



US 20170362608A1

(19) **United States**

(12) **Patent Application Publication**
McLaughlin et al.

(10) **Pub. No.: US 2017/0362608 A1**

(43) **Pub. Date: Dec. 21, 2017**

(54) **COMPOSITIONS AND METHODS FOR THE PRODUCTION OF SCAAV**

Publication Classification

(71) Applicant: **VOYAGER THERAPEUTICS, INC.**,
CAMBRIDGE, MA (US)

(72) Inventors: **James McLaughlin**, Cambridge, MA
(US); **Robert Kotin**, Boston, MA (US)

(21) Appl. No.: **15/535,389**

(22) PCT Filed: **Dec. 11, 2015**

(86) PCT No.: **PCT/US15/65218**

§ 371 (c)(1),

(2) Date: **Jun. 12, 2017**

(51) **Int. Cl.**

C12N 15/861 (2006.01)

C12N 7/02 (2006.01)

C12N 15/864 (2006.01)

(52) **U.S. Cl.**

CPC **C12N 15/861** (2013.01); **C12N 15/864**
(2013.01); **C12N 7/02** (2013.01); **C12N**
2710/14144 (2013.01); **C12N 2750/14143**
(2013.01); **C12N 2750/14151** (2013.01); **C12N**
2750/14122 (2013.01)

Related U.S. Application Data

(60) Provisional application No. 62/091,062, filed on Dec.
12, 2014.

(57)

ABSTRACT

The present invention is directed to viral vectors and meth-
ods of their production and use.

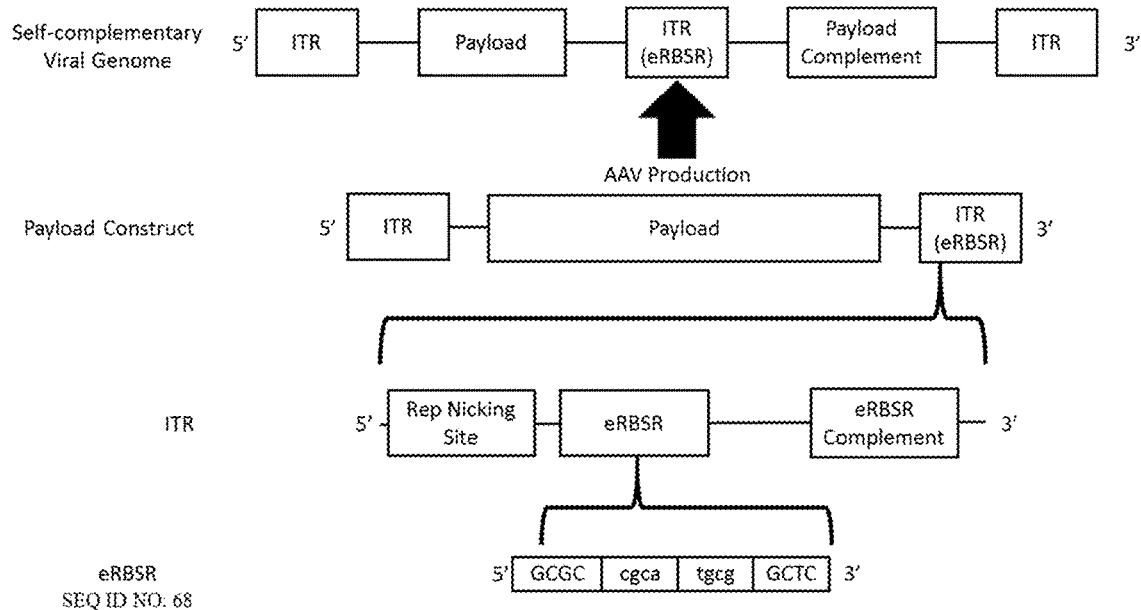
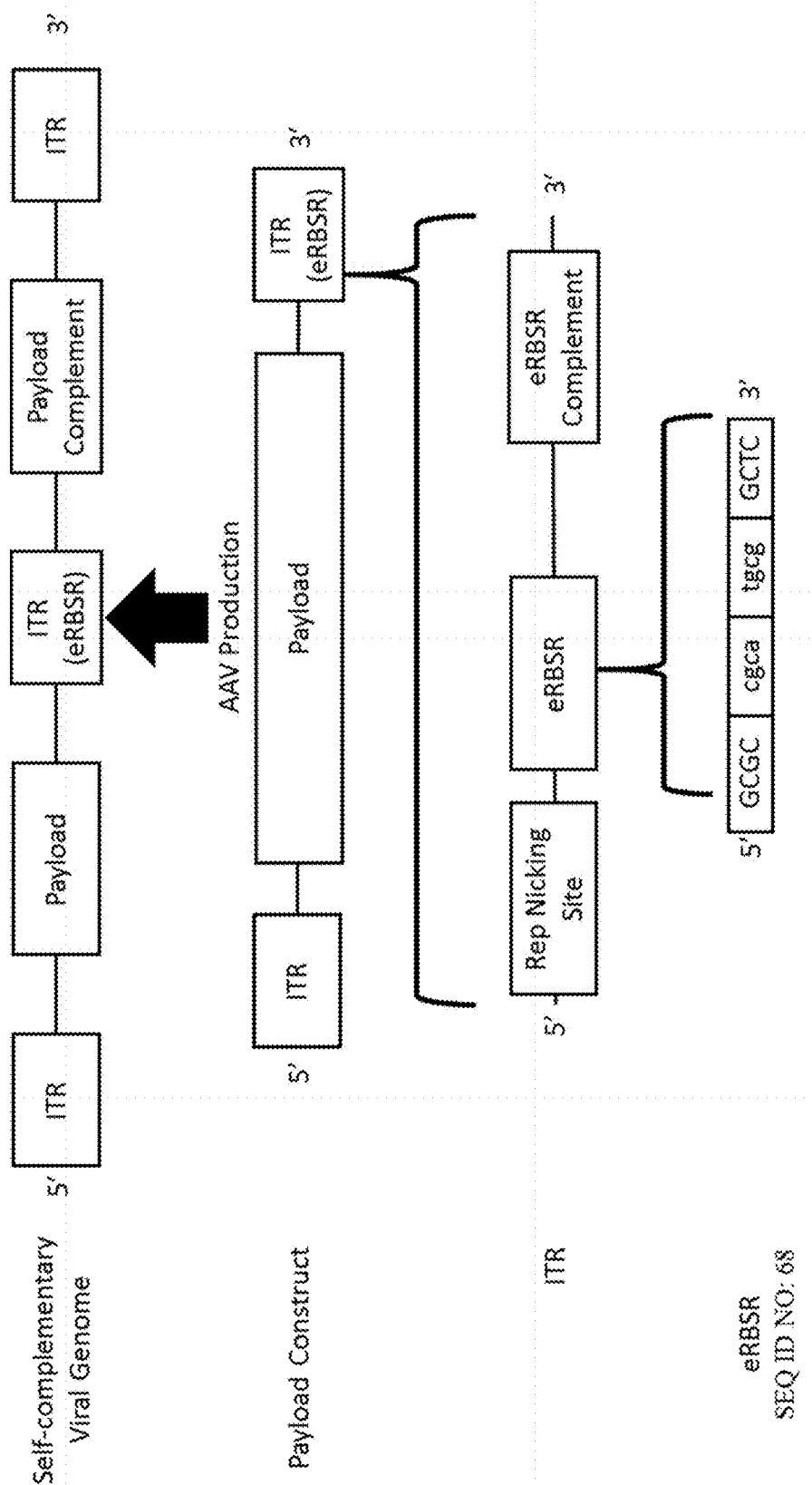
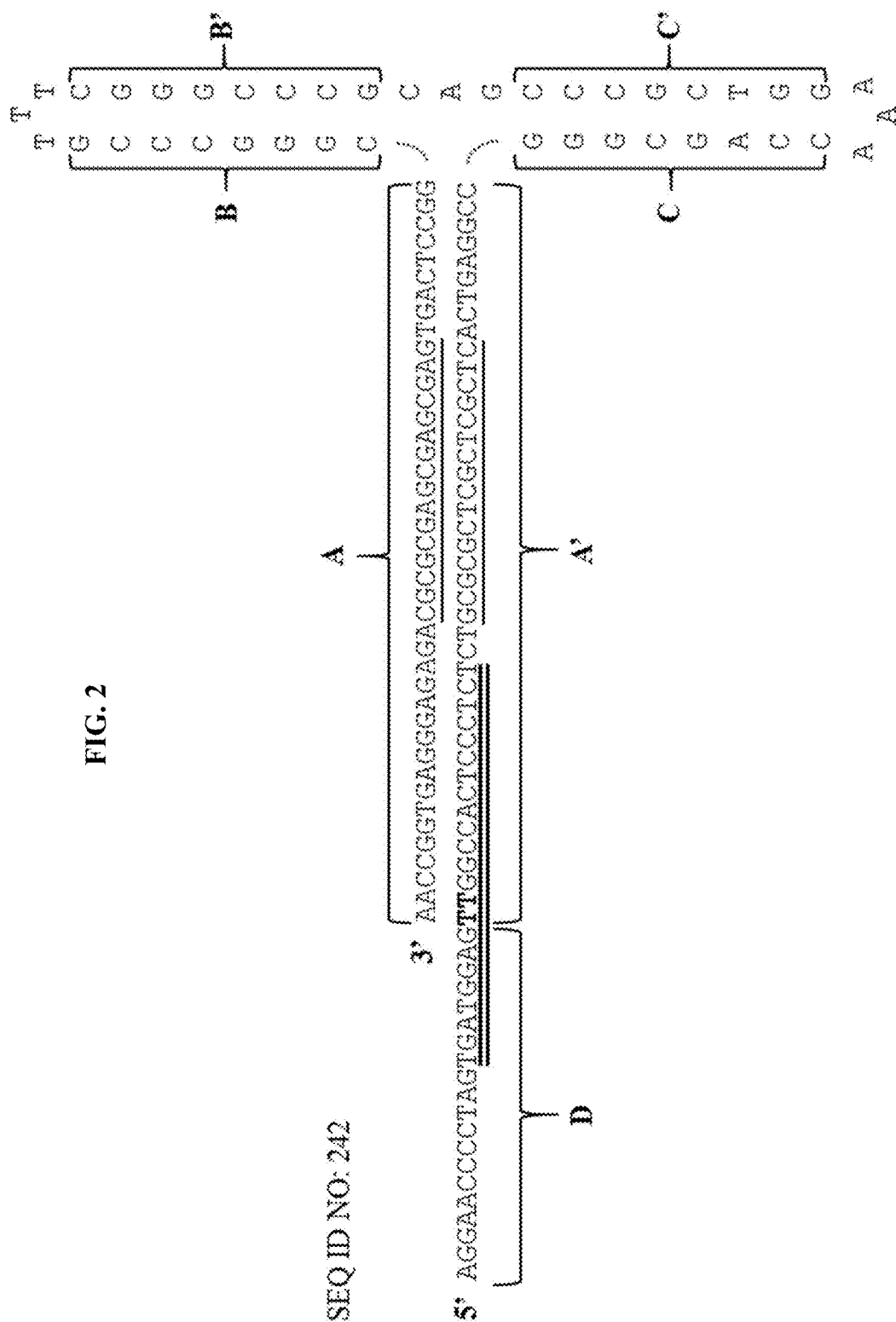


FIG. 1



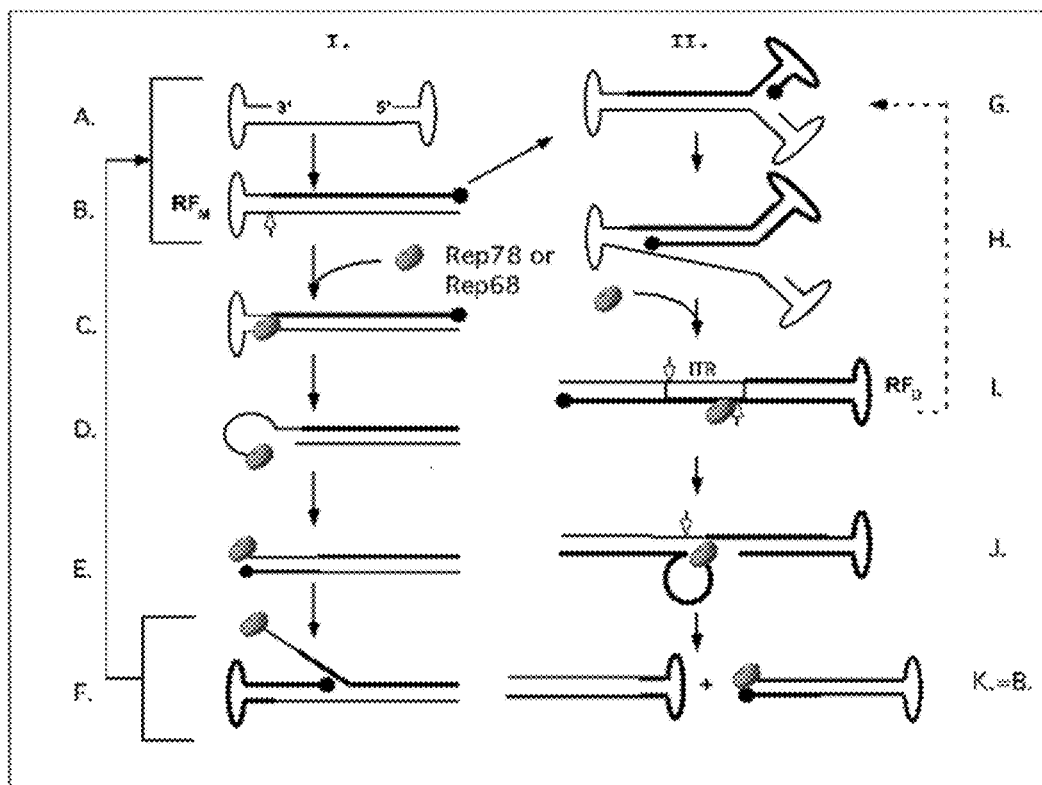


Left 5' TRA

Right 3' TR

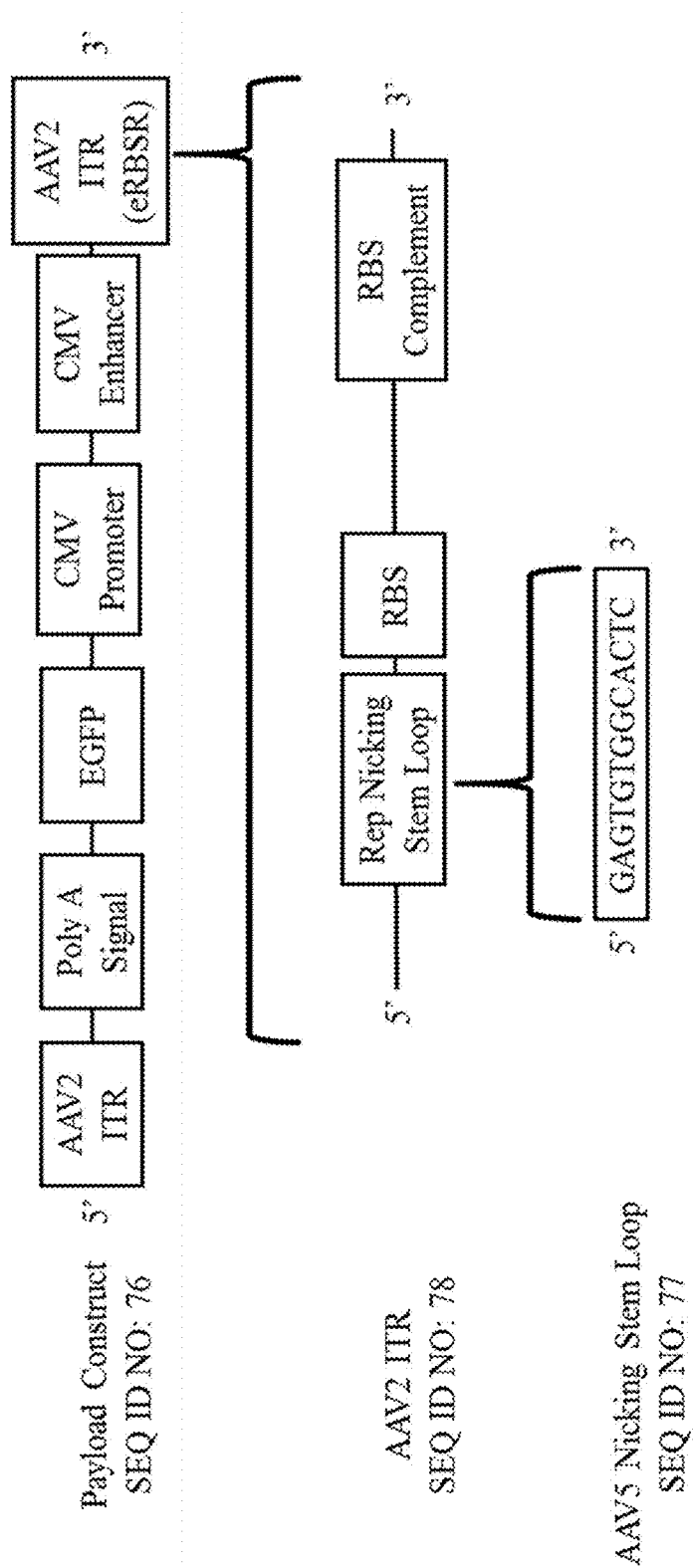
[illegible]

FIG. 5



PRIOR ART

FIG. 6A



82-0101078

3' AACCGTCAGGAGAGACGCCGAGCGAGTCACTCCGG
5' AGGAACCCCTAGTGAGATTGGCCACTCCCTCTCTGCGGCTCGCTCGCTCACTGAGGCC

GAGTGTGCATC

AAV5 Rep Nicking Stem Loop AAV2 Rep Nicking Stem Loop
SEQ ID NO: 77 SEQ ID NO: 312

T T C C G G C C C C C C A G C C C C G C T C G C
T G C C C C G G G C C C G G G C C G A C C C A A

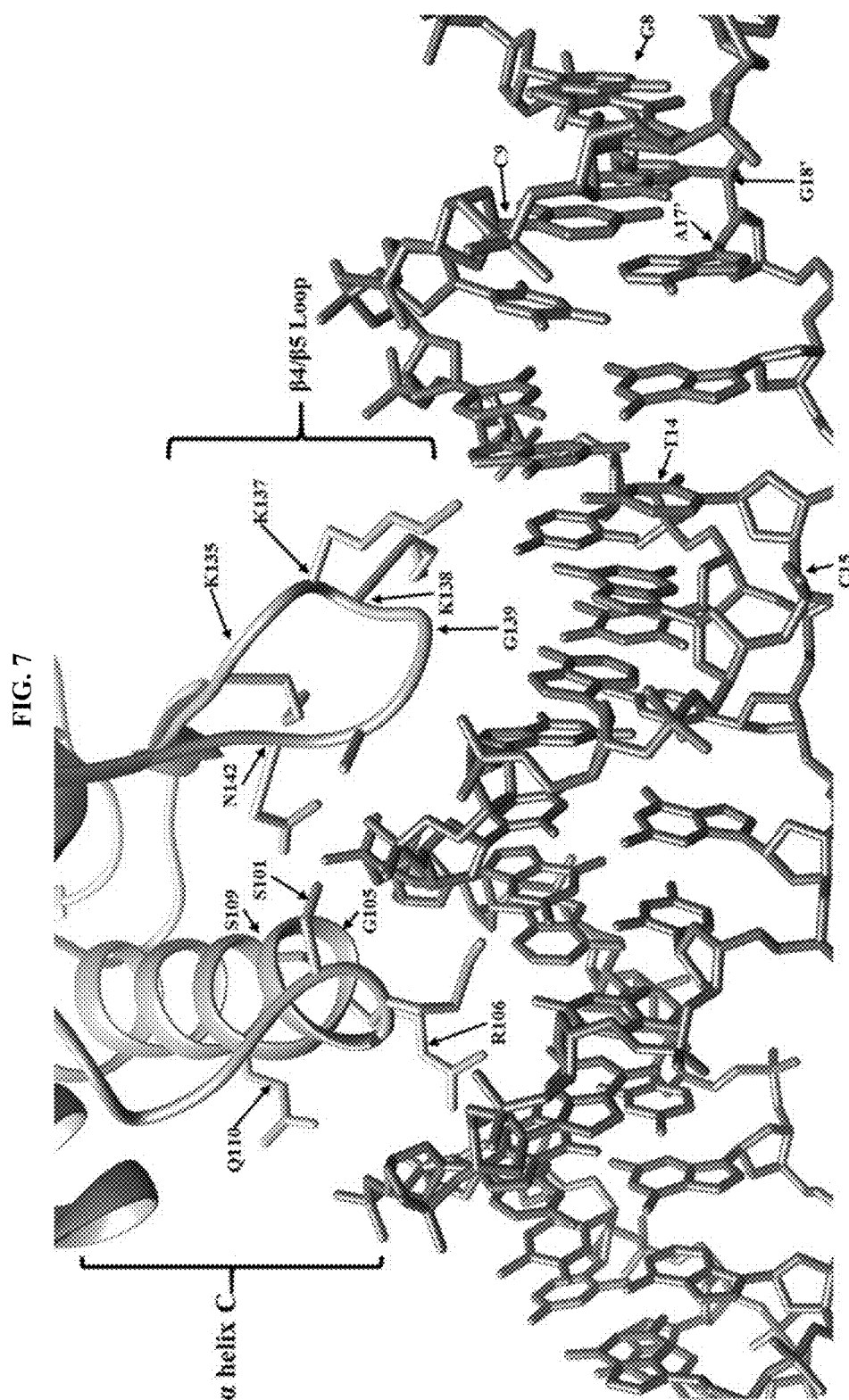


FIG. 8

AAV-5	(NC_006152.1)	SEQ ID NO: 269	MATFEYIVRVPEDVEEHLPGISDSFVDWVTGQIWELPPESDNLNLTVEQPQLTVADRIR	60
AAV-2_Rep68	(NC_001401.2)	SEQ ID NO: 265	MPGFEYIVKVPESDLDERLPGISDSFVNNVAERKEWELPPDSMDNLNIEQAPLTVAEKLQ	60
AAV-2_Rep78	(NC_001401.2)	SEQ ID NO: 266	MPGFEYIVKVPESDLDERLPGISDSFVNNVAERKEWELPPDSMDNLNIEQAPLTVAEKLQ	60
AAV-2_Rep40	(NC_001401.2)	SEQ ID NO: 267	-----	0
AAV-2_Rep52	(NC_001401.2)	SEQ ID NO: 268	-----	0
AAV-1	(NC_002077.1)	SEQ ID NO: 270	MPGFEYIVKVPESDLDERLPGISDSFVSWVAERKEWELPPDSMDNLNIEQAPLTVAEKLQ	60
AAV-3	(NC_001729.1)	SEQ ID NO: 271	MPGFEYIVKVPESDLDERLPGISDSFVNNVAERKEWELPPDSMDNLNIEQAPLTVAEKLQ	60
AAV-4	(NC_001829.1)	SEQ ID NO: 272	MPGFEYIVKVPESDLDERLPGISDSFVNNVAERKEWELPPDSMDNLNIEQAPLTVAEKLQ	60
AAV-7	(NC_006260.1)	SEQ ID NO: 273	MPGFEYIVKVPESDLDERLPGISDSFVNNVAERKEWELPPDSMDNLNIEQAPLTVAEKLQ	60
AAV-8	(NC_006261.1)	SEQ ID NO: 274	MPGFEYIVKVPESDLDERLPGISDSFVNNVAERKEWELPPDSMDNLNIEQAPLTVAEKLQ	60
Avian_AAV_DA-1	(NC_006263.1)	SEQ ID NO: 275	MRSYFEIVQLPNDSVSEQVPGISDSFVNNWITSREWTLPEADWDLDQVDQVQLTGDKIQ	60
Avian_AAV_VR-865	(NC_004828.1)	SEQ ID NO: 276	MRSYFEIVQLPNDSVSEQVPGISDSFVNNWITSREWTLPEADWDLDQVDQVQLTGDKIQ	60
Bat	(NC_014468.1)	SEQ ID NO: 278	-MEFYSIVLRPGDFDSEVPGIQDSFYKWSISGRRELPEWSDLPQGIESAYQILLADKLIV	59
Bovine	(NC_005889.1)	SEQ ID NO: 277	MATFEYIVRVPEDVEEHLPGISDSFVDWVTGQIWELPPESDNLNLTVEQPQLTVADRIR	60
AAV-5	(NC_006152.1)	SEQ ID NO: 269	RVFLYEWNKESEQ-ESKFFVQFEKGESEYFHLHLIVETSGISSMVLGRYVSQIRAQVVK-V	11
AAV-2_Rep68	(NC_001401.2)	SEQ ID NO: 265	RDELTWRRVSKAPEALFFVQFEKGESEYFHMHLIVETTGKSMVLGRFLSQIREKLIQ-R	11
AAV-2_Rep78	(NC_001401.2)	SEQ ID NO: 266	RDELTWRRVSKAPEALFFVQFEKGESEYFHMHLIVETTGKSMVLGRFLSQIREKLIQ-R	11
AAV-2_Rep40	(NC_001401.2)	SEQ ID NO: 267	-----	0
AAV-2_Rep52	(NC_001401.2)	SEQ ID NO: 268	-----	0
AAV-1	(NC_002077.1)	SEQ ID NO: 270	ROELVQWRRVSKAPEALFFVQFEKGESEYFHLHLIVETTGKSMVLGRFLSQIRDKLIVQ-T	11
AAV-3	(NC_001729.1)	SEQ ID NO: 271	REFLVEWRRVSKAPEALFFVQFEKGESEYFHLHLIVETTGKSMVVGGRYVSQIKERLVT-R	11
AAV-4	(NC_001829.1)	SEQ ID NO: 272	REFLVEWRRVSKAPEALFFVQFEKGESEYFHLHLIVETTGKSMVVGGRYVSQIKERLVT-R	11
AAV-7	(NC_006260.1)	SEQ ID NO: 273	RDELVQWRRVSKAPEALFFVQFEKGESEYFHLHLIVETTGKSMVLGRFLSQIREKLIQ-T	11
AAV-8	(NC_006261.1)	SEQ ID NO: 274	ROFLVQWRRVSKAPEALFFVQFEKGESEYFHLHLIVETTGKSMVLGRFLSQIREKLGPDH	12
Avian_AAV_DA-1	(NC_006263.1)	SEQ ID NO: 275	REIRNHGWTMAKEPDPFHIFIQLEQGEVFFHLHLIVLETCSVKPMVLGRYIRHHIQQRIVS-K	11
Avian_AAV_VR-865	(NC_004828.1)	SEQ ID NO: 276	REIRTHGWTMAKEPDPFHIFIQLEQGEVFFHLHLIVLETCSVKPMVLGRYIRHHIQQRIVS-K	11
Bat	(NC_014468.1)	SEQ ID NO: 278	REFAQKWAFFSEDPRAPTAQLEKGFENFHVHLASSKKVGSFVVGGRYVRKMRQHLVD-V	11
Bovine	(NC_005889.1)	SEQ ID NO: 277	RVFLYEWNKESEQ-ESKFFVQFEKGESEYFHLHLIVETSGISSMVLGRYVSQIRAQVVK-V	11

FIG. 8 cont.

