

$$N_{link_possible} = \sum_{J=1, L_{max}} 20^J$$

For $L_{max} = 25$

$$\begin{aligned} N_{link_possible} &= 353, 204, 547, 368, 421, 052, \\ &631, 578, 947, 368, 420 \\ &= 3.53 \times 10^{32} \end{aligned}$$

Using known peptide fragments thus reduces the possibilities by twenty-six orders of magnitude. Appropriate searching through the known peptide fragments reduces the possibilities a further five orders of magnitude.

Essentially, the present invention utilizes a selection strategy for reducing a list of possible candidates. This is done as explained below in a preferred form in a three step process. This three step process, as is illustrated in the explanation of the each of the three steps of the process, significantly reduces the computer time required to extract the most promising candidates from the data base of possible candidates. This should be contrasted with a serial search throughout the entire data base of candidates,

-43-

which would require all candidates to be examined in total. The present invention examines certain specific parameters of each candidate, and uses these parameters to produce subgroups of candidates that are then examined by using other parameters. In this way, the computer processing speed is significantly increased.

The best mode of the present invention uses a protein data base created and supplemented by the Brookhaven National Laboratory in Upton, Long Island, New York. This data base is called the Brookhaven Protein Data Base (BPDB). It provides the needed physical and chemical parameters that are needed by the present invention. It should be understood, that the candidate linkers can be taken from the Brookhaven Protein Data Base or any other source of three-dimensional protein structures. These sources must accurately represent the proteins. In the current embodiment, X-ray structures determined at resolution of 2.5A or higher and appropriately refined were used. Each peptide is replaced (by least-squares fit) by a standard planar peptide with standard bond lengths and angles. Peptides which do not accurately match a standard peptide (e.g. cis peptides) are not used to begin or end linkers, but may appear in the middle.

Each sequence up to some maximum number of amino acids (Lmax) is taken as a candidate. In the preferred embodiment, the maximum number of amino acids (Lmax) is set to 30. However, the present invention is not limited to this number, but can use any maximum number that is desired under the protein engineering circumstances involved.

-44-

1. Selecting Candidates with Proper Distance Between the N Terminal and the C Terminal.

The first step in the selection of candidates step is to select the candidate linkers with a proper distance between the N terminal and the C terminal from all of the candidate linkers that exist in the protein data base that is being used. Figure 9 shows in block diagram form the steps that make up this candidate selection process utilizing distance as the selection parameter.

Referring to Figure 9, a standard point relative to the peptide unit at the first site is selected, as shown by block 902.

A standard point relative to the peptide unit in the second site is also picked, as indicated by a block 904. Note that in the best mode the geometric centers of the peptide units of the first and second sites are used, but any other standard point can be utilized, if desired.

The distance between the standard points of the two peptides at the first and second sites defining the gap to be bridged by the linker is then calculated, as indicated by block 906. This scalar distance value is called the Span of the gap. Note that this scalar value does not include any directional information.

Next, as indicated by a step 908, the distance between the ends of the possible linker candidates are calculated. The distance between the ends of a particular candidate is called the span of the candidate. Note that each possible linker candidate has a span of the candidate scalar value.

-45-

The final step in the distance selection candidate selection process is that of a step 910. In step 910, candidates are discarded whose span of the candidate values differ from the span of the gap value by more than a preselected amount (this preselected amount is Max LSQFIT error). In the best mode of the present invention, the preselected amount for Max LSQFIT error is 0.50 Angstroms. However, any other suitable value can be used.

The preceding discussion has been for a single gap. In fact, the expert user often selects several gaps and the search uses all of them. The span of each candidate is compared to the span of each gap until it matches one, within the preset tolerance, or the list of gaps is exhausted. If the candidate matches none of the gaps, it is discarded. If it matches any gap it is carried to the next stage.

The inventors have determined that the use of the distance as the first parameter for discarding possible linker candidates results in a significant reduction in the number of possible candidates with a minimum amount of computer time that is needed. In terms of the amount of reduction, a representative example (using linkers up to 20 amino acids) starts out with 761,905 possible candidates that are in the protein data base. This selection of candidates using the proper distance parameter winnows this number down to approximately 63,727 possible candidates. As is discussed below, the distance selection operation requires much less computer time than is required by the other two steps which make up this selection step 304.

-46-

The result of this selection of candidates according to proper distance is a group (called a first group of candidates) which exhibit a proper length as compared to the gap that is to be bridged or linked. This first group of candidates is derived from the protein data base using the distance criteria only.

2. Selecting Candidates with Proper Direction from N Terminal to C Terminal

This substep essentially creates a second group of possible candidates from the first group of possible candidates which was produced by the distance selection substep discussed in connection with Figure 9. The second group of candidates is selected in accordance with the orientation of the C terminal residue (i.e. the final residue) of the linker with respect to the N terminal residue (i.e. the initial residue) which is compared to the orientation of the C terminal residue (i.e. the second site) of the gap with respect to the N terminal residue (i.e. the first site). See Figure 20B. In this way, this direction evaluation determines if the chain of the linker ends near the second site of the gap, when the amino terminal amino acid of the linker is superimposed on the first site of the gap so as to produce the minimum amount of unwanted molecular distortion.

Referring now to Figure 10, the first step used in producing the second group of possible candidates is a step 1002. In step 1002 a local coordinate system is established on the N terminal residue of one of the selected gaps. For example, one might take the local X-axis as running from the first alpha carbon of the N

-47-

terminal residue to the second alpha carbon of the N terminal residue, with the first alpha carbon at the origin - the second alpha carbon on the plus X-axis. The local Y-axis is selected so that the carbonyl oxygen lies in the xy plane with a positive y coordinate. The local Z-axis is generated by crossing X into Y. Next, as indicated by step 1004, a standard reference point in the C terminal residue of the gap is located and its spherical polar coordinates are calculated in the local system. The standard reference point could be any of the atoms in the C terminal peptide (throughout this application, peptide, residue, and amino acid are used interchangeably) or an average of their positions. Steps 1002 and 1004 are repeated for all gaps in the list of gaps. As indicated by step 1006, a local coordinate system is established on the N terminal residue of one of the candidates. This local coordinate system must be established in the same manner used for the local coordinate systems established on each of the gaps. Various local systems could be used, but one must use the same definition throughout. In step 1008, the standard reference point is found in the C terminal residue of the current candidate. This standard point must be chosen in the same manner used for the gaps. The spherical polar coordinates of the standard point are calculated in the local system of the candidate. (This use of local coordinate system is completely equivalent to rotating and translating all gaps and all candidates so that their initial peptide lies in a standard position at the origin.) In step 1010, the spherical polar coordinates of the gap vector (r , θ , ϕ) are compared

-48-

to the spherical polar coordinates of the candidate vector (r , θ , ϕ). In step 1012 a preset threshold is applied, if the two vectors agree closely enough, then one proceeds to step 1014 and enrolls the candidate in the second group of candidates. Currently, this preset threshold is set to 0.5 Å, but other values could be used. From step 1014, one skips forward to step 1022, vide infra. On the other hand, if the vectors compared in step 1012 are not close enough, one moves to the next gap vector in the list, in step 1016. If there are no more gaps, one goes to step 1018 where the candidate is rejected. If there are more gaps, step 1020 increments the gap counter and one returns to step 1010. From steps 1014 or 1018 one comes to step 1022 where one tests to see if all candidates have been examined. If not, step 1024 increments the candidate counter and one returns to step 1006. If all candidates have been examined, one has finished, step 1026.

Figure 11 shows the concept of comparing the direction of the gap to the direction of the candidate.

The inventors have determined that in the example discussed above where 761,905 possible candidates are in the protein data base, the winnowing process in this step reduces the approximate 63,727 candidates in the first group to approximately 50 candidates in the second group. The inventors have also determined that as referenced to the units of computer time referred to above in connection with the scalar distance parameter, it takes approximately 4 to 5 computer units of time to perform the selection of this step. Thus, it can be appreciated that it preserves computer time to

-49-

perform the distance selection first, and the direction selection second since the direction selection process takes more time than the distance selection process.

3. Selecting Candidates with Proper Orientation at Both Termini

In this step, the candidates in the second group of step 1016 of Figure 10 are winnowed down to produce a third group of plausible candidates using an evaluation of the relative orientation between the peptide groups at either end of the candidate, compared to the relative orientation between the peptide groups at either end of the gap. In a step 1201, (Figure 12) decide that a peptide will be represented by 3, 4, or 5 atoms (vide infra). Specifically, in a step 1202, one of the candidates in the second group (step 1014) is selected for testing. In a step 1204, three to five atoms in the first peptide are selected to define the orientation of the first peptide. So long as the atoms are not collinear, three atoms is enough, but using four or five atoms makes the least-squares procedure which follows over-determined and therefore compensates for errors in the coordinates. For example, assume selection of four atoms: C alpha, C, N, and C beta. Next, in a step 1206, one selects the corresponding 3, 4, or 5 atoms from the final peptide of the selected candidate. These 6, 8, or 10 atoms define a three-dimensional object. In a step 1208, select one of the gaps. Select the corresponding 6, 8, or 10 atoms from the gap. In a step 1210, least-squares fit the atoms from the candidate to the atoms from the gap. This least-squares fit allows degrees

-50-

of freedom to superimpose the two three-dimensional objects. Assume that one object is fixed and the other is free to move. Three degrees of freedom control the movement of the center of the free object. Three other degrees of freedom control the orientation of the free object. In a step 1212, the result of the least-square fit is examined. If the Root-Mean-Square (RMS) error is less than some preset threshold, the candidate is a good fit for the gap being considered and is enrolled in the third group in a step 1214. If, on the other hand, the RMS error is greater than the preset threshold, one checks to see if there is another gap in the list in a step 1216. If there is, one selects the next gap and returns to step 1208. If there are no more gaps in the list, then the current candidate from the second group is rejected in step 1218. In step 1220, one checks to see if there are more candidates in the second group; if so, a new candidate is selected and one returns to step 1201. If there are no more candidates, one is finished (step 1222). Again referring to a representative case, where linkers of length up to twenty amino acids were sought for a single gap with separation 12.7 Å, the protein data bank contained 761,905 potential linkers. Of these, 63,727 passed the distance test. The direction test removed all but 50 candidates. The orientation test passed only 1 candidate with RMS error less than or equal to 0.5 Å. There were two additional candidates with RMS error between 0.5 Å and 0.6 Å. Moreover, the inventors have determined that it takes about 25 units of computer time to evaluate each candidate in group 2 to decide whether they should be

-51-

selected for group 3. It can be appreciated now that the order selected by the inventors for the three steps of winnowing the candidates has been selected so that the early steps take less time per candidate than the following steps. The order of the steps used to select the candidate can be changed, however, and still produce the desired winnowing process. Logically, one might even omit steps one and two and pass all candidates through the least-squares process depicted in Figure 12 and achieve the same list of candidates, but at greater cost in computing. This may be done in the case of parallel processing where computer time is plentiful, but memory is in short supply.

Another approach (not illustrated) for determining whether the proper orientation exists between the ends of the candidate, is to examine only the atoms at the C terminal of the candidate as compared to the atoms at the final peptide of the gap. In step 2, the inventors aligned the first peptide of the candidate with the first peptide in the gap. Having done this, one could merely compare the atoms at the C terminal of the candidate with the atoms of the second peptide of the gap. This approach is inferior to that discussed above because all the error appears at the C terminus, while the least-squares method discussed above distributes the errors evenly.

C. Ranking and Eliminating Candidates.

As shown in Figure 3, the third general step in the present invention is that of ranking the plausible candidates from most plausible to least plausible, and eliminating those candidates that do not appear to be plausible based on criteria utilized by an expert operator and/or expert system.

-52-

In the best mode, the candidates in the third group (step 1214) are provided to the expert operator, who can sequentially display them in three dimensions utilizing the computer-graphics display system 116. The expert operator then can make decisions about the candidates based on knowledge concerning protein chemistry and the physical relationship of the plausible candidate with respect to the gap being bridged. This analysis can be used to rank the plausible candidates in the third group from most plausible to least plausible. Based on these rankings, the most plausible candidates can be selected for genetic engineering.

As noted above in connection with the illustrative example, there are typically few (under 100) candidates which make it to the third group of step 1214. Consequently, a moderately expert operator (one having a Bachelor of Science degree in chemistry, for example), can typically winnow down this number of plausible candidates to a group of 10 to 15. Thereafter, a more expert operator and/or expert system can further winnow down the number. In this way, only a very few of the plausible candidates needs to be tested in practice as compared to the hundreds, thousands or more of candidates that would have to be tested if no selection process like that of the present invention was used. This speeds up the process of engineering the single chain molecules by orders of magnitude, while reducing costs and other detriments by orders of magnitude as well.

In certain situations, however, automatic ranking in this third general step may be warranted. This could occur, for example, where the expert operator

-53-

was presented with quite a few candidates in the third group, or where it is desired to assist the expert operator in making the ranking selections and eliminating candidates based on prior experience that has been derived from previous engineering activities and/or actual genetic engineering experiments.

Referring now to Figure 13, a coordinate listing of the hypothetical molecule (candidate) is automatically constructed, as is indicated by a block 1302. The expert operator can then display using a first color the residues from domain 1 of the native protein. Color display 120 can provide a visual indication to the expert operator of where the residues lie in domain 1. This is indicated by a block 1304.

The expert operator then can display on color display 120 the residues from domain 2 of the native protein using a second color, as is indicated by a block 1306. The use of a second color provides a visual indication to the user which assists in distinguishing the residues from domain 1 from the residues from domain 2.

The linker (candidate) being ranked can be displayed in a selected color, which color can be different from the first color of step 1304 and/or the second color from step 1306. Again, by using this visual color indication, the expert operator can distinguish the residues of domain 1 and 2 of the native protein. This display of the linker candidate is indicated by a block 1308.

The initial picture on the color display 120 provided to the expert operator typically shows the alpha carbons for all of the residues. This is indicated by

-54-

a block 1310. In addition, the initial picture shows the main-chain and side-chains for residues and linkers and one residue before the linker and one residue after the linker. This is indicated by a block 1312.

The expert operator can also cause any of the other atoms in the native protein or linker candidate to be drawn at will. The molecule can be rotated, translated, and enlarged or reduced, by operator command, as was discussed generally in connection with the computer-graphics display system 116 above. The block diagram of Figure 13 indicates that each of the steps just discussed are accomplished in serial fashion. However, this is only for purposes of illustration. It should be understood that the operator can accomplish any one or more of these steps as well as other steps at will and in any sequence that is desired in connection with the ranking of the plausible candidates in group 3.

The expert operator and/or expert system utilized in this third general step in ranking the candidates from most plausible to least plausible and in eliminating the remaining candidates from group 3, can use a number of different rules or guidelines in this selection process. Representative of these rules and guidelines are the following which are discussed in connection with Figure 14. Note that the blocks in Figure 14 show the various rules and/or criteria, which are not necessarily utilized in the order in which the boxes appear. The order shown is only for purposes of illustration. Other rules and/or criteria can be utilized in the ranking process, as well.

-55-

As shown in step 1402, a candidate can be rejected if any atom of the linker comes closer than a minimum allowed separation to any retained atom of the native protein structure. In the best mode, the minimum allowed separation is set at 2.0 Angstroms. Note that any other value can be selected. This step can be automated, if desired, so that the expert operator does not have to manually perform this elimination process.

A candidate can be penalized if the hydrophobic residues have high exposure to solvent, as is indicated by a block 1404. The side chains of phenylalanine, tryptophan, tyrosine, leucine, isoleucine, methionine, and valine do not interact favorably with water and are called hydrophobic. Proteins normally exist in saline aqueous solution; the solvent consists of polar molecules (H_2O) and ions.

A candidate can be penalized when the hydrophilic residues have low exposure to solvent. The side chains of serine, threonine, aspartic acid, glutamic acid, asparagine, glutamine, lysine, arginine, and proline do interact favorably with water and are called hydrophilic. This penalization step for hydrophilic residues is indicated by a block 1406.

A candidate can be promoted when hydrophobic residues have low exposure to solvent, as is indicated by a block 1408.

A candidate can be promoted when hydrophilic residues have high exposure to solvent, as indicated by a block 1410.

A candidate can be penalized when the main chain fails to form hydrogen bonds, as is indicated by a block 1412.

-56-

A candidate can be penalized when the main chain makes useless excursions into the solvent region. Useless excursions are those which do not make any evident interaction with the retained native protein. This is indicated by a block 1414.

A candidate can be promoted when the main chain forms a helix, as is indicated by a block 1416. Helices are self-stabilizing. Thus a linker which is helical will be more stable because its main-chain polar atoms (O and N) will form hydrogen bonds within the linker.

As is indicated by a block 1418, a candidate can be promoted when the main chain forms a beta sheet which fits against existing beta sheets. The strands of beta sheets stabilize each other. If a linker were found which was in a beta-sheet conformation such that it would extend an existing beta sheet, this interaction would stabilize both the linker and the native protein.

Another expert design rule penalizes candidates which have sterically bulky side chains at undesirable positions along the main chain. Furthermore, it is possible to "save" a candidate with a bulky side chain by replacing the bulky side chain by a less bulky one. For example if a side chain carries a bulky substituent such as leucine or isoleucine, a possible design step replaces this amino acid by a glycine, which is the least bulky side chain.

Other rules and/or criteria can be utilized in the selection process of the third general step 306, and the present invention is not limited to the rules and/or criteria discussed. For example, once the

-57-

linker has been selected it is also possible to add, delete, or as stated, modify one or more amino acids therein, in order to accomplish an even better 3-D fit.

IV. Double and Multiple Linker Embodiments

Section III above described the single linker embodiment in accordance with the present invention. This section describes double linker and multiple linker embodiments in accordance with the present invention. For brevity purposes, only the significant differences between this embodiment and the single linker embodiment will be described here and/or illustrated in separate figures. Reference should therefore be made to the text and figures that are associated with the single linker embodiment

A. Plausible Site Selection.

The two main goals of minimizing distance between the sites to be linked and the least loss of native protein apply in the site selection in the double and multiple linker embodiments as they did apply in the single linker embodiment discussed above.

Figure 15A shows a simplified two dimensional representation of the use of two linkers to create the single polypeptide chain from the two naturally aggregated but chemically separate polypeptide chains. Figure 15B shows in two dimensions a three dimensional representation of the two chains of Figure 15A. Referring now to Figures 15A and B, the first step in de-

-58-

terminating suitable sites is to find a site in domain 1 which is close to either the C or N terminus of domain 2. For purposes of illustration, and as is shown in Figures 15A and 15B, it is assumed that the most promising location is the C terminus of domain 2. The residue in domain 1 is called Tau 1, while the residue in domain 2 is called Sigma 1.

Figures 16A and 16B are respectively two dimensional simplified plots of the two chains, and two dimensional plots of the three dimensional representation of the two chains. They are used in connection with the explanation of how plausible sites are selected for the second linker in the example situation.

The first step in connection with finding plausible sites for the second linker is to find a residue in domain 1 that is before Tau 1 in the light chain. This residue is called residue Tau 2. It is shown in the top portion in Figure 16A, and in the right middle portion in Figure 16B.

The next step in the site selection process for the second linker is to find a residue in domain 2 near the N terminus of domain 2. This residue is called residue Sigma 2. Reference again is made to Figures 16A and B to show the location of Sigma 2.

The second linker (linker 2) thus runs from Tau 2 to Sigma 2. This is shown in Figures 17A and 17B. Note that the chain that is formed by these two linkers has the proper direction throughout.

Figure 18 shows in two dimensional simplified form the single polypeptide chain that has been formed by the linking of the two independent chains using the two linkers. Note that the approach outlined above

-59-

resulted in the minimal loss of native protein. The completely designed protein is shown in Figure 17 and consists of domain 1 from the N terminal to Tau 2, linker 2, domain 2 from Sigma 2 to Sigma 1, linker 1, and domain 1 from Tau 1 to the C terminus. The arrows that are shown in Figure 17 indicate the direction of the chain.

Figure 17 shows that the residues lost by the utilization of the two linkers are: (a) from the N terminus of domain 2 up to the residue before Sigma 2; and (b) from the residue after Sigma 1 to the C terminus of domain 2; and (c) from the residue after Tau 2 to the residue before Tau 1 of domain 1.

If one of the linkers in the two linker case is very long, one could link from Tau 2 to a residue in domain 2 after Sigma 1. A third linker (not shown) would then be sought from a residue near the C terminal of domain 2 to a residue near the N terminal of domain 2.

Additionally, one could use two linkers to reconnect one of the domains in such a way that a single linker or a pair of linkers would weld the two domains into one chain.

B. Candidate Selection and Candidate Rejection Steps

Ranking of linkers in the multilinker cases follows the same steps as in the single linker case except there are some additional considerations.

(1) There may be a plurality of linkers for each of the two (or more) gaps to be closed. One must consider all combinations of each of the linkers for gap A with each of the linkers for gap B.

-60-

(2) One must consider the interactions between linkers.

As one must consider combinations of linkers, the ranking of individual linkers is used to cut down to a small number of very promising linkers for each gap. If one has only three candidates for each gap, there are nine possible constructs.

The process of examining interactions between linkers and discarding poor candidates can be automated by applying the rules discussed above.

V. Parallel Processing Embodiment

Figure 19 shows in block diagram form the parallel processing approach that can be utilized in the present invention.

As shown in Figure 19, a friendly serial processor 1902 is connected by a first bus 1904 to a plurality of data storage devices and input devices. Specifically, and only for purposes of illustration, a tape input stage 1906 is connected to bus 1904 so as to read into the system the parameters of the protein data base that is used. A high storage disk drive system 1908 (having, for example, 5 gigabits of storage) is also connected to bus 1904. Operationally, for even larger storage capabilities, an optical disk storage stage 1910 of conventional design can be connected to bus 1904.

The goal of the hypercube 1912 that is connected to the friendly serial processor 1902 via a bi-directional bus 1914 is twofold: to perform searching faster, and to throw out candidates more automatically.

-61-

The hypercube 1912, having for example, 2^{10} to 2^{16} nodes provides for parallel processing. There are computers currently available which have up to 1,024 computing nodes. Thus each node would need to hold only about 1400 candidate linkers and local memory of available machines would be sufficient. This is the concept of the hypercube 1912. Using the hypercube parallel processing approach, the protein data base can be divided into as many parts as there are computing nodes. Each node is assigned to a particular known protein structure.

The geometry of the gap that has to be bridged by a linker is sent by the friendly serial processor 1902 via bus 1914 to the hypercube stage 1912. Each of the nodes in the hypercube 1912 then processes the geometrical parameters with respect to the particular candidate linker to which it is assigned. Thus, all of the candidates can be examined in a parallel fashion, as opposed to the serial fashion that is done in the present mode of the present invention. This results in much faster location (the inventors believe that the processing speed can be brought down from 6 hours to 3 minutes using conventional technology) in locating the candidates that can be evaluated by the second step 304 of the present invention.

Another advantage for the parallel processing embodiment is that it will provide sufficient speed to allow candidates to be thrown out more automatically. This would be achieved using molecular dynamics and energy minimization. While this could be done currently on serial processing computers (of the super computer variety such as those manufactured by Cray

-62-

and Cyber) the parallel processing approach will perform the molecular dynamics and energy minimization much faster and cheaper than using the super computing approach.

In particular, hypercube computers exist which have inexpensive computing nodes which compare very favorably to supercomputers for scalar arithmetic. Molecular dynamics and energy minimization are only partly vectorizable because the potential functions used have numerous data-dependent branches.