AAV-5	(NC_006152.1)	SEQ ID NO: 269	VFQGIETPLNDWVAITKTKA--KGAGKVVDSGVIIPAYLLPKVQPELQWANTNIDBYKLA	175
AAV-2_Rep68	(NC_001401.2)	SEQ ID NO: 265	IYRGIEPTLPNWFVAVTKTKN--GAGGKVVWDECIIPNYLLKTKQPELQWANTNMEQYLSA	176
AAV-2_Rep78	(NC_001401.2)	SEQ ID NO: 266	IYRGIEPTLPNWFVAVTKTKN--GAGGKVVWDECIIPNYLLKTKQPELQWANTNMEQYLSA	178
AAV-2_Rep40	(NC_001401.2)	SEQ ID NO: 267	-----	0
AAV-2_Rep52	(NC_001401.2)	SEQ ID NO: 268	-----	0
AAV-1	(NC_002077.1)	SEQ ID NO: 270	IYRGIEPTLPNWFVAVTKTKN--NGAGGKVVWDECIIPNYLLKTKQPELQWANTNMEYISA	178
AAV-3	(NC_001729.1)	SEQ ID NO: 271	IYRGVEPQLPNWFVAVTKTKN--GAGGKVVWDDCIIPNYLLKTKQPELQWANTNDDYLSA	178
AAV-4	(NC_001829.1)	SEQ ID NO: 272	IYRGVEPQLPNWFVAVTKTKN--GAGGKVVWDDCIIPNYLLKTKQPELQWANTNDDYISA	178
AAV-7	(NC_006260.1)	SEQ ID NO: 273	IYRGVEPTLPNWFVAVTKTKA--NGAGGKVVWDECIIPNYLLKTKQPELQWANTNMEYISA	178
AAV-8	(NC_006261.1)	SEQ ID NO: 274	LEAGSSPTLPNWFVAVTKTKN--GAGGKVVWDECIIPNYLLKTKQPELQWANTNMEYISA	180
Avian_AAV_DA-1	(NC_006263.1)	SEQ ID NO: 275	YVCGHEPAMEGHWVTKTKN--FGGKVVRAESYIPAYLLIKQQPEVQWANTNVPEYIKA	177
Avian_AAV_VR-865	(NC_004828.1)	SEQ ID NO: 276	VYCALSIRNKDSCVTKTKN--FGGKVVRAESYIPAYLLIKQQPEVQWANTNVPEYIKA	177
Bat	(NC_014468.1)	SEQ ID NO: 278	VERKCEPVDADMLQVQKKG--NHSNFKQEGFIPAYLLIKRQSELQWANTNIEKVERA	175
Bovine	(NC_005889.1)	SEQ ID NO: 277	VFQNIETPRINDWVAITKTKA--KGAGKVVDSGVIIPAYLLPKVQPELQWANTNLEBYKLA	175
AAV-5	(NC_006152.1)	SEQ ID NO: 269	AINLEERKRLVAGFLAESQRS--QEAASQREFSADVPVKSQTSQKYMALVNWLVHVGITS	234
AAV-2_Rep68	(NC_001401.2)	SEQ ID NO: 265	CINLAEKRLVAGHLTHVSQTQEKQNKQNPNSDAPVIRSTSAARYMEIVGWLVDKGITS	238
AAV-2_Rep78	(NC_001401.2)	SEQ ID NO: 266	CINLAEKRLVAGHLTHVSQTQEKQNKQNPNSDAPVIRSTSAARYMEIVGWLVDKGITS	238
AAV-2_Rep40	(NC_001401.2)	SEQ ID NO: 267	-----	14
AAV-2_Rep52	(NC_001401.2)	SEQ ID NO: 268	-----	14
AAV-1	(NC_002077.1)	SEQ ID NO: 270	CINLAEKRLVAGHLTHVSQTQEKQNKQNPNSDAPVIRSTSAARYMEIVGWLVDKGITS	238
AAV-3	(NC_001729.1)	SEQ ID NO: 271	CINLAEKRLVAGHLTHVSQTQEKQNKQNPNSDAPVIRSTSAARYMEIVGWLVDKGITS	238
AAV-4	(NC_001829.1)	SEQ ID NO: 272	CINLAEKRLVAGHLTHVSQTQEKQNKQNPNSDAPVIRSTSAARYMEIVGWLVDKGITS	238
AAV-7	(NC_006260.1)	SEQ ID NO: 273	CINLAEKRLVAGHLTHVSQTQEKQNKQNPNSDAPVIRSTSAARYMEIVGWLVDKGITS	238
AAV-8	(NC_006261.1)	SEQ ID NO: 274	CINLAEKRLVAGHLTHVSQTQEKQNKQNPNSDAPVIRSTSAARYMEIVGWLVDKGITS	240
Avian_AAV_DA-1	(NC_006263.1)	SEQ ID NO: 275	CLHRLRASLARLHFEKAGLSQSK--ENLARTADGAPVIATRVSKRYMEIVDMLVKGIT	236
Avian_AAV_VR-865	(NC_004828.1)	SEQ ID NO: 276	CLHRLRASLARLHFEKAGLSQSK--ENLARTADGAPVMPTRVSKRYMEIVDMLVKGIT	236
Bat	(NC_014468.1)	SEQ ID NO: 278	TLSVAFERARLVEMKPSLAESDP----AEPERPRKSTKASEYMALVFNWLVNGIAT	231
Bovine	(NC_005889.1)	SEQ ID NO: 277	AINLEERKRLVAGFLAESQRS--QEAASQROVSADVPVKSQTSQKYMALVSWLVHVGITS	234

* * * * *

(NC_006152.1)	SEQ ID NO: 269	EKWIQEQQASYLSFSTGNSRSQIKKALDNATKMSLTKSAVPLVGSVPE---	DISKN	296
(NC_001401.2)	SEQ ID NO: 265	EKWIQEQQASYLSFSTGNSRSQIKKALDNATKMSLTKTAPDVLGQOPVE---	DISSN	296
(NC_001401.2)	SEQ ID NO: 266	EKWIQEQQASYLSFSTGNSRSQIKKALDNATKMSLTKTAPDVLGQOPVE---	DISSN	296
(NC_001401.2)	SEQ ID NO: 267	EKWIQEQQASYLSFSTGNSRSQIKKALDNATKMSLTKTAPDVLGQOPVE---	DISSN	72
(NC_001401.2)	SEQ ID NO: 268	EKWIQEQQASYLSFSTGNSRSQIKKALDNATKMSLTKTAPDVLGQOPVE---	DISSN	72
(NC_002077.1)	SEQ ID NO: 270	EKWIQEQQASYLSFSTGNSRSQIKKALDNATKMSLTKTAPDVLGQOPVE---	DIKTN	296
(NC_001729.1)	SEQ ID NO: 271	EKWIQEQQASYLSFSTGNSRSQIKKALDNATKMSLTKTAPDVLGQOPVE---	DIKTN	296
(NC_001829.1)	SEQ ID NO: 272	EKWIQEQQASYLSFSTGNSRSQIKKALDNATKMSLTKTAPDVLGQOPVE---	DISSN	296
(NC_006260.1)	SEQ ID NO: 273	EKWIQEQQASYLSFSTGNSRSQIKKALDNATKMSLTKTAPDVLGQOPVE---	DIKTN	296
(NC_006261.1)	SEQ ID NO: 274	EKWIQEQQASYLSFSTGNSRSQIKKALDNATKMSLTKTAPDVLGQOPVE---	DIKTN	296
(NC_006263.1)	SEQ ID NO: 275	EKWILENRESFERSFQSSNSRSQIKTALQATQEMLLTKTAEVDVLGKRPVSDDEIRQN	296	
(NC_004828.1)	SEQ ID NO: 276	EKWILENRESFERSFQSSNSRSQIKTALQATQEMLLTKTAEVDVLGKRPVSDDEIRQN	296	
(NC_014468.1)	SEQ ID NO: 278	ERENHREDSQVLSYATGATGATQIKKALDNATKMSLTKTAPDVLGQOPVE---	DVEDN	289
(NC_005889.1)	SEQ ID NO: 277	EKWIQEQQASYLSFSTGNSRSQIKKALDNATKMSLTKSAVPLVGSVPE---	DISEN	292
*::: * : : *::: : ***::: * : * *::: *****				
(NC_006152.1)	SEQ ID NO: 269	RINQIFEMNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	352	
(NC_001401.2)	SEQ ID NO: 265	RIYKILELNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	356	
(NC_001401.2)	SEQ ID NO: 266	RIYKILELNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	356	
(NC_001401.2)	SEQ ID NO: 267	RIYKILELNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	132	
(NC_001401.2)	SEQ ID NO: 268	RIYKILELNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	132	
(NC_002077.1)	SEQ ID NO: 270	RIYKILELNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	356	
(NC_001729.1)	SEQ ID NO: 271	RIYKILELNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	356	
(NC_001829.1)	SEQ ID NO: 272	RIYKILELNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	356	
(NC_006260.1)	SEQ ID NO: 273	RIYKILELNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	356	
(NC_006261.1)	SEQ ID NO: 274	RIYKILELNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	358	
(NC_006263.1)	SEQ ID NO: 275	RIYKILELNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	356	
(NC_004828.1)	SEQ ID NO: 276	RIYKILELNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	356	
(NC_014468.1)	SEQ ID NO: 278	RIYKILELNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	349	
(NC_005889.1)	SEQ ID NO: 277	RINQIFEMNGYDPQYAAVSFLGWAQKQKFKRNTIWLFGPATTGKTNIACAAIAHTVPFYGC	352	
*::: * : : *::: : ***::: * : * *::: *****				

AAV-5		(NC_006152.1)	SEQ ID NO: 269	VNWTNENFPFNDCCVDKXMIWNEEGKMTAKVVESAKAILGSSKVRVDQCKSSAQIDSTPEV	412
AAV-2_Rep68		(NC_001401.2)	SEQ ID NO: 265	VNWTNENFPFNDCCVDKXMIWNEEGKMTAKVVESAKAILGSSKVRVDQCKSSAQIDPTPEV	416
AAV-2_Rep78		(NC_001401.2)	SEQ ID NO: 266	VNWTNENFPFNDCCVDKXMIWNEEGKMTAKVVESAKAILGSSKVRVDQCKSSAQIDPTPEV	416
AAV-2_Rep40		(NC_001401.2)	SEQ ID NO: 267	VNWTNENFPFNDCCVDKXMIWNEEGKMTAKVVESAKAILGSSKVRVDQCKSSAQIDPTPEV	192
AAV-2_Rep52		(NC_001401.2)	SEQ ID NO: 268	VNWTNENFPFNDCCVDKXMIWNEEGKMTAKVVESAKAILGSSKVRVDQCKSSAQIDPTPEV	192
AAV-1		(NC_002077.1)	SEQ ID NO: 270	VNWTNENFPFNDCCVDKXMIWNEEGKMTAKVVESAKAILGSSKVRVDQCKSSAQIDPTPEV	416
AAV-3		(NC_001729.1)	SEQ ID NO: 271	VNWTNENFPFNDCCVDKXMIWNEEGKMTAKVVESAKAILGSSKVRVDQCKSSAQIEPTPEV	416
AAV-4		(NC_001829.1)	SEQ ID NO: 272	VNWTNENFPFNDCCVDKXMIWNEEGKMTAKVVESAKAILGSSKVRVDQCKSSAQIDPTPEV	416
AAV-7		(NC_006260.1)	SEQ ID NO: 273	VNWTNENFPFNDCCVDKXMIWNEEGKMTAKVVESAKAILGSSKVRVDQCKSSAQIDPTPEV	416
AAV-8		(NC_006261.1)	SEQ ID NO: 274	VNWTNENFPFNDCCVDKXMIWNEEGKMTAKVVESAKAILGSSKVRVDQCKSSAQIDPTPEV	418
Avian_AAV_DA-1		(NC_006263.1)	SEQ ID NO: 275	VNWTNENFPFNDCCVEKMIWNEEGKMTAKVVETAKAILGSSRVRVDQCKASVPTEPTPEV	416
Avian_AAV_VR-865		(NC_004828.1)	SEQ ID NO: 276	VNWTNENFPFNDCCVEKMIWNEEGKMTAKVVETAKAILGSSRVRVDQCKASVPTEPTPEV	416
Bat		(NC_014468.1)	SEQ ID NO: 278	VNWTNENFPFNDCCVDKXMIWNEEGKMTSKVVESAKAILGSSKVRVDQCKCNQQIEPTPEV	409
Bovine		(NC_005889.1)	SEQ ID NO: 277	VNWTNENFPFNDCCVEKMIWNEEGKMTSKVVESPAKAILGSSRVRVDQCKSSQVQDSTPEV	412
				*****;***;***** **** *****;***** * ;*	
AAV-5		(NC_006152.1)	SEQ ID NO: 269	IIVTSNTNMCVVDGNSTTFEHQQPLEDRMFFKFTLTKLPDFDGKTIKQEVKDFFAWAKVN	472
AAV-2_Rep68		(NC_001401.2)	SEQ ID NO: 265	IIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFFKFTLRLLDDHDFGKVTKQEVKDFFRWAKDH	476
AAV-2_Rep78		(NC_001401.2)	SEQ ID NO: 266	IIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFFKFTLRLLDDHDFGKVTKQEVKDFFRWAKDH	476
AAV-2_Rep40		(NC_001401.2)	SEQ ID NO: 267	IIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFFKFTLRLLDDHDFGKVTKQEVKDFFRWAKDH	252
AAV-2_Rep52		(NC_001401.2)	SEQ ID NO: 268	IIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFFKFTLRLLDDHDFGKVTKQEVKDFFRWAKDH	252
AAV-1		(NC_002077.1)	SEQ ID NO: 270	IIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFFKFTLRLLDDHDFGKVTKQEVKDFFRWAKDH	476
AAV-3		(NC_001729.1)	SEQ ID NO: 271	IIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFFKFTLRLLDDHDFGKVTKQEVKDFFRWAKDH	476
AAV-4		(NC_001829.1)	SEQ ID NO: 272	IIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFFKFTLRLLDDHDFGKVTKQEVKDFFRWAKDH	476
AAV-7		(NC_006260.1)	SEQ ID NO: 273	IIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFFKFTLRLLDDHDFGKVTKQEVKDFFRWAKDH	476
AAV-8		(NC_006261.1)	SEQ ID NO: 274	IIVTSNTNMCAVIDGNSTTFEHQQPLQDRMFFKFTLRLLDDHDFGKVTKQEVKDFFRWAKDH	478
Avian_AAV_DA-1		(NC_006263.1)	SEQ ID NO: 275	IIVTSNTNMCYVIDGNITTFEHKKQPLEDRMFLELLTLRPDDDFGKVTKQEVROFFRWSDH	476
Avian_AAV_VR-865		(NC_004828.1)	SEQ ID NO: 276	IIVTSNTNMCYVIDGNITTFEHKKQPLEDRMFLELLTLRPDDDFGKVTKQEVROFFRWSDH	476
Bat		(NC_014468.1)	SEQ ID NO: 278	IIVTSNTNMCEVVDGNSTTFEHRQPLEDRMFFKFTLVRLQPTFGKTIKQEVREFFTKWALIN	469
Bovine		(NC_005889.1)	SEQ ID NO: 277	IIVTSNTNMCVVDGNSTTFEHQQPLEDRMFFKFTLMRRLPDFDGKTIKQEVKDFFAWAKVN	472
				*;***** **;*****;*****;****;* ** *****;*****;*** *	

FIG. 8 cont.

AAV-5	(NC_006152.1)	SEQ ID NO: 269	QVPVTHEFKVPBELAGTCKAEKSLKPLGDVNTVTSYKSLKRRARLSFVPETPRSSDVTVD	532
AAV-2_Rep68	(NC_001401.2)	SEQ ID NO: 265	VVEVEHEFYVKKGG-----A--KRRPAPSD--ADISEPKRVRESV-----AQPSTS	518
AAV-2_Rep78	(NC_001401.2)	SEQ ID NO: 266	VVEVEHEFYVKKGG-----A--KRRPAPSD--ADISEPKRVRESV-----AQPSTS	518
AAV-2_Rep40	(NC_001401.2)	SEQ ID NO: 267	VVEVEHEFYVKKGG-----A--KRRPAPSD--ADISEPKRVRESV-----AQPSTS	294
AAV-2_Rep52	(NC_001401.2)	SEQ ID NO: 268	VVEVEHEFYVKKGG-----A--KRRPAPSD--ADISEPKRVRESV-----AQPSTS	294
AAV-1	(NC_002077.1)	SEQ ID NO: 270	VTEVAHEFYVKKGG-----A--NKRPAADD--ADKSEPKRACPSV-----ADPSTS	518
AAV-3	(NC_001729.1)	SEQ ID NO: 271	VTDVAHEFYVKKGG-----A--KRRPAPSD--ADKSEPKRACPSV-----ADPSTS	518
AAV-4	(NC_001829.1)	SEQ ID NO: 272	VTEVTHEFYVKKGG-----A--KRRPAPSD--ADKSEPKRACPSV-----ADPSTS	518
AAV-7	(NC_006260.1)	SEQ ID NO: 273	VTEVAHEFYVKKGG-----A--SKRPAPDD--ADISEPKRACPSV-----ADPSTS	518
AAV-8	(NC_006261.1)	SEQ ID NO: 274	VTEVAHEFYVKKGG-----A--SKRPAPDD--ADKSEPKRACPSV-----ADPSTS	520
Avian_AAV_DA-1	(NC_006263.1)	SEQ ID NO: 275	LTEVPEFLVRKAE-----SRKRPAPSG--EGYISPTKRPALAEQQ-----	515
Avian_AAV_VR-865	(NC_004828.1)	SEQ ID NO: 276	LTEVPEFLVRKAE-----SRKRPAPSG--EGYISPTKRPALAEQQ-----	515
Bat	(NC_014468.1)	SEQ ID NO: 278	AVDVYDFLVRKINGSDTGG--VVRGAAPT-----KDEPPAKRVTFYGAT--SEGEVREGA	523
Bovine	(NC_005889.1)	SEQ ID NO: 277	QVPVTHEFMPKRVAGTERAETSRRPLDDVNTVTSYKSEPKRARLSVWPETPRSSDVEVE	532
			*, * ; * ; *	
			**	
AAV-5	(NC_006152.1)	SEQ ID NO: 269	PAPLRPLNWSRYDCKDYHAQFDNISNKKODECEYLNRGKNGCICHNVTHQICHGIPPW	592
AAV-2_Rep68	(NC_001401.2)	SEQ ID NO: 265	DAE-ASINYADRLARGHSL-----	536
AAV-2_Rep78	(NC_001401.2)	SEQ ID NO: 266	DAE-ASINYADRYQNKCSRHVGMNMLFPCRCQCEMNQNSNICFTHGQKDCLECFPV-SE	576
AAV-2_Rep40	(NC_001401.2)	SEQ ID NO: 267	DAE-ASINYADRLARGHSL-----	312
AAV-2_Rep52	(NC_001401.2)	SEQ ID NO: 268	DAE-ASINYADRYQNKCSRHVGMNMLFPCRCQCEMNQNSNICFTHGQKDCLECFPV-SE	352
AAV-1	(NC_002077.1)	SEQ ID NO: 270	DAEGAPVDFADRYQNKCSRHAGMLQMLFCKTCERMNQNFNICFTHGTRDCECFPGVSE	578
AAV-3	(NC_001729.1)	SEQ ID NO: 271	DAE-APADYADRYQNKCSRHVGMNMLFCKTCERMNQISNVCFTHGORDCECFPGMSE	577
AAV-4	(NC_001829.1)	SEQ ID NO: 272	DAE-APVDYADRYQNKCSRHVGMNMLFPCRCQCEMNQNVDFICFTHGVMDCAECFPV-SE	576
AAV-7	(NC_006260.1)	SEQ ID NO: 273	DAEGAPVDFADRYQNKCSRHAGMIQMLFCKTCERMNQNFNICFTHGVRDCECFPGVSE	578
AAV-8	(NC_006261.1)	SEQ ID NO: 274	DAEGAPVDFADRYQNKCSRHAGMLQMLFCKTCERMNQNFNICFTHGVRDCECFPGVSE	580
Avian_AAV_DA-1	(NC_006263.1)	SEQ ID NO: 275	QASESAEPVPTYRIKCSRHCGMDRMLFPCQICE SMNRNINICALHKTTECECFPEYGD	575
Avian_AAV_VR-865	(NC_004828.1)	SEQ ID NO: 276	QASESADFVPTYRIKCSRHCGMDRMLFPCQICE SMNRDINICALHKTDCCECFPDYGD	575
Rat	(NC_014468.1)	SEQ ID NO: 278	PGESDSVNFALRYVSKSHLSWSNMRYPCRACRNMADVNVCTPHGCRDCECFPRPAP	583
Bovine	(NC_005889.1)	SEQ ID NO: 277	PAPLRPLNWSRYECRCDYHAKFDSVTGECDECEYLNRGKNGCIFHNATHQICHAVPPW	592

FIG. 8 cont.