VI. Preparation and Expression of Genetic Sequences, and Uses.

The polypeptide sequences generated by the methods described herein, give rise by application of the genetic code, to genetic sequences coding therefor. Given the degeneracy of the code, however, there are in many instances multiple possible codons for any one amino acid. Therefore, codon usage rules, which are also well understood by those of skill in the art, can be utilized for the preparation of optimized genetic sequences for coding in any desired organism. (See, for example, Ikemura, J. Mol. Biol. 151:389-409 (1981)).

Generally, it is possible to utilize the cDNA sequences obtained from the light and heavy chains of the variable region of the original antibody as a starting point. These sequences can then be joined by means of genetic linkers coding for the peptide linker candidates elucidated by the methods of the invention. The genetic sequence can be entirely synthesized de novo or fragments of cDNA can be linked together with the synthetic linkers, as described.

-63-

A large source of hybridomas and their corresponding monoclonal antibodies are available for the preparation of sequences coding for the H and L chains of the variable region. As indicated previously, it is well known that most "variable" regions of antibodies of a given class are in fact quite constant in their three dimensional folding pattern, except for certain specific hypervariable loops. Thus, in order to choose and determine the specific binding specificity of the single chain binding protein of the invention it becomes necessary only to define the protein sequence (and thus the underlying genetic sequence) of the hypervariable region. The hypervariable region will vary from binding molecule to molecule, but the remaining domains of the variable region will remain constant for a given class of antibody.

Source mRNA can be obtained from a wide range of hybridomas. See for example the catalogue ATCC Cell Lines and Hybridomas, December 1984, American Type Culture Collection, 20309 Parklawn Drive, Rockville, Maryland 20852, U.S.A., at pages 5-9. Hybridomas secreting monoclonal antibodies reactive with a wide variety of antigens are listed therein, are available from the collection, and usable in the invention. Of particular interest are hybridomas secreting antibodies which are reactive with viral antigens, tumor associated antigens, lymphocyte antigens, and the like. These cell lines and others of similar nature can be utilized to copy mRNA coding for the variable region or determine amino acid sequence from the monoclonal antibody itself. The specificity of the antibody to be engineered will be determined by the original se-

-64-

lection process. The class of antibody can be determined by criteria known to those skilled in the art. If the class is one for which there is a three-dimensional structure, one needs only to replace the sequences of the hyper-variable regions (or complementary determining regions). The replacement sequences will be derived from either the amino acid sequence or the nucleotide sequence of DNA copies of the mRNA.

It is to be specifically noted that it is not necessary to crystallize and determine the 3-D structure of each variable region prior to applying the method of the invention. As only the hypervariable loops change drastically from variable region to variable region (the remainder being constant in the 3-D structure of the variable region of antibodies of a given class), it is possible to generate many single chain 3-D structures from structures already known or to be determined for each class of antibody.

For example, linkers generated in the Examples in this application (e.g., TRY40, TRY61 or TRY59, see below) are for Fv regions of antibodies of the IgA class. They can be used universally for any antibody, having any desired specificity, especially if the antibody is of the IgA class.

Expression vehicles for production of the molecules of the invention include plasmids or other vectors. In general, such vectors containing replicon and control sequences which are derived from species compatible with a host cell are used in connection with the host. The vector ordinarily carries a replicon site, as well as specific genes which are capable of providing phenotypic selection in transformed cells. For example, E. coli is readily transformed

-65-

using pBR322, a plasmid derived from an E. coli species. pBR322 contains genes for ampicillin and tetracycline resistance, and thus provides easy means for identifying transformed cells. The pBR322 plasmid or other microbial plasmids must also contain, or be modified to contain, promoters which can be used by the microbial organism for expression of its own proteins. Those promoters most commonly used in recombinant DNA construction include the beta lactamase, lactose promoter systems, lambda phage promoters, and the tryptophan promoter systems. While these are the most commonly used, other microbial promoters have been discovered and can be utilized.

For example, a genetic construct for a single chain binding protein can be placed under the control of the leftward promoter of bacteriophage lambda. This promoter is one of the strongest known promoters which can be controlled. Control is exerted by the lambda repressor, and adjacent restriction sites are known.

The expression of the single chain antibody can also be placed under control of other regulatory sequences which may be homologous to the organism in its untransformed state. For example, lactose dependent E. coli chromosomal DNA comprises a lactose or lac operon which mediates lactose utilization by elaborating the enzyme beta-galactosidase. The lac control elements may be obtained from bacteriophage lambda plac5, which is infective for E. coli. The lac promoter-operator system can be induced by IPTG.

Other promoter/operator systems or portions thereof can be employed as well. For example, colicin E1,

-66-

galactose, alkaline phosphatase, tryptophan, xylose, tac, and the like can be used.

Of particular interest is the use of the O_L/P_R hybrid lambda promoter (see for example U.S. patent application Serial Number 534,982 filed September 3, 1983, and herein incorporated by reference).

Other preferred hosts are mammalian cells, grown in vitro in tissue culture, or in vivo in animals. Mammalian cells provide post translational modifications to immunoglobulin protein molecules including correct folding or glycosylation at correct sites.

Mammalian cells which may be useful as hosts include cells of fibroblast origin such as VERO or CHO-K1, or cells of lymphoid origin, such as the hybridoma SP2/0-AG14 or the myeloma P3x63Sg8, and their derivatives.

Several possible vector systems are available for the expression of cloned single chain binding proteins in mammalian cells. One class of vectors utilizes DNA elements which provide autonomously replicating extra-chromosomal plasmids, derived from animal viruses such as bovine papilloma virus, polyoma virus, or SV40 virus. A second class of vectors relies upon the integration of the desired gene sequences into the host cell chromosome. Cells which have stably integrated the introduced DNA into their chromosomes can be selected by also introducing drug resistance genes such as E. coli GPT or Tn5neo. The selectable marker gene can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transfection. Additional elements may also be needed for optimal synthesis of single chain binding

-67-

protein mRNA. These elements may include splice signals, as well as transcription promoters, enhancers, and termination signals. cDNA expression vectors incorporating such elements include those described by Okayama, H., Mol. Cel. Biol., 3:280 (1983), and others.

Another preferred host is yeast. Yeast provides substantial advantages in that it can also carry out post translational peptide modifications including glycosylation. A number of recombinant DNA strategies exist which utilize strong promoter sequences and high copy number of plasmids which can be utilized for production of the desired proteins in yeast. Yeast recognizes leader sequences on cloned mammalian gene products, and secretes peptides bearing leader sequences (i.e., pre-peptides).

Any of a series of yeast gene expression systems incorporating promoter and termination elements from the actively expressed genes coding for glycolytic enzymes produced in large quantities when yeasts are grown in mediums rich in glucose can be utilized. Known glycolytic genes can also provide very efficient transcription control signals. For example, the promoter and terminator signals of the phosphoglycerate kinase gene can be utilized.

Once the strain carrying the single chain building molecule gene has been constructed, the same can also be subjected to mutagenesis techniques using, chemical agents or radiation, as is well known in the art. From the colonies thus obtained, it is possible to search for those producing binding molecules with increased binding affinity. In fact, if the first lin-

-68-

ker designed with the aid of the computer fails to produce an active molecule, the host strain containing the same can be mutagenized. Mutant molecules capable of binding antigen can then be screened by means of a routine assay.

The expressed and refolded single chain binding proteins of the invention can be labelled with detectable labels such as radioactive atoms, enzymes, biotin/avidin labels, chromophores, chemiluminescent labels, and the like for carrying out standard immuno-diagnostic procedures. These procedures include competitive and immunometric (or sandwich) assays. These assays can be utilized for the detection of antigens in diagnostic samples. In competitive and/or sandwich assays, the binding proteins of the invention can also be immobilized on such insoluble solid phases as beads, test tubes, or other polymeric materials.

For imaging procedures, the binding molecules of the invention can be labelled with opacifying agents, such as NMR contrasting agents or X-ray contrasting agents. Methods of binding, labelling or imaging agents to proteins as well as binding the proteins to insoluble solid phases are well known in the art. The refolded protein can also be used for therapy when labelled or coupled to enzymes or toxins, and for purification of products, especially those produced by the biotechnology industry. The proteins can also be used in biosensors.

Having now generally described this invention the same will be better understood by reference to certain specific examples which are included for purposes of illustration and are not intended to be limiting unless otherwise specified.

-69-

EXAMPLES

In these experiments, the basic Fv 3-D structure used for the computer assisted design was that of the anti-phosphoryl choline myeloma antibody of the IgA class, MCPC-603. The X-ray structure of this antibody is publicly available from the Brookhaven data base.

The starting material for these examples was monoclonal antibody cell line 3C2 which produced a mouse anti-bovine growth hormone (BGH). This antibody is an IgG₁ with a gamma 1 heavy chain and kappa light chain. cDNA's for the heavy and light chain sequences were cloned and the DNA sequence determined. The nucleotide sequences and the translation of these sequences for the mature heavy and mature light chains are shown in Figures 21 and 22 respectively.

Plasmids which contain just the variable region of the heavy and light chain sequences were prepared. A ClaI site and an ATG initiation codon (ATCGATG) were introduced before the first codon of the mature sequences by site directed mutagenesis. A HindIII site and termination codon (TAAGCTT) were introduced after the codon 123 of the heavy chain and the codon 109 of the light chain. The plasmid containing the V_H sequences is pGX3772 and that containing the V_L is pGX3773 (Figure 23).

The examples below were constructed and produced by methods known to those skilled in the art.

-70-

EXAMPLE 1A. Computer Design

A two-linker example (referred to as TRY 40) was designed by the following steps.

First, it was observed that light chains were much easier to make in E. coli than were heavy chains. It was thus decided to start with light chain. (In the future, one could certainly make examples which begin with heavy chain because there is a very similar contact between a turn in the heavy chain and the exit strand of the light chain.)

Refer to stereo Figure 30A, which shows the light and heavy domains of the Fv from MOPC-603 antibody; the constant domains are discarded. A line joining the alpha carbons of the light chain is above and dashed. The amino terminus of the light chain is to the back and at about 10 o'clock from the picture center and is labeled "N." At the right edge of the picture, at about 2 o'clock is an arrow showing the path toward the constant domain. Below the light chain is a line joining the alpha carbons of the heavy chain. The amino terminus of the heavy chain is toward the viewer at about 7 o'clock and is also labeled "N." At about 4:30, one sees an arrow showing the heavy chain path to its constant domain.

The antigen-binding site is to the left, about 9 o'clock and between the two loops which project to the right above (light chain) and below (heavy chain).

In addition to the alpha carbon traces, there are three segments in which all non-hydrogen atoms have been drawn. These strands are roughly parallel and from upper right to lower left. They are

-71-

- (a) Proline 46 to Proline 50 of the light chain.
- (b) Valine 111 to Glycine 113 of the heavy chain.
- (c) Glutamic acid 1 to glycine 10 of the heavy chain.

The contact between tryptophan 112 of the heavy chain and proline 50 of the light chain seems very favorable. Thus it was decided that these two residues should be conserved. Several linkers were sought and found which would join a residue at or following Tryptophan 112 (heavy) to a residue at or following Proline 50 (light). Stereo figure 30B shows the region around TRP 112H in more detail. The letter "r" stands between the side-chain of TRP 112H and PRO 50L; it was wished to conserve this contact. The letter "q" labels the carboxy terminal strand which leads towards the constant domain. It is from this strand that a linker will be found which will connect to PRO 50L.

Once a linker is selected to connect 112H to 50L, one needs a linker to get from the first segment of the light chain into the beginning portion of the heavy chain. Note that PRO 46L turns the chain toward PRO 50L. This turning seemed very useful, so it was decided to keep PRO 46L. Thus the second linker had to begin after 46L and before 50L, in the stretch marked "s." A search for linkers was done beginning on any of the residues 46L, 47L, or 48L. Linkers beginning on residue 49L were not considered because the chain has already turned toward 50L and away from the amino terminal of the heavy chain. Linkers were sought which ended on any of the residues 1H to 10H.

-72-

Figure 30C shows the linked structure in detail. After TRP 112H and GLY 113H, was introduced the sequence PRO-GLY-SER, and then comes PRO 50L. A computer program was used to look for short contacts between atoms in the linker and atoms in the retained part of the Fv. There is one short contact between the beta carbon of the SER and PRO 50L, but small movements would relieve that. This first linker runs from the point labeled "x" to the point labeled "y." The second linker runs from "v" to "w." Note that most of the hydrophobic residues (ILE and VAL) are inside. There is a PHE on the outside. In addition, the two lysine residues and the asparagine residue are exposed to solvent as they ought to be. Figure 30D shows the overall molecule linked into a single chain.

B. Genetic Constructs

These constructs were prepared and the plasmids containing them using E. coli hosts. Once constructed, the sequences can be inserted into whichever expression vehicle used in the organism of choice.

The first construction was TRY40 (the two-linker construction) which produces a protein with the following sequence:

Met-[L-chain 1-41]-Ile-Ala-Lys-Ala-Phe-Lys-Asn-[H-chain 8-105]-Pro-Gly-Ser-[L-chain 45-109]. The nucleotide sequence and its translation are seen in Figure 24. The hypervariable regions in TRY40 (as in TRY61 59 and 104B, see below) correspond, as indicated, to an IgG1 anti BGH antibody, even though the 3-D analysis was done on the Fv region of MCPC-603 antibody, having a different specificity, (anti phosphoryl choline) but having a similar framework in the variable region.

-73-

The antibody sequences in the plasmids pGX3772 and pGX3773 were joined to give the sequence of TRY40 in the following manner. The plasmids used contained an M13 bacteriophage origin of DNA replication. When hosts containing these plasmids are superinfected with bacteriophage M13 two types of progeny are produced, one containing the single-strand genome and the other containing a specific circular single-strand of the plasmid DNA. This DNA provided template for the oligonucleotide directed site specific mutagenesis experiments that follow. Template DNA was prepared from the two plasmids. An EcoRI site was introduced before codon 8 of the V_H sequence in pGX3772, by site directed mutagenesis, producing pGX3772'. Template from this construction was prepared and an XbaI site was introduced after codon 105 of the V_H sequence producing pGX3772''.

An EcoRI and an XbaI site were introduced into pGX3773 between codons 41 and 45 of the V_L sequence by site directed mutagenesis producing pGX3773'.

To begin the assembly of the linker sequences plasmid pGX3773' (V_L) DNA was cleaved with EcoRI and XbaI and treated with calf alkaline phosphatase. This DNA was ligated to the EcoRI to XbaI fragment purified from plasmid pGX3772'' (V_H) which had been cleaved with the two restriction enzymes. The resulting plasmid pGX3774, contained the light and heavy chain sequences in the correct order linked by the EcoRI and XbaI restriction sites. To insert the correct linker sequences in frame, pGX3774 template DNA was prepared. The EcoRI junction was removed and the linker coding for the -Ile-Ala-Lys-Ala-Phe-Lys-Asn- inserted by site

-74-

directed mutagenesis, producing plasmid pGX3774'. Template DNA was prepared from this construction and the XbaI site corrected and the linker coding for -Pro-Gly-Ser- inserted by site directed mutagenesis producing plasmid pGX3775. The sequence was found to be correct as listed in Figure 24 by DNA sequencing.

In order to express the single-chain polypeptide, the sequence as a ClaI to HindIII fragment was inserted into a vector pGX3703. This placed the sequence under the control of the O_L/P_R hybrid lambda promoter (U.S. Patent Application 534,982, Sept. 23, 1983). The expression plasmid is pGX3776 (Figure 25). The plasmid pGX3776 was transformed into a host containing a heat sensitive lambda phage repressor; when grown at 30°C the synthesis of the TRY40 protein is repressed. Synthesis was induced by raising the temperature to 42°C, and incubating for 8-16 hours. The protein was produced at 7.2% of total cell protein, as estimated on polyacrylamide gel electropherograms stained with Coomassie blue.

EXAMPLE 2

A. Computer Design

A one-linker example (referred to as TRY 61) was designed by the following steps.

Refer to stereo Figure 31A which shows the light and heavy domains of the Fv; the constant domains are discarded. A line joining the alpha carbons of the light chain is dashed. The amino terminus of the light chain is to the back and at about the center of the picture and is labeled "N." At the right edge of the picture, at about 2 o'clock is an arrow showing

-75-

the path toward the constant domain of the light chain. Below the light chain is a line joining the alpha carbons of the heavy chain. The amino terminus of the heavy chain is toward the viewer at about 9 o'clock and is also labeled "N". At about 4:30, one sees an arrow showing the heavy chain path to its constant domain.

In addition to the alpha carbon traces, there are two segments in which all non-hydrogen atoms have been drawn. These segments are the last few residues in the light chain and the first ten in the heavy chain. Linkers were sought between all pairs of these residues, but only a few were found because these regions are widely separated.

Figure 31B shows the linker in place. Note that the molecule now proceeds from the amino terminal of the light chain to the carboxy terminal strand of the heavy chain. Note also that the antigen-binding region is to the left, on the other side of the molecule from the linker.

B. Genetic Constructs

The sequence of TRY61 (a single-linker embodiment) is Met-[L-chain 1-104]-Val-Arg-Gly-Ser-Pro-Ala-Ile-Asn-Val-Ala-Val-His-Val-Phe-[H-chain 7-123]. The nucleotide sequence and its translation are shown in Figure 26.

To construct TRY61, plasmid pGX3772' DNA was cleaved with ClaI and EcoRI and treated with calf alkaline phosphatase. This DNA was ligated with the ClaI to HindIII fragment from pGX3773 and two oligonucleotides which code for the linker sequence and have HindIII and EcoRI ends, so that the linker can

-76-

only be ligated in the correct orientation. The resulting plasmid, pGX3777, was used to prepare template DNA. This DNA was used for site directed mutagenesis to remove the HindIII site inside the antibody sequences. The correct construction, pGX3777', was used to make template DNA for a site directed mutagenesis to remove the EcoRI site. The ClaI to HindIII fragment from the final construction, pGX3778, containing the TRY61 coding sequence was confirmed by DNA sequencing. The ClaI to HindIII was inserted into the pGX3703 expression vector. This plasmid is called pGX4904 (Figure 27). This plasmid was transformed into an E. coli host. The strain containing this plasmid has been induced, and the single chain protein produced as >2% of total cell protein.

EXAMPLE 3

A. Computer Design

A one-linker example (referred to as TRY 59) was designed by the following steps.

Refer to stereo Figure 32A which shows the light and heavy domains of the Fv; the constant domains are discarded. A line joining the alpha carbons of the light chain is above and dashed. The amino terminus of the light chain is to the back and at about 10 o'clock from the center of the picture and is labeled "N". At the right edge of the picture, at about 2 o'clock is an arrow showing the path toward the constant domain of the light chain. Below the light chain is a line joining the alpha carbons of the heavy chain. The amino terminus of the heavy chain is toward the viewer at about 8 o'clock and is also labeled

-77-

"N". At about 4:30, one sees an arrow showing the heavy chain path to its constant domain.

In addition to the alpha carbon traces, there are two segments in which all non-hydrogen atoms have been drawn. These segments are the last few residues in the light chain and the first ten in the heavy chain. Linkers we sought between all pairs of these residues, but only a few were found because these regions are widely separated.

Figure 32B shows the linker in place. Note that the molecule now proceeds from the amino terminal of the light chain to the carboxy terminal strand of the heavy chain. Note also that the antigen-binding region is to the left, on the other side of the molecule from the linker.

The choice of end points in TRY59 is very similar to TRY61. Linkers of this length are rare. The tension between wanting short linkers that fit very well and which could be found for the two-linker case (TRY40) and the desire to have only one linker, (which is more likely to fold correctly) is evident in the acceptance of TRY59. The linker runs from the point marked "A" in Figure 32B to the point marked "J." After five residues, the linker becomes helical. At the point marked "x," however, the side-chain of an ILE residue collides with part of the light chain. Accordingly, that residue was converted to GLY in the actual construction.

B. Genetic Constructs

The sequence of TRY59 (the single linker construction) is Met-[L-chain 1-105]-Lys-Glu-Ser-Gly-Ser-Val-Ser-Ser-Glu-Gln-Leu-Ala-Gln-Phe-Arg-Ser-Leu-Asp-[H-

-78-

chain 2-1231. The nucleotide sequence coding for this amino acid sequence and its translation is shown in Figure 28. The BglI to HindIII fragment (read clockwise) from plasmid pGX3773 containing the V_L sequence and the ClaI to BglI fragment (clockwise) from pGX3772 has been ligated with two oligonucleotides which form a fragment containing the linker sequence for TRY59 and have ClaI and HindIII ends. The ClaI and HindIII junctions within this plasmid are corrected by two successive site directed mutageneses to yield the correct construction. The ClaI to HindIII fragment from this plasmid is inserted into the O_L/P_R expression vector as in Examples 1 and 2. The resulting plasmid, pGX4908 (Figure 29) is transformed into an E. coli host. This strain is induced to produce the protein coded by the sequence in Figure 28 (TRY59).

Example 4

A. Computer Design

In this design an alternative method of choosing a linker to connect the light and heavy variable regions was used. A helical segment from human hemoglobin was chosen to span the major distance between the carboxy terminus of the variable light chain and the amino terminus of the variable heavy chain. This alpha helix from human hemoglobin was positioned at the rear of the F_v model using the computer graphics system. Care was taken to position the helix with its ends near the respective amino and carboxyl termini of the heavy and light chains. Care was also taken to place hydrophobic side chains in toward the F_v and hydro-

-79-

philic side chains toward the solvent. The connections between the ends of the variable regions and the hemoglobin helix were selected by the previously described computer method (EXAMPLE 1-3).

B. Genetic Constructs

The sequence of TRY104b (a single linker construction) is Met-[L-chain 1-106]-Ala-Glu-Gly-Thr-[(Hemoglobin helix)Leu-Ser-Pro-Ala-Asp-Lys-Thr-Asn-Val-Lys-Ala-Ala-Trp-Gly-Lys-Val-]Met-Thr-[H-chain 3-123]. The nucleotide sequence coding for this amino acid sequence and its translation is shown in Figure 33. The BglI to HindIII fragment (read clockwise) from plasmid pGX3773 containing the V_L sequence and the ClaI to BglI fragment (clockwise) from pGX3772 has been ligated with two oligonucleotides which form a fragment containing the linker sequence for TRY104b and have ClaI and HindIII ends. The ClaI and HindIII junctions within this plasmid are corrected by two successive site directed mutageneses to yield the correct construction. The ClaI to HindIII fragment from this plasmid is inserted into the O_L/P_R expression vector as in Examples 1-3. The resulting plasmid, pGX4910 (Figure 34) is transformed into an E. coli host. This strain is induced to produce the protein coded by the sequence in Figure 33 (TRY104b).

EXAMPLE 5

Purification of the Proteins

The single-chain antigen binding proteins from TRY40, TRY61, TRY59 and TRY104b are insoluble, and cells induced to produce these proteins show refractile bodies called inclusions upon microscopic exami-

-80-

nation. Induced cells were collected by centrifugation. The wet pellet was frozen on dry ice, then stored at -20°C . The frozen pellet was suspended in a buffer and washed in the same buffer, and subsequently the cells were suspended in the same buffer. The cells were broken by passage through a French pressure cell, and the inclusion bodies containing the single-chain antigen binding protein (SCA) were purified by repeated centrifugation and washing. The pellet was solubilized in guanidine-HCl, and reduced with 2-mercaptoethanol. The solubilized material was passed through a gel filtration column, i.e., SephacrylTM S-300. Other methods such as ion exchange could be used.

EXAMPLE 6

Folding of the Proteins

Purified material was dialyzed against water, and the precipitate protein collected by centrifugation. The protein was solubilized in urea and reduced with 2-mercaptoethanol. This denatured and solubilized material was dialyzed against a buffer containing salt and reducing agents to establish the redox potential to form the intra domain (one each for the light and heavy chain variable region sequences) disulfide bridges (Saxena and Wetlanfer, Biochem 9:5015-5023 (1970)). The folded protein was assayed for BGH binding activity.

The TRY59 protein used in competition experiments was solubilized and renatured directly from inclusions. This material was subsequently purified by affinity to BGH-Sepharose.

-81-

EXAMPLE 7Binding Assay

BGH was immobilized on nitrocellulose strips along with non-specific proteins such as bovine serum albumin or lysozymes. Further non-specific protein binding was blocked with an immunologically inert protein, for example gelatin. Folded SCA was tested for its ability to bind to BGH. The SCA was detected by a rabbit anti-L chain (of the monoclonal) anti-serum. The rabbit antibodies were reacted with goat anti-rabbit IgG coupled to peroxidase. The strips were reacted with chemicals which react with the peroxidase to give a color reaction if the peroxidase is present.

Figure 35 shows the result of this spot assay for TRY61 (strip 1) and TRY40 (strip 2). Strip 3 was stained with amido black to show the presence of all three proteins. The other proteins, TRY59, TRY104b gave similar results in the spot assay. A competition assay with the SCA competing with the monoclonal can be used as well. The results of competing the F_{ab} of 3C2 monoclonal with 1 and 10 ug of TRY59 protein which had been affinity purified are shown in Figure 36 ($\blacktriangle$ F_{ab} alone, $\blacksquare$ F_{ab} + 1 ug TRY59, and $\bullet$ F_{ab} + 10 ug TRY59). The affinity estimated from the Ic_{50} of this experiment was approximately 10^6 . The data are summarized in Table 1.

Table 1
Properties of the Proteins Produced by the
Single-chain Constructions

Single-chain	Number of Linkers	Spot Test	Binding to BGH Column	K_a
TRY40	2	+	-	ND
TRY59	1	+	+	10^8
TRY61	1	+	-	ND
TRY104B	1	+	+	ND
3C2 monoclonal	NA	+	+	$10^8 - 10$

ND = not determined

NA = not applicable

-82-

WE CLAIM:

1. A single polypeptide chain binding molecule which has binding specificity substantially similar to the binding specificity of the light and heavy chain aggregate variable region of an antibody.

2. The molecule of claim 1 which comprises two peptide linkers joining said light and heavy chains into said single chain.

3. The molecule of claim 2 which comprises in sequence:

- (a) an N-terminal region derived from said light chain;
- (b) a peptide linker;
- (c) a peptide region derived from said heavy chain;
- (d) a second peptide linker; and
- (e) a C-terminal region derived from said light chain.

4. The molecule of claim 1 which comprises one peptide linker joining said light and heavy chains into said single chain.

5. The molecule of claim 4 which comprises, in sequence:

- (a) an N-terminal region derived from said light chain;
- (b) a peptide linker; and
- (c) a C-terminal region derived from said heavy chain.

-83-

6. The molecule of claim 4 which comprises in sequence:

- (a) an N-terminal region derived from said heavy chain;
- (b) a peptide linker; and
- (c) a C-terminal region derived from said light chain.

7. The molecule of claim 3, 5 or 6 which, prior to said N-terminal region (a), comprises a methionine residue.

8. The molecule of claim 1 which is detectably labeled.

9. The molecule of claim 1 which is in immobilized form.

10. The molecule of claim 1 which is conjugated to an imaging agent.

11. The molecule of claim 1 which is conjugated to a toxin.

12. A genetic sequence coding for the molecule of claim 1.

13. A recombinant DNA (rDNA) molecule comprising the sequence of claim 12.

14. The rDNA molecule of claim 13 which is a replicable cloning or expression vehicle.

-84-

15. The rDNA molecule of claim 14 wherein said vehicle is a plasmid.

16. A host cell transformed with the rDNA molecule of claim 13.

17. The host cell of claim 16 which is a bacterial cell, a yeast or other fungal cell or a mammalian cell line in vitro.

18. A method of producing a single polypeptide chain binding molecule which has binding specificity substantially similar to the binding specificity of the light and heavy chain aggregate variable region of an antibody, which comprises:

- (a) providing a genetic sequence coding for said molecule;
- (b) transforming a host cell with said sequence;
- (c) expressing said sequence in said host; and
- (d) recovering said molecule.

19. The method of claim 18 which further comprises purifying said recovered molecule.

20. The method of claim 18 wherein said host cell is a bacterial cell, yeast or other fungal cell, or a mammalian cell line.

21. The binding molecule produced by the method of claim 18 or 19.

-85-

22. In an immunoassay method which utilizes an antibody in labeled form, the improvement comprising using the molecule of claim 8 instead of said antibody.

23. In an immunoassay method which utilizes an antibody in immobilized form, the improvement comprising using the molecule of claim 9 instead of said antibody.

24. In the immunoassay of claim 21 or 22 wherein said immunoassay is a competitive immunoassay.

25. In the immunoassay of claim 21 or 22 wherein said immunoassay is a sandwich immunoassay.

26. In an immunotherapeutic method which utilizes an antibody conjugated to a therapeutic agent, the improvement comprising using the molecule of claim 1 instead of said antibody.

27. In a method of immunoaffinity purification which utilizes an antibody therefor, the improvement which comprises using the molecule of claim 1 instead of said antibody.

1/34

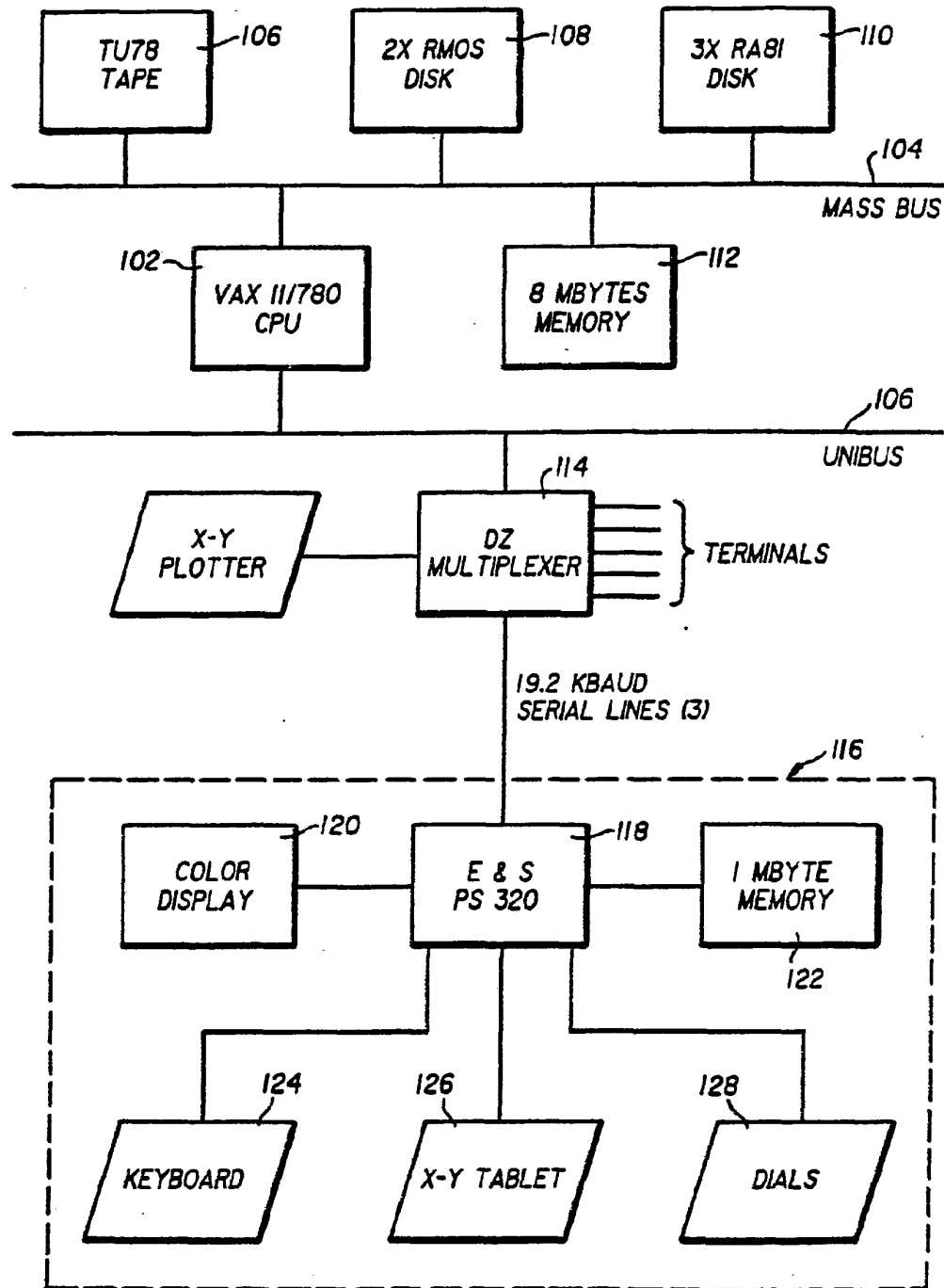


FIG. 1

SUBSTITUTE SHEET

2/34

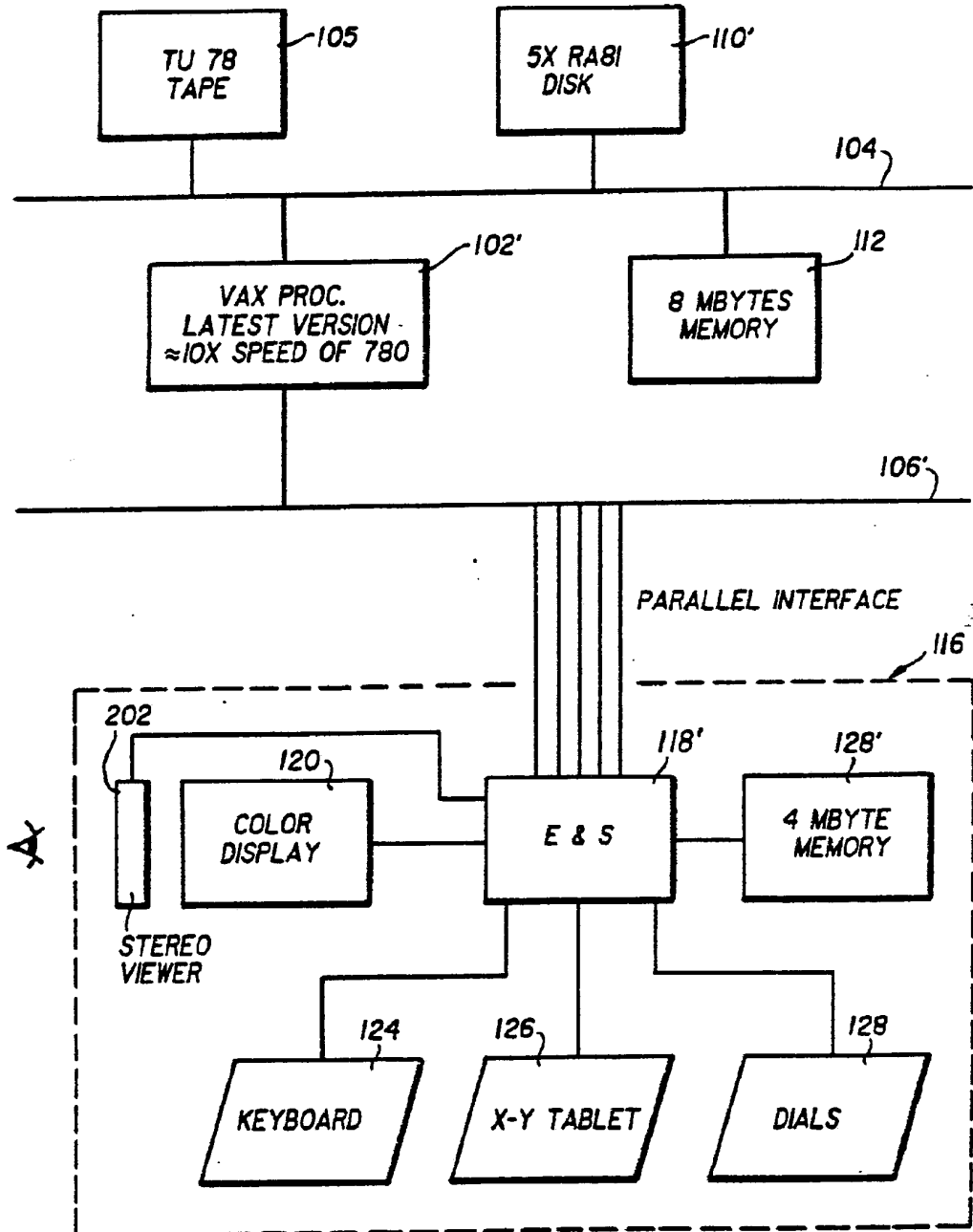
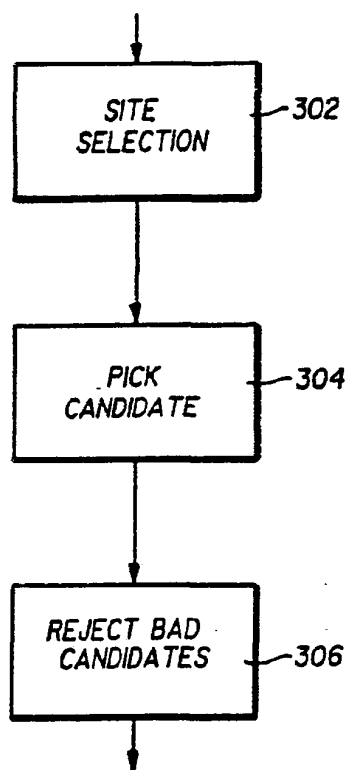
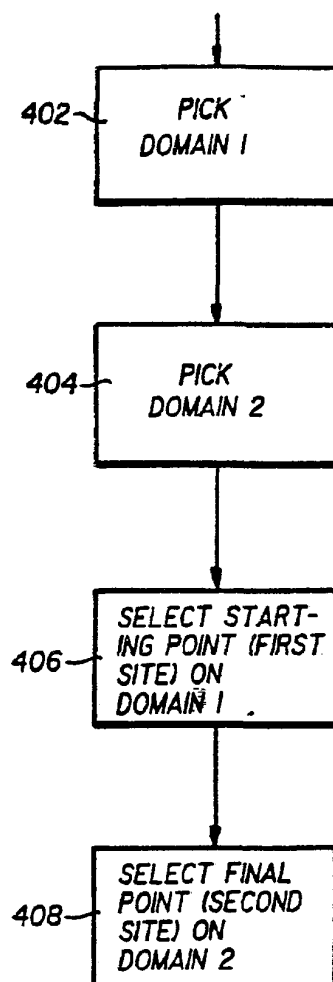
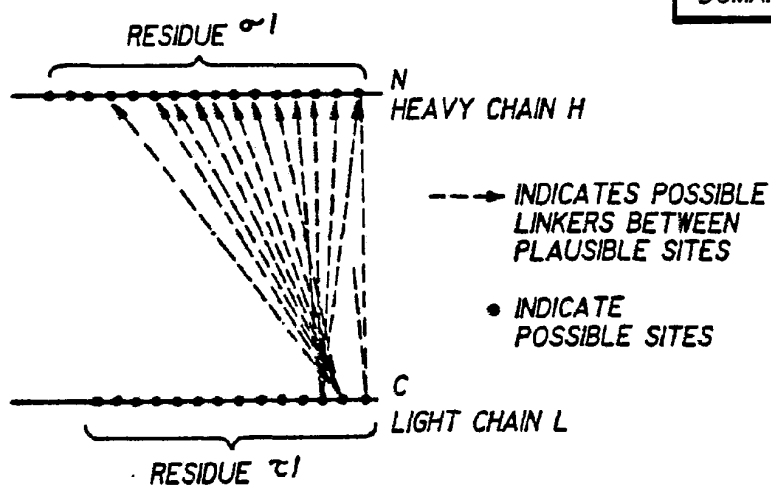


FIG. 2

SUBSTITUTE SHEET

3/34

**FIG. 3****FIG. 4****FIG. 7**

SUBSTITUTE SHEET

4/34

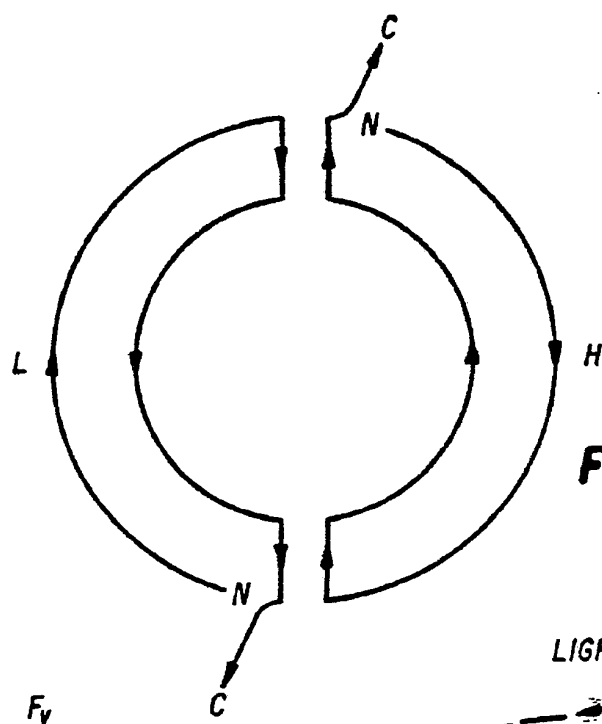


FIG. 5A

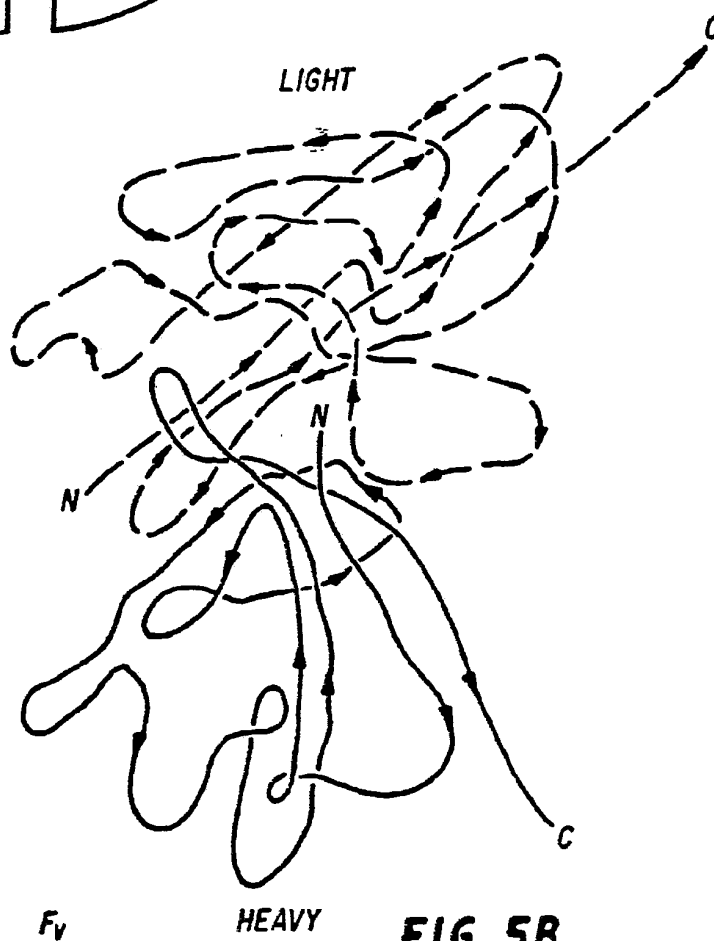
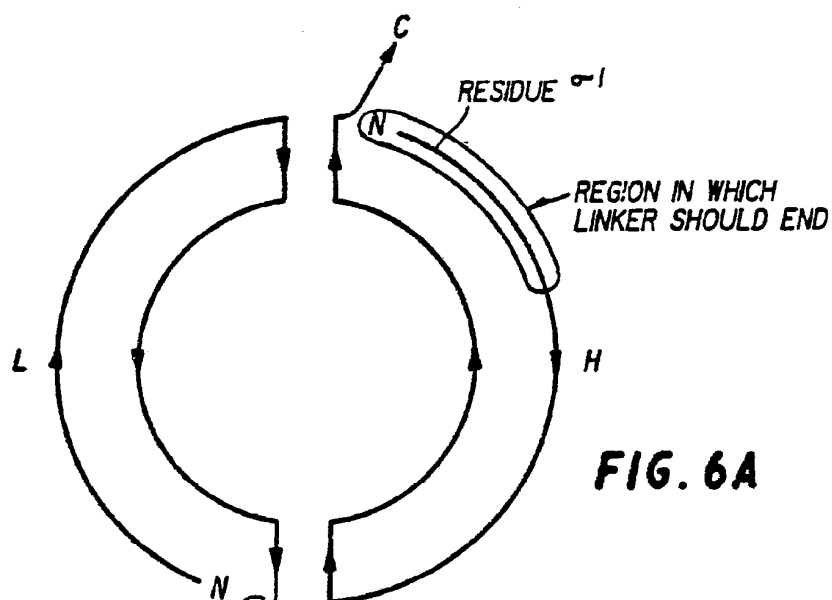
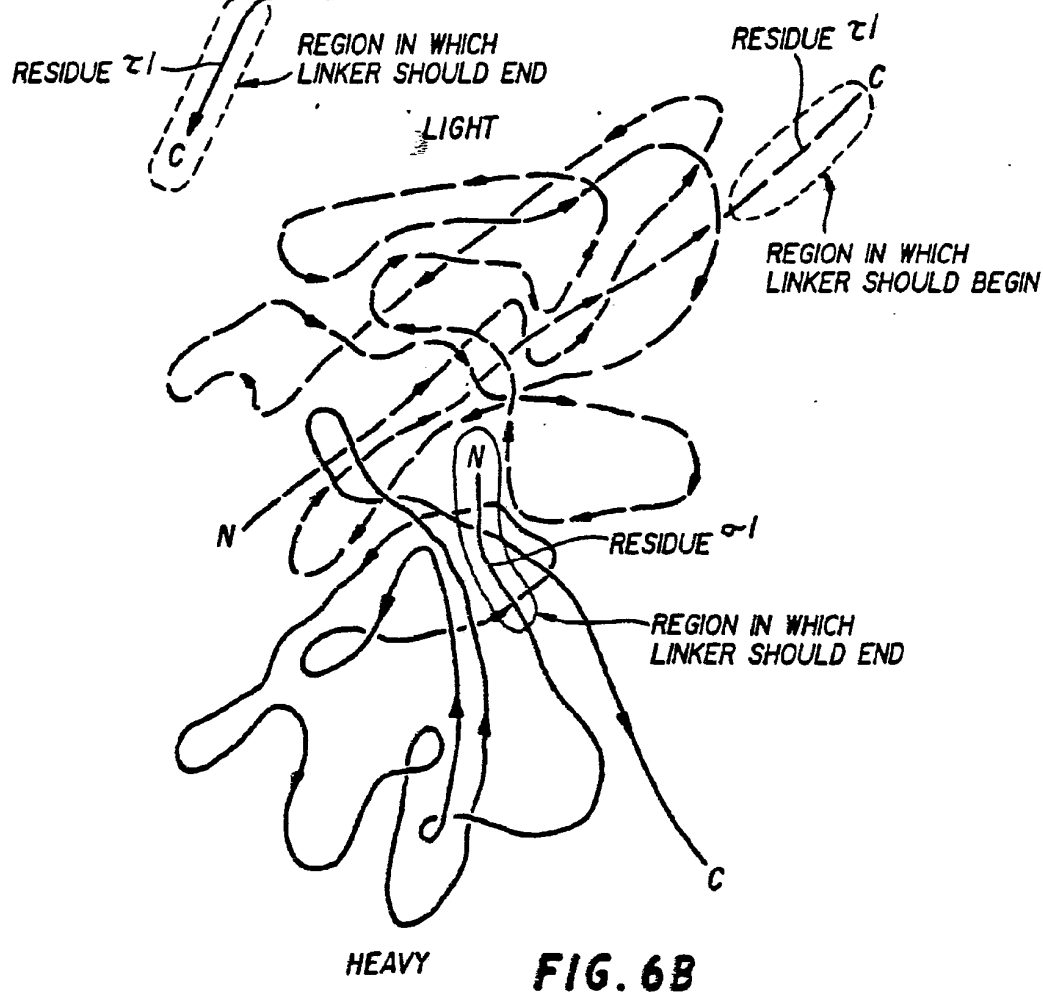