AAV-5	(NC_006152.1)	SEQ ID NO: 269	EK-----ENLSD-----	599
AAV-2_Rep68	(NC_001401.2)	SEQ ID NO: 265	-----	536
AAV-2_Rep78	(NC_001401.2)	SEQ ID NO: 266	SQPVSVV-----KKAYQKLCYIHHIMGK	599
AAV-2_Rep40	(NC_001401.2)	SEQ ID NO: 267	-----	312
AAV-2_Rep52	(NC_001401.2)	SEQ ID NO: 268	SQPVSVV-----KKAYQKLCYIHHIMGK	375
AAV-1	(NC_002077.1)	SEQ ID NO: 270	SQPV---VR-----RTYRKLCIAIHHLLGR	600
AAV-3	(NC_001729.1)	SEQ ID NO: 271	SQPVSVK-----KTYQKLCPIHHILGR	601
AAV-4	(NC_001829.1)	SEQ ID NO: 272	SQPVSVR-----KTYQKLCPIHHIMGR	600
AAV-7	(NC_006260.1)	SEQ ID NO: 273	SQPV---VR-----KTYRKLCIAIHHLLGR	600
AAV-8	(NC_006261.1)	SEQ ID NO: 274	SQPV---VR-----KTYRKLCIAIHHLLGR	602
Avian_AAV_DA-1	(NC_006263.1)	SEQ ID NO: 275	KDPVPELPCTEHNVSRCYQCHSGELYRVTSDEKPAPESEDEGTEPSYAPCTIHHIMGK	635
Avian_AAV_VR-865	(NC_004828.1)	SEQ ID NO: 276	KDDV-ELPPCTEHNVSRCYQCHSGELYRVTSDEKPAPESEDEGTEPSYAPCTIHHIMGK	634
Bat	(NC_014468.1)	SEQ ID NO: 278	---VPIAE-----HDLCLAPI-EDS	599
Bovine	(NC_005889.1)	SEQ ID NO: 277	EK-----ENVSD-----	599
AAV-5	(NC_006152.1)	SEQ ID NO: 269	-----FGDFD-----DANKEQ	610
AAV-2_Rep68	(NC_001401.2)	SEQ ID NO: 265	-----	536
AAV-2_Rep78	(NC_001401.2)	SEQ ID NO: 266	VP-D-ACTACDLVNVLDL-----DCIFEQ	621
AAV-2_Rep40	(NC_001401.2)	SEQ ID NO: 267	-----	312
AAV-2_Rep52	(NC_001401.2)	SEQ ID NO: 268	VP-D-ACTACDLVNVLDL-----DCIFEQ	397
AAV-1	(NC_002077.1)	SEQ ID NO: 270	AP-EIACSACDLVNVLDL-----DCVSEQ	623
AAV-3	(NC_001729.1)	SEQ ID NO: 271	AP-EIACSACDLANVLDL-----DCVSEQ	624
AAV-4	(NC_001829.1)	SEQ ID NO: 272	AP-EVACSACELANVLDL-----DCMEQ	623
AAV-7	(NC_006260.1)	SEQ ID NO: 273	AP-EIACSACDLVNVLDL-----DCVSEQ	623
AAV-8	(NC_006261.1)	SEQ ID NO: 274	AP-EIACSACDLVNVLDL-----DCVSEQ	625
Avian_AAV_DA-1	(NC_006263.1)	SEQ ID NO: 275	SRGLVSCAACLKNSLHDDLDGDLQ	663
Avian_AAV_VR-865	(NC_004828.1)	SEQ ID NO: 276	SRGLVTCACRLKNSLHDDLDGDLQ	662
Bat	(NC_014468.1)	SEQ ID NO: 278	DF-YVGC-----ID-----DVNKEQ	613
Bovine	(NC_005889.1)	SEQ ID NO: 277	-----FNFED-----DCNKEQ	610

FIG. 9

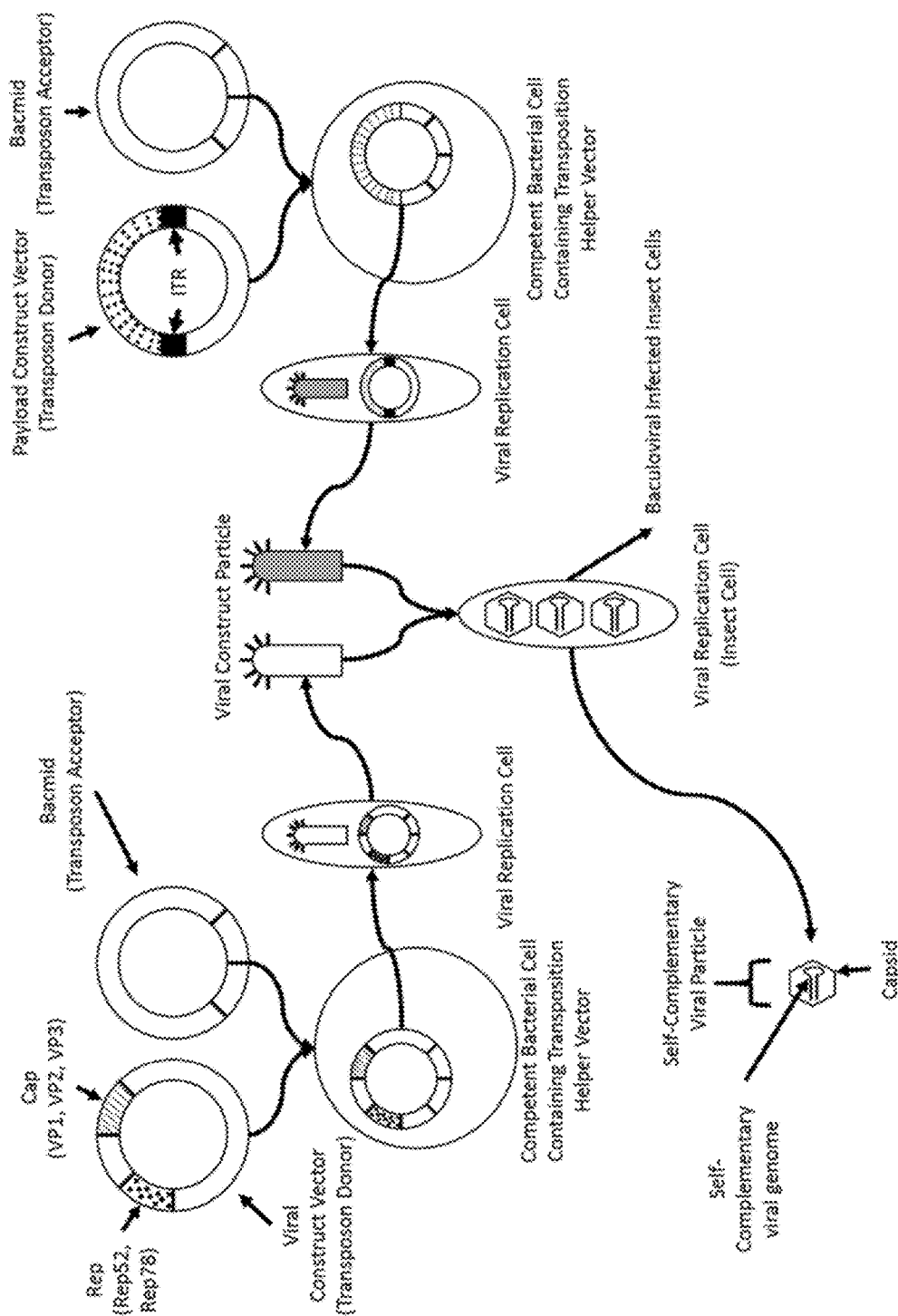


FIG. 10

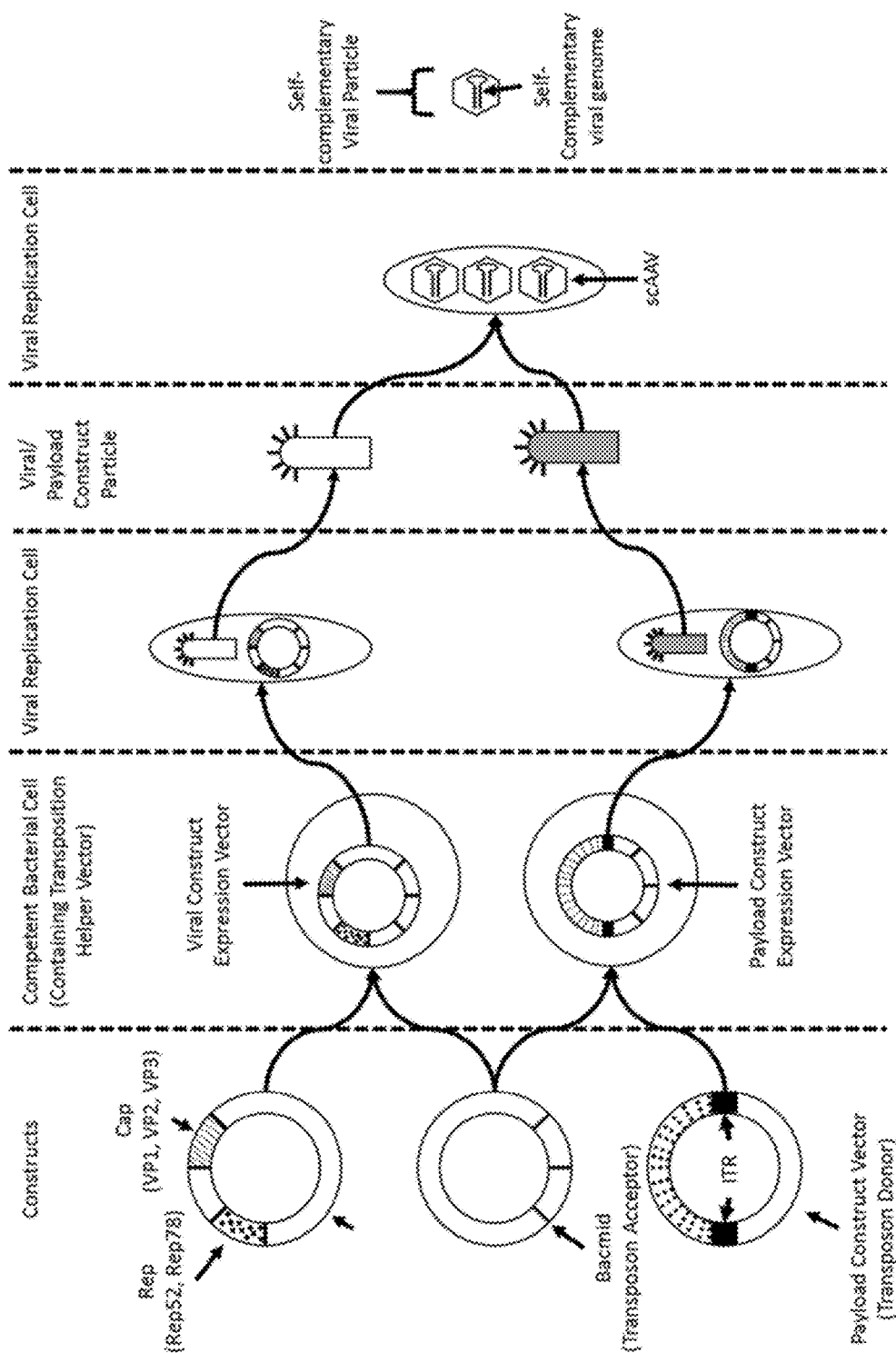


FIG. 11A

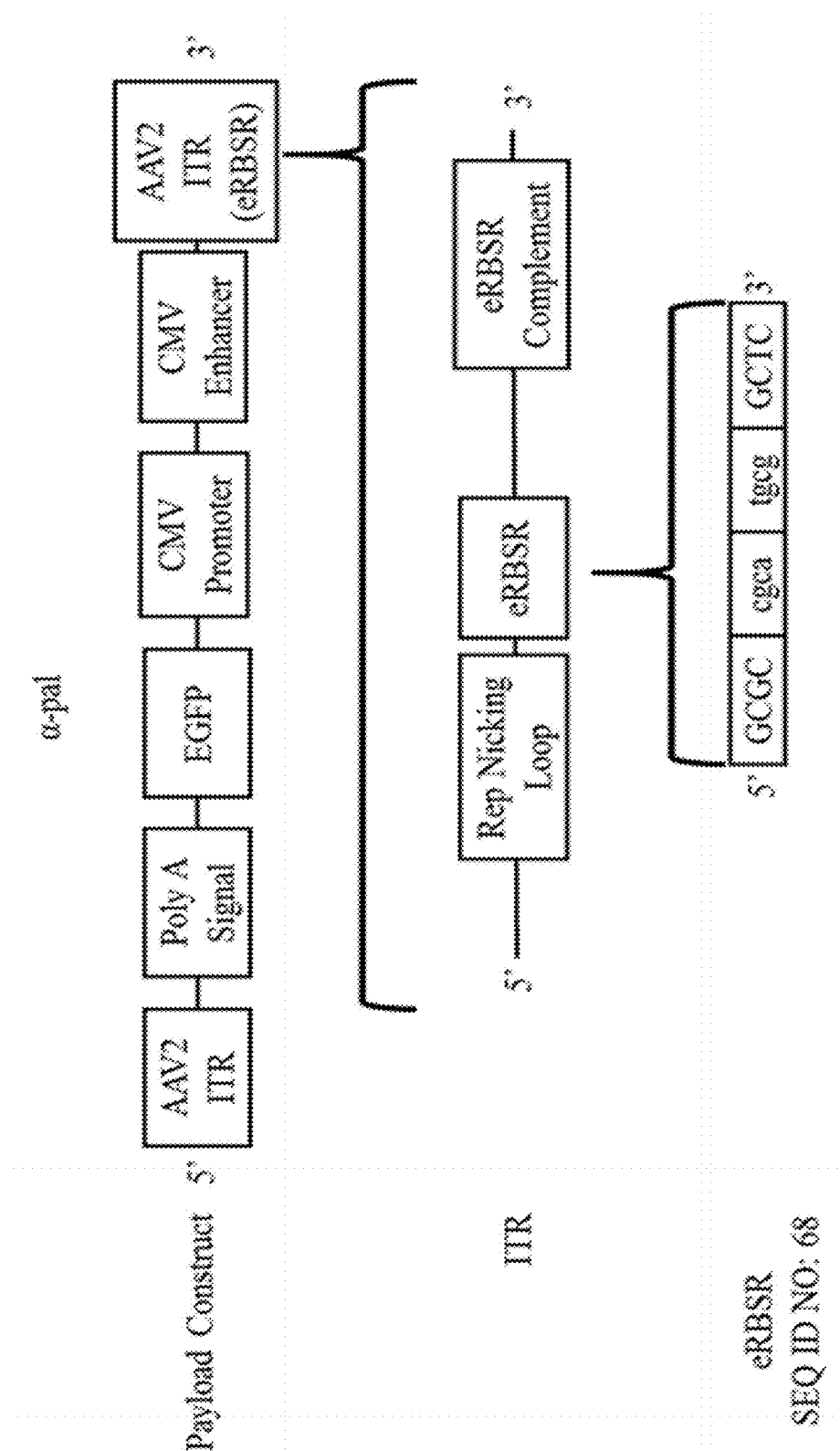


FIG. 11C

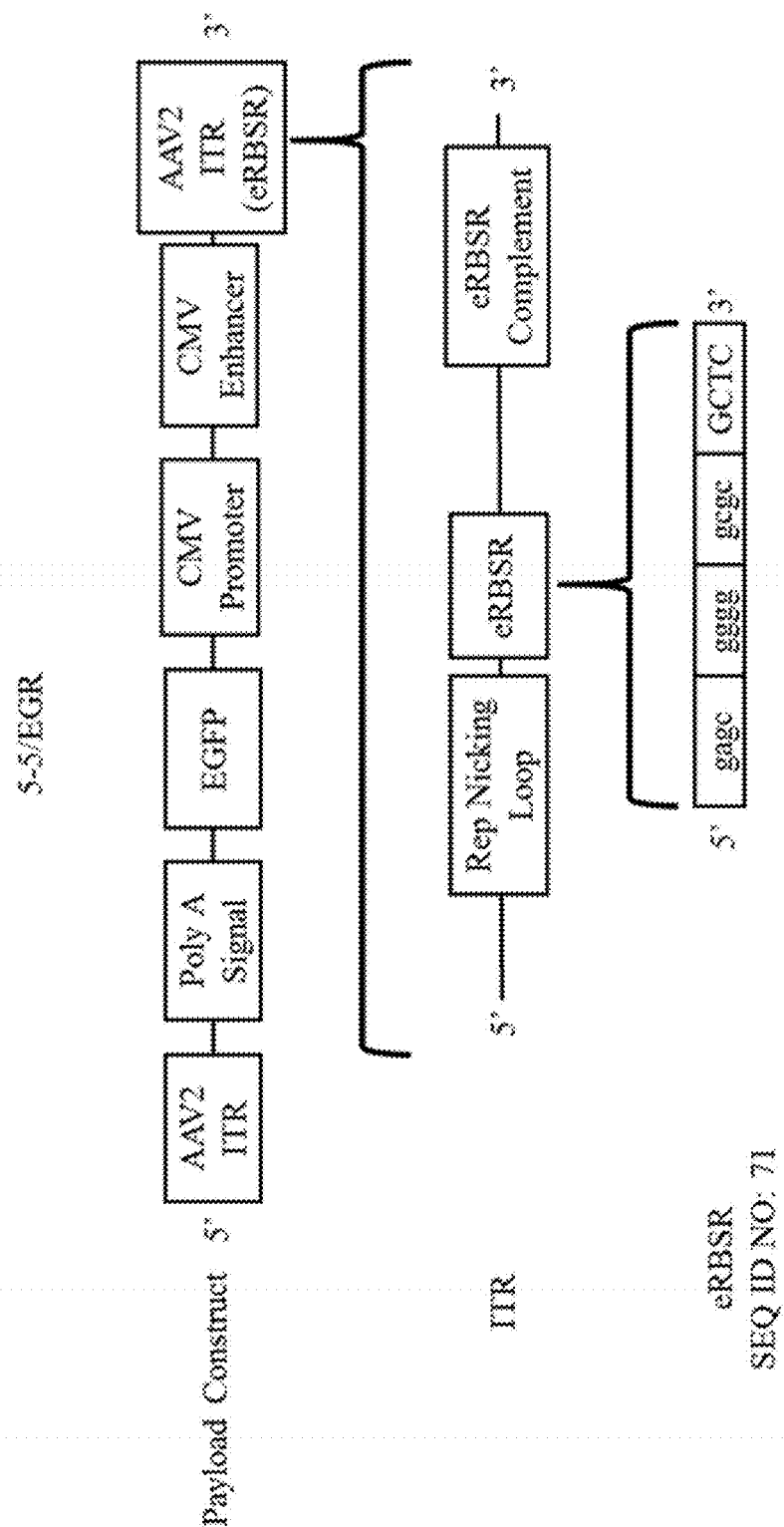


FIG. 11D

5-5/EGR

SEQ ID NO: 72

3' AACCGGTGAGGGAGAGAGctggcccccgcgcgCGAGTCACTCCGG
5' AGGAACCCCTAGTGAGTTGGCCACTCCCTCTCTgagcgggggcgcgcgCGCTCACTGAGGCC

T T
G C
C G
C G
C G
G C
G C
G C
C G
C G
G C
G C
G C
C G
G C
A T
C G
C G
A A

FIG. 11E

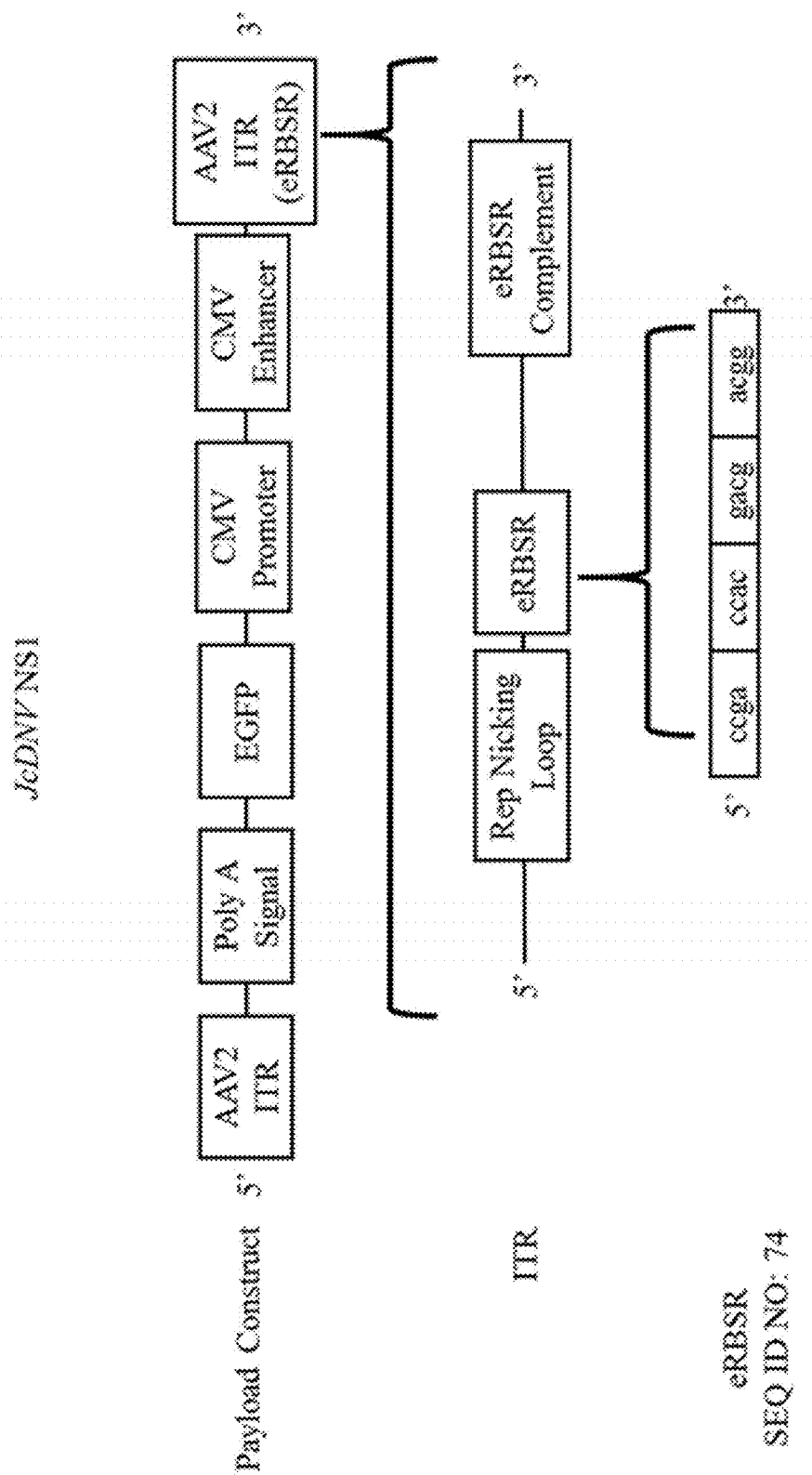


FIG. 11F

jcDNV/NS1

52-0170-75

3' AACCGGTGAGGGAGAGAggctgtgtgctgcctTGACTCCGG
5' AGGAACCCCTAGTGGAGTTGGCCACTCCCTCTCTccgaccacgacggACTGAGGCC

G C C C C G C T C G C A A

A A

COMPOSITIONS AND METHODS FOR THE PRODUCTION OF SCAAV

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 62/091,062, entitled Compositions and Method for the Production of scAAV, filed Dec. 12, 2014, the contents of which are herein incorporated by reference in their entirety.

REFERENCE TO THE SEQUENCE LISTING

[0002] The present application is being filed along with a Sequence Listing in electronic format. The Sequence Listing file, entitled 1500PCTSEQLST.txt, was created on Dec. 10, 2015 and is 241,930 bytes in size. The information in electronic format of the Sequence Listing is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

[0003] The present invention relates to compositions and methods for the production of self-complementary genome-containing viral particles with improved transduction efficiencies and gene expression and therapeutic uses of these scAAV,

BACKGROUND OF THE INVENTION

[0004] Viruses of the Parvoviridae family are small non-enveloped icosahedral capsid viruses characterized by a single stranded DNA genome. Parvoviridae family viruses consist of two subfamilies: Parvoviridae, which infect vertebrates, and Densovirinae, which infect invertebrates.

[0005] Viruses of the Parvoviridae family are used as biological tools due to a relatively simple structure that may be easily manipulated with standard molecular biology techniques. The genome of the virus may be modified to contain a minimum of components for the assembly of a functional recombinant virus, or viral particle, which is loaded with, or engineered to deliver and/or express, a desired payload nucleic acid construct, e.g., a transgene, genome-editing sequence, miR, polypeptide-encoding polynucleotide, or a modulatory nucleic acid.

[0006] Preclinical studies have demonstrated the efficacy of recombinant adeno-associated virus (rAAV) as payload delivery viral particles, and recently rAAVs have been successfully used in clinical trials of gene therapy. Transduction efficiencies generally range from 25 to several hundred viral genome-containing particles (VGP) per transducing unit, depending on the cell type. However, the transduction efficiency of these viral genomes, in terms of the number of VGP required for transduction, is hindered by the need to convert the single-stranded DNA (ssDNA) genome into double-stranded DNA (dsDNA) prior to expression.

[0007] This rate-limiting step can be entirely circumvented through the use of scAAV, which package an inverted repeat genome that can fold into dsDNA without the requirement for DNA synthesis or base-pairing between multiple viral genomes.

[0008] Like all viral particle-based approaches to gene therapy, in addition to the production of large quantities of highly concentrated virus, another obstacle in translating therapies from pre-clinical trials into a human clinical appli-

cation is the efficient transduction and delivery of the therapeutic payload nucleic acid construct.

[0009] In view of these issues there remains a need for alternative and improved methods of efficiently, safely, and economically producing viral particles with improved transduction efficiencies. The present invention provides compositions and methods for the production of self-complementary parvoviral genome-containing particles, e.g. scAAV particles, with improved transduction efficiencies and consequently more efficacious therapeutic outcomes.

SUMMARY OF THE INVENTION

[0010] The present disclosure provides compositions and methods for the production of viral particles with improved transduction efficiencies and increased gene expression useful as therapeutic modalities which employ viral delivery methods. Also provided are engineered nucleic acid and vector constructs for viral vector design and production which are useful components for viral particle production methods. Further provided are modified regulatory proteins engineered to improve viral particle transduction efficiencies and increase gene expression used alone or in conjunction with the engineered nucleic acid and vector constructs. The viral particles of the invention comprise one or more engineered polynucleotides or polypeptides as described herein. In addition, the present disclosure provides methods for producing viral particles which comprise one or more engineered polynucleotides or polypeptides, e.g., engineered genomes and/or engineered ITRs and/or engineered regulatory proteins and/or engineered payloads or any combination thereof.

[0011] The single-stranded nature of the parvoviral genome requires the use of cellular mechanisms to provide a complementary-strand for gene expression, the recruitment of which is considered to be a rate-limiting factor in the efficiency of transduction and gene expression in parvoviruses and parvoviral particles. This problem can be circumvented by packaging both strands as a single duplex DNA molecule. Parvoviral particles comprising such duplexed DNA have been shown to have increased transduction efficiency and a higher level of transgene expression than their single-stranded counterparts.

[0012] The present disclosure provides parvoviral particles having a duplexed genome resulting from the presence of a self-complementary viral genome sequence which may be packaged in a viral particle. The present disclosure thus provides a parvoviral particle comprising a parvovirus capsid and a parvoviral genome encoding a heterologous nucleotide sequence, e.g., a payload sequence, in which the parvoviral genome may be self-complementary, i.e., forms a dimeric inverted repeat via intra-strand base-pairing. In this manner, a double-stranded sequence may be formed by the base-pairing between complementary heterologous nucleotide sequences, thus producing the requisite structure for gene expression in a host cell without the need for host cell machinery to convert the viral genome into a double-stranded form.

[0013] Such a parvoviral genome may be formed by employing the use of one or more parvoviral genome sequences that are altered in a manner that results in the formation of a self-complementary viral genome during replication.

[0014] In one embodiment, the altered parvoviral genome sequence comprises at least one parvoviral inverted terminal

repeat (ITR) sequence having an engineered. Rep binding sequence region (eRBSR). In this embodiment, the Rep binding sequence region may be altered such that binding with the Rep protein is decreased. The decrease in Rep binding affects the Rep nicking activity in a manner that promotes the formation of a self-complementary viral genome during replication.

[0015] In another embodiment, the altered parvoviral genome sequence comprises at least one parvoviral inverted terminal repeat sequence having a chimeric nicking stem loop, which abolishes Rep mediated nicking such that a self-complementary viral genome is formed during replication.

[0016] In another embodiment, the Rep protein, or polynucleotide sequence encoding the Rep protein, is engineered such that Rep protein binding to its cognate Rep binding sequence region is decreased in a manner that likewise promotes the formation of a self-complementary viral genome during replication.