FIG. 5B

SUBSTITUTE SHEET

5/34



SUBSTITUTE SHEET

6/34

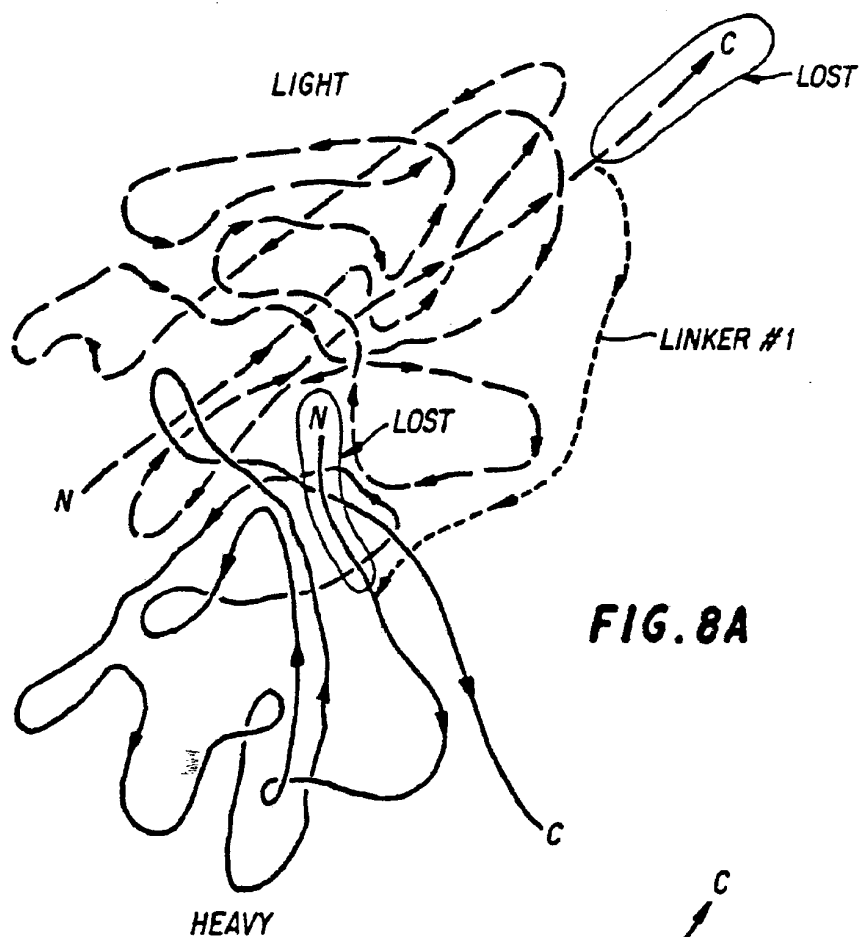


FIG. 8A

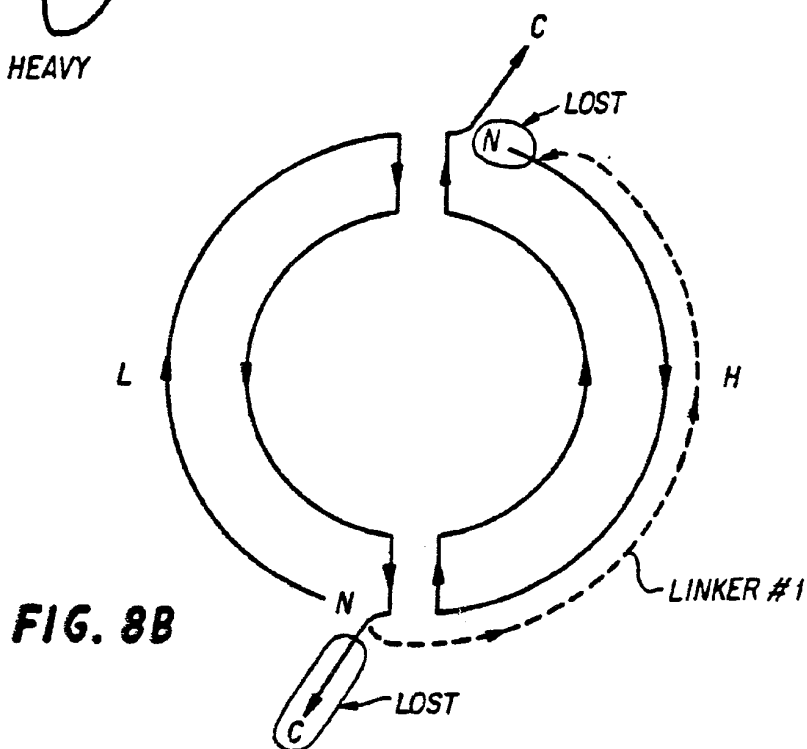
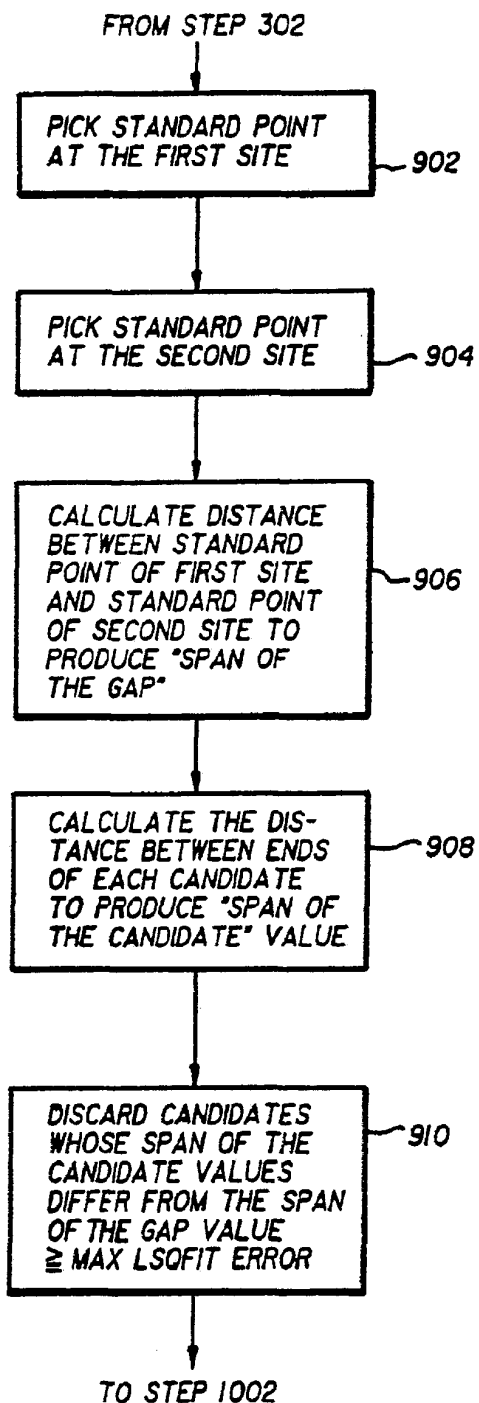


FIG. 8B

SUBSTITUTE SHEET

7/34

**FIG. 9**

SUBSTITUTE SHEET

8/34

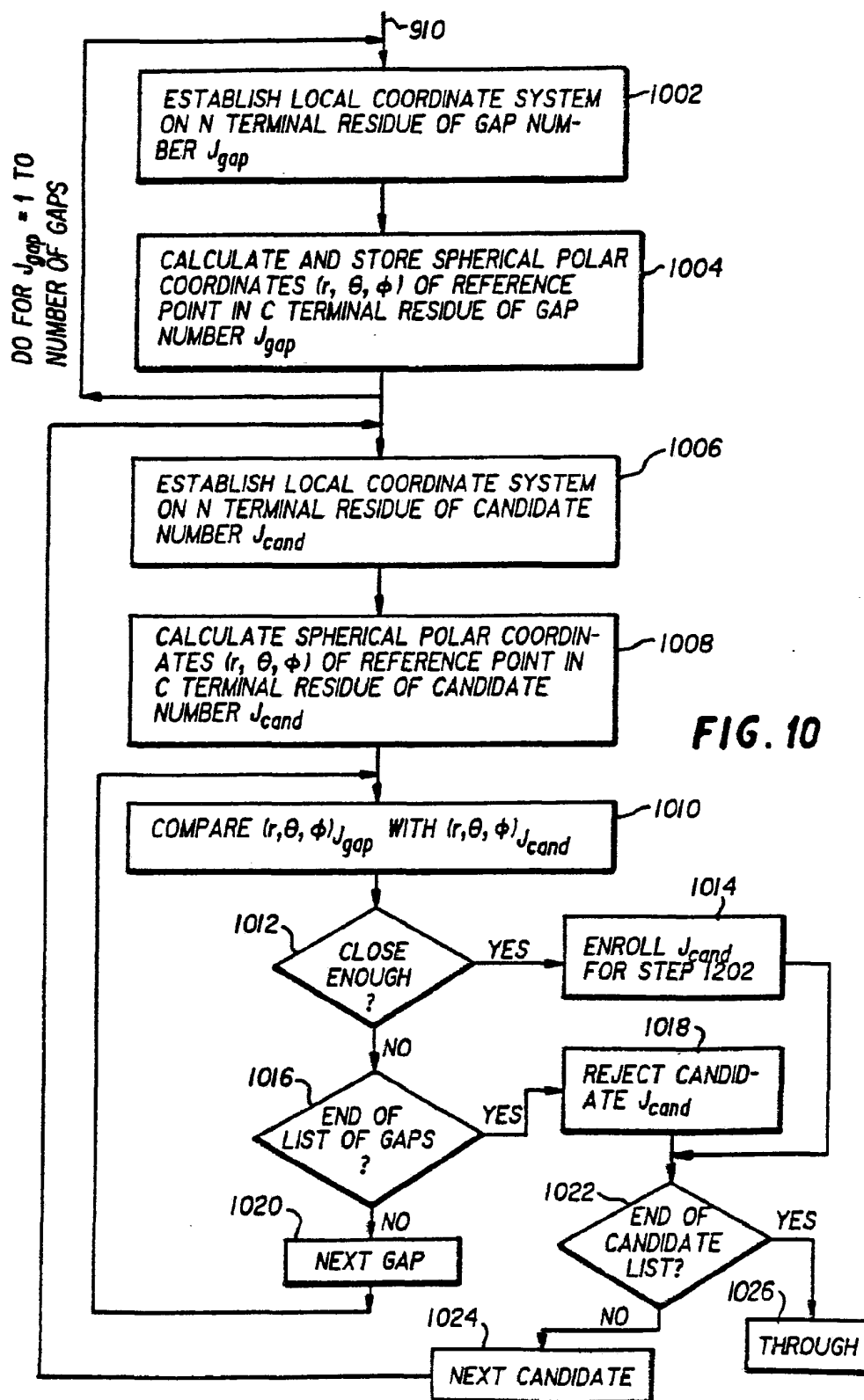


FIG. 10

SUBSTITUTE SHEET

9/34

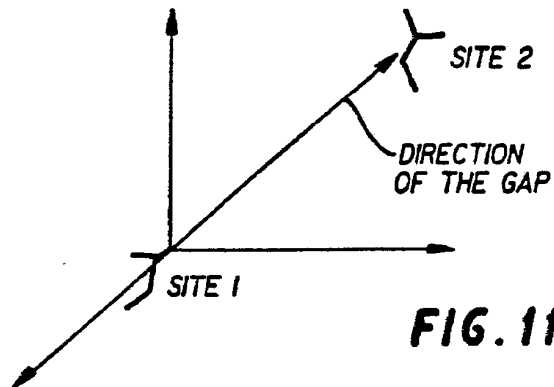


FIG. 11A

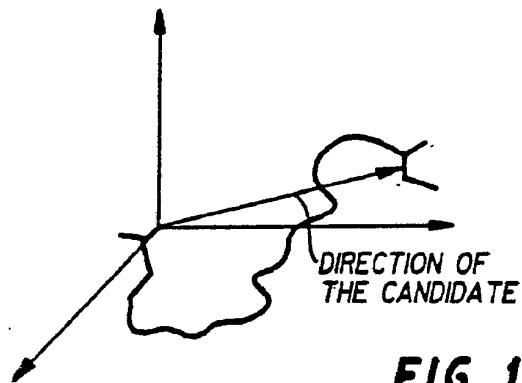


FIG. 11B

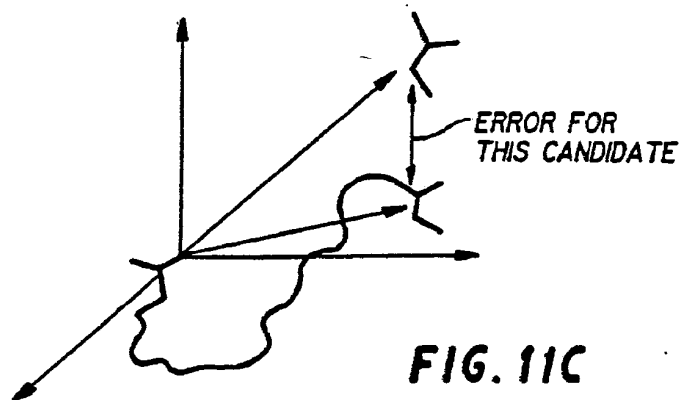


FIG. 11C

SUBSTITUTE SHEET

10/34

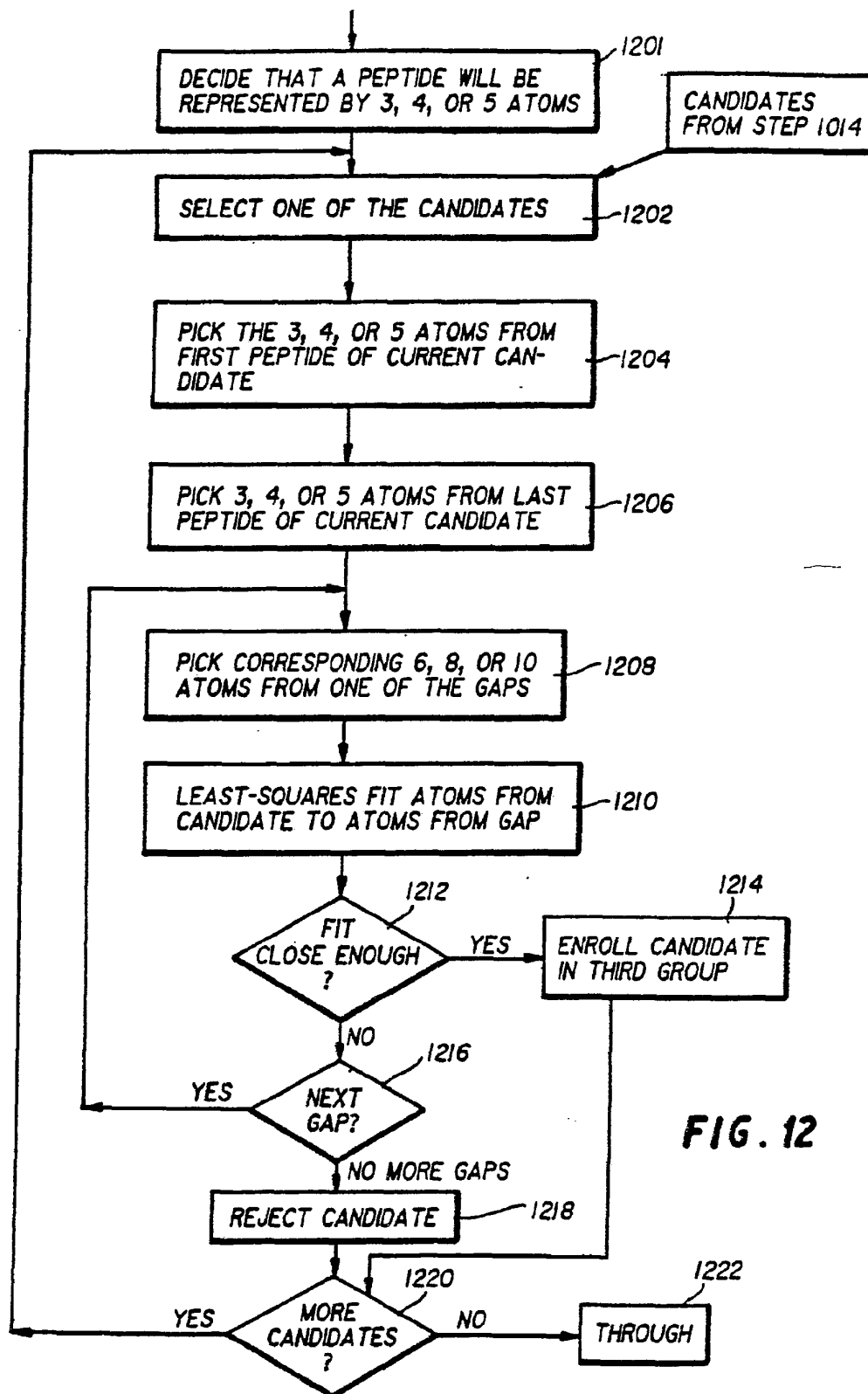
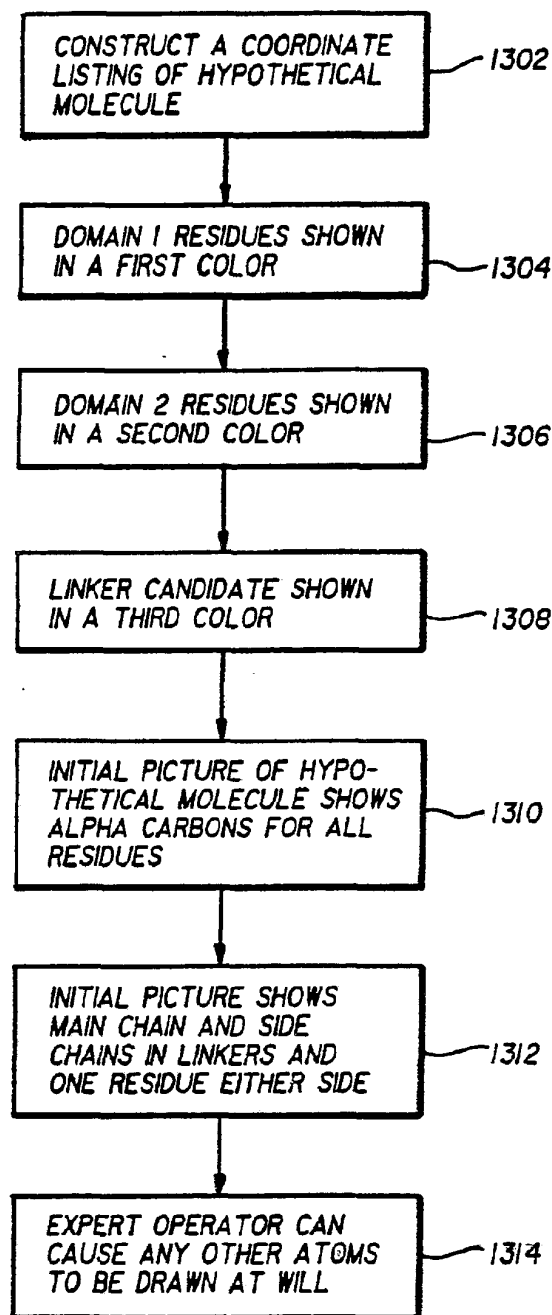


FIG. 12

SUBSTITUTE SHEET

11/34

**FIG. 13**

SUBSTITUTE SHEET

12/34

REJECT CANDIDATE IF ANY ATOM OF
LINKER COMES CLOSER THAN MINIMUM
ALLOWED SEPARATION TO ANY RETAINED
ATOM OF NATIVE STRUCTURE

1402

PENALIZE CANDIDATE WHEN HYDROPHOBIC
RESIDUES HAVE HIGH EXPOSURE TO SOLVENT

1404

PENALIZE CANDIDATE WHEN HYDROPHYLIC
RESIDUES HAVE LOW EXPOSURE TO SOLVENT

1406

PROMOTE CANDIDATE WHEN HYDROPHOBIC RES-
IDUES HAVE LOW EXPOSURE TO SOLVENT

1408

PROMOTE CANDIDATE WHEN HYDROPHYLIC RES-
IDUES HAVE HIGH EXPOSURE TO SOLVENT

1410

PENALIZE CANDIDATE WHEN MAIN CHAIN FAILS
TO FORM HYDROGEN BOND

1412

PENALIZE CANDIDATE WHEN MAIN CHAIN MAKES
USELESS EXCURSIONS INTO THE SOLVENT
REGION

1414

PROMOTE CANDIDATE WHEN MAIN CHAIN
FORMS A HELIX

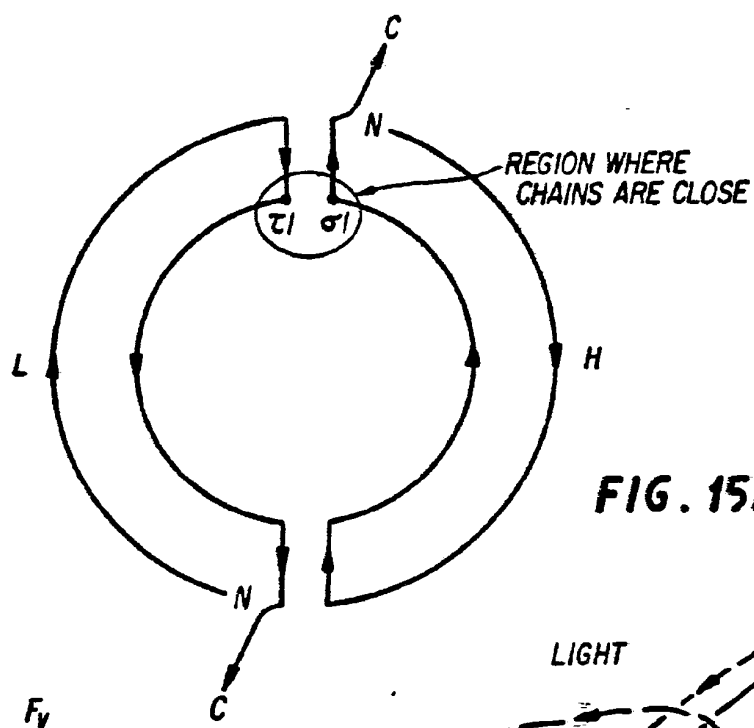
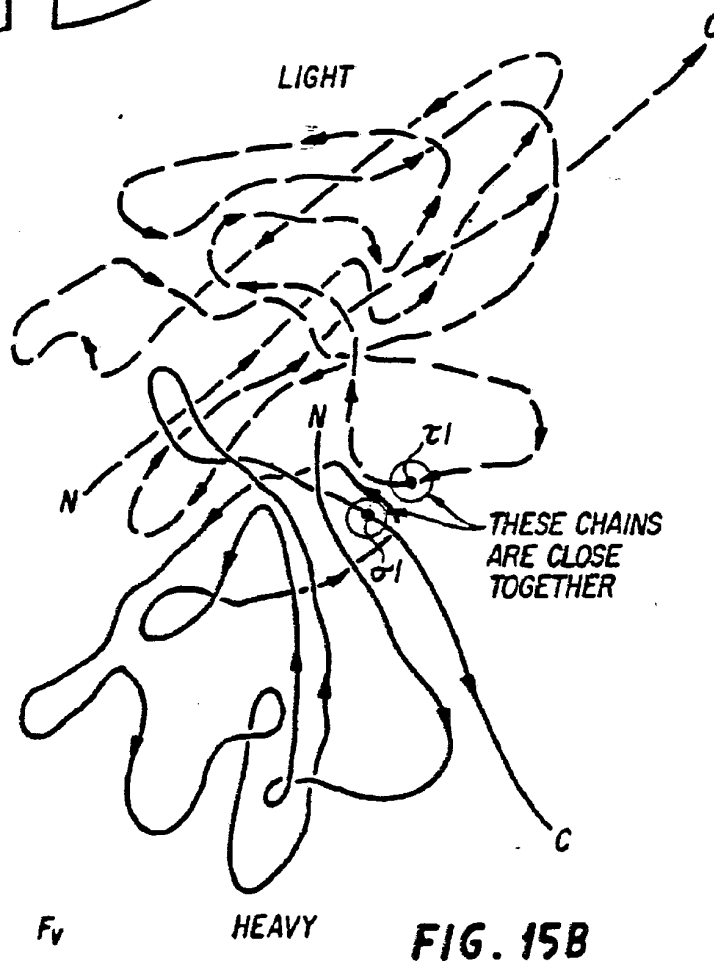
1416