[0017] In one embodiment, a parvoviral particle comprising a self-complementary viral genome is provided. The parvoviral particle may comprise a parvoviral capsid and a self-complementary parvoviral genome comprising: a first parvoviral inverted terminal repeat sequence, a first heterologous sequence, a parvoviral inverted terminal repeat nucleotide sequence comprising an engineered Rep binding sequence region, a second heterologous sequence, wherein the second heterologous sequence is complementary to the first heterologous sequence, and a second parvoviral inverted terminal repeat sequence. The viral genome may be capable of intra-strand base-pairing between the heterologous nucleotide sequences.

[0018] In another embodiment, the parvoviral particle comprises a parvoviral capsid and a self-complementary parvoviral genome comprising: a first parvoviral inverted terminal repeat sequence, a first heterologous sequence, a parvoviral inverted terminal repeat nucleotide sequence comprising a chimeric nicking stem loop, a second heterologous sequence, wherein the second heterologous sequence is complementary to the first heterologous sequence, and a second parvoviral inverted terminal repeat sequence. The viral genome may be capable of intra-strand base-pairing between the heterologous nucleotide sequences.

[0019] In another embodiment, the parvoviral particle comprises a parvoviral capsid and a self-complementary parvoviral genome comprising: a first parvoviral inverted terminal repeat sequence, a first heterologous sequence, a parvoviral inverted terminal repeat nucleotide sequence comprising an engineered Rep binding sequence region and a chimeric nicking stem loop, a second heterologous sequence, wherein the second heterologous sequence is complementary to the first heterologous sequence, and a second parvoviral inverted terminal repeat sequence. The viral genome may be capable of intra-strand base-pairing between the heterologous nucleotide sequences.

[0020] In any of the embodiments where the parvoviral particle comprises an engineered Rep binding sequence region and/or a chimeric nicking stem loop, production of the parvoviral particle can optionally further comprise a sequence encoding an engineered Rep protein which binds to the Rep binding sequence region and/or engineered Rep binding sequence region with decreased affinity.

[0021] The present disclosure additionally provides compositions and methods for improving the transduction efficiency and heterologous gene expression of viral particles having parvoviral polynucleotide sequences (e.g., a parvoviral genome encoding heterologous nucleotide sequence or a payload sequence or a fragment thereof) by altering one or more of several Rep binding activities to effect the production of viral particles having a self-complementary or duplexed genome. As a non-limiting example, the transduction efficiency and heterologous gene expression of viral particles having parvoviral polynucleotide sequences may be increased by altering the Rep binding sequence region and/or by altering the Rep nicking sequence found in parvoviral ITR sequences, and/or altering one or more Rep proteins. Thus, the present disclosure provides, among other things, a parvoviral polynucleotide, such as an adeno-associated viral (AAV) polynucleotide, comprising an engineered Rep binding sequence region and/or an altered Rep nicking sequence. In addition, the present disclosure provides an engineered Rep protein, and a polynucleotide sequence encoding the engineered Rep protein, which can be used independently in the production of a viral particle having a self-complementary or duplexed genome (i.e., with wild-type Rep binding sequence region and wild-type Rep nicking stem loop) or can be used in conjunction with an engineered Rep binding sequence region and/or an altered Rep nicking sequence in the production of a viral particle having a self-complementary or duplexed genome.

[0022] The present disclosure provides a polynucleotide (e.g., a parvoviral polynucleotide) comprising an engineered Rep binding sequence region (eRBSR). In one embodiment, the eRBSR comprises a sequence in which one or more of the nucleotides are altered or modified as compared to the native or reference Rep binding sequence region. In one embodiment, the eRBSR is altered as a result of one or more nucleotide insertions, deletions, substitutions, or any combination thereof as compared to the native Rep binding sequence region. In another embodiment, the eRBSR is altered by swapping the native Rep binding sequence region with the Rep binding sequence region of a different parvoviral serotype or species. In one embodiment, the entire Rep binding sequence region is swapped. In another embodiment, a portion or partial sequence of the Rep binding sequence region is swapped. In one embodiment, the engineered Rep binding sequence region can comprise a non-native Rep binding sequence region from an AAV viral genome, including, but not limited to, the Rep binding sequence regions of AAV1, AAV2, AAV3, AAV4, AAV5, AAV 5, AAV7, AAV8, AAV9, AAV10, AAV 11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8.

[0023] In one embodiment, one or more of the nucleotides in the eRBSR are chemically modified as compared with the native Rep binding sequence region. In individual embodiments, the chemical modification(s) is on the nucleobase and/or the sugar and/or the phosphate backbone, or any combination thereof. In another embodiment, the eRBSR has one or more altered nucleotides and one or more chemically-modified nucleotides. In one embodiment, the Rep binding to the sequence is decreased.

[0024] Studies have shown that Rep protein recognizes and binds to a series of about four to five GCTC consensus motifs. Thus, decreased binding with Rep protein can be affected by altering the number and/or position of the GCTC motifs present in the Rep binding sequence region and/or by

altering the sequence of the individual GCTC motif(s). Thus, in one embodiment, the eRBSR comprises one to three GCTC consensus motifs. In another embodiment, the eRBSR comprises about one to about five GCTC consensus motifs in which one or more nucleotides in at least one of the GCTC consensus motifs is altered or modified.

[0025] The present disclosure also provides polynucleotides comprising any of the eRBSRs described herein. In one embodiment, the polynucleotide comprises one or more parvoviral inverted terminal repeat (ITR) sequences wherein one of the ITR sequences comprises an eRBSR. In one embodiment, the polynucleotide comprising at least one parvoviral ITR sequence with an eRBSR is capable of forming a double-stranded Rep binding sequence region. In one embodiment, the engineered Rep binding sequence region is located between a 5' parvoviral inverted terminal repeat sequence and a 3' parvoviral inverted terminal repeat sequence. In another embodiment, the engineered Rep binding sequence region is located within a parvoviral inverted terminal repeat sequence.

[0026] In one embodiment, the polynucleotides described herein comprising the eRBSR can further comprise a heterologous or payload sequence, which can be, for example, a regulatory sequence, a sequence encoding a nucleic acid (e.g., such as a miR or an antisense sequence), or a sequence encoding a polypeptide (e.g., such as a transgene). In one embodiment, the heterologous sequence may be operatively linked to one or more regulatory sequence(s) that allows or controls expression of the heterologous sequence. In one embodiment, the polynucleotide comprises a first ITR sequence comprising an eRBSR and a second ITR sequence, wherein the heterologous sequence is flanked by the first and second ITR sequences.

[0027] In one embodiment, a parvoviral polynucleotide (e.g., a parvoviral genome) comprises in the 5' to 3' direction: (a) a 5' parvoviral inverted terminal repeat sequence; (b) a first payload encoding region; (c) an engineered Rep binding sequence region; (d) a second payload encoding region; and (e) a 3' parvoviral inverted terminal repeat sequence. In this embodiment, the first and second payload encoding regions are essentially self-complementary and may form a hairpin structure that comprises the parvoviral polynucleotide.

[0028] In one embodiment, the parvoviral ITR sequence (s) can be an adeno-associated virus (AAV) ITR or derived from an adeno-associated virus (AAV) ITR. AAV ITR sequences include, but are not limited to, any of those having an AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 ITR sequence or a sequence derived therefrom.

[0029] Another way to promote the generation of self-complementary parvoviral genome sequences is to alter or modify the Rep nicking stem loop such that nicking by the Rep protein is decreased. This alteration or modification can promote the generation of self-complementary parvoviral genome sequences. The present disclosure thus provides a polynucleotide comprising a chimeric nicking stem loop, in which the native nicking stem loop sequence of one parvoviral serotype or species is substituted or swapped with the nicking stem loop sequence of a different parvoviral serotype or species. In one embodiment, the entire nicking stem loop sequence is swapped. In another embodiment, a portion or partial sequence of the nicking stem loop sequence is

swapped. In one embodiment, the chimeric nicking stem loop can comprise a non-native nicking stem loop from an AAV viral genome, including the Rep binding sequence region of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV 10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8.

[0030] The present disclosure also provides a polynucleotide comprising the chimeric nicking stem loop. The polynucleotide comprising the chimeric nicking stem loop can further comprise a heterologous or payload sequence, which can be, for example, a regulatory sequence, a sequence encoding a nucleic acid, or a sequence encoding a polypeptide. In one embodiment, the heterologous sequence is operatively linked to one or more regulatory sequence(s) that allows expression of the heterologous sequence. In one embodiment, the polynucleotide comprises a first ITR sequence comprising a chimeric nicking stem loop and a second ITR sequence, wherein the heterologous sequence is flanked by the first and second ITR sequences.

[0031] In one embodiment, the disclosure provides a parvoviral polynucleotide (e.g., a parvoviral genome) comprising in the 5' to 3' direction: (a) a 5' parvoviral inverted terminal repeat sequence; (h) a first payload encoding region; (c) a chimeric nicking stem loop sequence; (d) a second payload encoding region; and (e) a 3' parvoviral inverted terminal repeat sequence. In this embodiment, the first and second payload encoding regions are essentially self-complementary and may form a hairpin structure that comprises the parvoviral polynucleotide.

[0032] The present disclosure also provides a polypeptide comprising an engineered Rep protein and a polynucleotide encoding the polypeptide comprising the engineered Rep protein. The engineered Rep protein comprises at least one amino acid that may be altered or modified as compared to the corresponding wild-type Rep protein. In one embodiment, the engineered Rep protein comprises one or more amino acids involved with DNA binding that are altered or modified as compared to the corresponding wild-type Rep protein. The engineered Rep protein has decreased binding to its cognate Rep binding sequence region and, as a consequence, promotes the formation of a self-complementary viral genome during replication.

[0033] The engineered Rep binding sequence region, chimeric nicking stem loop, and engineered Rep protein are each altered and/or modified, and used individually or in combination, such that binding to and/or nicking of the parvoviral ITR sequence is decreased. Thus, the present disclosure provides a parvoviral polynucleotide (e.g., a parvoviral genome) comprising an engineered Rep binding sequence region in which the binding affinity between a Rep protein and the engineered Rep binding sequence region is decreased relative to a parvoviral polynucleotide (e.g., a parvoviral genome) comprising the native Rep binding sequence region. The present disclosure provides a parvoviral polynucleotide (e.g., a parvoviral genome) comprising an engineered Rep binding sequence region in which the binding affinity between an engineered Rep protein and the engineered Rep binding sequence region is decreased relative to a parvoviral polynucleotide (e.g., a parvoviral genome) comprising the native Rep binding sequence region.

[0034] The present disclosure also provides a parvoviral polynucleotide (e.g., a parvoviral genome) encoding an engineered Rep protein in which the binding affinity

between the engineered Rep protein and a Rep binding sequence region is decreased relative to a parvoviral polynucleotide (a parvoviral genome) encoding the native Rep protein. The present disclosure additionally provides a parvoviral polynucleotide (e.g., a parvoviral genome) encoding an engineered Rep protein in which the binding affinity between the engineered Rep protein and an engineered Rep binding sequence region is decreased relative to a parvoviral polynucleotide (e.g., a parvoviral genome) encoding the native Rep protein.

[0035] The present disclosure further provides a parvoviral polynucleotide (e.g., a parvoviral genome) comprising an engineered Rep binding sequence region and/or a chimeric nicking stem loop in which the nicking by a Rep protein is decreased relative to a parvoviral polynucleotide (e.g., a parvoviral genome) comprising the native Rep binding sequence region and/or native Rep nicking stem loop. The present disclosure also provides a parvoviral polynucleotide (e.g., a parvoviral genome) comprising an engineered Rep binding sequence region and/or chimeric nicking stem loop in which the nicking by an engineered Rep protein is decreased relative to a parvoviral polynucleotide (e.g., a parvoviral genome) comprising the native Rep binding sequence region and/or native Rep nicking stem loop.

[0036] The present disclosure also provides a parvoviral polynucleotide (e.g., a parvoviral genome) encoding an engineered Rep protein in which the nicking of the native Rep nicking stem loop is decreased relative to a parvoviral polynucleotide (e.g., a parvoviral genome) encoding the native Rep protein. The present disclosure additionally provides a parvoviral polynucleotide (e.g., a parvoviral genome) encoding an engineered Rep protein in which the nicking of the chimeric nicking stem loop Rep is decreased relative to a parvoviral polynucleotide (e.g., a parvoviral genome) encoding the native Rep protein.

[0037] The present disclosure also provides a parvoviral polynucleotide (e.g., a parvoviral genome) comprising an engineered Rep binding sequence region and/or a chimeric nicking stem loop in which a self-complementary (SC) parvoviral polynucleotide (e.g., a parvoviral genome) is produced during genome replication. The present disclosure also provides a parvoviral polynucleotide (e.g., a parvoviral genome) encoding an engineered Rep protein in which a self-complementary (SC) parvoviral polynucleotide (e.g., a parvoviral genome) is produced during genome replication.

[0038] The present disclosure provides a method of delivering to a cell or tissues a parvoviral particle comprising a payload, which method comprises contacting the cell or tissue with the parvoviral particle. The method of delivering the parvoviral particle to a cell or tissue can be accomplished using in vitro, ex vivo, or in vivo methods.

[0039] The present disclosure additionally provides a method of delivering to a subject, including a mammalian subject, parvoviral particle described herein, which method comprises administering to the subject the parvoviral particle.

[0040] The present disclosure additionally provides a method for treating a disease, disorder and/or condition in a mammalian subject, including a human subject, comprising administering to the subject parvoviral payload particle described herein. In one embodiment, the disease, disorder and/or condition is a neurological disease, disorder and/or condition.

BRIEF DESCRIPTION OF THE FIGURES

[0041] The foregoing and other objects, features and advantages will be apparent from the following description of particular embodiments of the invention, as illustrated in the accompanying drawings. The drawings are not necessarily to scale, emphasis instead being placed upon illustrating the principles of various embodiments of the invention.

[0042] FIG. 1 is a schematic showing the structure of the self-complementary parvoviral genome that is produced by a parvoviral payload construct comprising an engineered Rep binding sequence region (eRBSR).

[0043] FIG. 2 is a schematic showing the structure of the 3' ITR from AAV2 (SEQ ID NO: 242). The strands of the double stranded Rep binding sequence region are underlined (SEQ ID NO: 1 and SEQ ID NO: 2), the terminal resolution site is bolded; and the nicking stem loop (SEQ ID NO: 199) is double-underlined.

[0044] FIG. 3 shows an alignment of 5' and 3' ITR sequences from selected AAV sequences aligned to AAV2 ITR sequence. Rep binding sequence regions are underlined and the nicking stem loop sequences are boxed.

[0045] FIG. 4 shows an alignment of 5' and 3' ITR sequences from selected AAV sequences aligned to AAV5 ITR sequence, Rep binding sequence regions are underlined and the nicking stem loop sequences are boxed.

[0046] FIG. 5 is a PRIOR ART schematic depicting an example of parvovirus (AAV) replication in which single-stranded genomes are derived and showing the various unprocessed replicative intermediates that produce self-complementary genomes. The figure is from N. L. Craig et al. *Mobile DNA II*. ASM Press, Washington, D.C. 2002, the contents of which are herein incorporated by reference in their entirety.

[0047] FIG. 6A is a schematic diagram of a polynucleotide having a swapped nicking stem loop payload construct (SEQ ID NO: 76). FIG. 6B is a schematic diagram comprising the AAV5 Rep nicking stem loop sequence (SEQ ID NO: 77) and the AAV2 Rep nicking stem loop sequence (SEQ ID NO: 312) inserted into an AAV2 3' ITR sequence (SEQ ID NO: 78).

[0048] FIG. 7 is a schematic depicting a 3-dimensional model showing Rep protein binding to the Rep binding domain of the AAV5 ITR. Arrows indicate amino acid and nucleic acid residues involved in binding affinity.

[0049] FIG. 8 shows an alignment of selected AAV Rep protein sequences aligned to AAV5 Rep protein. Beta sheet 4 and Beta sheet 5 sequence regions are underlined; alpha helix C sequences are bold, residues involved in binding to the Rep binding sequence region are boxed.

[0050] FIG. 9 is a diagram depicting the steps involved in scAAV viral particle production.

[0051] FIG. 10 is a diagram depicting the steps involved in scAAV viral particle production divided by the stages of production.

[0052] FIG. 11A-FIG. 11F are schematic diagrams of polynucleotides comprising an engineered Rep binding sequence region and in which the eRBSR sequences are underlined. FIG. 11A is a schematic diagram comprising an α -pal payload construct (SEQ ID NO: 67). FIG. 11B is a schematic diagram comprising α -pal sequence (SEQ ID NO: 68) inserted into the Rep binding sequence region of an AAV2. ITR sequence (SEQ ID NO: 69). FIG. 11C is a schematic diagram comprising the 5-5/EGR payload con-

struct (SEQ ID NO: 70). FIG. 11D is a schematic diagram comprising 5-5/EGR sequence (SEQ ID NO: 71) inserted into the Rep binding sequence region of an AAV2. ITR sequence (SEQ ID NO: 72). FIG. 11E is a schematic diagram comprising the JcDNV NS1 payload construct (SEQ ID NO: 73). FIG. 11F is a schematic diagram comprising JcDNV NS1 sequence (SEQ ID NO: 74) inserted into the Rep binding sequence region of an AAV2 ITR sequence (SEQ ID NO: 75).

DETAILED DESCRIPTION

[0053] Like all parvoviruses, the adeno-associated (AAV) genome is packaged as a linear ssDNA molecule with palindromic inverted terminal repeat (ITR) sequences forming dsDNA hairpin structures at each end. The wild-type AAV2 genome encodes four replication proteins from a single Rep gene open reading frame, three capsid proteins from a single Cap gene open reading frame and a viral assembly protein (APP). The self-complementary ITR sequences serve as priming sites for host-cell DNA polymerase to begin synthesis of the second (complementary) strand and as replication origins during productive infection. Thus, the wild-type replication scheme of AAV requires de novo synthesis of the complementary DNA strand before subsequent steps of genome replication and genome transcription can begin. The second strand synthesis is considered to be one of several blocks to efficient infection.

[0054] Transduction of a target cell by wild-type AAV is dependent on a stepwise series of events including, but not limited to, cell surface binding, endocytic uptake, endosomal escape, nuclear entry, capsid uncoating, release of the ssDNA genome, second strand synthesis, transcription of the genome, and replication of the genome (Murlidharan et al. 2014, *Frontiers in Molecular Neuroscience*, 2014 Sep. 19; 7:76; the contents of which are herein incorporated by reference in their entirety). In addition to the few gene products of the AAV genome, productive wild-type AAV infection is dependent on the presence of adenovirus co-infection which supplies helper genes E1a, E1b, E2a, E4, and VA. These helper genes provide transactivation activity, aid in transcription of the AAV genome, and facilitate AAV mRNA processing.

[0055] Replication of wild-type ssDNA AAV genomes begins at the 3' end of the genome where the self-complementary portion of the ITR forms a double stranded region that acts as a primer for the initiation of DNA synthesis. Replication continues in the 5' to 3' direction until a double stranded hairpin structure has been formed with a loop at one end. Rep protein binds to the Rep binding sequence region near the end of the hairpin comprising the loop and nicks the DNA, allowing the resolution of the loop and formation of a linear double stranded DNA molecule. The complementary strands are then separated and the replication cycle continues on both resultant strands of ssDNA.

[0056] Self-complementary or scAAV is formed in the event that Rep protein does not nick the DNA. Replication of the genome may still proceed by an alternative pathway wherein the complementary strands of the hairpin are separated, forming a replication species that comprises an ITR on either end and an ITR region in the center of the ssDNA strand. An AAV genome comprising scAAV is efficiently packaged within a viral capsid and transduces target cells wherein the double stranded genome that is delivered does

not require the rate limiting second strand synthesis step before expression of the payload.

[0057] Recombinant adeno-associated virus (rAAV) comprises the minimal number of components to produce a non-replicative virus designed to deliver a payload to a target cell. The genome of the rAAV is comprised of ITRs flanking a payload sequence that replaces the wild-type Rep and Cap genes. Genes provided in trans for AAV replication comprise Rep and Cap genes expressing the three capsid proteins VP1, VP2, and VP3 and the non-structural protein, Rep78. The ITR sequences are the only required cis acting element required for AAV replication.

I. SELF-COMPLEMENTARY ADENO-ASSOCIATED VIRUS (SCAAV)

[0058] The present disclosure provides compositions and methods for the production of viral particles (e.g., parvoviral particles) with improved transduction efficiencies and gene expression.

[0059] The single-stranded nature of the parvoviral genome requires the use of cellular mechanisms to provide a complementary strand for gene expression. Recruitment of cellular factors and second strand synthesis are considered to be the rate-limiting factors in the efficiency of transduction and gene expression in parvoviral particles. This problem can be circumvented by packaging both strands as a single duplex DNA molecule.

[0060] The scAAV duplex molecule of the invention is produced by engineering certain components of the AAV genome to promote the generation of scAAV during the replication process. In one non-limiting example of the invention, the payload construct polynucleotide sequences encoding an ITR are altered to comprise an engineered Rep binding sequence region (eRBSR) that promotes production of self-complementary genomes during genome replication (FIG. 1). The final self-complementary ssDNA genome comprises in the 5' to 3' direction an ITR, a first complementary payload sequence, an ITR comprising an engineered Rep binding sequence region, a second complementary payload sequence, and an ITR. The ssDNA spontaneously folds to form a double stranded hairpin comprising an engineered Rep binding sequence region forming a closed loop at one end of the molecule, a double stranded DNA payload encoding region, and at least two ITR hairpins at the end of the molecule opposite the closed end.

[0061] The present disclosure provides viral particles, e.g., parvoviral particles, comprising a duplexed genome resulting from the presence of a self-complementary viral genomic sequence which is packaged in a single viral particle. In general, the parvovirus particle comprises a parvovirus capsid and a parvoviral genome encoding a heterologous nucleotide sequence, e.g. a payload sequence, in which the parvoviral genome is self-complementary. Such parvoviral particles are produced by employing the use of different parvoviral genome sequences that are altered in a manner that result in the formation of a self-complementary viral genome during replication. For example, a parvoviral particle comprising a parvoviral capsid and a duplexed genome can be produced using a parvoviral inverted terminal repeat nucleotide sequence comprising an engineered Rep binding sequence region which has decreased binding with its cognate Rep protein during viral replication. Similarly, a parvoviral particle comprising a parvoviral capsid

and a duplexed genome can be produced using a parvoviral inverted terminal repeat nucleotide sequence comprising a chimeric nicking stem loop which has decreased nicking activity during viral replication. Parvoviral particles having a duplexed genome can also be produced using a sequence encoding an engineered Rep protein which binds to the Rep binding sequence region with decreased affinity. The parvoviral particles of the invention can further be produced using a combination of these nucleotide sequence components which can function individually or in combination to produce the parvoviral particles having duplexed genome.

[0062] Parvoviral ITRs are palindromic sequences, comprising complementary, symmetrically arranged sequences referred to as “A,” “B,” “C” and regions (FIG. 2). The ITR functions as an origin of replication comprising recognition sites for replication proteins Rep 78 and Rep68. The “D” region of the ITR is an asymmetrical region of the ITR where DNA replication initiates and provides directionality to the nucleic acid replication step.

[0063] Rep78 and Rep68 function in two distinct roles as part of a replication mechanism that is characteristic of Rep proteins from parvoviruses. Rep78 and Rep68 comprise both exonuclease activity and DNA helicase activity. Rep exonuclease activity comprises the steps of binding to a specific site within the ITR and nicking the DNA phosphodiester backbone at a specific location. Either Rep78 and/or Rep68 bind to unique and known sites within the ‘A’ region of the ITR (FIG.2). The Rep binding sequence region comprises a sequence of repeated consensus ‘GCTC’ motifs located on the ‘A’ region of the ITR hairpin. The ‘A’ region of the AAV genome is a dsDNA region formed by base pairing of two complementary ssDNA sequence regions encoded in the ITR. The nicking stem loop region is comprised of a sequence that spans the border of the dsDNA ‘A’ and ssDNA ‘D’ regions that further comprises a specific nicking stem loop of the DNA backbone at the border of the ‘A’ and ‘D’ regions. In a second mode of activity, Rep78 or Rep68 exerts an ATP-dependent helicase activity for unwinding double-stranded DNA. Consequently, Rep78 and Rep68 act to break and unwind the hairpin structures on the end of the parvoviral genome, thereby providing access to replication machinery of the viral replication cell.

[0064] A single ITR may be engineered with Rep binding sites or sequence regions on both strands of the A regions and two symmetrical I regions on each side of the ITR palindrome. Such an engineered construct on a double-stranded circular DNA template allows Rep78 or Rep68 initiated nucleic acid replication that proceeds in both directions. A single ITR is sufficient for AAV replication of a circular vector.

[0065] Although similar to other AAV serotypes in many respects, AAV5 differs from other human and simian AAV serotypes more than other known human and simian serotypes. For example, AAV5 Rep and ITR sequences are unable to efficiently cross-complement corresponding Rep and ITR sequences from AAV2 in mammalian cells. See, e.g., Chiorini et al., *J. Virol.*, 73(5):4293-4298 (1999) and Chiorini et al., *J. Virol.*, 73(2):1309-1.319 (1999) the contents of which are herein incorporated by reference in their entirety. This lack of functional homology in AAV5 Rep and ITR sequences may be due to the relatively significant differences in the nucleotide and amino acid sequences of AAV5 from the corresponding sequences of other AAV serotypes. See, e.g., Bantel-Schaal et al., *J. Virol.*, 73(2):

939-947 (1999) the contents of which are herein incorporated by reference in their entirety.

[0066] To identify the Rep binding sequence regions at the genomic sequence level, the genome sequences of several different AAV serotypes and species including AAV1, AAV2, AAV3, AAV4, AAV5, AAV7, AAV8, Bat AAV, Bovine AAV, Snake AAV, Avian AAV DA-1, and Avian AAV VR865 (SEQ ID NO: 280-291) were obtained from the NCBI Genome database. The 5' and 3' inverted terminal repeat sequences of selected genomes were aligned with the Clustal Omega multiple sequence alignment program using either AAV2 or AAV5 as the parent sequence, as shown in FIG. 3 and FIG. 4, respectively. Additional Rep binding sequence regions found in the ITR sequences of other parvoviral serotypes and species, including AAV8 5' ITR (SEQ ID NO: 253), AAV8 3' ITR (SEQ ID NO: 254), Bat AAV 5' ITR (SEQ ID NO: 261), Bat AAV 3' ITR (SEQ ID NO: 262) are obtained in the same manner using this alignment program.

[0067] Based on the genomic sequence alignments, two groups of Rep binding sequence regions, “AAV2-like” and “AAV5-like,” were identified as regions of interest. FIG. 3 shows the sequences having the AAV2-like Rep binding sequence region and FIG. 4 shows the sequences having the AAV5-like Rep binding sequence region, in which the underlined regions represent the complementary Rep binding sequence in the 5' and 3' ITR regions. The Rep binding sequence regions share a consensus GCTC motif that is repeated four to five times. In total, the identified consensus Rep binding sequence regions are about 16-20 nucleotides in length.

[0068] In one embodiment, the altered parvoviral genome sequence comprises an engineered Rep binding sequence region. In this embodiment, the Rep binding sequence region is altered such that binding with its cognate Rep protein is decreased. Decreasing Rep binding to the Rep binding sequence region reduces the ability of the Rep protein exonuclease to successfully nick the DNA at the Rep nicking stem loop. Failure of the Rep protein to successfully nick DNA promotes an alternative pathway of genome replication and formation of a self-complementary viral genome during replication.

[0069] The affinity of Rep protein for a Rep binding sequence region is described as the affinity constant or K_D . The kinetics of binding between a Rep protein and a polynucleotide comprising a Rep binding sequence region is a dynamic process defined by the rate of association (k_a) and the rate of dissociation (k_d). The affinity constant (K_D) is determined by the following formula: $K_D = k_d/k_a$. At equilibrium, the rate of protein-DNA association is equal to the rate of dissociation wherein the measurement of reaction rate constants are used to define an equilibrium constant $1/K_D$. Due to the inverse nature of this constant, a smaller affinity constant represents a greater binding affinity. The affinity constant describing the binding of an engineered Rep binding sequence region and a Rep protein will be larger than the affinity constant describing the binding of a wild-type Rep binding sequence region and a Rep protein.

[0070] Replication of ssDNA AAV genomes begins at the 3' end of the genome where the self-complementary portion of the ITR forms a double stranded region that acts as a primer for the initiation of DNA synthesis. Replication continues in the 5' to 3' direction until a double stranded hairpin structure called the replicative form monomer (RF_M)

has been formed with a loop at one end (FIG. 5). Rep protein binds to the Rep binding sequence region near the end of the hairpin comprising the loop, and nicks the DNA, allowing the polymerase to replicate the loop and form a linear double stranded DNA molecule. The complementary strands are then separated by Rep protein to allow reformation of the ITR complementary loop and the replication cycle continues on both resultant strands of ssDNA.

[0071] In the event that the RF_M species is not nicked by Rep protein, an alternative replication process is used. The 3' end of the molecule reforms the ITR complementary loop and replication continues along the length of the molecule, forming the replicative form dimer duplex (RF_D). The RF_D species may be resolved by two different pathways. In the first pathway, productive nicking of the RF_D species on both strands results in two copies of the RF_M species. In the second pathway, the RF_D species reforms the ITR complementary loop allowing replication and formation of a $2 \times RF_D$ species. Productive nicking of the $2 \times RF_D$ species results in two copies of the RF_D species.

[0072] An AAV genome comprising one engineered Rep binding sequence region with decreased affinity for Rep protein and one wild-type Rep binding sequence region prevents productive nicking. A lack of binding to the engineered Rep binding sequence region reduces nicking and promotes replication of the RF_M species by the alternative pathway. In the resultant RF_D species, the position of the engineered Rep binding sequence region reduces Rep binding and promotes productive replicative cycling of the RF_D species. In the second cycle of RF_D species replication a wild-type Rep binding sequence region is positioned to promote nicking and results in two copies of the RF_D species.

[0073] Viral particles. e.g., AAV particles, produced from viral genomes comprising wild-type ITR regions are observed to contain both single stranded and self-complementary viral genomes. Self-complementary AAV genomes are derived from unprocessed replicative intermediates. The nucleic acid content of recombinant AAV analyzed by analytical ultracentrifugation has shown that although the Raleigh interference curves for DNA species with sedimentation coefficients of 84S and 106S indicate similar particle numbers, the 106S absorbance at 260 nm is 1.6 times greater than the 84S peak. The difference in absorbance peak area indicates greater nucleic acid content in the 106S peak. Thus, approximately half of the viral particles contain a self-complementary genome that is 1.6 times as large as the single stranded genome.

[0074] Intermediate replicative species are successfully packaged in virus particles. AAV viral particles produced in vitro are collected by ultracentrifugation and the DNA genomes isolated by standard molecular biology techniques known in the art. The genomes are separated by denaturing (alkaline) gel electrophoresis. Analysis of the DNA content banding pattern has shown multiple sizes of viral genomes including scAAV, scAAV intermediate species, ssAAV monomer, and an ssAAV repaired from scAAV.

[0075] In accordance with the present disclosure, a parvoviral particle comprising a duplexed genome is provided. In one embodiment, the parvoviral particle comprises a parvoviral capsid and a viral genome comprising: a first parvoviral inverted terminal repeat sequence, a first heterologous sequence, a parvoviral inverted terminal repeat nucleotide sequence comprising an engineered Rep binding

sequence region, a second heterologous sequence, wherein the second heterologous sequence is complementary to the first heterologous sequence, and a second parvoviral inverted terminal repeat sequence. The viral genome is capable of intra-strand base-pairing between the heterologous nucleotide sequences upon release from the parvoviral capsid. The parvoviral particle can optionally further comprise a sequence encoding an engineered Rep protein which binds to the Rep binding sequence region and/or engineered Rep binding sequence region with decreased affinity.

[0076] In another embodiment, the parvoviral particle comprises a parvoviral capsid and a parvoviral genome comprising: a first parvoviral inverted terminal repeat sequence, a first heterologous sequence, a parvoviral inverted terminal repeat nucleotide sequence comprising a chimeric nicking stem loop, a second heterologous sequence, wherein the second heterologous sequence is complementary to the first heterologous sequence, and a second parvoviral inverted terminal repeat sequence. The parvoviral particle can further comprise a sequence encoding an engineered Rep protein which binds to the Rep binding sequence region and/or engineered Rep binding sequence region with decreased affinity.

[0077] The present disclosure additionally provides compositions and methods for improving the transduction efficiency and heterologous gene expression of viral particles having parvoviral polynucleotide sequences by altering one or more of several Rep binding activities to effect the production of viral particles having a self-complementary or duplexed genome, for example, by altering the Rep binding sequence region, and/or by altering one or more Rep proteins, and/or by altering the Rep nicking sequence found in parvoviral ITR sequences. Thus, the present disclosure provides, among other things, a parvoviral polynucleotide, such as an adeno-associated viral (AAV) polynucleotide, comprising an engineered Rep binding sequence region and/or an altered Rep nicking sequence. In addition, the present disclosure provides an engineered Rep protein, and a polynucleotide sequence encoding the engineered Rep protein, which can be used independently in the production of a viral particle having a self-complementary or duplexed genome (i.e., with wild-type Rep binding sequence region and wild-type Rep nicking stem loop) or can be used in conjunction with an engineered Rep binding sequence region and/or an altered Rep nicking sequence in the production of a viral particle having a self-complementary or duplexed genome.

II. ENGINEERED REP BINDING SEQUENCE REGION (ERBSR) AND POLYNUCLEOTIDES

[0078] One way to promote the generation of self-complementary parvoviral genome sequences is to alter or modify the Rep binding sequence region such that Rep binding is decreased, thus affecting the nicking activity of the Rep protein and consequently promoting the generation of self-complementary parvoviral genome sequences. Accordingly, the present disclosure provides a nucleic acid comprising an eRBSR in which one or more of the nucleotides in the engineered Rep binding sequence region are altered or modified as compared to the native or wild-type Rep binding sequence region.

[0079] As used herein, the term "altered nucleotide" or the term "altered" as it is used to describe a nucleotide, refers to a nucleotide that differs from the native nucleotide found at

the same sequence position in a nucleic acid or polynucleotide sequence, including an eRBSR or polynucleotide comprising an eRBSR. The altered nucleotide can contain any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine, as long as it differs from the native nucleoside. Any known method can be used to produce an eRBSR having one or more of the altered nucleotides. In one embodiment, the altered nucleotide is designed in silica and manufactured by methods standard in the art including solid phase synthesis. In one embodiment, the altered nucleotide is designed in silico and manufactured by the phosphoramidite method of solid state synthesis.

[0080] As used herein, the term “modified nucleotide” or the term “modified” as it is used to describe a nucleotide, refers to a nucleotide that has been chemically modified. The present disclosure provides for modified nucleosides and nucleotides. As described herein “nucleoside” is defined as a compound containing a sugar molecule (e.g., a pentose or ribose) or a derivative thereof in combination with an organic base (e.g., a purine or pyrimidine) or a derivative thereof (also referred to herein as “nucleobase”). As described herein, “nucleotide” is defined as a nucleoside including a phosphate group. The modified nucleotides may be synthesized by any useful method, as described herein (e.g., chemically, enzymatically, or recombinantly to include one or more modified or non-natural nucleosides). The chemical modification can be on the nucleobase, and/or the sugar, and/or on the backbone of the nucleotide. The chemical modification can be any chemical modification used to modify nucleic acid.

[0081] The modified nucleotides can be modified on the sugar of the nucleic acid. In some embodiments, chemical modifications include for example, the 2' hydroxyl group (OH) can be modified or replaced with a number of different substituents. Exemplary substitutions at the 2'-position include, but are not limited to, H, halo, optionally substituted C1-6 alkyl; optionally substituted C1-6 alkoxy; optionally substituted C6-10 aryloxy; optionally substituted C3-8 cycloalkyl; optionally substituted C3-8 cycloalkoxy; optionally substituted C6-10 aryloxy; optionally substituted C6-10 aryl-C1-6 alkoxy, optionally substituted C1-12 (heterocycl)oxy, a sugar (e.g., ribose, pentose, or any described herein); a polyethyleneglycol (PEG), $-\text{O}(\text{CH}_2\text{CH}_2\text{O})_n\text{CH}_2\text{CH}_2\text{OR}$, where R is H or optionally substituted alkyl, and n is an integer from 0 to 20 (e.g., from 0 to 4, from 0 to 8, from 0 to 10, from 0 to 16, from 1 to 4, from 1 to 8, from 1 to 10, from 1 to 16, from 1 to 20, from 2 to 4, from 2 to 8, from 2 to 10, from 2 to 16, from 2 to 20, from 4 to 8, from 4 to 10, from 4 to 16, and from 4 to 20); “locked” nucleic acids (LNA) in which the 2'-hydroxyl is connected by a C1-6 alkylene or C1-6 heteroalkylene bridge to the 4'-carbon of the same ribose sugar, where exemplary bridges included methylene, propylene, ether, or amino bridges; aminoalkyl, aminoalkoxy, amino, and amino acid.

[0082] Further examples include replacement of the oxygen in ribose (e.g., with S, Se, or alkylene, such as methylene or ethylene); addition of a double bond (e.g., to replace ribose with cyclopentenyl or cyclohexenyl); ring contraction of ribose (e.g., to form a 4-membered ring of cyclobutane or oxetane); ring expansion of ribose (e.g., to form a 6- or 7-membered ring having an additional carbon or heteroatom, such as for anhydrohexitol, altritol mannitol, cyclohexanyl, cyclohexenyl, and morpholino that also has a

phosphoramidate backbone); multicyclic forms (e.g., tricyclo; and “unlocked” forms, such as glycol nucleic acid (GNA) (e.g., R-GNA or S-GNA, where ribose is replaced by glycol units attached to phosphodiester bonds), threose nucleic acid (TNA, where ribose is replaced with α -L-threofuranosyl-(3'→2')), and peptide nucleic acid (PNA, where 2-amino-ethyl-glycine linkages replace the ribose and phosphodiester backbone) The sugar group can also contain one or more carbons that possess the opposite stereochemical configuration than that of the corresponding carbon in ribose.

[0083] The modified nucleotide base pairing encompasses not only the standard adenosine-thymine, adenosine-uracil, or guanosine-cytosine base pairs, but also base pairs formed between nucleotides and/or modified nucleotides comprising non-standard or modified bases, wherein the arrangement of hydrogen bond donors and hydrogen bond acceptors permits hydrogen bonding between a non-standard base and a standard base or between two complementary non-standard base structures. One example of such non-standard base pairing is the base pairing between the modified nucleotide inosine and adenine, cytosine or uracil.

[0084] The modified nucleotides can be modified on the nucleobase of the nucleic acid. Examples of nucleobases include, but are not limited to, adenine, guanine, cytosine, thymine, and uracil.

[0085] In some embodiments, the modified nucleobase is a modified uracil. Exemplary nucleobases and nucleosides having a modified uracil include pseudouridine (ψ), pyridin-4-one ribonucleoside, 5-aza-uridine, 6-aza-uridine, 2-thio-5-aza-uridine, 2-thio-uridine (s_2U), 4-thio-uridine (s_4U), 4-thio-pseudouridine, 2-thio-pseudouridine, 5-hydroxy-uridine ($hoSU$), 5-aminoallyl-uridine, 5-halo-uridine (e.g., 5-iodo-uridine or 5-bromo-uridine), 3-methyl-uridine (m_3U), 5-methoxy-uridine (mo_5U), uridine 5-oxyacetic acid (cmo_5U), uridine 5-oxyacetic acid methyl ester ($mcmo_5U$), 5-carboxymethyl-uridine (cm_5U), 1-carboxymethyl-pseudouridine, 5-carboxyhydroxymethyl-uridine (chm_5U), 5-carboxyhydroxymethyl-uridine methyl ester ($mchm_5U$), 5-methoxycarbonylmethyl-uridine (mcm_5U), 5-methoxycarbonylmethyl-2-thio-uridine (mcm_5s_2U), 5-aminomethyl-2-thio-uridine (nm_5s_2U), 5-methylaminomethyl-uridine (mnm_5U), 5-methylaminomethyl-2-thio-uridine (mnm_5s_2U), 5-methylaminomethyl-2-seleno-uridine (mnm_5se_2U), 5-carbamoylmethyl-uridine (ncm_5U), 5-carboxymethylaminomethyl-uridine ($cmnm_5U$), 5-carboxymethylaminomethyl-2-thio-uridine ($cmnm_5s_2U$), 5-propynyl-uridine, 1-propynyl-pseudouridine, 5-taurinomethyl-uridine (τm_5U), 1-taurinomethyl-pseudouridine, 5-taurinomethyl-2-thio-uridine (τm_5s_2U), 1-taurinomethyl-4-thio-pseudouridine, 5-methyl-uridine (m_5U , i.e., having the nucleobase deoxythymine), 1-methylpseudouridine ($m_1\psi$), 5-methyl-2-thio-uridine (m_5s_2U), 1-methyl-4-thio-pseudouridine ($m_1s_4\psi$), 4-thio-1-methyl-pseudouridine, 3-methyl-pseudouridine ($m_3\psi$), 2-thio-1-methyl-pseudouridine, 1-methyl-1-deaza-pseudouridine, 2-thio-1-methyl-1-deaza-pseudouridine, dihydrouridine (D), dihydropseudouridine, 5,6-dihydrouridine, 5-methyl-dihydrouridine (m_5D), 2-thio-dihydrouridine, 2-thio-dihydropseudouridine, 2-methoxy-uridine, 2-methoxy-4-thio-uridine, 4-methoxy-pseudouridine, 4-methoxy-2-thio-pseudouridine, N1-methyl-pseudouridine (also known as 1-methylpseudouridine ($m_1\psi$)), 3-(3-amino-3-carboxypropyl)uridine (acp_3U), 1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine (acp_3

ψ), 5-(isopentenylaminomethyl)uridine (inm5U), 5-(isopentenylaminomethyl)-2-thio-uridine (inm5s2U), α -thio-uridine, 2'-O-methyl-uridine (Um), 5,2'-O-dimethyl-uridine (m5Um), 2'-O-methyl-pseudouridine (ψ m), 2-thio-2'-O-methyl-uridine (s2Um), 5-methoxycarbonylmethyl-2'-O-methyl-uridine (mcm5Um), 5-carbamoylmethyl-2'-O-methyl-uridine (ncm5Um), 5-carboxymethylaminomethyl-2'-O-methyl-uridine (cmnm5Um), 3,2'-O-dimethyl-uridine (m3Um), 5-(isopentenylaminomethyl)-2'-O-methyl-uridine (inm5Um), 1-thio-uridine, deoxythymidine, 2'-F-ara-uridine, 2'-F-uridine, 2'-OH-ara-uridine, 5-(2-carbomethoxyvinyl) uridine, and 5-[3-(1-E-propenylamino)uridine.

[0086] In some embodiments, the modified nucleobase is a modified cytosine. Exemplary nucleobases and nucleosides having a modified cytosine include 5-aza-cytidine, 6-aza-cytidine, pseudoisocytidine, 3-methyl-cytidine (m3C), N4-acetyl-cytidine (ac4C), 5-formyl-cytidine (f5C), N4-methyl-cytidine (m4C), 5-methyl-cytidine (m5C), 5-halo-cytidine (e.g., 5-iodo-cytidine), 5-hydroxymethyl-cytidine (hm5C), 1-methyl-pseudoisocytidine, pyrrolo-cytidine, pyrrolo-pseudoisocytidine, 2-thio-cytidine (s2C), 2-thio-5-methyl-cytidine, 4-thio-pseudoisocytidine, 4-thio-1-methyl-pseudoisocytidine, 4-thio-1-methyl-1-deaza-pseudoisocytidine, 1-methyl-1-deaza-pseudoisocytidine, zebularine, 5-aza-zebularine, 5-methyl-zebularine, 5-aza-2-thio-zebularine, 2-thio-zebularine, 2-methoxy-cytidine, 2-methoxy-5-methyl-cytidine, 4-methoxy-pseudoisocytidine, 4-methoxy-1-methyl-pseudoisocytidine, lysidine (k2C), α -thio-cytidine, 2'-O-methyl-cytidine Cm), 5,2'-O-dimethyl-cytidine (m5Cm), N4-acetyl-2'-O-methyl-cytidine ac4Cm), N4,2'-O-dimethyl-cytidine (m4Cm), 5-formyl-2'-O-methyl-cytidine (f5Cm), N4,N4,2'-O-trimethyl-cytidine (m42C), 1-thio-cytidine, 2'-F-ara-cytidine, 2'-F-cytidine, and 2'-OH-ara-cytidine.

[0087] In some embodiments, the modified nucleobase is a modified adenine. Exemplary nucleobases and nucleosides having a modified adenine include 2-amino-purine, 2,6-diaminopurine, 2-amino-6-halo-purine (e.g., 2-amino-6-chloro-purine), 6-halo-purine (e.g., 6-chloro-purine), 2-amino-6-methyl-purine, 8-azido-adenosine, 7-deaza-adenine, 7-deaza-8-aza-adenine, 7-deaza-2-amino-purine, 7-deaza-8-aza-2-amino-purine, 7-deaza-2,6-diaminopurine, 7-deaza-8-aza-2,6-diaminopurine, 1-methyl-adenosine (m1A), 2-methyl-adenine (m2A), N6-methyl-adenosine (m6A), 2-methylthio-N6-methyl-adenosine (ms2m6A), N6-isopentenyl-adenosine (i6A), 2-methylthio-N6-isopentenyl-adenosine (ms2i6A), N6-(cis-hydroxyisopentenyl)adenosine (io6A), 2-methylthio-N6-(cis-hydroxyisopentenyl)adenosine (ms2io6A), N6-glycylcarbamoyl-adenosine (g6A), N6-threonylcarbamoyl-adenosine (t6A), N6-methyl-N6-threonylcarbamoyl-adenosine (m6t6A), 2-methylthio-N6-threonylcarbamoyl-adenosine (ms2g6A), N6,N6-dimethyl-adenosine (m62A), N6-hydroxynorvalylcarbamoyl-adenosine (hn6A), 2-methylthio-N6-hydroxynorvalylcarbamoyl-adenosine (ms2hn6A), N6-acetyl-adenosine (ac6A), 7-methyl-adenine, 2-methylthio-adenine, 2-methoxy-adenine, α -thio-adenosine, 2'-O-methyl-adenosine (Am), N6,2'-O-dimethyl-adenosine (m6Am), N6,N6,2'-O-trimethyl-adenosine (m62Am), 1,2'-O-dimethyl-adenosine (m1Am), 2'-O-ribosyladenosine (phosphate) (Ar(p)), 2-amino-N6-methyl-purine, 1-thio-adenosine, 8-azido-adenosine, 2c-F-ara-adenosine, 2'-F-adenosine, 2'-OH-ara-adenosine, and N6-(19-amino-penta-oxanonadecyl)-adenosine.

[0088] In some embodiments, the modified nucleobase is a modified guanine. Exemplary nucleobases and nucleosides having a modified guanine include inosine (I), 1-methyl-inosine (m1I), wyosine (imG), methylwyosine (mimG), 4-demethyl-wyosine (imG-14), isowyosine (imG2), wybutosine (yW), peroxywybutosine (o2yW), hydroxywybutosine (OHyW), undermodified hydroxywybutosine (OHyW*), 7-deaza-guanosine, queuosine (Q), epoxyqueuosine (oQ), galactosyl-queuosine (galQ), mannosyl-queuosine (manQ), 7-cyano-7-deaza-guanosine (preQ0), 7-aminomethyl-7-deaza-guanosine (preQ1), archaeosine (G+), 7-deaza-8-aza-guanosine, 6-thio-guanosine, 6-thio-7-deaza-guanosine, 6-thio-7-deaza-8-aza-guanosine, 7-methyl-guanosine (m7G), 6-thio-7-methyl-guanosine, 7-methyl-inosine, 6-methoxy-guanosine, 1-methyl-guanosine (m1G), N2-methyl-guanosine (m2G), N2,N2-dimethyl-guanosine (m22G), N2,7-dimethyl-guanosine (m2,7G), N2, N2,7-dimethyl-guanosine (2,2,7G), 8-oxo-guanosine, 7-methyl-8-oxo-guanosine, 1-methyl-6-thio-guanosine, N2-methyl-6-thio-guanosine, N2,N2-dimethyl-6-thio-guanosine, a-thio-guanosine, 2'-O-methyl-guanosine (Gm), N2-methyl-2'-O-methyl-guanosine (m2Gm), N2,N2-dimethyl-2'-O-methyl-guanosine (m22Gm), 1-methyl-2'-O-methyl-guanosine (m1Gm), N2,7-dimethyl-2'-O-methyl-guanosine (m2,7Gm), 2'-O-methyl-inosine (Im), 1,2'-O-dimethyl-inosine (m1Im), and 2'-O-ribosylguanosine (phosphate) (Gr(p)).

[0089] The nucleobase of the nucleotide can be independently selected from a purine, a pyrimidine, or a purine or pyrimidine analog. For example, the nucleobase can be independently selected from adenine, cytosine, guanine, uracil, or hypoxanthine. In another embodiment, the nucleobase can also include, for example, naturally-occurring and synthetic derivatives of a base, including pyrazolo[3,4-d]pyrimidines, 5-methylcytosine (5-me-C), 5-hydroxy methyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo (e.g., 8-bromo), 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 8-azaguanine and 8-azaadenine, deazaguanine, 7-deazaguanine, 3-deazaguanine, deazaadenine, 7-deazaadenine, 3-deazaadenine, pyrazolo[3,4-d]pyrimidine, imidazo[1,5-a]1,3,5 triazinones, 9-deazapurines, imidazo[4,5-d]pyrazines, thiazolo[4,5-d]pyrimidines, pyrazin-2-ones, 1,2,4-triazine, pyridazine; and 1,3,5 triazine. When the nucleotides are depicted using the shorthand A, G, C, T or U, each letter refers to the representative base and/or derivatives thereof, e.g., A includes adenine or adenine analogs, e.g., 7-deaza adenine).

[0090] The modified nucleotides can be modified on the internucleoside linkage (e.g., phosphate backbone). Herein, in the context of the polynucleotide backbone, the phrases "phosphate" and "phosphodiester" are used interchangeably. Backbone phosphate groups can be modified by replacing one or more of the oxygen atoms with a different substituent. Further, the modified nucleosides and nucleotides can include the wholesale replacement of an unmodified phosphate moiety with another internucleoside linkage as described herein. Examples of modified phosphate groups include, but are not limited to, phosphorothioate, phospho-

roselenates, boranophosphates, boranophosphate esters, hydrogen phosphonates, phosphoramidates, phosphorodiamidates, alkyl or aryl phosphonates, and phosphotriesters. Phosphorodithioates have both non-linking oxygens replaced by sulfur. The phosphate linker can also be modified by the replacement of a linking oxygen with nitrogen (bridged phosphoramidates), sulfur (bridged phosphorothioates), and carbon (bridged methylene-phosphonates).

[0091] The (t-thio substituted phosphate moiety is provided to confer stability to RNA and DNA polymers through the unnatural phosphorothioate backbone linkages. Phosphorothioate DNA and RNA have increased nuclease resistance and subsequently a longer half-life in a cellular environment.