PROMOTE CANDIDATE WHEN MAIN CHAIN FORMS
A BETA SHEET WHICH FITS AGAINST EXISTING
BETA SHEETS

1418

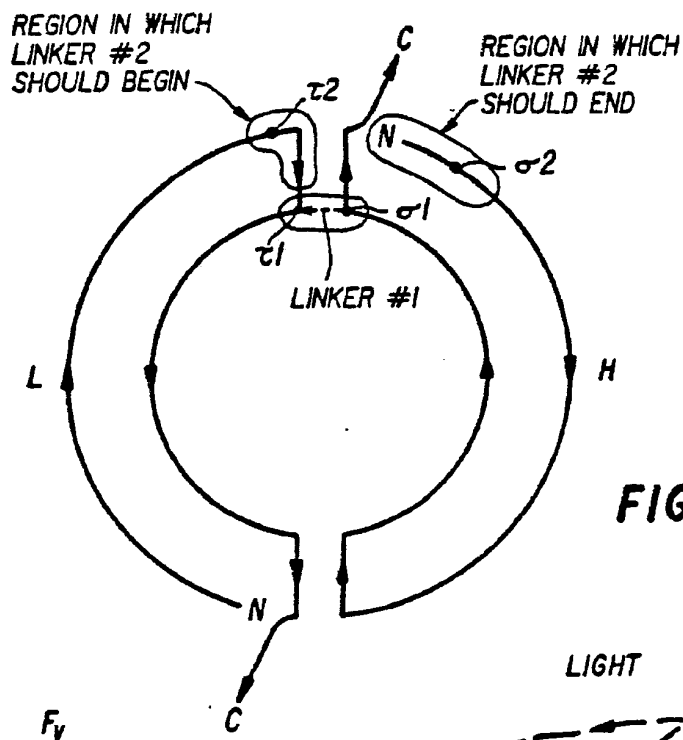
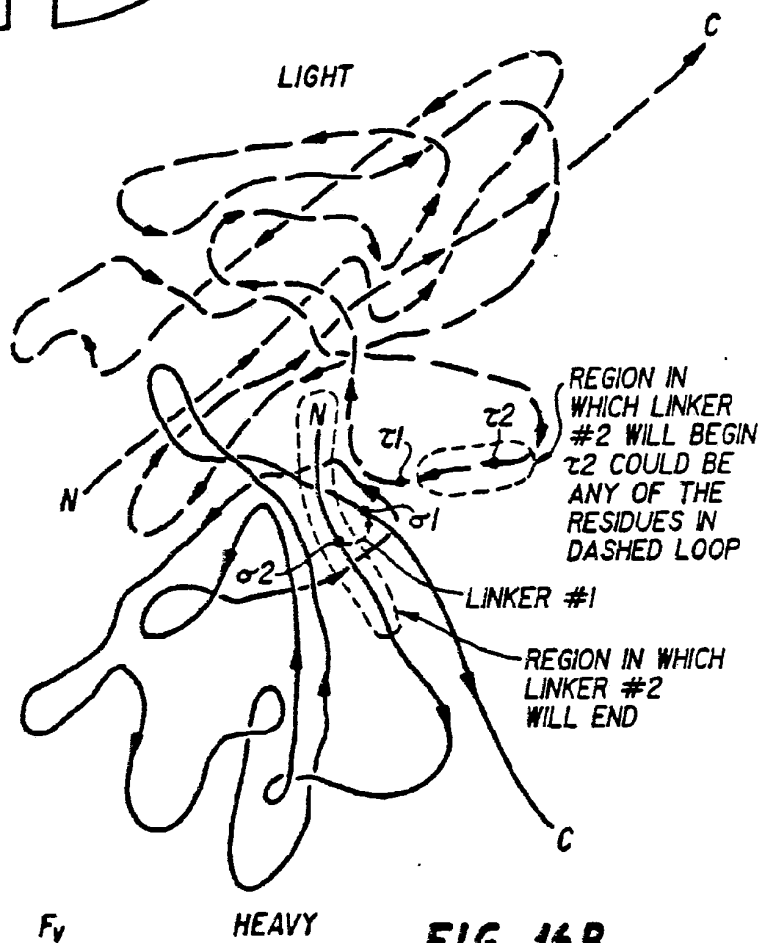
FIG. 14

13/34



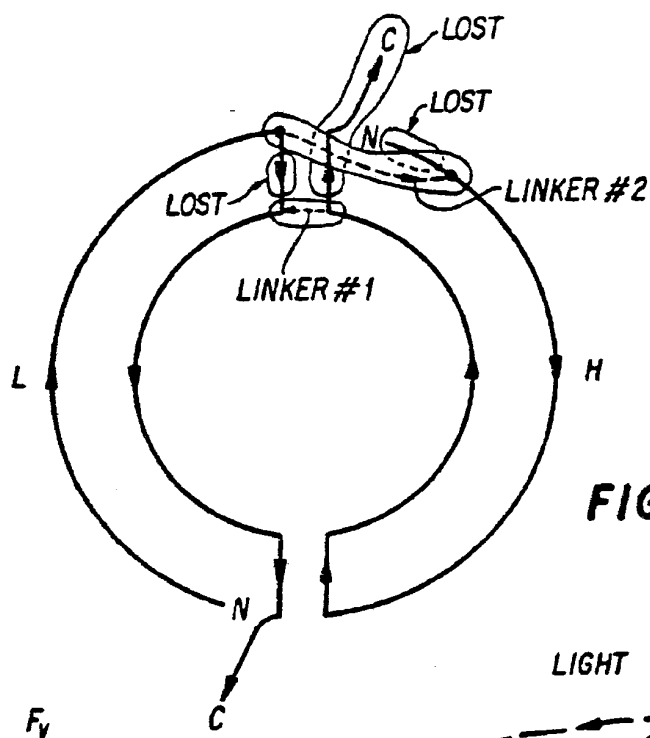
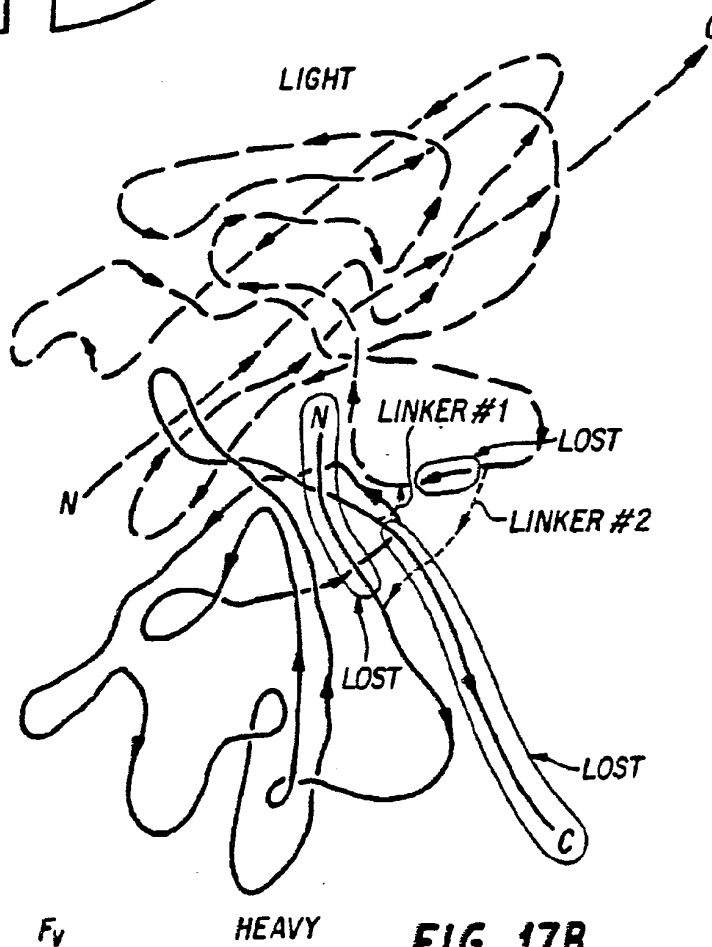
SUBSTITUTE SHEET