[0092] In specific embodiments, a modified nucleoside includes an alpha-thio-nucleoside (e.g., 5'-O-(1-thiophosphate)-adenosine, 5'-O-(1-thiophosphate)-cytidine (α -thio-cytidine), 5'-O-(1-thiophosphate)-guanosine, 5'-O-(1-thiophosphate)-uridine, or 5'-O-(1-thiophosphate)-pseudouridine).

[0093] In some embodiments, a combination of modifications to the sugar, the nucleobase, and/or the internucleoside linkage may be used.

[0094] Methods for chemically modifying nucleic acids are known in the art and include, for example, any available technique including, but not limited to, chemical synthesis, enzymatic synthesis, and enzymatic or chemical cleavage of a longer precursor. Methods of synthesizing nucleotides are known in the art (see, e.g., Gait, M. J. (ed.) *Oligonucleotide synthesis: a practical approach*, Oxford [Oxfordshire], Washington, D.C.: IRL Press, 1984; and Herdewijn, P. (ed.) *Oligonucleotide synthesis: methods and applications*, Methods in Molecular Biology, v. 288 (Clifton, N.J.) Totowa, N.J.: Humana Press, 2005; both of which are incorporated herein by reference in their entirety).

[0095] In one embodiment, the engineered Rep binding sequence region or eRBSR has a single altered or modified nucleotide as compared to the native Rep binding sequence region. In another embodiment, the eRBSR has two or more altered or modified nucleotides as compared to the native Rep binding sequence region. In certain embodiments, the eRBSR has two to twelve altered or modified nucleotides, including, for example, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 altered or modified nucleotides as compared to the native Rep binding sequence region. In one embodiment, all of the nucleotides of the eRBSR are altered, e.g., such as in a swapped sequence (discussed below), as long as the eRBSR retains some minimal binding with the Rep protein.

[0096] In another embodiment, the eRBSR has one or more nucleotide insertions, deletions, substitutions, or any combination thereof as compared to the corresponding native Rep binding sequence region. For example, the engineered Rep binding sequence region can comprise a sequence having one or more nucleotides that differ from the Rep binding sequence region of an adeno-associated virus (AAV), including the Rep binding sequence region of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8.

[0097] In another embodiment, the present disclosure provides a nucleic acid comprising an eRBSR in which the Rep binding sequence region is not native to the parvoviral genome. The eRBSR can comprise a Rep binding sequence region that is found in a different parvoviral serotype or

species genome. In one embodiment, the eRBSR is a Rep binding sequence region from a parvoviral serotype or species that differs from the serotype or species of the native parvoviral ITR sequence. In one embodiment, the eRBSR is a Rep binding sequence region from a parvoviral serotype or species that differs from the serotype or species of the native Rep protein or sequence encoding the Rep protein. Thus, in certain embodiments, the eRBSR is created by swapping the native Rep binding sequence region with all or most of the Rep binding sequence region of a different serotype or species. In one embodiment, the entire Rep binding sequence region is swapped. In another embodiment, a portion or partial sequence of the Rep binding sequence region is swapped. In one embodiment, the engineered Rep binding sequence region can comprise a non-native or swapped Rep binding sequence region from an AAV viral particle, including the Rep binding sequence region of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8. In one embodiment, the eRBSR for an AAV2 polynucleotide, genome, or viral particle (i.e., comprising an AAV2 ITR and/or AAV2 Rep protein coding sequence) is the Rep binding sequence region found in AAV-5 (Table 1; SEQ ID NO: 3, SEQ ID NO: 4).

[0098] In other embodiments, the engineered Rep binding sequence region can comprise a sequence having both alterations and chemical modifications. Thus, in one embodiment, the eRBSR comprises one or more nucleotides that are altered and one or more nucleotides that are chemically modified. Examples of suitable chemical modifications are provided above.

[0099] The engineered Rep binding sequence region must have minimal length sufficient to bind a Rep protein and is only constrained in its maximal length by the length of any ITR packaged in the viral particle. In one embodiment, the eRBSR can approximate the length of a typical parvoviral ITR sequence. In one embodiment, the eRBSR is from about four to about 145 nucleotides in length. In one embodiment, the engineered Rep binding sequence region is from four to 145 nucleotides in length. In another embodiment, the engineered Rep binding sequence region is from about four to about sixty-four nucleotides in length, including any length within that size range, for example 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, and 64 nucleotides in length. In other embodiments, the engineered Rep binding sequence region is from about four to about forty-eight, about four to about thirty-six, about four to about twenty-four, about four to about twenty, about four to about sixteen, and about four to about eight nucleotides in length. In another embodiment, the engineered Rep binding sequence region is from about eight to about twenty, about eight to about sixteen, and about eight to twelve nucleotides in length. In another embodiment, the engineered Rep binding sequence region is four, eight, twelve, sixteen, or twenty nucleotides in length. In one embodiment, the eRBSR is sixteen nucleotides in length.

[0100] In one embodiment, the eRBSR is a nucleic acid comprising one or more consensus motifs having the sequence GCTC. In one embodiment, the eRBSR comprises one to three GCTC consensus motifs. In one embodiment, the eRBSR comprises one GCTC consensus motif. In one embodiment, the eRBSR comprises two GCTC consensus

motifs. In one embodiment, the eRBSR comprises three GCTC consensus motifs. In any of these embodiments, the one to three GCTC consensus motifs can function as an eRBSR when the consensus motif(s) sequence is substituted for (swapped with) the native Rep binding sequence region. Thus, the eRBSR is created by swapping the native Rep binding sequence region with one, two, or three GCTC consensus motif(s). In one embodiment, the entire native Rep binding sequence region is swapped. In another embodiment, a portion or partial sequence of the native Rep binding sequence region is swapped. In any of these embodiments, the eRBSR is about 4-20 nucleotides in length. In one embodiment, the eRBSR has a length of about 4-20 nucleotides and comprises one GCTC motif. In another embodiment, the eRBSR has a length of 4-20 nucleotides and comprises one GCTC motif. In another embodiment, the eRBSR has a length of 4-16 nucleotides and comprises one GCTC motif. In any of these variations, the eRBSR can have a sequence selected from: [GCTC]-(N)x (SEQ ID NO: 292) or (N)x-[GCTC] (SEQ ID NO: 293), wherein $x=0-16$ and N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine. In one variation, $x=0-12$. In another embodiment, the eRBSR has a length of about 8-20 nucleotides and comprises two GCTC motifs. In another embodiment, the eRBSR has a length of 8-20 nucleotides and comprises two GCTC motifs. In another embodiment, the eRBSR has a length of 8-16 nucleotides and comprises two GCTC motifs. In any of these variations, the eRBSR can have a sequence selected from: (1) 1GCTC12-(N)x (SEQ ID NO: 294) or (N)x-1GCTC12 (SEQ ID NO: 295), wherein x is 0-12 and N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine; (2) [GCTC]N x-[GCTC]-(N)y (SEQ ID NO: 296) or (N)x-[GCTC]-(N)y-[GCTC] (SEQ ID NO: 297), wherein x and y are independently 0-12, with the proviso that $x+y<12$ and N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine; and (3) (N)x-[GCTC]12-(N)y (SEQ ID NO: 298) or [GCTC]1-(N)x-[GCTC] (SEQ ID NO: 299), wherein x and y are independently 0-12, with the proviso that $x+y<12$ and N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine. In one variation, $x=0-8$ and $y=0-8$, with the proviso that $x+y<8$. In another embodiment, the eRBSR has a length of 12-20 nucleotides and comprises three GCTC motifs. In one embodiment, the eRBSR has a length of 12-20 nucleotides and comprises three GCTC motifs. In another embodiment, the eRBSR has a length of 12-16 nucleotides and comprises three GCTC motifs. In any of these variations, the eRBSR can have a sequence selected from: (1) [GCTC]3-(N)x (SEQ ID NO: 300) or (N)x-[GCTC]3 (SEQ ID NO: 301), wherein x is 0-8 and N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine; (2) [GCTC]-(N)x-[GCTC]2-(N)y (SEQ ID NO: 302) or (N)x-[GCTC]-(N)y-[GCTC]2 (SEQ ID NO: 303), wherein x and y are independently 0-8, with the proviso that $x+y<8$ and N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine; (3) [GCTC]2-(N)x-[GCTC]N y (SEQ ID NO: 304) or Nx-[GCTC]2-Ny-[GCTC] (SEQ ID NO: 305), wherein x and y are independently 0-8, with the proviso that $x+y<8$ and N is any naturally or non-naturally occurring nucleoside selected

from adenosine, guanosine, uridine, cytidine, thymidine; (4) (N)x-[GCTC]3-(N)y (SEQ ID NO: 306) or [GCTC]-(N)x-[GCTC]-(N)y-[GCTC] (SEQ ID NO: 307), wherein x and y are independently 0-8, with the proviso that $x+y<8$ and N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine. In one variation, x and y are independently 0-4, with the proviso that $x+y<4$. In any of these embodiments, the engineered Rep binding sequence region comprising one, two, or three GCTC consensus motifs can be swapped with a Rep binding sequence region from an AAV viral genome, including the Rep binding sequence region of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8. Also, in any of these embodiments, one or more of the nucleotides in at least one of the GCTC consensus motifs is altered or modified.

[0101] In one embodiment, the eRBSR is a nucleic acid comprising one or more consensus motifs having the sequence GCTC, in which one or more of the nucleotides in at least one of the GCTC consensus motifs is altered or modified. In one embodiment, one or more of the GCTC consensus motif(s) is altered or modified to a motif sequence selected from the group consisting of (1) NCTC, (2) GNTC, (3) GCNC, (4) GCTN, and (5) in the case of two or more altered or modified motif sequences, any combination of these altered or modified motifs, wherein N is either any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine with the proviso that N differs from the consensus nucleotide in that position or N is a chemically modified nucleotide, in which the chemical modification(s) can be on the nucleobase, and/or on the sugar and/or on the backbone. In another embodiment, one or more of the GCTC consensus motif(s) is altered or modified to a motif sequence selected from the group consisting of (1) NNCTC, (2) GNCTC, (3) GCNN, (4) GNTN, (5) NCNC, (6) NCN, and (7) in the case of two or more altered or modified GCTC: motif sequences, any combination of these altered or modified motifs, wherein N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine with the proviso that N differs from the consensus nucleotide in that position or N is a chemically modified nucleotide, in which the chemical modification(s) can be on the nucleobase, and/or on the sugar and/or on the backbone. The individual GCTC consensus motifs can be contiguous with one another or can be separated by intervening sequence, in certain of these embodiments having two or more GCTC consensus motifs, the consensus GCTC motifs are contiguous with one another. In other of these embodiments having one or more GCTC consensus motifs none of the consensus motifs are contiguous with one another. In other individual embodiments, two, three, four, five or more GCTC consensus motifs are contiguous with one another.

[0102] In one embodiment, the eRBSR is a nucleic acid comprising one or more consensus motifs having the sequence GCTC, in which one or more of the nucleotides in at least one of the GCTC consensus motifs is altered or modified. In one embodiment, one or more of the GCTC consensus motif(s) is altered to a motif sequence selected from the group consisting of (1) NCTC, (2) GNTC, (3) GCNC, (4) GCTN, and (5) in the case of two or more altered motif sequences, any combination of these modified altered

motifs, wherein N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine. In one embodiment, one or more of the GCTC consensus motifs) contains at least one nucleotide which is chemically modified. The chemical modification(s) can be on the nucleobase, and/or on the sugar and/or on the backbone. The individual GCTC consensus motifs can be contiguous with one another or can be separated by intervening sequence. In certain of these embodiments having two or more GCTC consensus motifs, the consensus GCTC motifs are contiguous with one another. In other of these embodiments having one or more GCTC consensus motifs none of the consensus motifs are contiguous with one another. In other individual embodiments, two, three, four, five or more GCTC consensus motifs are contiguous with one another.

[0103] in one embodiment, the eRBSR has the formula (GCTC)_x (SEQ ID NO: 308), where x=1-4, in which one or more of the nucleotides in at least one of the GCTC consensus motifs is altered or modified. In one embodiment, one nucleotide in at least one of the GCTC consensus motifs is altered or modified. In one variation of this embodiment, one nucleotide in one of the GCTC consensus motifs is altered or modified. In another variation, one nucleotide in two of the GCTC consensus motifs is altered or modified. In yet another variation, one nucleotide in three of the GCTC consensus motifs is altered or modified. In still another variation, one nucleotide in four of the GCTC consensus motifs is altered or modified. In one embodiment, two nucleotides in at least one of the GCTC consensus motifs is altered or modified. In variations of this embodiment, two nucleotides in one, two, three, or four of the GCTC consensus motifs are altered or modified.

[0104] In one embodiment, the eRBSR has the formula (GCTC)_x-N_y-(GCTC)_z (SEQ ID NO: 309), wherein N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine; x and z are 0-2 with the proviso that when x=0, z=1 or 2, and when z=0, x=1 or 2; and y=0-8, in which one or more of the nucleotides in at least one of the GCTC consensus motifs is modified. In one embodiment, the eRBSR has the formula (GCTC)_x-N_y-(GCTC)_z-N_w-(GCTC)_w (SEQ ID NO: 310), wherein N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine; w, x, and z are 0-2 with the proviso that when x=0, z=1 or 2 and w=1 or 2, when z=0, x=1 or 2 and w=1 or 2; when w=0, x=1 or 2 and z=1 or 2, and y=0-8. In any of the embodiments described here, when y=4 the N's can comprise a GCTC consensus motif in which at least one of the nucleotides in the consensus motif is altered or modified. In one variation of this embodiment, one of the nucleotides in the consensus motif is altered or modified. In another variation of this embodiment, two of the nucleotides in the consensus motif is altered or modified. Also, in any of the embodiments described here, when y=8 the N's can comprise two GCTC consensus motif in which at least one of the nucleotides in at least one of the consensus motifs is altered or modified. In one variation of this embodiment, one of the nucleotides in one or both of the consensus motifs is altered or modified. In another variation of this embodiment, two of the nucleotides in one or both of the consensus motifs is altered or modified.

[0105] Table 1 provides Rep binding sequence regions found in various AAV serotypes and species which can function as an eRBSR (SEQ ID NOs: 1-20).

TABLE 1

Native or Wild-Type Rep Binding Sequence Regions		
Rep-binding sequence	Sequence	SEQ ID NO
AAV2 5'	GCGCGCTCGCTCGCTC	1
AAV2 3'	GAGCGAGCGAGCGCGC	2
AAV5 5'	GCTCGCTCGCTGGCTC	3
AAV5 3'	GAGCCAGCGAGCGAGC	4
AAV1 5'	GCGCGCTCGCTCGCTC	5
AAV1 3'	GAGCGAGCGAGCGCGC	6
AAV3 5'	GCGCACTCGCTCGCTC	7
AAV3 3'	GAGCGAGCGAGTGC GC	8
AAV4 5'	GCGCGCTCGCTCACTC	9
AAV4 3'	GAGTGAGCGAGCGCGC	10
AAV7 5'	GCGCGCTCGCTCGCTC	11
AAV7 3'	GAGCGAGCGAGCGCGC	12
Avian AAV Strain DA1 5'	GCTCGCTCGCTC	13
Avian AAV Strain DA1 3'	GAGCGAGCGAGC	14
Avian AAV ATCC VR-865 5'	GCTCGCTCGCTC	15
Avian AAV ATCC VR-865 3'	GAGCGAGCGAGC	16
Bovine AAV5'	GCTCGTTGCTGGCTC	17
Bovine AAV3'	GAGCCAGCGAACGAGC	18
Snake AAV5'	GCGCGCGCTC	19
Snake AAV3'	GAGCGCGCGC	20

[0106] Table 2 provides engineered Rep binding sequence regions which can function as an eRBSR (SEQ ID NOs: 21-66).

TABLE 2

Engineered Rep Binding Sequence Regions	
Sequence	SEQ ID NO
GCTC N N N N N N N N N N N N N N N N	21
GCTC N N N N N N N N N N N N N N	22
GCTC N N N N N N N N N	23
GCTC N N N N N	24
N N N N N N N N N N N N N N N GCTC	25
N N N N N N N N N N N N GCTC	26
N N N N N N N N GCTC	27
N N N N GCTC	28

TABLE 2-continued

Engineered Rep Binding Sequence Regions	
Sequence	SEQ ID NO
GCTCNNNNNNNNNNNGCTC	29
GCTCNNNNNNNNNGCTC	30
GCTCNNNNNGCTC	31
GCTCGCTCNNNNNNNNNN	32
GCTCGCTCNNNNNNNN	33
GCTCGCTCNNNN	34
GCTCGCTCGCTCNNNNNNNN	35
GCTCGCTCGCTCNNNN	36
NNNNNNNNNNNGCTCGCTC	37
NNNNNNNGCTCGCTC	38
NNNGCTCGCTC	39
NNNNNNNNNGCTCGCTCGCTC	40
NNNGCTCGCTCGCTC	41
GCTCGCTCNNNNNNNNNGCTC	42
GCTCGCTCNNNNNGCTC	43
GCTCNNNNNNNNNGCTCGCTC	44
GCTCNNNNNGCTCGCTC	45
NNNGCTCGCTCGCTCNNNN	46
NNNNNNNNNGCTCGCTCNNNN	47
NNNGCTCGCTCNNNNNNNN	48
NNNGCTCGCTCNNNN	49
NNNNNNNNNNNGCTCNNNN	50
NNNNNNNNNGCTCNNNNNNNN	51
NNNNNNNNNGCTCNNNN	52
NNNGCTCNNNN	53
NNNGCTCNNNNNNNNNNNN	54
NNNGCTCNNNNNNNN	55
GCTCNNNNNGCTCNNNN	56
GCTCNNNNNNNNNGCTCNNNN	57
GCTCNNNNNGCTCNNNNNNNN	58
NNNGCTCNNNNNGCTC	59
NNNGCTCNNNNNNNNNGCTC	60
NNNNNNNNNGCTCNNNNNGCTC	61
GCTCGGGG	62
GCTCGGGG	63
GCTCGGGGG	64

TABLE 2-continued

Engineered Rep Binding Sequence Regions	
Sequence	SEQ ID NO
GCTCGGGGGG	65
GCTCGGGGGGG	66

[0107] Table 3 provides oligonucleotide sequences that comprise Rep protein binding properties which may therefore function as an eRBSR (SEQ NOs: 79-197) (Chiorini et al. 1995 Journal of Virology 69(11) 7334-7338, the contents of which are incorporated herein by reference in their entirety).

TABLE 3

Rep Binding Sequence Region Oligonucleotides	
Sequence	SEQ ID NO
ATACGCCGCTCGCGCTCAG	79
ATCTGTCTGCTCGTCCGGCTA	80
TCCGCGCTGGCTCATCGTCC	81
TCCGCGCTGGCTCATCGTCC	82
TCCCCGCCCCGCTCATTCT	83
GTGCCCCGCTCAGAGTCCA	84
TTTACCGCCGCTCAGATAGA	85
GACCCCAAGGCGCTCCTATG	86
ATCTCGCTCATGCCCTTAG	87
TCCCCGGTCAGGGGCTCACT	88
CCGCCGCTCTATCCACTGGT	89
TGCTCTGCTGCTACTGTTC	90
CATATCTCCGCTTAGTTGCC	91
TCGTTAAGAACCTTCCTCAT	92
CCCCGCATCCTCCGCTTC	93
CATATCTCCGCTTAGTTGCC	94
CACAGTTCTCGCCTACCCGT	95
CTCTCCTTCAGGGCCTCAGC	96
GCTGCCCGCTACTACCCG	97
GCTCGAAGGAAGCGGGAAC	98
GTCAGTTCGCTGGGTGATTC	99
ATCGGACGGCTTCGTGTGC	100
TCGCTGACCAAGCCGATGC	101
CCAGTATTCTGCGCAGCTGG	102
GGCGTCCCTTTCTTTTCG	103
TAATCGTATGCATCGTCGTG	104
TGAACGGTCGCGACGCAGCA	105

TABLE 3-continued

Rep Binding Sequence Region Oligonucleotides	
Sequence	SEQ ID NO
ACGTTCAACCCGCCGCTCG	106
ATCGGACGGCTTCGTTGTGC	107
CCATGTGTTGCGCCGTCGC	108
GGTCCCCCATGCACTGCC	109
CCCGTATCCACGCCCCACGC	110
TGCCCCACGCGCGCTCGTAC	111
ATGCTTTCTCGCTCAGTCC	112
ACACTTCTCACTCGCTGCCT	113
CCTTCGCGCTCGTTCAATA	114
TGCTTTCTCGCTCAGTCC	115
GCGATTGCCTCGGTGGCTCA	116
CTTCACTTACTCGCGACCCC	117
TGCCCTTGGCTTGCTCAGTG	118
GTCTCGCTCGGTCACTACT	119
ATCCGTTCACTCGTTCGCCT	120
ACGCACTCACTCGCCGGCGC	121
CTCTTGCTCGCTCAATTGCT	122
GGGTGCTCGCCTGGTTGCG	123
CCAGTGCTCACTCAGCTCGC	124
TCGGCGCTCGCTCGGTCCCTC	125
CGCGGGTTGTTTCACTCAC	126
TGTCATCTGCTCGCTCACTT	127
AGATCCTTGTGGCTCACTCG	128
ACTGGCTGGTTGCTCAGAC	129
CCACAGATGTAGCTCACTCA	130
CTGGTATCGCTCACTTGACC	131
TCGCGCTCGTTCGCTCTGCT	132
AGATCCTTGTGGCTCACTCG	133
CAATCAGTTAACCACCTTCAA	134
CTCGCTTGCACTCACTCACT	135
GAGTTTATCTCATGTTCTGC	136
CAATTCAGCCAACCACTCAA	137
GCTGACTCGTTGACTCATCT	138
TCTCCGCTGGTTACCGTAC	139
TGTGCGGCCCACTGAGACGT	140
CCACGTGCTCGCTCAACCTT	141

TABLE 3-continued

Rep Binding Sequence Region Oligonucleotides	
Sequence	SEQ ID NO
TCCTACAGCTTGCTCACTCT	142
CACAGTTACTGGCTCACTGA	143
GGGGTTGGTTGGTTCACTCC	144
GGGTGCTCGCCTGGGTGCG	145
GCCACTTAGCGTACTGGTTC	146
CCCTGTTGCTTGCTCGTTC	147
TTAACATGGCGCGTTCACTG	148
CCATCTTATTGGTTCACTGG	149
CTCTCCCGCGCACTGACTC	150
CTAAGATGCCCCGCTCGCTC	151
GTATGGCACTCGCTCGCTGG	152
TGCATCTGCCCTTGCTCACT	153
ACTCACTCACTGAGTTGTCC	154
TTCGTTGGCTCACTACCCC	155
GTGCCCGTTTCTACTCACTG	156
TGGTCGCTCGCCCCAGCCGC	157
TTGCTTGCTCACTCCGCGCT	158
CGCCAGCTCCCTCCTCATGC	159
CCTCCACTCGCTGGGTGAGC	160
TCGTGCTCGCTCTGCATCC	161
TACCGCGATCTCGCTCTCTG	162
CCCACGCCCTCGCTCACTGC	163
TCTCTGTCACTCGTCTCCGG	164
CTCGTGCCCCGGTGCGCTCA	165
CGTGTACTTCGTGCTCAGCC	166
CCCGTGTAGGTGCGCTCGCT	167
TCCACCCTACTCCGTCGGCC	168
CTCACCGCACTCACTGGCCC	169
GCCCCCTCCTTTCGCTCGTTT	170
CCTCTCCGCGTGCCCACTCG	171
CCTTCTTTTCCGCTGGCTCA	172
CTCGTGCCCCGGTGCGCTCA	173
GGCGATTCACTCATCTGACC	174
ACTGCCCGCGCTTTCGCTC	175
CCTCTCCGCGTGCCCACTCG	176
GCTTGCTCGACCCAGCCACG	177
CCTGGTTCGGTCCTTCTCCC	178

TABLE 3-continued

Rep Binding Sequence Region Oligonucleotides	
Sequence	SEQ ID NO
ATCCTCTATCTCACTCGCT	179
CCCGGGTACTGACTCGCGCT	180
TCCATGCCACTTGGTTCACCT	181
CTGCTATCGTTAACTCACCT	182
TGCACTCATTTCGACTGCCTC	183
CGAGCGAGCTTAGTGAACGT	184
AGCTGGTCGGTGTTCACCTGC	185
ATTCACTCGCGGTTCGCGCAC	186
TCTCGCTGGTTCAGTCCTTC	187
GCATTTCTGGGTAGCTCGCT	188
TAGCTGAATCGCCAGGCTTG	189
TGGCTCATTCATTGAGTCCA	190
CATTGGCTGTTGCTGACT	191
TCCATGCCACTTGGTTCACCT	192
GCCCCCTCCTTCGCTCGTTT	193
CGGCCCTTCCCACTGGCTC	194
CGCCAGCTCCCTCCTCATGC	195
GTCGCTCCCTCCTTACCGCG	196
CCAGCGAACGCCCTCCCGCA	197

[0108] In some embodiments, the present disclosure provides an eRBSR comprising any of SEQ ID NOs: 1-20, 21-66, and/or 79-197. Any one of the Rep binding sequence regions listed in Table 1 can function as an eRBSR when substituted into a non-native ITR sequence in place of the native Rep binding sequence regions. In one embodiment, SEQ ID NOs: 3-4 (AAV-5) are inserted into an AAV1, AAV2, AAV3, AAV4, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 in place of the native AAV5 Rep binding sequence regions. In other embodiments, SEQ ID NOs: 7-8 (AAV3) are inserted into an AAV1, AAV2, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 in place of the native AAV3 Rep binding sequence regions. In other embodiments, SEQ ID NOs: 9-10 (AAV4) are inserted into an AAV 1, AAV2, AAV3, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 in place of the native AAV4 Rep binding sequence regions. In other embodiments, SEQ ID NOs: 1-2 (AAV2/AAV1/AAV7) are inserted into an AAV3, AAV4, AAV5, AAV6, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 in place of the native AAV2 Rep binding sequence regions. In other embodiments, SEQ ID NOs: 17-18 (bovine) are inserted into an AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and

AAV-DJ8 in place of the native Rep binding sequence regions. In other embodiments, SEQ ID NOs: 19-20 (snake) are inserted into an AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 in place of the native Rep binding sequence regions. In other embodiments, SEQ ID NOs: 13-14 (avian) are inserted into an AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 in place of the native Rep binding sequence regions.