14/34



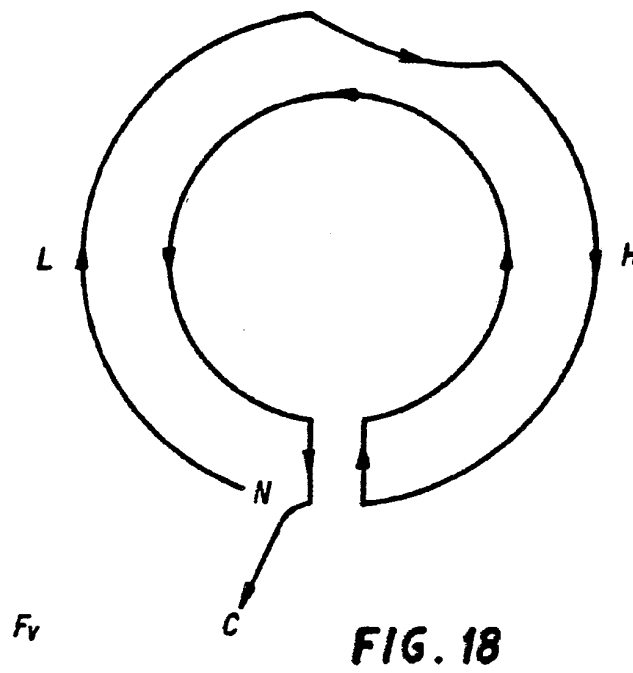
SUBSTITUTE SHEET

15/34



SUBSTITUTE SHEET

16/34



SUBSTITUTE SHEET

17/34

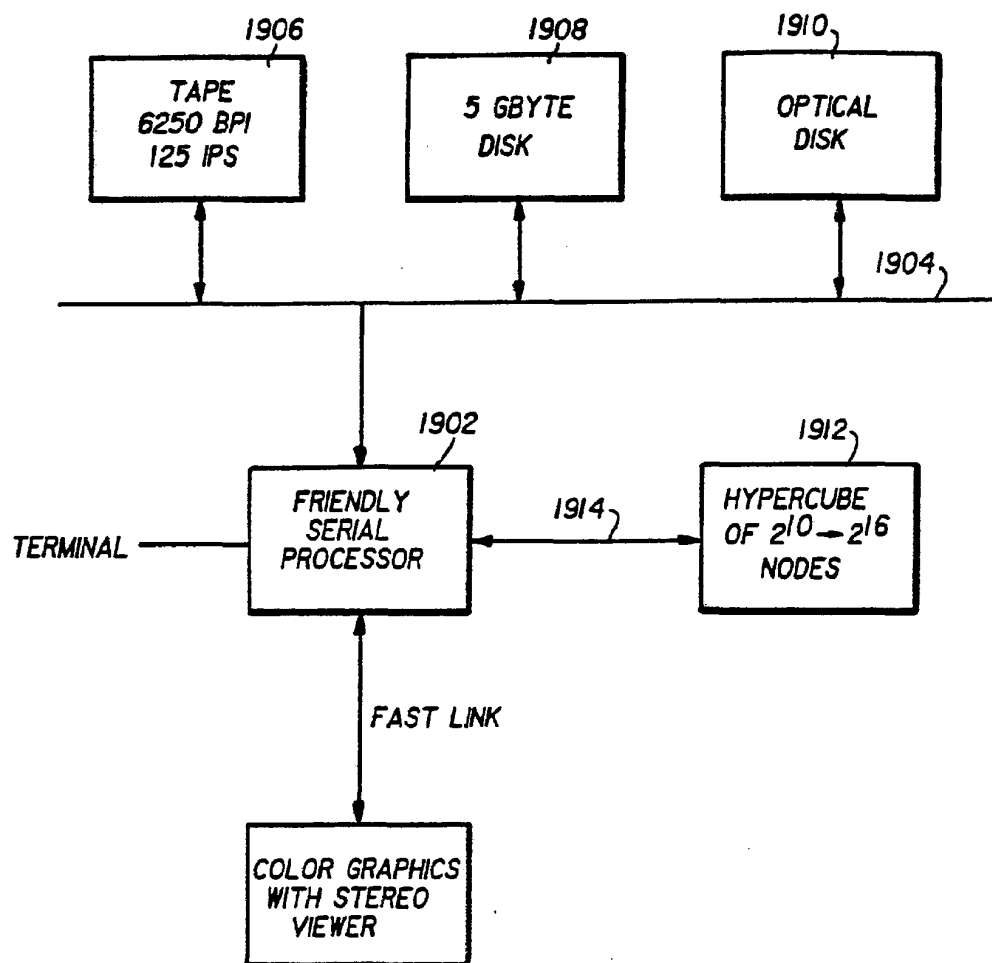


FIG. 19

SUBSTITUTE SHEET

18/34

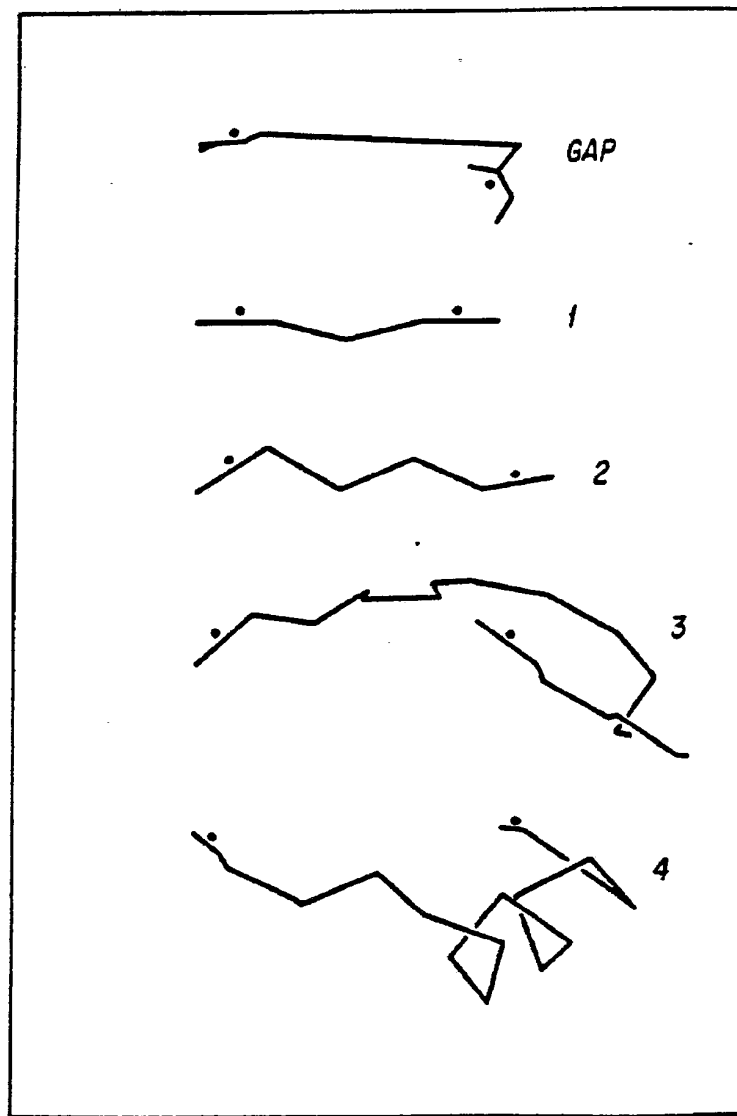


FIG. 20A

19/34

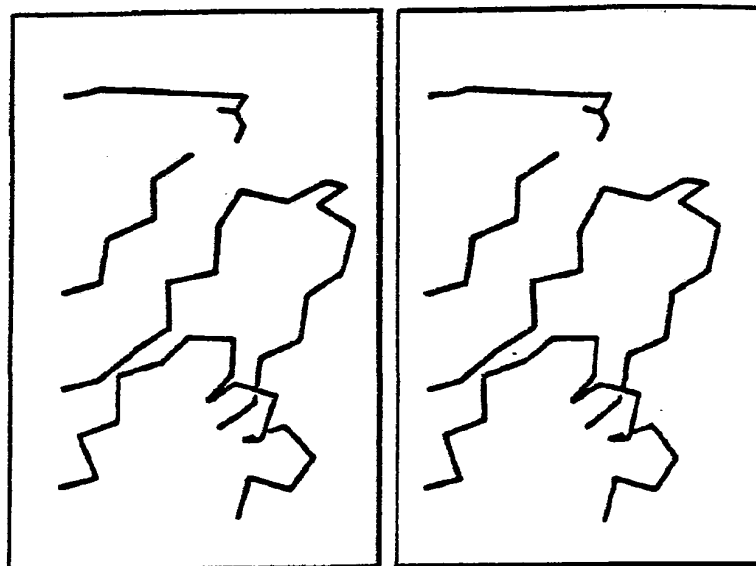


FIG. 20B

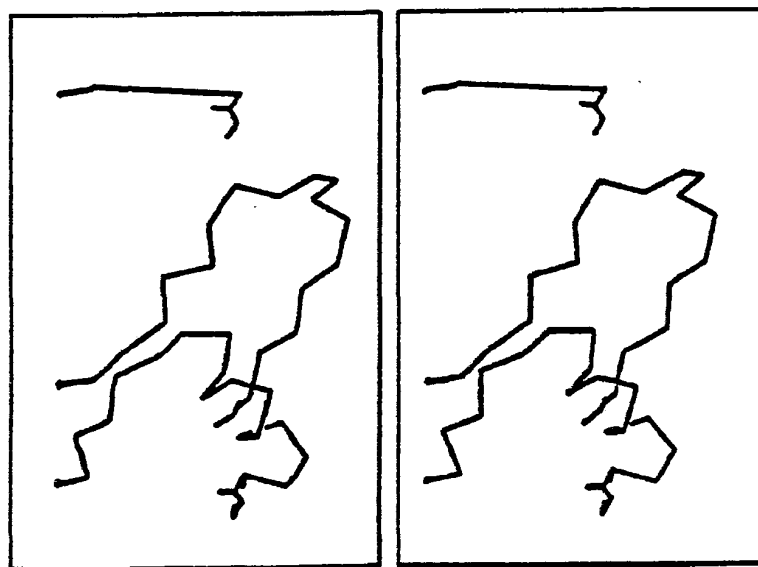


FIG. 20C

SUBSTITUTE SHEET

20/34

Glu	Val	Gln	Leu	Val	Glu	Ser	Gly	Gly	Asp	Leu	Val	Lys	Pro	Gly	Gly	Ser	Leu	Lys	Leu	5	10	15	20
GAG	GTG	CAC	CTG	GTG	GAG	TCT	GGG	GGA	GAC	TTA	GTG	AAG	CCT	GGA	GGG	TCC	CTG	AAA	CTC				
Ser	Cys	Ala	Ala	Ser	Gly	Phe	Thr	Phe	Ile	Ser	Tyr	Gly	Met	Ser	Trp	Val	Arg	Gln	Thr	25	30	35	40
TCC	TGT	GCA	GCC	TCT	GGA	TTC	ACT	TTC	ATT	AGC	TAT	GGC	ATG	TCT	TGG	GTT	CGC	CAG	ACT				
Pro	Asp	Lys	Arg	Leu	Glu	Trp	Val	Ala	Thr	Ile	Ser	Ser	Gly	Ser	Thr	Tyr	Thr	Tyr	Tyr	45	50	55	60
CCA	GAC	AAG	AGG	CTG	GAG	TGG	GTC	GCA	ACC	ATT	AGT	AGT	GGT	AGT	ACT	TAC	ACC	TAC	TAT				
Pro	Asp	Ser	Val	Lys	Gly	Arg	Phe	Thr	Ile	Ser	Arg	Asp	Asn	Ala	Lys	Asn	Thr	Leu	Tyr	65	70	75	80
CCA	GAC	AGT	GTG	AAG	GGG	CGA	TTC	ACC	ATC	TCC	AGA	GAC	AAT	GCC	AAG	AAC	ACC	CTG	TAC				
Leu	Gln	Met	Ser	Gly	Leu	Lys	Ser	Glu	Asp	Thr	Ala	Met	Tyr	Tyr	Cys	Ala	Arg	Arg	Ile	85	90	95	100
CTG	CAA	ATG	AGC	GGT	CTG	AAG	TCT	GAG	GAC	ACA	GCC	ATG	TAT	TAC	TGT	GCA	AGA	CGG	ATT				
Thr	Thr	Val	Val	Leu	Thr	Asp	Tyr	Tyr	Ala	Met	Asp	Tyr	Trp	Gly	Gln	Gly	Thr	Ser	Val	105	110	115	120
AGT	ACG	GTA	GTA	CTT	ACG	GAT	TAC	TAT	GCT	ATG	GAC	TAC	TGG	GGT	CAA	GGA	ACC	TCA	GTC				
Thr	Val	Ser	Ser	Ala	Lys	Thr	Thr	Pro	Ser	Val	Tyr	Pro	Leu	Ala	Pro	Gly	Ser	Ala	Ala	125	130	135	140
ACC	GTC	TCC	TCA	GCC	AAA	ACG	ACA	CCC	CCA	TCT	GTC	TAT	CCA	CTG	GCC	CCT	GGA	TCT	GCT				
Ala	Gln	Thr	Asn	Ser	Met	Val	Thr	Leu	Gly	Cys	Leu	Val	Lys	Gly	Tyr	Phe	Pro	Glu	Pro	145	150	155	160
GCC	CAA	ACT	AAC	TCG	ATG	GTG	ACC	CTG	GGA	TGC	CTG	CTC	AAG	GGC	TAT	TTC	CCT	GAG	CCA				
Val	Thr	Val	Thr	Trp	Asn	Ser	Gly	Ser	Leu	Ser	Ser	Gly	Val	His	Thr	Phe	Pro	Ala	Val	165	170	175	180
GTG	ACA	GTG	ACC	TGG	AAC	TCT	GGA	TCC	CTG	TCC	AGC	GGT	GTG	CAC	ACC	TTC	CCA	GCT	GTC				
Leu	Gln	Ser	Asp	Leu	Tyr	Thr	Leu	Ser	Ser	Ser	Val	Thr	Val	Pro	Ser	Ser	Thr	Trp	Pro	185	190	195	200
CTG	CAG	TCT	GAC	CTC	TAC	ACT	CTG	AGC	AGC	TCA	GTG	ACT	GTG	CCC	TCC	AGC	ACC	TGG	CCC				
Ser	Glu	Thr	Val	Thr	Cys	Asn	Val	Ala	His	Pro	Ala	Ser	Ser	Thr	Lys	Val	Asp	Lys	Lys	205	210	215	220
AGC	GAG	ACC	GTC	ACC	TGC	AAC	GTT	GCC	CAC	CCG	GCG	AGC	ACC	ACC	AAG	GTG	GAC	AAG	AAA				
Ile	Val	Pro	Arg	Asp	Cys	Gly	Cys	Lys	Pro	Cys	Ile	Cys	Thr	Val	Pro	Glu	Val	Ser	Ser	225	230	235	240
ATT	GTG	CCC	AGG	GAT	TGT	GGT	TGT	AAG	CCT	TGC	ATA	TGT	ACA	GTG	CCA	GAA	GTA	TCA	TCT				
Val	Phe	Ile	Phe	Pro	Pro	Lys	Pro	Lys	Asp	Val	Leu	Thr	Ile	Thr	Leu	Thr	Pro	Lys	Val	245	250	255	260
GTC	TTC	ATC	TTC	CCC	CCA	AAG	CCC	AAG	GAT	GTG	CTC	ACC	ATT	ACT	CTG	ACT	CCT	AAG	GTC				
Thr	Cys	Val	Val	Val	Asp	Ile	Ser	Lys	Asp	Asp	Pro	Glu	Val	Gln	Phe	Ser	Trp	Phe	Val	265	270	275	280
ACG	TGT	GTT	GTG	GTA	GAC	ATC	AGC	AAG	GAT	GAT	CCC	GAG	GTG	CAG	TTC	AGC	TGG	TTT	GTA				
Asp	Asp	Val	Glu	Val	His	Thr	Ala	Gln	Thr	Gln	Pro	Arg	Glu	Gln	Gln	Phe	Asn	Ser	Thr	285	290	295	300
GAT	GAT	GTG	GAG	CTG	CAC	ACA	GCT	CAG	ACG	CAA	CCC	CGG	GAG	GAG	CAG	TTC	AAC	AGC	ACT				
Ser	Arg	Ser	Val	Ser	Glu	Leu	Pro	Ile	Met	His	Gln	Asp	Trp	Leu	Asn	Gly	Lys	Glu	Phe	305	310	315	320
TCC	CGC	TCA	GTC	AGT	GAA	CTT	CCC	ATC	ATG	CAC	CAG	GAC	TGG	CTC	AAT	GGC	AAG	GAG	TTC				
Lys	Cys	Arg	Val	Asn	Ser	Ala	Ala	Phe	Pro	Ala	Pro	Ile	Glu	Lys	Thr	Ile	Ser	Lys	Thr	325	330	335	340
AAA	TGC	AGG	GTC	AAC	AGT	GCA	GCT	TTC	CCT	GCC	CCC	ATC	GAG	AAA	ACC	ATC	TCC	AAA	ACC				
Lys	Gly	Arg	Pro	Lys	Ala	Pro	Gln	Val	Tyr	Thr	Ile	Pro	Pro	Pro	Lys	Glu	Gln	Met	Ala	345	350	355	360
AAA	GGC	AGA	CCG	AAG	GCT	CCA	CAG	GTG	TAC	ACC	ATT	CCA	CCT	CCC	AAG	GAG	CAG	ATG	GCC				
Lys	Asp	Lys	Val	Ser	Leu	Thr	Cys	Met	Ile	Thr	Asp	Phe	Phe	Pro	Glu	Asp	Ile	Thr	Val	365	370	375	380
AAG	GAT	AAA	GTC	AGT	CTG	ACC	TGC	ATG	ATA	ACA	GAC	TTC	TTC	CCT	GAA	GAC	ATT	ACT	GTG				
Glu	Trp	Gln	Trp	Asn	Gly	Gln	Pro	Ala	Glu	Asn	Tyr	Lys	Asn	Thr	Gln	Arg	Ile	Met	Asn	385	390	395	400
GAG	TGG	CAG	TGG	AAT	GGG	CAG	CCA	GCG	GAG	AAC	TAC	AAG	AAC	ACT	CAG	CGC	ATC	ATG	AAC				
Thr	Asn	Gly	Ser	Tyr	Phe	Val	Tyr	Ser	Lys	Leu	Asn	Val	Gln	Lys	Ser	Asn	Trp	Glu	Ala	405	410	415	420
ACG	AAT	GGC	TCT	TAC	TTC	GTC	TAC	AGC	AAG	CTC	AAT	GTG	CAG	AAG	AGC	AAC	TGG	GAG	GCA				
Gly	Asn	Thr	Phe	Thr	Cys	Ser	Val	Leu	His	Glu	Gly	Leu	His	Asn	His	His	Thr	Glu	Lys	425	430	435	440
GGA	AAT	ACT	TTC	ACC	TGC	TCT	GTG	TTA	CAT	GAG	GGC	CTG	CAC	AAC	CAC	CAT	ACT	GAG	AAG				
Ser	Leu	Ser	His	Ser	Pro	Gly	Lys	***												445			
AGC	CTC	TCC	CAC	TCT	CCT	GGT	AAA	TGA															

FIG. 21

SUBSTITUTE SHEET

21/34

Glu	Asn	Val	Leu	Thr	5	Gln	Ser	Pro	Ala	Ile	10	Met	Ser	Ala	Ser	15	Gly	Glu	Lys	VAL	Thr	20
GAA	AAT	GTG	CTC	ACC		CAG	TCT	CCA	GCA	ATC		ATG	TCT	GCA	TCT	CCA	GGG	GAA	AAG	GTC	ACC	
					25						30					35						40
Met	Thr	Cys	Arg	Ala	Ser	Ser	Ser	Val	Ser	Ser	Ser	Tyr	Leu	His	Trp	Phe	Gln	Gln	Lys			
ATG	ACC	TGC	AGG	GCC	AGC	TCA	AGT	GTA	AGT	TCC	AGT	TAC	TTG	CAC	TGG	TTC	CAG	CAG	AAG			
					45						50					55						60
Ser	Gly	Ala	Ser	Pro	Lys	Leu	Trp	Ile	Tyr	Ser	Thr	Ser	Asn	Leu	Ala	Ser	Gly	Val	Pro			
TCA	GGT	GCC	TCC	CCC	AAA	CTC	TGG	ATT	TAT	AGC	ACA	TCC	AAC	TTG	GCT	TCT	GGA	GTC	CCT			
					65						70					75						80
Ala	Arg	Phe	Ser	Gly	Ser	Gly	Ser	Gly	Thr	Ser	Tyr	Ser	Leu	Thr	Ile	Ser	Ser	Val	Glu			
GCT	CGC	TTC	AGT	GGC	AGT	GGG	TCT	GGG	ACC	TCT	TAC	TCT	CTC	ACA	ATC	AGC	AGT	GTG	GAG			
					85						90					95						100
Ala	Glu	Asp	Ala	Ala	Thr	Tyr	Tyr	Cys	Gln	Gln	Tyr	Ser	Gly	Tyr	Pro	Leu	Thr	Phe	Gly			
GCT	GAA	GAT	GCT	GCC	ACT	TAT	TAC	TGC	CAG	CAG	TAC	AGT	GGT	TAC	CCA	CTC	ACG	TTC	GGT			
					105						110					115						120
Ala	Gly	Thr	Lys	Leu	Glu	Leu	Lys	Arg	Ala	Asp	Ala	Ala	Pro	Thr	Val	Ser	Ile	Phe	Pro			
GCT	GGG	ACC	AAG	CTG	GAG	CTG	AAA	CGG	GCT	GAT	GCT	GCA	CCA	ACT	GTA	TCC	ATC	TTC	CCA			
					125						130					135						140
Pro	Ser	Ser	Glu	Gln	Leu	Thr	Ser	Gly	Gly	Ala	Ser	Val	Val	Cys	Phe	Leu	Asn	Asn	Phe			
CCA	TCC	AGT	GAG	CAG	TTA	ACA	TCT	GGA	GGT	GCC	TCA	GTC	GTG	TGC	TTC	TTG	AAC	AAC	TTC			
					145						150					155						160
Tyr	Pro	Lys	Asp	Ile	Asn	Val	Lys	Trp	Lys	Ile	Asp	Gly	Ser	Glu	Arg	Gln	Asp	Gly	Val			
TAC	CCC	AAA	GAC	ATC	AAT	GTC	AAG	TGG	AAG	ATT	GAT	GGC	AGT	GAA	CGA	CAA	AAT	GGC	GTC			
					165						170					175						180
Leu	Asn	Ser	Trp	Thr	Asp	Gln	Asp	Ser	Lys	Asp	Ser	Thr	Tyr	Ser	Met	Ser	Ser	Thr	Leu			
CTG	AAC	AGT	TGG	ACT	GAT	CAG	GAC	AGC	AAA	GAC	AGC	ACC	TAC	AGC	ATG	AGC	AGC	ACC	CTC			
					185						190					195						200
Thr	Leu	Thr	Lys	Asp	Glu	Tyr	Glu	Arg	His	Asn	Ser	Tyr	Thr	Cys	Glu	Ala	Thr	His	Lys			
ATG	TTG	ACC	AAG	GAC	GAG	TAT	GAA	CGA	CAT	AAC	AGC	TAT	ACC	TGT	GAG	GCC	ACT	CAC	AAG			
					205						210					215						
Thr	Ser	Thr	Ser	Pro	Ile	Val	Lys	Ser	Phe	Asn	Arg	Asn	Glu	Cys	***							
ACA	TCA	ACT	TCA	CCC	ATT	GTC	AAG	AGC	TTC	AAC	AGG	AAT	GAG	TGT	TAG							

FIG. 22

SUBSTITUTE SHEET

22/34

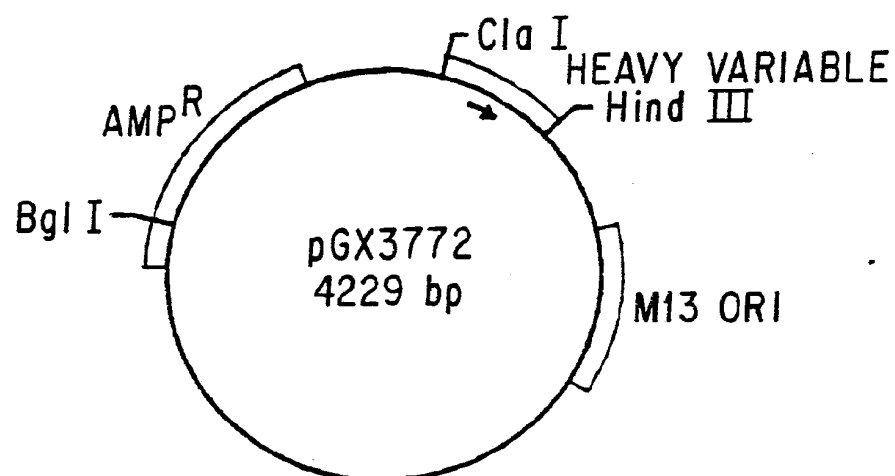


FIG. 23A

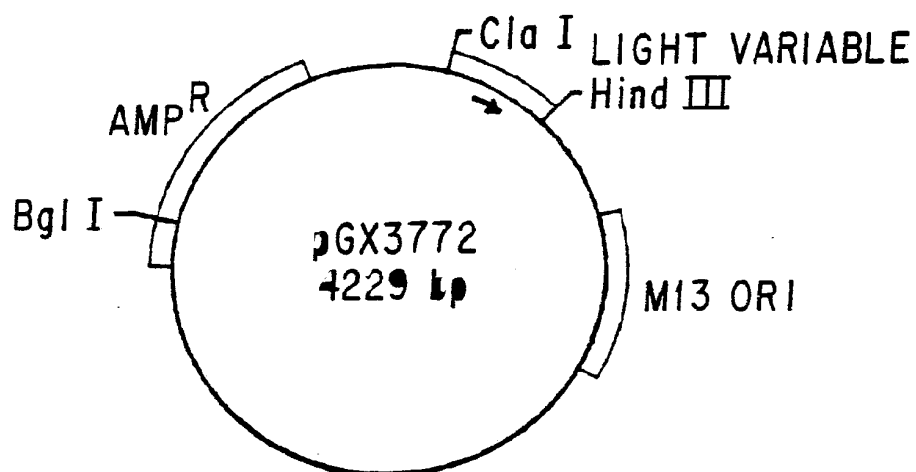


FIG. 23B

SUBSTITUTE SHEET

23/34

Met	Glu	Asn	Val	Leu	Thr	Gln	Ser	Pro	Ala	Ile	Met	Ser	Ala	Ser	Pro	Gly	Glu	Lys	Val	5	10	15	20
ATG	GAA	AAT	GTG	CTC	ACC	CAG	TCT	CCA	GCA	ATC	ATG	TCT	GCA	TCT	CCA	GGG	GAA	AAG	GTC				
Thr	Met	Thr	Cys	Arg	Ala	Ser	Ser	Ser	Val	Ser	Ser	Ser	Tyr	Leu	His	Trp	Phe	Gln	Gln	25	30	35	40
ACC	ATG	ACC	TGC	AGG	GCC	AGC	TCA	AGT	GTA	AGT	TCC	AGT	TAC	TTG	CAC	TGG	TTC	CAG	CAG				
Lys	Ser	Ile	Ala	Lys	Ala	Phe	Lys	Asn	Gly	Gly	Asp	Leu	Val	Lys	Pro	Gly	Gly	Ser	Leu	45	50	55	60
AAG	TCA	ATC	GCG	AAA	GCG	TTC	AAA	AAC	GGG	GGG	GAC	TTA	GTG	AAG	CCT	GGA	GGG	TCC	CTG				
Lys	Leu	Ser	Cys	Ala	Ala	Ser	Gly	Phe	Thr	Phe	Ile	Ser	Tyr	Gly	Met	Ser	Trp	Val	Arg	65	70	75	80
AAA	CTC	TCC	TGT	GCA	GCC	TCT	GGA	TTC	ACT	TTC	ATT	AGC	TAT	GGC	ATG	TCT	TGG	GTT	CGC				
Gln	Thr	Pro	Asp	Lys	Arg	Leu	Glu	Trp	Val	Ala	Thr	Ile	Ser	Ser	Gly	Ser	Thr	Tyr	Thr	85	90	95	100
CAG	ACT	CCA	GAC	AAG	AGG	CTG	GAG	TGG	GTC	GCA	ACC	ATT	AGT	GGT	GGT	AGT	ACT	TAC	ACC				
Tyr	Tyr	Pro	Asp	Ser	Val	Lys	Gly	Arg	Phe	Thr	Ile	Ser	Arg	Asp	Asn	Ala	Lys	Asn	Thr	105	110	115	120
TAC	TAT	CCA	GAC	AGT	GTG	AAG	GGG	CGA	TTC	ACC	ATC	TCC	AGA	GAC	AAT	GCC	AAG	AAC	ACC				
Leu	Tyr	Leu	Gln	Met	Ser	Gly	Leu	Lys	Ser	Glu	Asp	Thr	Ala	Met	Tyr	Tyr	Cys	Ala	Arg	125	130	135	140
CTG	TAC	CTG	CAA	ATG	AGC	GGT	CTG	AAG	TCT	GAG	GAC	ACA	GCC	ATG	TAT	TAC	TGT	GCA	AGA				
Arg	Ile	Thr	Thr	Val	Val	Leu	Thr	Asp	Tyr	Tyr	Ala	Met	Asp	Tyr	Trp	Gly	Pro	Gly	Ser	145	150	155	160
CGG	ATT	ACT	ACG	GTA	GTA	CTT	ACG	GAT	TAC	TAT	GCT	ATG	GAC	TAC	TGG	GGT	CCG	GGT	TCT				
Pro	Lys	Leu	Trp	Ile	Tyr	Ser	Thr	Ser	Asn	Leu	Ala	Ser	Gly	Val	Pro	Ala	Arg	Phe	Ser	165	170	175	180
CCC	AAA	CTC	TGG	ATT	TAT	AGC	ACA	TCC	AAC	TTG	GCT	TCT	GGG	GTC	CCT	GCT	CGC	TTC	AGT				
Gly	Ser	Gly	Ser	Gly	Thr	Ser	Tyr	Ser	Leu	Thr	Ile	Ser	Ser	Val	Glu	Ala	Glu	Asp	Ala	185	190	195	200
GGC	AGT	GGG	TCT	GGG	ACC	TCT	TAC	TCT	CTC	ACA	ATC	AGC	AGT	GTG	GAG	GCT	GAA	GAT	GCT				
Ala	Thr	Tyr	Tyr	Cys	Gln	Gln	Tyr	Ser	Gly	Tyr	Pro	Leu	Thr	Phe	Gly	Ala	Gly	Thr	Lys	205	210	215	220
GCC	ACT	TAT	TAC	TGC	CAG	CAG	TAC	AGT	GGT	TAC	CCA	CTC	ACG	TTC	GGT	GCT	GGG	ACC	AAG				
Leu	Glu	Leu	Lys	Arg	***															225			
CTG	GAG	CTG	AAA	CGG	TAA																		

FIG. 24

SUBSTITUTE SHEET

24/34

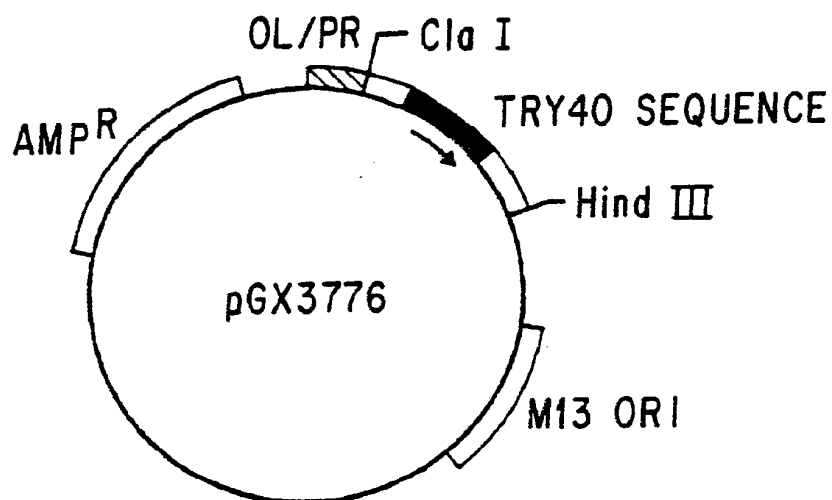


FIG. 25

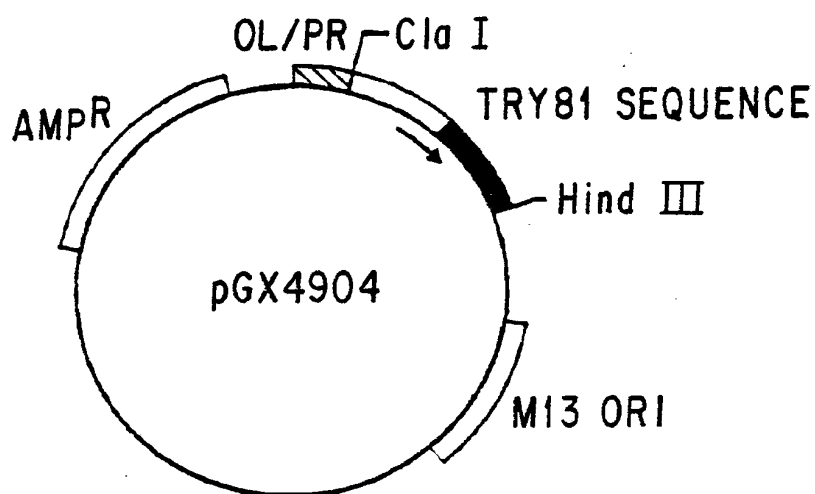


FIG. 27

SUBSTITUTE SHEET

25/34

TRY61
sca try 61

				5					10					15				20	
Met	Glu	Asn	Val	Leu	Thr	Gln	Ser	Pro	Ala	Ile	Met	Ser	Ala	Ser	Pro	Gly	Glu	Lys	Val
ATG	GAA	AAT	GTG	CTC	ACC	CAG	TCT	CGA	GCA	ATC	ATG	TCT	GCA	TCT	CCA	GGG	GAA	AAG	GTC
				25					30					35				40	
Thr	Met	Thr	Cys	Arg	Ala	Ser	Ser	Ser	Val	Ser	Ser	Ser	Tyr	Leu	His	Trp	Phe	Gln	Gln
ACC	ATG	ACC	TGC	AGG	GCC	AGC	TCA	AGT	GTA	AGT	TCC	AGT	TAC	TTG	CAC	TGG	TTC	CAG	CAG
				45					50					55				60	
Lys	Ser	Gly	Ala	Ser	Pro	Lys	Leu	Trp	Ile	Tyr	Ser	Thr	Ser	Asn	Leu	Ala	Ser	Gly	Val
AAG	TCA	GGT	GCC	TCC	CCC	AAA	CTC	TGG	ATT	TAT	AGC	ACA	TCC	AAC	TTC	GCT	TCT	GGA	GTC
				65					70					75				80	
Pro	Ala	Arg	Phe	Ser	Gly	Ser	Gly	Ser	Gly	Thr	Ser	Tyr	Ser	Leu	Thr	Ile	Ser	Ser	Val
CCT	GCT	CGC	TTC	AGT	GGC	AGT	GGG	TCT	GGG	ACC	TCT	TAC	TCT	CTC	ACA	ATC	AGC	AGT	GTG
				85					90					95				100	
Glu	Ala	Glu	Asp	Ala	Ala	Thr	Tyr	Tyr	Cys	Gln	Gln	Tyr	Ser	Gly	Tyr	Pro	Leu	Thr	Phe
GAG	GCT	GAA	GAT	GCT	GCC	ACT	TAT	TAC	TGC	CAG	TAC	AGT	GGT	GGT	TAC	CCA	CTC	ACG	TTC
				105					110					115				120	
Gly	Ala	Gly	Thr	Lys	Val	Arg	Gly	Ser	Pro	Ala	Ile	Asn	Val	Ala	Val	His	Val	Phe	Ser
GGT	GCT	GGG	ACC	AAG	GTT	CGT	GGT	TCT	CCG	GCA	ATC	AAC	GTA	GCT	GTA	CAC	GTA	TTC	TCT
				125					130					135				140	
Gly	Gly	Asp	Leu	Val	Lys	Pro	Gly	Gly	Ser	Leu	Lys	Leu	Ser	Cys	Ala	Ala	Ser	Gly	Phe
GGG	GGA	GAC	TTA	GTG	AAG	CCT	GGA	GGG	TCC	CTG	AAA	CTC	TCC	TGT	GCA	GCC	TCT	GGA	TTC
				145					150					155				160	
Thr	Phe	Ile	Ser	Tyr	Gly	Met	Ser	Trp	Val	Arg	Gln	Thr	Pro	Asp	Lys	Arg	Leu	Glu	Trp
ACT	TTC	ATT	AGC	TAT	GGC	ATG	TCT	TGG	GTT	CGC	CAG	ACT	CCA	GAC	AAG	AGG	CTG	GAG	TGG
				165					170					175				180	
Val	Ala	Thr	Ile	Ser	Ser	Gly	Ser	Thr	Tyr	Thr	Tyr	Tyr	Pro	Asp	Ser	Val	Lys	Gly	Arg
GTC	GCA	ACC	ATT	AGT	AGT	GGT	AGT	ACT	TAC	ACC	TAC	TAT	CCA	GAC	AGT	GTG	AAG	GGG	CGA
				185					190					195				200	
Phe	Thr	Ile	Ser	Arg	Asp	Asn	Ala	Lys	Asn	Thr	Leu	Tyr	Leu	Gln	Met	Ser	Gly	Leu	Lys
TTC	ACC	ATC	TCC	AGA	GAC	AAT	GCC	AAG	AAC	ACC	CTG	TAC	CTG	CAA	ATG	AGC	GGT	CTG	AAG
				205					210					215				220	
Ser	Glu	Asp	Thr	Ala	Met	Tyr	Tyr	Cys	Ala	Arg	Arg	Ile	Thr	Thr	Val	Val	Leu	Thr	Asp
TCT	GAG	GAC	ACA	GCC	ATG	TAT	TAC	TGT	GCA	AGA	CGG	ATT	ACT	ACG	GTA	GTA	CTT	ACG	GAT
				225					230					235					
Tyr	Tyr	Ala	Met	Asp	Tyr	Trp	Gly	Thr	Gly	Thr	Ser	Val	Thr	Val	Ser	***			
TAC	TAT	GCT	ATG	GAC	TAC	TGG	GGT	CAA	GGA	ACC	TCA	GTC	ACC	GTC	TCC	TAA			

FIG. 26

SUBSTITUTE SHEET

26/34

				5					10					15					20
Met	Glu	Asn	Val	Leu	Thr	Gln	Ser	Pro	Ala	Ile	Met	Ser	Ala	Ser	Pro	Gly	Glu	Lys	Val
ATG	GAA	AAT	GTG	CTC	ACC	CAG	TCT	CGA	GCA	ATC	ATG	TCT	GCA	TCT	CCA	GGG	GAA	AAG	GTC
				25					30					35					40
Thr	Met	Thr	Cys	Arg	Ala	Ser	Ser	Ser	Val	Ser	Ser	Ser	Tyr	Leu	His	Trp	Phe	Gln	Gln
ACC	ATG	ACC	TGC	AGG	GCC	AGC	TCA	AGT	GTA	AGT	TCC	AGT	TAC	TTG	CAC	TGG	TTC	CAG	CAG
				45					50					55					60
Lys	Ser	Gly	Ala	Ser	Pro	Lys	Leu	Trp	Ile	Tyr	Ser	Thr	Ser	Asn	Leu	Ala	Ser	Gly	Val
AAG	TCA	GGT	GCC	TCC	CCC	AAA	CTC	TGG	ATT	TAT	AGC	ACA	TCC	AAC	TTC	GCT	TCT	GGA	GTC
				65					70					75					80
Pro	Ala	Arg	Phe	Ser	Gly	Ser	Gly	Ser	Gly	Thr	Ser	Tyr	Ser	Leu	Thr	Ile	Ser	Ser	Val
CCT	GCT	CGC	TTC	AGT	GGC	AGT	GGG	TCT	GGG	ACC	TCT	TAC	TCT	CTC	ACA	ATC	AGC	AGT	GTG
				85					90					95					100
Glu	Ala	Glu	Asp	Ala	Ala	Thr	Tyr	Tyr	Cys	Gln	Gln	Tyr	Ser	Gly	Tyr	Pro	Leu	Thr	Phe
GAG	GCT	GAA	GAT	GCC	ACT	TAT	TAC	TGC	CAG	TAC	AGT	GGT	GGT	TAC	CCA	CTC	ACG	TTC	TTC
				105					110					115					120
Gly	Ala	Gly	Thr	Lys	Leu	Lys	Glu	Ser	Gly	Ser	Val	Ser	Ser	Glu	Gln	Leu	Ala	Gln	Phe
GGT	GCT	GGG	ACC	AAG	CTG	AAA	GAA	TCT	GGT	TCT	GTT	TCT	TCT	GAA	CAG	CTG	GCT	CAG	TTT
				125					130					135					140
Arg	Ser	Leu	Asp	Val	Gln	Leu	Val	Glu	Ser	Gly	Gly	Asp	Leu	Val	Lys	Pro	Gly	Gly	Ser
CGT	TCT	CTG	GAT	GTG	CAG	CTG	GTG	GAG	TCT	GGG	GGG	GAC	TTA	GTG	AAG	CCT	GGA	GGG	TCC
				145					150					155					160
Leu	Lys	Leu	Ser	Cys	Ala	Ala	Ser	Gly	Phe	Thr	Phe	Ile	Ser	Tyr	Gly	Met	Ser	Trp	Val
CTG	AAA	CTC	TCC	TGT	GCA	GCC	TCT	GGA	TTC	ACT	TTC	ATT	AGC	TAT	GGC	ATG	TCT	TGG	GTT
				165					170					175					180
Arg	Gln	Thr	Pro	Asp	Lys	Arg	Leu	Glu	Trp	Val	Ala	Thr	Ile	Ser	Ser	Gly	Ser	Thr	Tyr
GGC	CGA	ACT	CCA	GAC	AAG	AGG	CTG	GAG	TGG	GTC	GCA	ACC	ATT	AGT	AGT	GGT	AGT	ACT	TAC
				185					190					195					200
Thr	Tyr	Tyr	Pro	Asp	Ser	Val	Lys	Gly	Arg	Phe	Thr	Ile	Ser	Arg	Asp	Asn	Ala	Lys	Asn
ACC	TAC	TAT	CCA	GAC	AGT	GTG	AAG	GGG	CGA	TTC	ACC	ATC	TCC	AGA	GAC	AAT	GCC	AAG	AAC
				205					210					215					220
Thr	Leu	Tyr	Leu	Gln	Met	Ser	Gly	Leu	Lys	Ser	Glu	Asp	Thr	Ala	Met	Tyr	Tyr	Cys	Ala
ACC	CTG	TAC	CTG	CAA	ATG	AGC	GGT	CTG	AAG	TCT	GAG	GAC	ACA	GCC	ATG	TAT	TAC	TGT	GCA
				225					230					235					240
Arg	Arg	Ile	Thr	Thr	Val	Val	Leu	Thr	Asp	Tyr	Tyr	Ala	Met	Asp	Tyr	Trp	Gly	Gln	Gly
AGA	CGG	ATT	ACT	ACG	GTA	GTA	CTT	ACG	GAT	TAC	TAT	GCT	ATG	GAC	TAC	TGG	GGT	CAA	GGA
				245															
Thr	Ser	Val	Thr	Val	Ser	***													
ACC	TCA	GTC	ACC	GTC	TCC	TAA													

FIG. 28

SUBSTITUTE SHEET

27/34

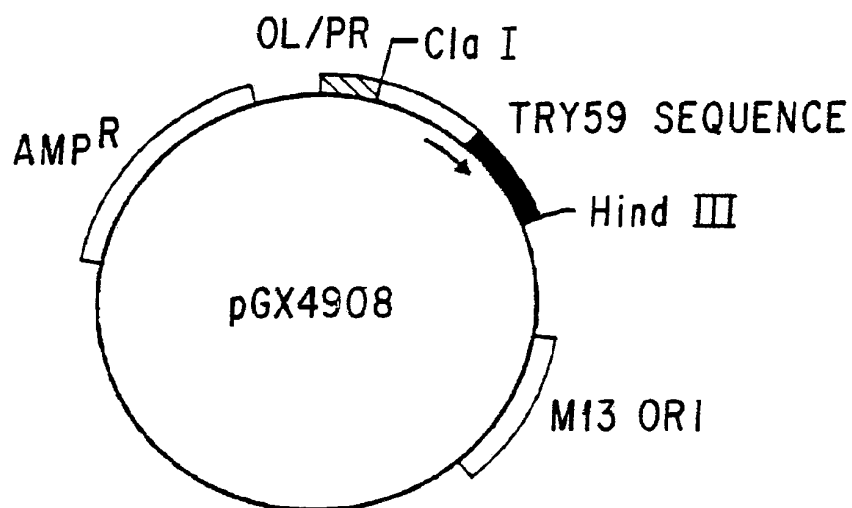


FIG. 29

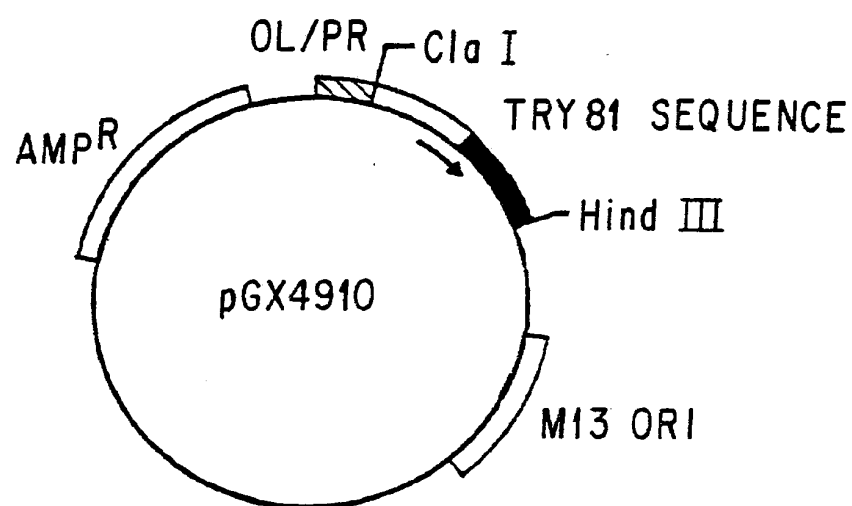


FIG. 34

SUBSTITUTE SHEET

28/34

FIG. 30A

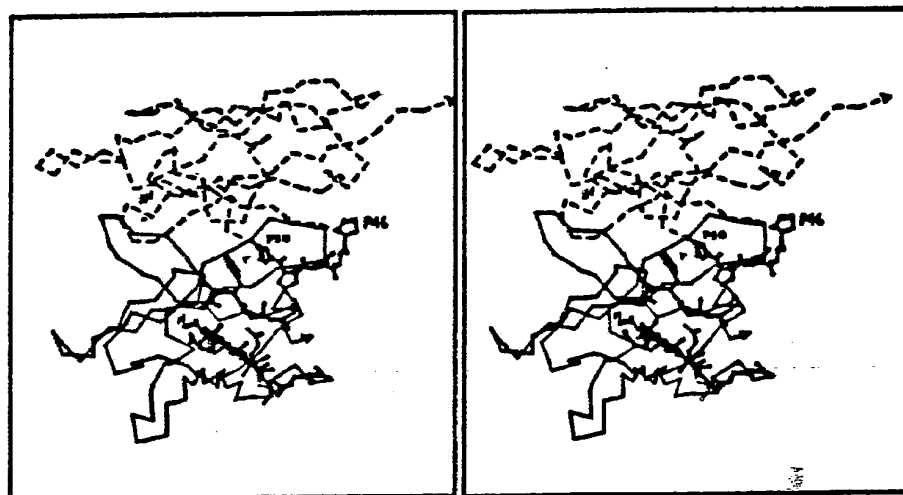
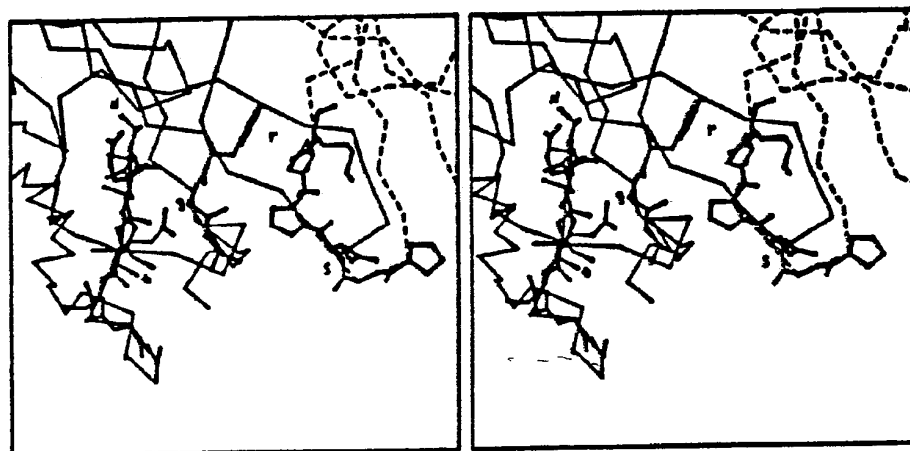


FIG. 30B



SUBSTITUTE SHEET

29/34

FIG. 30C

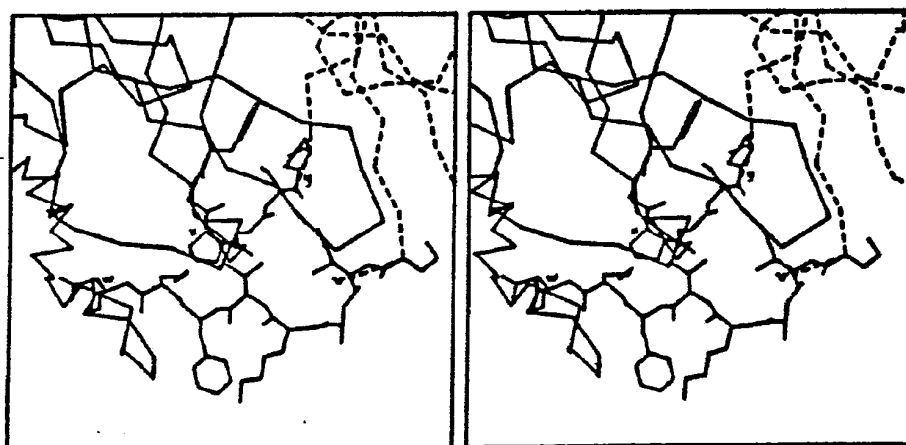
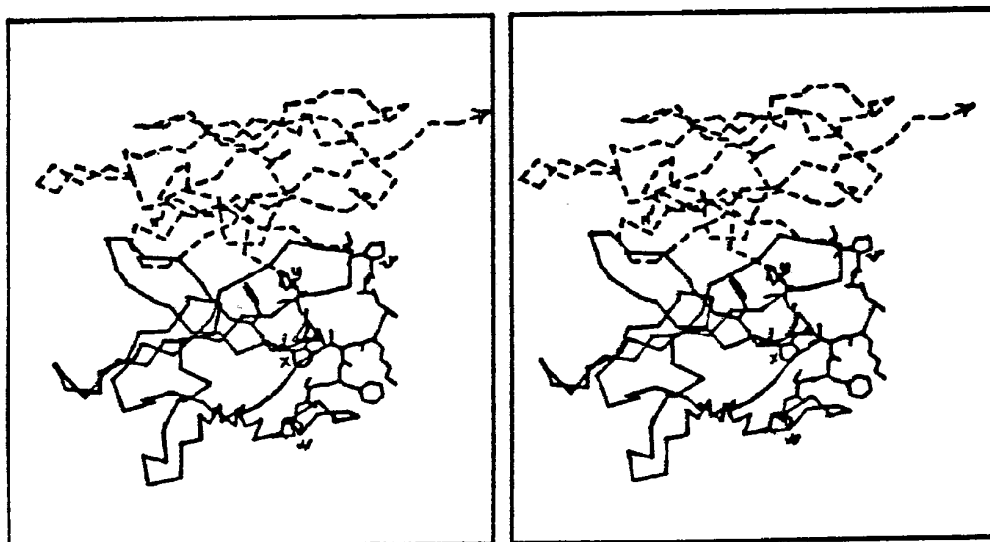


FIG. 30D



SUBSTITUTE SHEET

30/34

FIG. 3IA

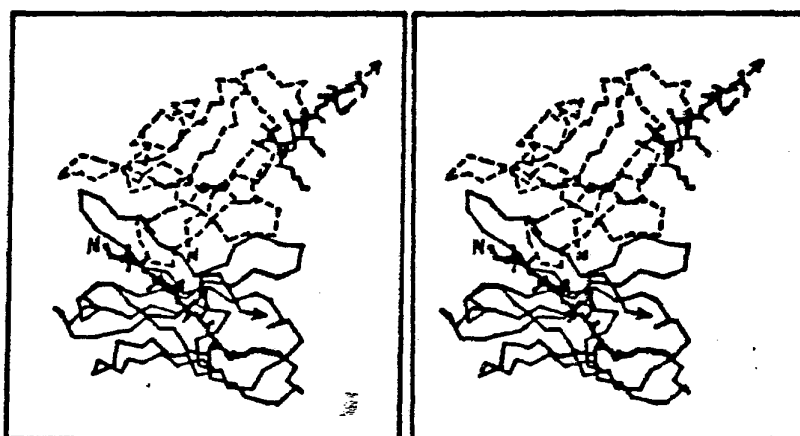
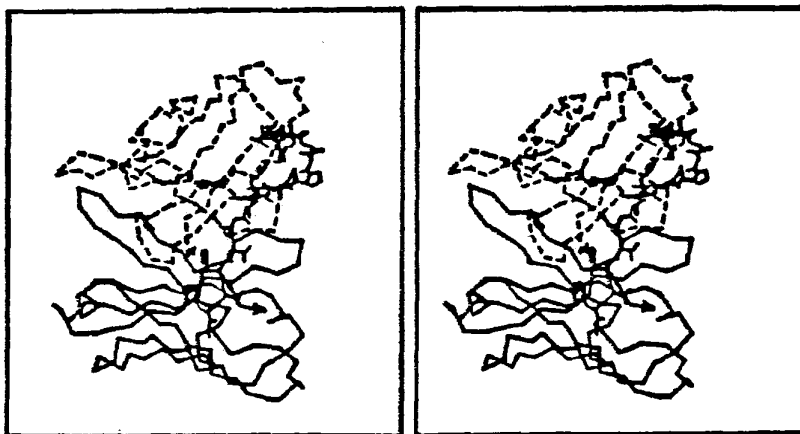


FIG. 3IB



SUBSTITUTE SHEET

31/34

FIG. 32A



FIG. 32B



SUBSTITUTE SHEET

32/34

TRY048

SCA TRY104B, ALL OF VL AND VH

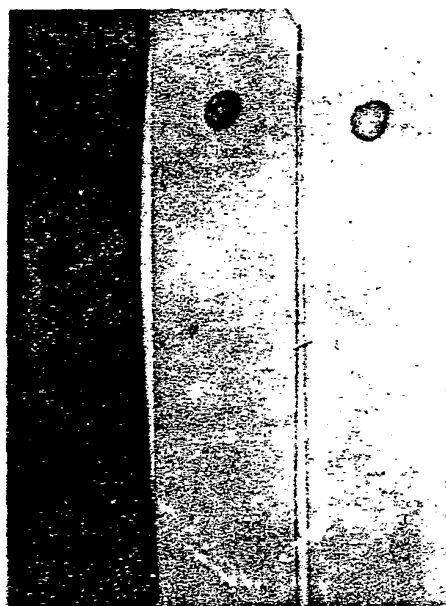
Met	Glu	Asn	Val	Leu	Thr	Gln	Ser	Pro	Ala	Ile	Met	Ser	Ala	Ser	Pro	Gly	Glu	Lys	Val	5	10	15	20
ATG	GAA	AAT	GTG	CTC	ACC	CAG	TCT	CCA	GCA	ATC	ATG	TCT	GCA	TCT	CCA	GGG	GAA	AAG	GTC				
Thr	Met	Thr	Cys	Arg	Ala	Ser	Ser	Ser	Val	Ser	Ser	Ser	Tyr	Leu	His	Trp	Phe	Gln	Gln	25	30	35	40
ACC	ATG	ACC	TGC	AGG	GCC	AGC	TCA	AGT	GTA	AGT	TCC	AGT	TCC	AGT	TAC	TGG	TTC	CAG	CAG				
Lys	Ser	Gly	Ala	Ser	Pro	Lys	Leu	Trp	Ile	Tyr	Ser	Thr	Ser	Asn	Leu	Ala	Ser	Gly	Val	45	50	55	60
AAG	TCA	GGT	GCC	TCC	CCC	AAA	CTC	TGG	ATT	TAT	AGC	ACA	TCC	AAC	TTG	GCT	TCT	GGA	GTC				
Pro	Ala	Arg	Phe	Ser	Gly	Ser	Gly	Ser	Gly	Thr	Ser	Tyr	Ser	Leu	Thr	Ile	Ser	Ser	Val	65	70	75	80
CCT	GCT	CGC	TTC	AGT	GGC	AGT	GGG	TCT	GGG	ACC	TCT	TAC	TCT	CTC	ACA	ATC	AGC	AGT	GTC				
Glu	Ala	Glu	Asp	Ala	Ala	Thr	Tyr	Tyr	Cys	Gln	Gln	Tyr	Ser	Gly	Tyr	Pro	Leu	Thr	Phe	85	90	95	100
GAG	GCT	GAA	GAT	GCT	GCC	ACT	TAT	TAC	TGC	CAG	CAG	TAC	AGT	GGT	TAC	CCA	CTC	ACG	TTC				
Gly	Ala	Gly	Thr	Lys	Leu	Glu	Ala	Glu	Gly	Thr	Leu	Ser	Pro	Ala	Asp	Lys	Thr	Asn	Val	105	110	115	120
GGT	GCT	GGG	ACC	AAG	CTG	GAG	GCA	GAA	GGC	ACT	CTG	TCT	CCA	GCA	GAT	AAA	ACT	AAC	GTT				
Lys	Ala	Ala	Trp	Gly	Lys	Val	Met	Thr	Gln	Leu	Val	Glu	Ser	Gly	Gly	Asp	Leu	Val	Lys	125	130	135	140
AAA	GCA	GCA	TGG	GGC	AAA	GTT	ATG	ACT	CAG	CTG	GTG	GAG	TCT	GGG	GGG	GAC	TTA	GTG	AAG				
Pro	Gly	Gly	Ser	Leu	Lys	Leu	Ser	Cys	Ala	Ala	Ser	Gly	Phe	Thr	Phe	Ile	Ser	Tyr	Gly	145	150	155	160
CCT	GGA	GGG	TCC	CTG	AAA	CTC	TCC	TGT	GCA	GCA	TCT	GGA	TTC	ACT	TTC	ATT	AGC	TAT	GGC				
Met	Ser	Trp	Val	Arg	Gln	Thr	Pro	Asp	Lys	Arg	Leu	Glu	Trp	Val	Ala	Thr	Ile	Ser	Ser	165	170	175	180
ATG	TCT	TGG	GTT	CGC	CAG	ACT	CCA	GAC	AAG	AGG	CTG	GAG	TGG	GTC	GCA	ACC	ATT	AGT	AGT				
Gly	Ser	Thr	Tyr	Thr	Tyr	Tyr	Pro	Asp	Ser	Val	Lys	Gly	Arg	Phe	Thr	Ile	Ser	Arg	Asp	185	190	195	200
GGT	AGT	ACT	TAC	ACC	TAC	TAT	CCA	GAC	AGT	GTG	AAG	GGG	CGA	TTC	ACC	ATC	TCC	AGA	GAC				
Asn	Ala	Lys	Asn	Thr	Leu	Tyr	Leu	Gln	Met	Ser	Gly	Leu	Lys	Ser	Glu	Asp	Thr	Ala	Met	205	210	215	220
AAT	GCC	AAG	AAC	ACC	CTG	TAC	CTG	CAA	ATG	AGC	GGT	CTG	AAG	TCT	GAG	GAC	ACA	GCC	ATG				
Tyr	Tyr	Cys	Ala	Arg	Arg	Ile	Thr	Thr	Val	Val	Leu	Thr	Asp	Tyr	Tyr	Ala	Met	Asp	Tyr	225	230	235	240
TAT	TAC	TGT	GCA	AGA	CGG	ATT	ACT	ACG	GTA	GTA	CTT	ACG	GAT	TAC	TAT	GCT	ATG	GAC	TAC				
Trp	Gly	Gln	Gly	Thr	Ser	Val	Thr	Val	Ser	***										245	250		
TGG	GGT	CAA	GGA	ACC	TCA	GTC	ACC	GTC	TCC	TAA													