[0109] In another embodiment, the present disclosure provides an engineered Rep binding sequence region in which one or more nucleotides in any of SEQ ID NOs: 1-20 listed in Table 1 is altered or modified as discussed elsewhere herein. For example, any of the Rep binding sequence regions listed in Table 1 (SEQ ID NOs: 1-20) can be altered by having one or more nucleotide insertions, deletions, substitutions, or any combination thereof as compared to the corresponding native Rep binding sequence region. **[00110]** In certain embodiments, the engineered Rep binding sequence region is selected from the sequences of SEQ ID NO: 68 (GCGCCGCATGCGGCTC) (α Pal, SEQ ID NO: 71 (GAGCGGGGGCGCGCTC) (5-5/EGR) and SEQ ID NO: 74 (CCGACCACGACGACGG) (JcDNVNS1).

[0110] The eRBSRs provided here can be made by methods commonly known and practiced by those skilled in the art. For example, the eRBSRs provided here can be designed in silico and synthesized using molecular biology techniques well known to those of skill in the art. See, e.g., Sambrook et al, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press (Cold Spring Harbor, N.Y.).

[0111] In any of SEQ ID NOs: 1-291, one or more nucleotides can be altered or modified as provided in the present disclosure.

[0112] As discussed herein, the Rep binding sequence region of the parvoviral genome is altered and/or modified to decrease the binding of the Rep protein so as to promote the production of a parvovirus having a duplexed or self-complementary genome. Thus, the engineered Rep binding sequence regions (eRBSRs) provided herein comprise at least one nucleic acid sequence that has decreased binding with wild-type and/or engineered Rep protein. In one embodiment, the engineered Rep binding sequence region has decreased binding with a wild-type Rep protein as compared to the binding with the corresponding wild-type or native Rep binding sequence region. In another embodiment, the engineered Rep binding sequence region has decreased binding with an engineered Rep protein as compared to the binding with the corresponding wild-type or native Rep binding sequence region. In another embodiment, the engineered Rep binding sequence region has decreased binding with both wild-type and engineered Rep protein as compared to the respective binding with the corresponding wild-type or native Rep binding sequence region. In one embodiment, the binding affinity of an eRBSR with a wild-type and/or engineered Rep protein is decreased 1-10 fold relative to the binding affinity of the corresponding wild-type or native Rep binding sequence region with the same Rep protein. In one embodiment, the binding affinity is decreased by at least 2-fold. In another embodiment, the binding affinity is decreased by at least 5-fold. In a further embodiment, the binding affinity is decreased by at least 10-fold. In one embodiment, the wild-type Rep protein with

which the eRBSR has decreased binding affinity is an AAV Rep protein. In another embodiment, the engineered Rep protein with which the eRBSR has decreased affinity is derived from an AAV Rep protein that has been altered or modified, e.g., having one or more altered or modified nucleotides. In certain embodiments, the wild-type AAV Rep protein is selected from Rep22, Rep68, Rep78, Rep4-0, and Rep52. In certain embodiments, the engineered AAV Rep protein is selected from a Rep 22, Rep68, Rep78, Rep40, and Rep52 that has one or more altered or modified nucleotides.

[0113] The binding affinity can be measured by several methods well-known in the art, including electro-mobility shift assay and surface plasmon resonance, as described in the Examples section. In one embodiment, the binding affinities of the following polypeptide-polynucleotide pairs are determined using an Octet surface plasmon resonance instrument (ForteBio, Menlo Park, Calif.): wild-type Rep protein and a wild-type Rep binding sequence region, an engineered Rep protein and a wild-type Rep binding sequence region, a wild-type Rep protein and an eRBSR, an engineered Rep protein and an eRBSR. The polynucleotide encoding either a wild-type Rep binding sequence region or an engineered Rep binding sequence region is immobilized on the biosensor layer. A solution is prepared comprising either wild-type Rep protein or engineered Rep protein that is then passed over the biosensor to determine the rate of association (K_a). A second solution devoid of wild-type or Rep protein is passed over the saturated biosensor layer to determine the rate of dissociation (K_d). 1001151 The present disclosure also provides polynucleotides comprising any of the engineered Rep binding sequence regions (eRBSRs) described herein. The polynucleotide can be any deoxyribonucleotide or ribonucleotide polymer or sequence of nucleotide bases, in linear or circular conformation, and in either single- or double-stranded form. The polynucleotide can include natural nucleotides, known mutants of natural nucleotides, as well as nucleotides that are modified in the base, sugar, and/or phosphate moieties (e.g., the phosphorothioate backbone). Thus, the polynucleotide can be RNA, DNA or DNA-RNA hybrid sequences that are naturally occurring or non-naturally occurring nucleotides.

[0114] The polynucleotide comprising the eRBSR can further comprise one or more heterologous sequences. The heterologous sequence can be any sequence that is not native to the viral genome. Heterologous sequences include regulatory sequences, sequences encoding a nucleic acid, such as but not limited to an siRNA, dsRNA, miRNA, antisense, aptamer, and the like, genome editing sequences, sequence encoding a polypeptide (e.g., such as a transgene), functional domain sequences, marker, tag, stuffer sequences, linker sequences. In one embodiment, the heterologous sequence is operatively linked to one or more regulatory sequence(s) that allows expression of the heterologous sequence.

[0115] In one embodiment, the polynucleotide comprises one or more parvoviral inverted terminal repeat (ITR) sequences wherein one of the ITR sequences comprises an eRBSR. Typically, the ITR sequence comprising the eRBSR also contains a reverse complementary sequence of the eRBSR downstream of the eRBSR such that the eRBSR and the reverse complementary sequence are capable of intra-strand base-pairing. Thus, in one embodiment, the polynucleotide comprises a parvoviral ITR nucleotide sequence

comprising an eRBSR wherein the parvoviral ITR nucleotide sequence is capable of forming a double-stranded engineered Rep binding sequence region. In one embodiment, the engineered Rep binding sequence region is located between a 5' parvoviral inverted terminal repeat sequence and a 3' parvoviral inverted terminal repeat sequence. In another embodiment, the engineered Rep binding sequence region is located within a parvoviral inverted terminal repeat sequence.

[0116] The parvoviral inverted terminal repeat sequence(s) can be a full-length ITR sequence, a partial ITR sequence, or an ITR with an additional sequence, for example, one or more regulatory sequences, one or more functional domain sequence(s) and/or one or more stuffer or linker sequences. The parvoviral inverted terminal repeat sequence(s) can comprise naturally-occurring or wild-type sequence(s) or variant sequence(s), including natural variants and artificial variants, as well as truncated and elongated ITR sequences. The parvoviral ITR sequence can alternatively be modified to include one or more insertions, deletions and/or substitutions.

[0117] In one embodiment, in a polynucleotide comprising an eRBSR which comprises an ITR sequence, the polynucleotide can further comprise a regulatory sequence or other heterologous sequence, such as a payload sequence. In one embodiment, the heterologous sequence is operatively linked to one or more regulatory sequence(s) that allows expression of the heterologous sequence. In one embodiment, the polynucleotide comprises a first ITR sequence comprising an eRBSR and a second ITR sequence, wherein the heterologous sequence is flanked by the first and second ITR sequences. In any of these embodiments, the ITR sequence can be an adeno-associated virus (AAV) ITR or derived from an adeno-associated virus (AAV) ITR. AAV ITR sequences include any of those having an AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 ITR sequence or a sequence derived therefrom.

[0118] In one embodiment, the parvoviral polynucleotide comprises in the 5' to 3' direction: (a) a 5' parvoviral inverted terminal repeat sequence; (b) a first payload encoding region; (c) an engineered Rep binding sequence region; (d) a second payload encoding region; and (e) a 3' parvoviral inverted terminal repeat sequence. In this embodiment, the first and second payload encoding regions are essentially self-complementary and may form a hairpin structure that comprises the parvoviral polynucleotide. In one embodiment, the parvoviral polynucleotide (e.g., a parvoviral genome) is an AAV polynucleotide, genome or parvoviral particle.

[0119] The designations of 5' and 3' with respect to the ITR sequences are also referred to as "left" and "right" ITR. The 5' and 3' wild-type ITR sequences encoded by ssDNA and dsDNA are reverse complements of each other that encode identical secondary structures. The AAV2 ssDNA genome, for example, comprises a TT dimer Rep nicking site within the nicking stem loop sequence of one ITR and a complementary AA dimer within the nicking stem loop sequence of the second ITR. As discussed herein, Rep protein only nicks AAV2 ssDNA at the TT dimer Rep nicking site. As used herein, the term 3' or "right" designates the ITR comprising the Rep nicking site.

[0120] In one embodiment, the transgene or payload sequence encodes human aromatic L-amino acid decarboxy-

lase. In another embodiment, the transgene or payload sequence encodes human SOD1. In another embodiment, the transgene or payload sequence encodes human frataxin (FXN).

[0121] The present disclosure provides a polynucleotide comprising an eRBSR in which the binding affinity of a Rep protein (wild-type or engineered) to the eRBSR is decreased relative to the binding of a polynucleotide comprising a wild-type or native Rep binding sequence region. In one embodiment, the binding affinity is decreased 1-10 fold relative to a polynucleotide comprising a wild-type Rep binding sequence region. In one embodiment, the binding affinity is decreased at least 2-fold. In another embodiment, the binding affinity is decreased at least 5-fold. In a further embodiment, the binding affinity is decreased at least 10-fold. In one embodiment, the Rep protein is Rep68 or Rep78. In one embodiment, the wild-type AAV Rep protein is selected from Rep 22, Rep68, Rep78, Rep40, and Rep52. In another embodiment, the engineered AAV Rep protein is selected from a Rep 22, Rep68, Rep78, Rep40, and Rep52 that has one or more altered or modified nucleotides.

[0122] The present disclosure also provides a polynucleotide comprising an engineered Rep binding sequence region in which the nicking by a Rep protein (wild-type or engineered) decreased relative to the nicking of a polynucleotide comprising a wild-type Rep binding sequence region. In one embodiment, the nicking is decreased 1-10 fold relative to the nicking of a polynucleotide comprising a wild-type Rep binding sequence region. In one embodiment, the nicking is decreased by at least 2-fold. In another embodiment, the nicking is decreased by at least 5-fold. In a further embodiment, the nicking is decreased by at least 10-fold. In one embodiment, the Rep protein is Rep68 or Rep78. In one embodiment, the Rep protein is Rep68 or Rep78. In one embodiment, the wild-type AAV Rep protein is selected from Rep 22, Rep68, Rep78, Rep40, and Rep52. In another embodiment, the engineered AAV Rep protein is selected from a Rep 22, Rep68, Rep78, Rep40, and Rep52 that has one or more altered or modified nucleotides.

[0123] The present disclosure also provides polynucleotide comprising an engineered Rep binding sequence region in which the presence of the engineered Rep binding sequence region results in an increase in the generation of self-complementary AAV genomes during viral genome replication.

[0124] In any of the above-described embodiments, the polynucleotide is an AAV polynucleotide or derived from an AAV polynucleotide. AAV polynucleotides include those of AAV1, AAV2, AAV3, AAV4, AAV 5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8.

[0125] In one embodiment, the serotype which may be useful in the present invention may be AAV-DJ8. The amino acid sequence of AAV-DJ8 may comprise two or more mutations in order to remove the heparin binding domain (HBD). As a non-limiting example, the AAV-DJ sequence described as SEQ ID NO: 1 in U.S. Pat. No. 7,588,772, the contents of which are herein incorporated by reference in their entirety, may comprise two mutations: (1) R587Q where arginine (R; arg) at amino acid 587 is changed to glutamine (Q; gln) and (2) R590T where arginine (R; arg) at amino acid 590 is changed to threonine (T; thr). As another non-limiting example, the AAV-DJ sequence may comprise three mutations: (1) K406R where lysine (K; lys) at amino

acid 406 is changed to arginine (R; arg), (2) R587Q where arginine (R; arg) at amino acid 587 is changed to glutamine (Q; gln) and (3) 85901 where arginine (R; arg) at amino acid 590 is changed to threonine (T; thr).

Swapped Nicking Stem Loop

[0126] Parvoviral ITR sequences comprise a short self-complementary region that forms a secondary structure loop required for nicking single stranded DNA during the course of replication as discussed previously: The nicking stem loop sequence varies between parvoviral serotypes and may be between about 20-25 nucleotides in length spanning the border between the A and D regions of the parvoviral genome. Complementary sequences of about 6-10 nucleotides form the stem of the secondary structure. The loop at the top of the stem is a non-complementary sequence of about 3-6 nucleotides that further comprises the terminal resolution site where the nicking of the phosphodiester backbone occurs. The terminal resolution site, or nicking site, of the AAV2 genome is a TT dimer that is located at the end of the ITR A region adjacent to the D region. This TT dimer is also present in the AAV1, AAV3, AAV4, and AAV7 genomes. An alternative nicking stem loop and nicking site is exemplified by AAV5, wherein the nicking site is a GT dimer. Nicking of the parvoviral genome only occurs at the nicking site, TT or GT, comprised within the nicking stem loop sequence that spans the A and D regions. Nicking of the parvoviral genome does not occur at the reverse complementary nicking site, AA or AC, comprised within a reverse complementary nicking stem loop sequence that spans the A and D regions.

[0127] Parvoviral serotypes with closely related genomes encode Rep proteins that are complementary. For example, An AAV2 Rep protein may productively promote replication of an AAV3 genome by binding to the AAV3 genome Rep binding sequence and cutting the nicking stem loop at the nicking site. However, some parvoviral serotypes encode Rep proteins that recognize heterologous Rep binding sequence regions, nicking stem loops, and nicking sites that are not cross-complementary. In one example, an AAV2 Rep protein may not nick the nicking stem loop of an AAV5 genome due to the lack of the TT dimer.

[0128] Accordingly, one way to promote the generation of self-complementary parvoviral genome sequences is to alter or modify the Rep nicking stem loop such that nicking by a heterologous Rep protein is abolished, consequently promoting the generation of self-complementary parvoviral genome sequences as discussed above in the previous section. The present disclosure thus provides a chimeric, altered, or modified Rep nicking stem loop sequence designed to decrease or abolish the nicking activity of the Rep protein.

[0129] Typically, the Rep nicking stem loop sequence of a parvoviral genome is from about 20 to about 2.5 nucleotides in length. The present disclosure accordingly provides a chimeric, altered, or modified Rep nicking stem loop sequence that is from about 20 to about 25 nucleotides in length. In one embodiment, the chimeric, altered, or modified Rep nicking sequence is from 20 to 25 nucleotides in length.

[0130] The present disclosure provides a nucleic acid comprising a Rep nicking stem loop in which one or more of the nucleotides in the nicking stem loop are altered or modified as compared to the native nicking stem loop. An

“altered” nucleotide can be any nucleotide having a naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine, as long as it differs from the native nucleoside. Methods for producing nucleic acid sequences with altered nucleosides have been discussed herein in the previous section. A modified nucleotide refers to a nucleotide that has been chemically modified, in which the modification can be on the nucleobase, and/or the sugar, and/or on the backbone of the nucleotide. The chemical modification can be any chemical modification used to modify nucleic acids. Examples of suitable chemical modifications have been provided herein in the previous section, along with methods for chemically modifying a nucleic acid.

[0131] In one embodiment, the Rep nicking stem loop sequence has a single altered or modified nucleotide as compared to the native Rep nicking stem loop sequence. In another embodiment, the Rep nicking stem loop sequence has two or more altered or modified nucleotides as compared to the native Rep nicking stem loop sequence. In certain embodiments, the Rep nicking stem loop sequence has two to twenty five altered or modified nucleotides, including, for example, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 altered or modified nucleotides as compared to the native Rep nicking stem loop sequence. In one embodiment, all of the nucleotides of the Rep nicking stem loop sequence are altered or modified.

[0132] In another embodiment, the Rep nicking sequence has one or more nucleotide insertions, deletions, substitutions, or any combination thereof as compared to the corresponding native Rep nicking sequence. In one embodiment, the Rep nicking sequence can comprise a sequence having one or more nucleotides that differ from the Rep nicking sequence of an adeno-associated virus (AAV), including the Rep nicking sequence of AAV 1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8.

[0133] In other embodiments, the Rep nicking stem loop can comprise a sequence having both alterations and chemical modifications. Thus, in one embodiment, the Rep nicking stem loop comprises one or more nucleotides that are altered and one or more nucleotides that are chemically modified.

[0134] The present disclosure further provides a chimeric nicking stem loop, in which the Rep nicking stem loop is not native to the parvoviral genome. In one embodiment, the invention provides a chimeric nicking stem loop in which the chimeric nicking stem loop is from a parvoviral serotype or species that differs from the serotype or species of the native parvoviral ITR sequence. In one embodiment, the chimeric nicking stem loop is from a parvoviral serotype or species that differs from the serotype or species of the native Rep protein or sequence encoding the Rep protein. The chimeric nicking stem loop is created by substituting or swapping the native Rep nicking stem loop with the Rep nicking stem loop of a different serotype or species. In one embodiment, the entire Rep nicking stem loop is swapped. In another embodiment, a portion of the Rep nicking stem loop is swapped. In another embodiment, the Rep nicking stem loop comprises a non-native or swapped Rep nicking stem loop from an AAV viral genome, including but not limited to the Rep nicking stem loop of AAV1, AAV2,

AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8.

[0135] The Rep nicking stem loops found in various AAV serotypes and species are shown in FIG. 3 and FIG. 4 and provided in Table 4. The genome sequences of several different AAV serotypes and species including AAV1, AAV2, AAV3, AAV4, AAV5, AAV7, AAV8, Bat AAV, Bovine AAV, Snake AAV, Avian AAV DA-1, and Avian AAV VR865 (SEQ ID NO: 280-291) were obtained from the NCBI Genome database. The 5' and 3' inverted terminal repeat sequences of selected genomes were aligned with the Clustal Omega multiple sequence alignment program using either AAV2 or AAV5 as the parent sequence, as shown in FIG. 3 and FIG. 4, respectively. Additional Rep nicking stem loops found in the ITR sequences of other parvoviral serotypes and species, including AAV8 5' ITR (SEQ ID NO: 253), AAV8 3' ITR (SEQ ID NO: 254), Bat AAV 5' ITR (SEQ ID NO: 261), Bat AAV 3' ITR (SEQ ID NO: 262) are obtained in the same manner using this alignment program.

TABLE 4

Native or Wild-Type Rep Nicking Stem Loop Sequences		
Name	Sequence	SEQ ID NO
AAV2 5'	AGAGGGAGTGGCCAACCTCCATCA	198
AAV2 3'	TGATGGAGTTGGCCACTCCCTCT	199
AAV5 5'	GGGGGAGAGTGCCCACTCTCA	200
AAV5 3'	TGAGAGTGTGGCACTCTCCCC	201
AAV1 5'	AGAGGGAGTGGCCAACCTCCATCA	202
AAV1 3'	TGATGGAGTTGGCCACTCCCTCT	203
AAV3 5'	AGAGGGAGTGGCCAACCTCCATCA	204
AAV3 3'	TGATGGAGTTGGCCACTCCCTCT	205
AAV4 5'	AGAGGGAGTGGCCAACCTCCATCA	206
AAV4 3'	TGATGGAGTTGGCCACATTAGCT	207
AAV7 5'	AGAGGGAGTGGCCAACCTCCATCA	208
AAV7 3'	AATGGAGTTGGCCACTCCCTCT	209
Avian AAV Strain DA1 5'	ACTGGCCAGCACTCCGGTGA	210
Avian AAV Strain DA1 3'	TCACCGAGTGCTGGCCAGT	211
Avian AAV ATCC VR-865 5'	ACTGGCCAGCACTCCGGTGA	212
Avian AAV ATCC VR-865 3'	TCACCGAGTGCTGGCCAGT	213
Bovine AAV 5'	GGGGGGAGTGCCCACTCTCT	214
Bovine AAV 3'	AGAGAGTGTGGCACTCCCCCCC	215
Snake AAV 5'	TGGGCGAGTGCCCTGCTC	216
Snake AAV 3'	GAGCAGGGCACTCGCCCCA	217

[0136] Based on the sequence alignments, several different Rep nicking stem loops were identified (see boxed sequence in FIG. 3 and FIG. 4). The Rep nicking stem loops for AAV2, AAV1, AAV3, and AAV7 share the same nicking stem loop sequence: AGAGGGAGTGGCCAACTCCATCA (SEQ ID NO: 198) on the 5' ITR and its reverse complement TGATGGAGTGGCCACTCCCTCT (SEQ ID NO: 199) on the 3' ITR. AAV-4 has the same nicking sequence on the 5' ITR and a slightly different sequence on the 3' ITR (TGATGGAGTTGGCCACATTAGCT; SEQ ID NO: 207). The Rep nicking stem loop for AAV5 is GGGGGAGAGTGCCAACTCTCA (SEQ ID NO: 200) on the 5' ITR and its reverse complement TGAGAGTGIGGCACTCTCCCC (SEQ ID NO: 201) on the 3' ITR. The Rep nicking stem loop for bovine AAV is GGGGGGAGTGCCAACTCTCT (SEQ ID NO: 214) on the 5' ITR and its reverse complement is AGAGAGTGTGGCACTCCCCC (SEQ ID NO: 215) on the 3' ITR. The Rep nicking stem loop for snake AAV is TGGGGCGAGTGCCCTGCTC (SEQ ID NO: 216) on the 5' ITR and its reverse complement is GAGCAGGGCACTCGCCCCA (SEQ ID NO: 217) on the 3' ITR. The Rep nicking stem loop for avian AAV (AAVDA1 and VR865) is ACTGGCCAGCACTCCGGTGA (SEQ ID NO: 210) on the 5' ITR and its reverse complement is TCACCGGAGTGCTGGCCAGT (SEQ ID NO: 211) on the 3' ITR. In one embodiment, the Rep nicking stem loop comprises any of SEQ ID NOs: 21-66. In another embodiment, the Rep nicking stem loop comprises any of SEQ ID NOs: 198-217, in which one or more nucleotides are altered or modified.

[0137] The chimeric nicking stem loop is achieved by swapping the nicking stem loop sequence of one parvoviral serotype or species with the native nicking stem loop sequence of a different parvoviral serotype or species. In one embodiment, the nicking stem loop sequence of any of SEQ ID NOs: 198-217 (listed in above paragraph) is swapped with the native nicking stem loop sequence of a different parvoviral serotype or species. In one embodiment, the nicking stem loop for AAV2 (SEQ ID NO: 199) (which is the same nicking stem loop for AAV1, AAV3, and AAV7) is swapped into (replaces all or a portion of) the native nicking stem loop sequence in any of the following ITR sequences: AAV4, AAV5, AAV6, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 ITR. In another embodiment, the nicking stem loop for AAV4 (SEQ ID NO: 207) is swapped into (replaces all or a portion of) the native nicking stem loop sequence in any of the following ITR sequences: AAV1, AAV2, AAV3, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 ITR. In another embodiment, the nicking stem loop for AAV5 (SEQ ID NO: 201) is swapped into (replaces all or a portion of) the native nicking stem loop sequence in any of the following ITR sequences: AAV1, AAV2, AAV3, AAV4, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 ITR. In yet another embodiment, the nicking stem loop for snake AAV (SEQ ID NO: 217) is swapped into (replaces all or a portion of) the native nicking

stem loop sequence in any of the following ITR sequences: AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 ITR. In still another embodiment, the nicking stem loop for avian AAVDA1 or avian AAVVR865 (SEQ ID NO: 211) is swapped into (replaces all or a portion of) the native nicking stem loop sequence in any of the following ITR sequences: AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 ITR.

[0138] In other embodiments, the polynucleotide comprising the chimeric nicking stem loop comprises any of SEQ ID NOs: 218-240. (Table 5). In one embodiment, the nicking stem loop for AAV2 is swapped with (replaces all of) the native nicking stem loop sequence of the AAV5 ITR sequence (SEQ ID NO: 218). In one embodiment, the nicking stem loop for AAV4 is swapped with (replaces all of) the native nicking stem loop sequence of the AAV5 ITR sequence (SEQ ID NO: 219). In one embodiment, the nicking stem loop for AAV2 is swapped with (replaces all of) the native nicking stem loop sequence of the AAV4 ITR sequence (SEQ ID NO: 220). In one embodiment, the nicking stem loop for AAV5 is swapped with (replaces all of) the native nicking stem loop sequence of the AAV2 ITR sequence (SEQ ID NO: 221). In one embodiment, the nicking stem loop for AAV5 is swapped with (replaces all of) the native nicking stem loop sequence of the AAV1 ITR sequence (SEQ ID NO: 222). In one embodiment, the nicking stem loop for AAV5 is swapped with (replaces all of) the native nicking stem loop sequence of the AAV3 ITR sequence (SEQ ID NO: 223). In one embodiment, the nicking stem loop for AAV5 is swapped with (replaces all of) the native nicking stem loop sequence of the AAV4 ITR sequence (SEQ ID NO: 224). In one embodiment, the nicking stem loop for AAV5 is swapped with (replaces all of) the native nicking stem loop sequence of the AAV7 ITR sequence (SEQ ID NO: 225). In one embodiment, the nicking stem loop for Bovine AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV2 ITR sequence (SEQ ID NO: 226). In one embodiment, the nicking stem loop for Bovine AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV1 ITR sequence (SEQ ID NO: 227). In one embodiment, the nicking stem loop for Bovine AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV3 ITR sequence (SEQ ID NO: 228). In one embodiment, the nicking stem loop for Bovine AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV4 ITR sequence (SEQ ID NO: 229). In one embodiment, the nicking stem loop for Bovine AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV7 ITR sequence (SEQ ID NO: 230). In one embodiment, the nicking stem loop for Snake AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV2 ITR sequence (SEQ ID NO: 231). In one embodiment, the nicking stem loop for Snake AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV1 ITR sequence (SEQ ID NO: 232). In one embodiment, the nicking stem loop for Snake AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV3 ITR sequence (SEQ ID NO: 233). In one embodiment, the nicking stem loop for Snake AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV4 ITR sequence (SEQ ID NO: 234). In one

embodiment, the nicking stem loop for Snake AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV7 ITR sequence (SEQ ID NO: 235). In one embodiment, the nicking stem loop for Avian AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV2 ITR sequence (SEQ NO: 236). In one embodiment, the nicking stem loop for Avian AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV1 ITR sequence (SEQ ID NO: 237). In

one embodiment, the nicking stem loop for Avian AAV is swapped with (replaces all of) the native nicking stem loop sequence of the _AAV3 ITR sequence (SEQ ID NO: 238). In one embodiment, the nicking stem loop for Avian AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV4 ITR sequence (SEQ ID NO: 239). In one embodiment, the nicking stem loop for Avian AAV is swapped with (replaces all of) the native nicking stem loop sequence of the AAV7 ITR sequence (SEQ ID NO: 240).