FIG. 33

SUBSTITUTE SHEET

33/34

FIG. 35



BGH

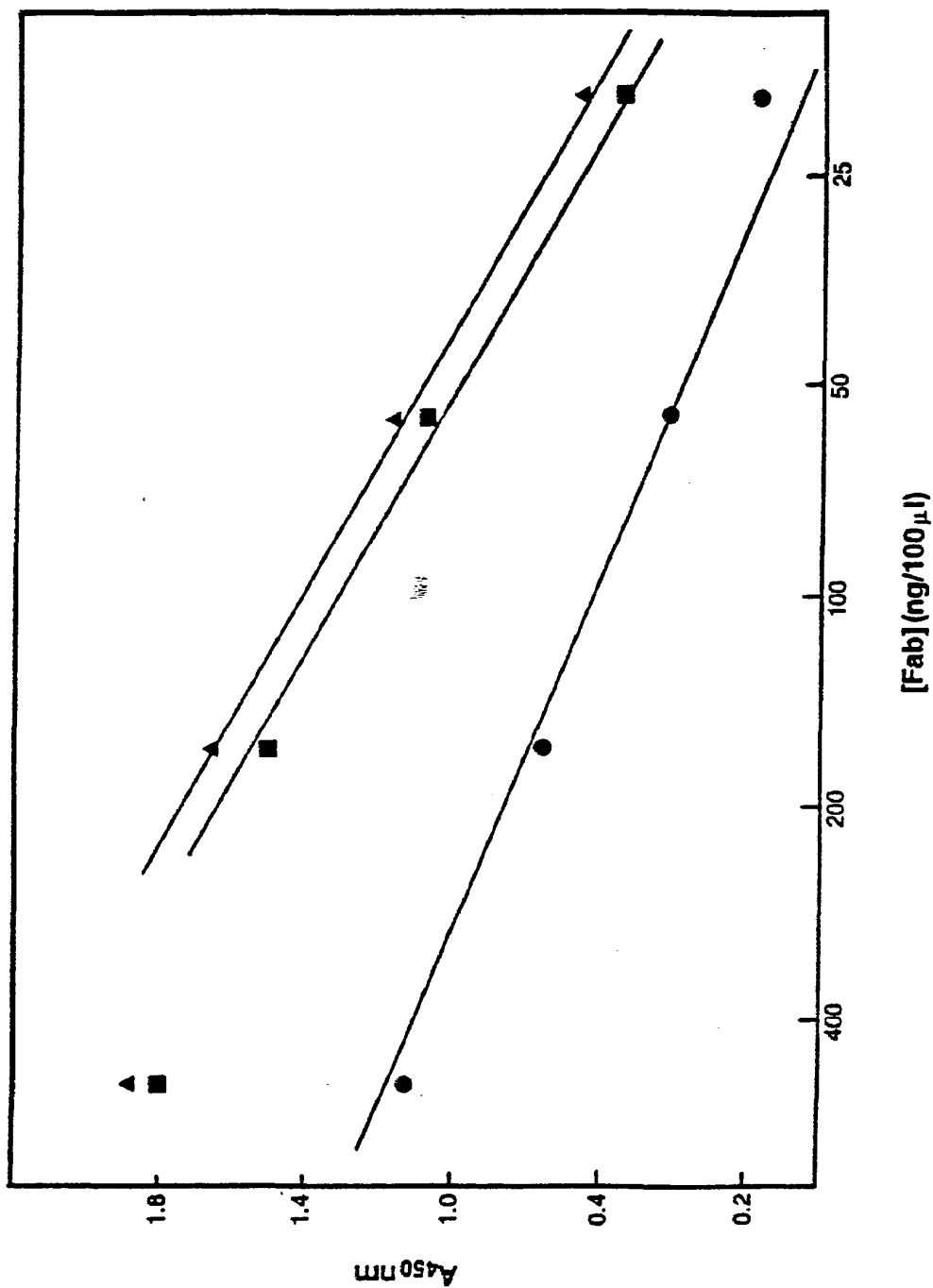
BSA

LYSOZYME

SUBSTITUTE SHEET

34/34

FIG. 36
Competitive Elisa



SUBSTITUTE SHEET

INTERNATIONAL SEARCH REPORT

International Application No. PCT/US87/02208

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all) ¹		
According to International Patent Classification (IPC) or to both National Classification and IPC		
U.S.C1.: 435/68 536/27 530/387 424/85		
IPC(4): C12P 21/00 C07H 17/00 C07K 15/04 A61K 39/395		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁴		
Classification System	Classification Symbols	
U.S.	435/7, 68, 91, 172.3, 240.2, 253, 254, 255, 320 530/387, 388, 389, 390, 391, 413 536/27 424/85	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched ⁵		
Chemical Abstracts Data Base 1967-1987 Keywords: Antibod?, Variable Region, Conformation, FOLD?, Peptide, Linker, Heavy chain, Light chain, chimeric antibody?, monospecific antibody?		
III. DOCUMENTS CONSIDERED TO BE RELEVANT ¹⁴		
Category ⁶	Citation of Document, ¹⁶ with indication, where appropriate, of the relevant passages ¹⁷	Relevant to Claim No. ¹⁸
Y	GB, A, 2,137,631A (BOSS ET AL) 10 October 1984. See pages 3-6, 9-16 and claims 1-7 and 19-33.	1-27
Y	EP, A, 0,125,023 (CABILLY ET AL) 14 November 1984. See pages 3, 4, 15-18, and 21-29.	1-27
P, A	SCIENCE, Volume 81, issued August 1987, (Washington, D.C., USA) (R. HUBER) "Structural Basis for Antigen-Antibody Recognition", see pages 702-703.	1-11 and 21
A	BIOTECHNOLOGY, Volume 4, issued April 1986, (New York, New York, U.S.A.) (J. VAN BRUNT) "Protein Architecture: Designing From the Ground Up", See pages 277-283.	1-11 and 21
<p>¹⁹ Special categories of cited documents: ¹⁵</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such docu- ments, such combination being obvious to a person skilled in the art.</p> <p>"Δ" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search ²	Date of Mailing of this International Search Report ³	
19 NOVEMBER 1987	10 DEC 1987	
International Searching Authority ¹	Signature of Authorized Officer ¹⁰	
ISA/US	Jayme A. Huleatt	

Form PCT/ISA/210 (second sheet) (May 1986)

PCT/US87/02208

VI. OBSERVATIONS WHERE UNITY OF INVENTION
IS LACKING (ATTACHMENT)

Group I, claims 1-11 and 21

Group II, claims 12-20

Group III, claim 26

Group IV, claim 27

A single inventive concept was not presented by
these claims for the reasons outlined in Form
PCT/ISA/206 mailed Oct. 22, 1987.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : C12N 15/13, C07K 15/28, A61K 39/395		A1	(11) International Publication Number: WO 94/25591
			(43) International Publication Date: 10 November 1994 (10.11.94)
(21) International Application Number: PCT/EP94/01442			
(22) International Filing Date: 28 April 1994 (28.04.94)			
(30) Priority Data: 93201239.6 29 April 1993 (29.04.93) EP (34) Countries for which the regional or international application was filed: NL et al. 93201454.1 19 May 1993 (19.05.93) EP (34) Countries for which the regional or international application was filed: NL et al. 93202079.5 15 July 1993 (15.07.93) EP (34) Countries for which the regional or international application was filed: NL et al.		(72) Inventors; and (75) Inventors/Applicants (for US only): FRENKEN, Leon, Gerardus, J. [NL/NL]; Geldersestraat 90, NL-3011 MP Rotterdam (NL). VERRIPS, Cornelis, Theodorus [NL/NL]; Hagedoorn 18, NL-3142 KB Maassluis (NL).	
(71) Applicant (for all designated States except AU BB CA GB IE LK MN MW NZ SD US): UNILEVER N.V. [NL/NL]; Weena 455, NL-3013 AL Rotterdam (NL).		(74) Common Representative: UNILEVER N.V.; Patent Division, P.O. Box 137, NL-3130 AC Vlaardingen (NL).	
(71) Applicant (for AU BB CA GB IE LK MN MW NZ SD only): UNILEVER PLC [GB/GB]; Unilever House, Blackfriars, London EC4 4BQ (GB).			
(71)(72) Applicants and Inventors: HAMERS, Raymond [BE/BE]; Vijversweg 15, B-1640 Sint-Genesius-Rode (BE). HAMERS-CASTERMAN, Cécile [BE/BE]; Vijversweg		(81) Designated States: AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, ES, FI, GB, GE, HU, JP, KG, KP, KR, KZ, LK, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
Published With international search report.			
(54) Title: PRODUCTION OF ANTIBODIES OR (FUNCTIONALIZED) FRAGMENTS THEREOF DERIVED FROM HEAVY CHAIN IMMUNOGLOBULINS OF CAMELIDAE			
(57) Abstract			
<p>A process is provided for the production of an antibody or a fragment or functionalized fragment thereof using a transformed lower eukaryotic host containing an expressible DNA sequence encoding the antibody or (functionalized) fragment thereof, wherein the antibody or (functionalized) fragment thereof is derived from a heavy chain immunoglobulin of <i>Camelidae</i> and is devoid of light chains, and wherein the lower eukaryotic host is a mould, preferably belonging to the genera <i>Aspergillus</i> or <i>Trichoderma</i>, or a yeast, preferably belonging to the yeast genera <i>Saccharomyces</i>, <i>Kluyveromyces</i>, <i>Hansenula</i>, or <i>Pichia</i>. The heavy chain fragment can contain at least the whole variable domain. A complementary determining region (CDR) different from the CDR belonging to the natural antibody ex <i>Camelidae</i> can be grafted on the framework of the variable domain of the heavy chain immunoglobulin. The catalytic antibodies can be raised in <i>Camelidae</i> against transition state molecules. The functionalized antibody or fragment thereof can comprise a fusion protein of both a heavy chain immunoglobulin from <i>Camelidae</i> or a fragment thereof and another polypeptide, e.g., an enzyme, preferably an oxido-reductase. Also provided are new products obtainable by a process as described, and compositions containing a product produced by a process as described, which composition may contain a new product as provided.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

Title: **Production of antibodies or (functionalized) fragments thereof derived from heavy chain immunoglobulins of *Camelidae***

The present invention relates to a process for the production of antibodies or
5 (functionalized) fragments thereof derived from heavy chain immunoglobulins of
Camelidae and is partly based on research investigations carried out at the Free
University of Brussels. A draft publication thereon already submitted to the
periodical Nature and communicated to the present applicants by Prof. R. Hamers
reads as follows.

10

FUNCTIONAL HEAVY CHAIN IMMUNOGLOBULINS IN THE CAMELIDS

Random association of V_L and V_H repertoires contributes considerably to antibody
diversity (1). The diversity and the affinity are then increased by hypermutation in
15 B-cells located in germinal centres (2). Except in the heavy chain disease (3),
naturally occurring heavy chain antibodies have not been described, although
antigen binding has been demonstrated for separated heavy chains (4) or cloned V_H
domains (5). The presence of considerable amounts IgG like material of 100 Kd in
the serum of the camel (*Camelus dromedarius*) (6) was confirmed. These molecules
20 are composed of heavy chain dimers and are devoid of light chains. Nevertheless
they bear an extensive antigen binding repertoire, a finding which questions the role
of the light chains in the camel. Camel heavy chain IgGs lack the C_H1 , which in one
IgG class might be structurally replaced by an extended hinge. Heavy chain IgGs are
a feature of all camelids. These findings open perspectives in engineering of
25 antibodies.

By a combination of affinity chromatography on Protein A and Protein G, three
quantitatively important fractions corresponding to subclasses of IgG can be isolated
from the serum of camels (*Camelus dromedarius*) (Fig. 1A, lanes c-f).
30 One fraction (IgG₁) contains molecules of 170 Kd (Fig. 1B, lane 2) which upon
reduction yield 50 Kd heavy chains and large 30 kD light chains (Fig. 1C, lane 2).
The two other immunoglobulin fractions contain molecules of approximately 100 Kd

(Fig. 1B, lanes 1 and 3) which upon reduction yield only heavy chains of respectively 46 Kd (IgG₂ fraction binding only to Protein A) (Fig. 1C, lane 3) and 43 Kd (IgG₃ fraction binding to Protein A and Protein G) (Fig. 1C, lane 1). These two IgG classes appear to lack the light chain completely.

5

To exclude the possibility that the light chains were only weakly associated with the heavy chains and lost during the selective purification, whole serum was size fractionated by gel filtration. Coomassie blue staining of unreduced fractions revealed the sequential elution of the 170 Kd IgG₁ followed by the incompletely
10 resolved isotypes IgG₂ and IgG₃ (90 Kd) (Fig. 1D, upper inset). Immunostaining of the same fractions after reduction confirmed that the light chains were present solely in the 50 Kd heavy chain containing fractions (Fig. 1D, lower inset).

A comparative study of old world camelids (*Camelus bactrianus* and *Camelus*
15 *dromedarius*) and new world camelids (*Lama pacos*, *Lama glama* and *Lama vicugna*) showed that heavy chain immunoglobulins are abundant in the sera of all species examined (data not shown) and total up to 75% of the molecules binding to protein A.

The abundance of the heavy chain immunoglobulins in the serum of camelids raises
20 the question as to whether they bear an extensive antigen binding repertoire. This question could be answered by examining the IgG₁, IgG₂ and IgG₃ fractions from the serum of camels (*Camelus dromedarius*) with a high antitrypanosome titer (7). In radio-immunoprecipitation, purified fractions of IgG₁, IgG₂ and IgG₃ derived from infected camels were shown to bind a large number of antigens present in a ³⁵S
25 methionine labelled trypanosome lysate (Fig. 2A), indicating an extensive repertoire complexity for the three IgG classes. Conversely, in blotting experiments, ³⁵S methionine labelled trypanosome lysate binds to SDS-PAGE separated IgG₁, IgG₂ and IgG₃ obtained from infected animals (Fig. 2B). These findings indicate that the heavy chains alone can generate an extensive repertoire and question the obligatory
30 contribution of the light chain to the useful antibody repertoire in the camelids.

The camelid $\gamma 2$ and $\gamma 3$ chains are considerably shorter than the normal mammalian γ or camel $\gamma 1$ chains. This would suggest that, as in the case of heavy chain disease (3), deletions have occurred in the C_H1 protein domain (8,9). To address this question, cDNA was synthesized from camel spleen mRNA and the sequences
5 between the 5' end of the V_H and the C_H2 were amplified by a Polymerase Chain Reaction (PCR), and cloned. Seventeen clones presenting a different V_H sequence were isolated and sequenced. Their most striking feature was the complete lack of the C_H1 domain, the last framework (FR4) residues of the V_H region being immediately followed by the hinge (Fig. 3, lower part). The absence of the C_H1 domain
10 clarifies two important dilemmas.

First, immunoglobulin heavy chains are normally not secreted unless the heavy chain chaperoning protein or BIP (10) has been replaced by the L chain (11), or alternatively the C_H1 domain has been deleted (3,8,9). Secondly, isolated heavy
15 chains from mammalian immunoglobulins tend to aggregate, but are only solubilized by light chains (8,12) which bind to the C_H1 and the V_H domains (13).

14 of the 17 clones were characterized by a short hinge sequence with a length equal to that of human IgG₂ and IgG₄ (14) (Fig. 3). The other 3 had a long hinge
20 sequence containing the 'EPK' hinge motif found in human IgG₁ and IgG₃ (14). They possess the C_H2 'APELL/P' motif also found in human IgG₁ and IgG₃ (see SEQ. ID. NO: 1-2), and which is associated with mammary transport of bovine IgG₁ (15). On basis of molecular weight, we expect the "short hinge" clones to correspond to IgG₃ and the "long hinge" clones to IgG₂.

25

In the short hinge containing antibody, the extreme distance between the extremities of the V_H regions will be of the order of 80 Å corresponding to twice the size of a single domain of 40 Å ($2 \times V_H$) (16). This could be a severe limitation for agglutinating, cross linking or complement fixation (17,18). In the long hinge containing
30 immunoglobulin the absence of C_H1 might be compensated by the extremely long hinge itself, composed of a 12 fold repeat of the sequence Pro-X (X=Gln, Glu, Lys) (Fig. 3 & 4). NMR (19) and molecular modelling (20) of Pro-X repeats present in

the TonB protein of *E. coli* (X=Glu, Lys) and the membrane procyclin of trypanosomes (X=Asp, Glu) indicate that these repeated sequences function as rigid rodlike spacers with a diameter of 8 Å and a rise of 2.9 Å per residue. Assuming the same geometry, the long hinge would be 70 Å which compensates for the absence of the C_H1 domain.

The binding site of heavy chain antibodies cannot form the pocket resulting from adjoining light and heavy chain V regions and the residues of the V_H which normally interact with V_L will be exposed to solvent (3,5,13). It was found that leucine at position 45 conserved in 98% of human and murine V_H sequences (14), and crucial in the V_H-V_L association (13), can be replaced by an arginine (Fig. 3, upper part). This substitution is in accordance with both the lost contact with a V_L domain and an increased solubility.

Unlike myeloma heavy chains which result mainly from C_H1 deletion in a single antibody producing cell (21) the camelid heavy chain antibodies have emerged in a normal immunological environment and it is expected that they will have undergone the selective refinement in specificity and affinity accompanying B cell maturation (1, 2). The obtention of camelid heavy chain antibodies could therefore be an invaluable asset in the development and engineering of soluble V_H domains (5) or of new immunologicals for diagnostic, therapeutic or biochemical purposes.

REFERENCES

1. Tonegawa, S. *Nature* **302**, 575-581 (1983).
- 25 2. Jacob, J., Kelsoe, G., Rajewski, K., & Weiss, U. *Nature* **354**, 389-392 (1991).
3. Fleischman J.B., Pain R.H. & Porter R.R. *Arch.Biochem.Biophys* **Suppl. 1**, 174-180 (1962).
4. Utsumi, S. & Karush, F. *Biochemistry* **3**, 1329-1338 (1964).
5. Ward, E.S., Güssow, D., Griffiths, A.d., Jones, P.T. & Winter G. *Nature* **341**, 544-546 (1989).
- 30 6. Ungar-Waron H., Eliase E., Gluckman A. and Trainin Z. *Isr. J. Vet. Med.* **43**, 198-203 (1987).

7. Bajyana Songa, E., & Hamers R. *Ann.Soc.Belge Méd.Trop.* **68**, 233-240 (1988).
8. Seligmann M., Mihaesco E., Preud'homme J.-L., Danon F. & Brouet J.-C. *Immun.Rev.* **48**, 145-167 (1979).
9. Traunecker, A., Schneider, J., Kiefer, H., Karjalainen, K., *Nature* **339**, 68-70
5 (1989).
10. Hendershot L.M., Bole D., Köhler, G. & Kearney, J.F. *J. Cell Biol.* **104**, 761-767 (1987).
11. Hendershot L.M. *J.Cell Biol.* **111**, 829-837 (1990).
12. Roholt O., Onoue K. & Pressman D. *Proc.Natn.Acad.Sci. USA* **51**, 173-178
10 (1964).
13. Chothia, C., Novotny, J., Bruccoleri, R., Karplus, M. *J. Mol.Biol.* **186**, 651-663 (1985).
14. Kabat E.A., Wu, T.T., Reid-Miller, M., Perry H.M. & Gottesman, K.S. *Sequences of Proteins of Immunological Interest* 511 (U.S. Dept of Health and Human Services,
15 US Public Health Service, National Institutes of Health, Bethesda, 1987).
15. Jackson, T., Morris, B.A, Sanders, P.G. *Molec.Immun.* **29**, 667-676 (1992).
16. Poljak R.J. *et al. Proc.Natn.Acad.Sci. USA* **70**, 3305-3310 (1973).
17. Dangi J.L., *et al. EMBO J.* **7**, 1989-1994 (1988).
18. Schneider W.P. *et al. Proc.Natn.Acad.Sci USA* **85**, 2509-2513 (1988).
- 20 19. Evans, J.S. *et al. FEBS Lett.* **208**, 211-216 (1986).
20. Roditi, I. *et al. J.Cell Biol.* **108**, 737-746 (1989).
21. Dunnick, W., Rabbits, T.H., Milstein, C. *Nucl.Acids Res.*, **8**, 1475-1484 (1980).
22. Bülow, R., Nonnengässer, C., Overath, P. *Mol.Biochem.Parasitol.* **32**, 85-92 (1989).
- 25 23. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd Edn (Cold Spring Harbor Laboratory Press, New York, 1989).
24. Sastry, L *et al. Proc.Natn.Acad.Sci. USA* **86**, 5728-5732 (1989).
25. Sanger, F., Nicklen, S. & Coulson, A.R. *Proc.Natn.Acad.Sci. USA* **74**, 5463-5467 (1977).
- 30 26. Klein, J. *Immunology* (Blackwell Scientific Publications, London, 1990).

Figure 1 Characterisation and purification of camel IgG classes on Protein A, Protein G and gel filtration.

- (A) The fraction of *C. dromedarius* serum adsorbed on Protein A shows upon reduction on SDS-PAGE three heavy chain components of respectively 50, 46, and 43 Kd (bands between dots), absent in the non adsorbed fraction (lane d), and light chain components of around 30 Kd (lane c) considerably larger than rabbit light chain (lane a, rabbit IgG). The fractions adsorbed on Protein G (lane e) lack the 46 Kd heavy chain which remains in the non adsorbed fraction (lane f). Lane b contains a size marker.
- 5
- (B and C) By differential adsorption and elution on Protein G and Protein A, the IgG fractions containing 43 Kd (lane 1), 46 Kd (lane 3) and 50 Kd (lanes 2) heavy chains were purified and analysed on SDS-PAGE in absence (B) or presence (C) of DTT.
- 10
- (D) Whole camel serum (0.1 ml) was fractionated by gel filtration on a Superdex 200 column using 150 mM NaCl, 50 mM sodium phosphate buffer pH 7.0 as eluent. Affinity purified IgG₂ and IgG₃ elute at the positions indicated by arrows. The fractions of interest were further analysed by SDS-PAGE with or without prior reduction. The protein contents as visualized by Coomassie blue (without reduction, upper inset) are compared with the immunoglobulins from the same fractions (after reduction with DTT, lower inset) as revealed by Western blotting with a rabbit anti-camel-IgG (lower inset).
- 15
- 20

METHODS. 5 ml of *C. dromedarius* serum is adsorbed onto a 5 ml Protein G Sepharose (Pharmacia) column, and washed with 20 mM phosphate buffer, pH 7.0.

25

Upon elution with 0.15 M NaCl, 0.58 % acetic acid (pH 3.5), IgG₃ of 100 Kd is eluted which upon reduction yields heavy chains of 43 Kd (lane 1, B and C). IgG₁ of 170 Kd can subsequently be eluted with pH 2.7 buffer (0.1 M Gly-HCl). This fraction, upon reduction, yields a 50 Kd heavy chain and a broad light chain band (lane 2, C). The fraction not adsorbed on Protein G is brought on a 5 ml Protein A

30

Sepharose column. After washing and elution with 0.15 M NaCl, 0.58% acetic acid (pH 4.5) IgG₂ of 100 Kd is obtained which consists solely of 46 Kd heavy chains (lane 3, C).