TABLE 5

Chimeric Rep Stem Loop Sequences			
Acceptor ITR	Donor Nicking Stem Loop	Chimeric Sequence	SEQ ID NO
AAV5	AAV2	TACAAAACCCCTTGCTTGATGGAGTTGGCCACTCCCTCTCTGTGCGCTTCGCTCGCTCGCTGGCTCGTTTGGGGGGGCGACGCGCCAGAGGGCCGTCGTCGTCGACGCTCTTTGAGCTGCCACCCCCCAACGAGCCAGCGAGCGAGCGAACGCGACAGGGGGAGAG	218
AAV5	AAV4	TACAAAACCCCTTGCTTGATGGAGTTGGCCACATTACTCTCTGTGCGCTTCGCTCGCTCGCTGGCTCGTTTGGGGGGGCGACGCGCCAGAGGGCCGTCGTCGTCGACGCTCTTTGAGCTGCCACCCCCCAACGAGCCAGCGAGCGAGCGAACGCGACAGGGGGAGAG	219
AAV4	AAV2	TGGCAAACCCAGATGATGGAGTTGGCCACTCCCTCTATGCGCGCTCGCTCACTCACTCGGCCCTGGAGACCAAGGTCTCCAGACTGCGGCCCTCTGGCCGGCAGGGCCGAGTGA GTGAGCGAGCGCGCATAGAGGGAGTGGCCAA	220
AAV2	AAV5	AGGAACCCCTAGTGAGAGTGTGGCACTCTCCCCCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCCGGGCAAAGCCC GGGCGTCGGGCGACCTTTGGTCGCCCGGCCCTCAGTGAG CGAGCGAGCGCGCAGAGAGGGAGTGGCCAA	221
AAV1	AAV5	TTACCCCTAGTGAGACTTGTGGCACTCTCCCCCTGCGCGCTCGCTCGCTCGGTGGGGCCGGCAGAGCAGAGCTCTG CCGTCTGCGGACCTTTGGTCCGCGAGGCCCAACCGAGCG AGCGAGCGCGCAGAGAGGGAGTGGCCAA	222
AAV3	AAV5	GCCATACCTCTAGTGAGAGTGTGGCACTCTCCCCCATGCGCACTCGCTCGCTCGGTGGGGCCGGCAGTGCACAAAGCA CGTCCGTCTGGCGACCTTTGGTCGCCAGGCCCAACCGA GCGAGCGAGTGCATAGAGGGAGTGGCCAA	223
AAV4	AAV5	TGGCAAACCCAGATGAGAGTGTGGCACTCTCCCCCATGCGCGCTCGCTCACTCACTCGGCCCTGGAGACCAAGGTCTCCAGACTGCGGCCCTCTGGCCGGCAGGGCCGAGTGAG TGAGCGAGCGCGCATAGAGGGAGTGGCCAA	224
AAV7	AAV5	CGCGGTACCCCTAGTGAGAGTGTGGCACTCTCCCCCATGCGCGCTCGCTCGCTCGGTGGGGCCGGCAGAGCAGAG CTCTGCCGTCTGCGGACCTTTGGTCCGCGAGGCCCAAC GAGCGAGCGAGCGCGCATAGAGGGAGTGGCCAA	225
AAV2	Bovine	AGGAACCCCTAGAGAGTGTGGCACTCCCCCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCCGGGCAAAGCCC GGGCGTCGGGCGACCTTTGGTCGCCCGGCCCTCAGTGAG CGAGCGAGCGCGCAGAGAGGGAGTGGCCAA	226
AAV1	Bovine	TTACCCCTAGAGAGTGTGGCACTCCCCCTGCGCGCTCGCTCGCTCGGTGGGGCCGGCAGAGCAGAGCTCTG CCGTCTGCGGACCTTTGGTCCGCGAGGCCCAACCGAGCG AGCGAGCGCGCAGAGAGGGAGTGGCCAA	227
AAV3	Bovine	GCCATACCTCTAGAGAGTGTGGCACTCCCCCATGCGCACTCGCTCGCTCGGTGGGGCCGGCAGTGCACAAAGCA CGTCCGTCTGGCGACCTTTGGTCGCCAGGCCCAACCGA GCGAGCGAGTGCATAGAGGGAGTGGCCAA	228
AAV4	Bovine	TGGCAAACCCAGAGAGTGTGGCACTCCCCCATGCGCGCTCGCTCACTCACTCGGCCCTGGAGACCAAGGTCT	229

TABLE 5-continued

Chimeric Rep Stem Loop Sequences			
Acceptor ITR	Donor Nicking Stem Loop	Chimeric Sequence	SEQ ID NO
		TCCAGACTGCCGGCCTCTGGCCGGCAGGGCCGAGTGAG TGAGCGAGCGCGCATAGAGGGAGTGGCCAA	
AAV7	Bovine	CGCGGTACCCCTAGAGAGAGTGTGGCACTCCCCCAT GCGCGCTCGCTCGCTCGGTGGGGCCGGCAGAGCAGAG CTCTGCCGTCTGCGGACCTTTGGTCCGCAGGCCCCACC GAGCGAGCGAGCGCGCATAGAGGGAGTGGCCAA	230
AAV2	Snake	AGGAACCCCTAGGAGCAGGGCACTCGCCCCA CTGCGC GCTCGCTCGCTCACTGAGGCCGCGCCGGGCAAAGCCCGG GCGTCGGGCGACCTTTGGTCCGCCGCGCTCAGTGAGCG AGCGAGCGCGCAGAGAGGGAGTGGCCAA	231
AAV1	Snake	TTACCCCTTAGAGAGAGTCTGGCACTCCCCCCTGCGC GCTCGCTCGCTCGGTGGGGCCGGCAGAGCAGAGCTCTG CCGTCTGCGGACCTTTGGTCCGCAGGCCCCACCGAGCG AGCGAGCGCGCAGAGAGGGAGTGGCCAA	232
AAV3	Snake	GCCATACCTCTAGGAGCAGGGCACTCGCCCCAATGCGC ACTCGCTCGCTCGGTGGGGCCGGACGTGCAAGCACGT CCGTCTGGCGACCTTTGGTCCGCAGGCCCCACCGAGCG AGCGAGTGCGCATAGAGGGAGTGGCCAA	233
AAV4	Snake	TGGCAAACCAGAGAGCAGGGCACTCGCCCCAATGCGC GCTCGCTCACTCACTCGGCCCTGGAGACCAAAGGTCTC CAGACTGCCGGCCTCTGGCCGGCAGGGCCGAGTGAGT GAGCGAGCGCGCATAGAGGGAGTGGCCAA	234
AAV7	Snake	CGCGGTACCCCTAGGAGCAGGGCACTCGCCCCAATGCGC CGCTCGCTCGCTCGGTGGGGCCGGCAGAGCAGAGCTCT GCCGTCTGCGGACCTTTGGTCCGCAGGCCCCACCGAGC GAGCGAGCGCGCATAGAGGGAGTGGCCAA	235
AAV2	Avian	AGGAACCCCTAGTCACCGGAGTGCTGGCCAGTCTGCGC GCTCGCTCGCTCACTGAGGCCGCGCCGGGCAAAGCCCGG GCGTCGGGCGACCTTTGGTCCGCCGCGCTCAGTGAGCG AGCGAGCGCGCAGAGAGGGAGTGGCCAA	236
AAV1	Avian	TTACCCCTAGTCACCGGAGTGCTGGCCAGTCTGCGCGC TCGCTCGCTCGGTGGGGCCGGCAGAGCAGAGCTCTGCC GTCTGCGGACCTTTGGTCCGCAGGCCCCACCGAGCGAG CGAGCGCGCAGAGAGGGAGTGGCCAA	237
AAV3	Avian	GCCATACCTCTAGTCACCGGAGTGCTGGCCAGTATGCGC CACTCGCTCGCTCGGTGGGGCCGGACGTGCAAGCACG TCCGTCTGGCGACCTTTGGTCCGCAGGCCCCACCGAGC GAGCGAGTGCGCATAGAGGGAGTGGCCAA	238
AAV4	Avian	TGGCAAACCAGATCACCGGAGTGCTGGCCAGTATGCGC GCTCGCTCACTCACTCGGCCCTGGAGACCAAAGGTCTC CAGACTGCCGGCCTCTGGCCGGCAGGGCCGAGTGAGT GAGCGAGCGCGCATAGAGGGAGTGGCCAA	239
AAV7	Avian	CGCGGTACCCCTAGTCACCGGAGTGCTGGCCAGTATGC GCGCTCGCTCGCTTCGGTGGGGCCGGCAGAGCAGAGCTC TGCCGTCTGCGGACCTTTGGTCCGCAGGCCCCACCGAG CGAGCGAGCGCGCATAGAGGGAGTGGCCAA	240

IV. ENGINEERED REP PROTEIN AND POLYNUCLEOTIDES

[0139] The present disclosure also provides a polypeptide comprising an engineered Rep protein and a polynucleotide encoding the polypeptide comprising the engineered Rep protein. The engineered Rep protein of the present disclosure comprises at least one amino acid that is altered or modified as compared to the corresponding wild-type Rep protein. In one embodiment, the engineered Rep protein

comprises one or more amino acids involved with DNA binding that are altered or modified as compared to the corresponding wild-type Rep protein. The engineered Rep protein has decreased binding to its cognate Rep binding sequence region and, as a consequence, promotes the formation of a self-complementary viral genome during replication.

[0140] Rep protein is an essential viral protein that catalyzes several reactions during viral replication. The N-terminal domain of Rep protein, consisting of about the first

200 amino acids, has site-specific endonuclease and sequence-specific DNA binding activities. The C-terminal region of Rep protein has ATPase and 3'-5' helicase activities. The Rep protein interacts with the viral ITRs, which serve as viral origins of replication, in the performance of these various functions. Within the ITRs are two sequences required for replication: a Rep binding sequence region consisting of several direct repeats of a 5'-GCTC-3' motif and a terminal resolution site, or nicking site. Viral replication requires Rep binding at the Rep binding sequence region and subsequent cleavage of the ssDNA at the nicking site to generate the 3'-OH group which allows conversion of the viral ends into linear duplex DNA. An additional Rep binding sequence comprising a CTTTG motif, located in the loop region between the B and B' sequences of the ITR hairpin, aids in determining the directionality of parvoviral genome replication. The endonuclease domain of the Rep protein recognizes the nicking site substrate in the context of an ssDNA stem loop secondary structure generated by the Rep helicase activity.

[0141] The Rep protein is altered so as to decrease its binding with its cognate Rep binding sequence region by altering or modifying amino acid residues involved in DNA binding. Previous studies analyzing the crystal structure of AAV5 Rep protein bound to the AAV5 Rep binding sequence region indicate which residues play a role in DNA binding. (Hickman et al 2004 The Nuclease Domain of Adeno-Associated Virus Rep Coordinates Replication Initiation Using Two Distinct DNA Recognition Interfaces, Molecular Cell 13:403-414, the contents of which are referenced herein in their entirety). The N-terminal domain of Rep protein has determinants for ITR binding and has been shown to specifically bind ITR sequences (Owens 1993, Yoon, 2001, Hickman). In the Hickman studies, a crystal structure of an AAV5 N-terminal Rep protein fragment (AA 1-197) bound to a 26 by double-stranded sequence containing the AAV5 Rep binding sequence region was analyzed. Three-dimensional modeling showed that the stable protein-DNA complex contained five Rep protein monomers bound independently to one molecule of the AAV5 Rep binding sequence region, which adopted an overall β -DNA structure. Further examination of the crystal structure showed that the five monomers bind to the Rep binding sequence region and spiral around the DNA axis, off-set from one another by four base pairs. The studies further revealed that the structural elements of the Rep protein integral to recognition and binding of the repeated GCTC consensus motif in the Rep binding sequence region are the surface loop sequence between the β 4 and β 5 strands (referred to as the j34/05 loop: residues 135-144) and the α -helix C sequence (residues 101-118), which are located along one edge of the central 13 sheet.

[0142] FIG. 7 is a schematic depicting a three-dimensional model showing Rep protein binding to the Rep binding domain of the AAV5 ITR. As shown in FIG. 7, several amino acid residues in the β 4/ β 5 loop (amino acid residues K135, K137, and N142) and in the α -helix C (S101, G105, S109, and Q110) contact the phosphodiester backbone in the major groove structure of the Rep binding sequence region. FIG. 7 also shows the amino acid residues in the α -helix C (R106, K137, and K138) and an amino acid residue in the β 4/ β 5 loop (K137) that contact one or more nucleosides in the Rep binding sequence region. Amino acid residue K137 contacts both the phosphodiester backbone and one or more nucleosides

in the Rep binding sequence region. The identified amino acid residues contribute to Rep protein binding to its cognate Rep binding sequence region, with those residues contacting the phosphodiester backbone contributing to stabilization of the protein-DNA interaction and those residues contacting one or more nucleosides contributing to the specificity of the interaction.

[0143] An alignment of the polynucleotide sequences encoding Rep protein from different AAV serotypes (AAV5, AAV2Rep68, AAV2Rep78, AAV2Rep40, AAV2Rep52, AAV1, AAV3, AAV4, AAV7, and AAV8) and AAV species (Avian AAVDA-1, Avian AAVVR865, Bat, and Bovine), shown in FIG. 8, shows that the identified amino acids discussed above are conserved across AAV serotypes and AAV species. An analysis of the correlation between structural features of the Rep binding sequence region in the ITR and the degree of conservation of the Rep protein sequence was performed in silico. The combination of sequence and structural analysis of Rep proteins confirms that the identified conserved residues of the α -helix C and β -sheet 4/ β -sheet 5 loop are involved in DNA binding and are the appropriate target residues for alteration and/or modification. Table 6 provides a list of target residues within structural elements of the Rep protein that contact the phosphodiester backbone contributing to stabilized protein-DNA interaction as shown in FIG. 7. Table 7 provides a list of target residues in the Rep protein that contact nucleosides of either DNA strand (the complementary strand is noted with a ') contributing to specific protein-DNA interaction as shown in FIG. 7. The residue numbering used here and in Tables 4 and 5 are that of AAV5. Corresponding residues in other AAV serotypes and species can be obtained using the sequence alignment shown in FIG. 8. Of the ten residues that form direct side chain interactions with the Rep binding sequence region, seven are strictly conserved among serotypes AAV2-6. There are three consecutive amino acids in the β 4/ β 5 loop that are highly conserved- Gly-139, Gly-140, and Ala-141 (in AAV5), which correlate to three sequential Glycines in AAV2-6. Other conserved residues in the β 4/ β 5 loop are Asp-142, Lys-135, and Lys-137. Conserved residues in the α -helix C are Ser101, Met-102, Gly105, Arg-106, Ser-109, Gln-110, Lys137, and Lys138. Arg-106 and Lys-137 provide important base contacts and Lys-135 forms a salt bridge to the phosphate backbone.

TABLE 6

Residues Contacting the Phosphodiester Backbone Contributing to Stabilized Protein-DNA Interaction	
Residue	Structural Element
S101	α helix C
G105	α helix C
S109	α helix C
Q110	α helix C
K135	β 4/ β 5 loop
K137	β 4/ β 5 loop
N142	β 4/ β 5 loop

TABLE 7

Residues Contacting Nucleosides Contributing to Specific Protein-DNA Interaction	
Residue	Nucleoside
R106	T14, C15
K137	G18', C9, G8
K138	G8
G139	A17'

[0144] The engineered Rep protein of the present disclosure comprises one or more amino acids that are altered or modified as compared to the corresponding wild-type Rep protein, in which modification(s) and/or alteration(s) results in decreased binding to the Rep binding sequence region. The engineered Rep protein can be derived from an AAV Rep protein. In one embodiment, the engineered Rep protein is derived from an AAV Rep protein selected from an AAV1, AAV2, AAV3, AAV4, AAV5, AAV7, AAV8, Avian AAVDA-1, Avian AAV VR865, Bat, Snake, and Bovine AAV Rep protein (SEQ :ID NO: 265-279). In one embodiment, the engineered Rep protein is derived from an AAV Rep protein selected from art AAV6, AANT9, AAV10, AAV12, AAVrh8, AAVrh10, and AAV-DJ Rep protein. In one embodiment, the AAV Rep protein from which the engineered Rep protein is derived is selected from AAV2 Rep68, AAV2 Rep78, AAV2 Rep40, AAV2 Rep52, AAV1 Rep, AAV3 Rep, AAV4 Rep, AAV7 Rep, AAV 8 Rep, Avian Rep, Bat Rep, or Bovine Rep. In some embodiments, the AAV Rep protein from which the engineered Rep protein is derived is Rep protein 68 or Rep protein 78. AAV Rep78 proteins are described in International Patent Publication No. WO2001032711, WO2007130519, and WO2007148971, the contents of which are herein incorporated by reference in their entirety.

[0145] The engineered Rep protein can be derived from a naturally-occurring parvoviral Rep protein or a naturally-occurring variant parvoviral Rep protein. Alternatively, the Rep protein can be derived from a non-naturally Rep protein, including, for example, a synthetic or artificial Rep protein. The Rep protein can be altered by insertions, deletions, and/or substitutions of amino acid residues. In one embodiment, the engineered Rep protein has at least 85% sequence identity with any of the AAV Rep proteins disclosed herewith. In another embodiment, the engineered Rep protein has at least 90% sequence identity with any of the AAV Rep proteins disclosed herewith, including 90% 91%, 92%, 93%, 94%, 95%, 96%, 97% 98%, and 99% identity with the AAV Rep proteins.

[0146] In one embodiment, the one or more altered or modified amino acid residues in the engineered Rep protein is an amino acid residue that in its wild-type form (i.e., unaltered or unmodified form) contacts a Rep binding sequence region. In one embodiment, one of the altered or modified amino acid residues in the engineered Rep protein is an amino acid residue that in its wild-type (unaltered or unmodified) form contacts a Rep binding sequence region. In another embodiment two of the altered or modified amino acid residues are amino acid residues that in their wild-type form contact a Rep binding sequence region. In another embodiment, three of the altered or modified amino acid residues are amino acid residues that in their wild-type form contact a Rep binding sequence region. In another embodiment, four of the modified amino acid residues are amino acid residues that in their wild-type form contact a Rep

binding sequence region. In another embodiment, about 10% to about 50% of the amino acid residues that in their wild-type form (unaltered or unmodified form) contact a Rep binding sequence region are altered or modified in the engineered Rep protein. In one embodiment, at least 10% of the amino acid residues that in their wild-type form contact a Rep binding sequence region are altered or modified. In one embodiment, at least 20%, 30%, 40% or 50% of the amino acid residues that in their wild-type form contact a Rep binding sequence region are altered or modified.

[0147] In one embodiment, the one or more altered or modified amino acid residue(s) of the engineered Rep protein is an amino acid residue(s) that in its wild-type form (e.e., unaltered or unmodified) contacts the phosphodiester backbone of a nucleotide sequence comprising the Rep binding sequence region. In another embodiment, the one or more altered or modified amino acid residue(s) of the engineered Rep protein is an amino acid residue(s) that in its wild-type form contacts at least one nucleoside of a nucleotide sequence comprising a Rep binding sequence region. In another embodiment, the one or more altered or modified amino acid residue(s) of the engineered Rep protein is an amino acid residue(s) that in its wild-type form contacts the phosphodiester backbone and at least one nucleoside of a nucleotide sequence comprising a Rep binding sequence region. In one embodiment, the one or more altered or modified amino acid residue(s) of the engineered Rep protein is an amino acid residue that in its wild-type form stabilizes the binding of Rep protein to a Rep binding sequence region. In one embodiment, the one or more altered or modified amino acid residue(s) in the engineered Rep protein is an amino acid residue(s) that in its wild-type form is located in the α -helix C of the Rep protein. In one variation of this embodiment, the altered or modified amino acid residue(s) further contacts the phosphodiester backbone and/or at least one nucleoside of a nucleotide sequence comprising a Rep binding sequence region. In another embodiment, the one or more altered or modified amino acid residue(s) in the engineered Rep protein is an amino acid residue(s) that in its wild-type form is located in the β 4/ β 5 loop of the Rep protein. In one variation of this embodiment, the altered or modified amino acid residue(s) further contacts the phosphodiester backbone and/or at least one nucleoside of a nucleotide sequence comprising a Rep binding sequence region. In another embodiment, the one or more altered or modified amino acid residue(s) is selected from the amino acid residue that is or corresponds with: (1) Serine-101; (2) Glycine-105; (3) Arginine-106; (4) Serine-109; (5) Glutamine-110; (6) Lysine-135; (7) Lysine-137; (8) Lysine-138; (9) Glycine-139; and (10) Asparagine-142. In another embodiment, the one or more altered or modified amino acid residue(s) is selected from the amino acid residue that is or corresponds with: (1) Glycine-140; (2) Alanine-141; and (3) Met-102.

[0148] In one embodiment, the one or more altered or modified amino acid residue(s) of the engineered Rep protein is modified by conservative substitution with a different amino acid having similar biochemical properties as the wild-type residue. For example, acidic amino acid residues, e.g., Aspartic acid and Glutamic acid can be substituted with one another. Basic amino acids, such as Lysine, Arginine, and Histidine, can be substituted with one another. Hydrophobic aliphatic amino acids, such as Valine, Isoleucine, Leucine, and Alanine, can be substituted with one another.

Hydrophobic aromatic amino acids, such as Phenylalanine, Tyrosine, and Tryptophan, can be substituted with one another. Neutral polar amino acids, such as Asparagine, Cysteine, Glutamine, Methionine, Serine, and Threonine, can be substituted with one another, Hydrophilic amino acids, such as Alanine, Proline, Glycine, Glutamic acid, Aspartic acid, Glutamine, Asparagine, Serine, and Threonine, can be substituted with one another. These are non-limiting examples of conservative amino acid substitution and are not meant to exclude other combinations of amino acid substitutions.

[0149] In one embodiment, the engineered Rep protein comprises one or more alterations in amino acid residue(s) corresponding to the following amino acid residues in wild-type AAVS Rep protein sequence selected from: (1) amino acid corresponding to Glycine-139 substituted with a Proline, Alanine, or Serine residue; (2) amino acid corresponding to Glycine-140 substituted with a Proline, Alanine, or Serine residue; (3) amino acid corresponding to Alanine-141 substituted with a Serine, Glycine, Threonine, Cysteine, or Valine residue; (4) amino acid corresponding to Lysine-138 substituted with a Arginine, Glutamine, Glutamic acid, Asparagine, or Serine residue; (5) amino acid corresponding to Methionine-102 substituted with a Leucine, Isoleucine, Glutamine, Valine, or Phenylalanine residue; (6) amino acid corresponding to Serine-109 substituted with a Threonine, Alanine, Asparagine, Aspartic acid, Glutamine, Glutamic acid, Glycine, Lysine, or Threonine residue; (7) amino acid corresponding to Glutamine-110 substituted with an Arginine, Asparagine, Aspartic acid, Histidine, Lysine, or Serine residue; (8) amino acid corresponding with Asparagine-142 substituted with an Aspartic acid, Serine, Glutamine, or Glutamic acid residue; (9) amino acid corresponding to Lysine-137 substituted with a Arginine, Glutamine, Glutamic acid, Asparagine, or Serine residue; (10) amino acid corresponding to Lysine-135 substituted with a Arginine, Glutamine, Glutamic acid, Asparagine, or Serine residue; (11) amino acid corresponding to Ser-101 substituted with a Threonine, Alanine, Asparagine, Aspartic acid, Glutamine, Glutamic acid, Glycine, Lysine, or Threonine residue; (12) amino acid corresponding to Glycine-105 substituted with a Proline, Alanine, or Serine residue; (13) amino acid corresponding to Arginine 106 substituted with a Lysine or Histidine residue; and (14) any combination of these alterations. In one embodiment, the engineered Rep protein comprises one or more alterations in amino acid residue(s) corresponding to the following amino acid residues in wild-type AAV5 Rep protein sequence selected from: (1) amino acid corresponding to Lysine-135 substituted with Glycine; (2) amino acid corresponding to Lysine-135 substituted with Threonine. (3) amino acid corresponding to Asparagine 142 substituted with Glycine, (4) amino acid corresponding to Asparagine 142 substituted with Threonine, (5) amino acid corresponding to Lysine-135 substituted with Glycine and amino acid corresponding to Asparagine 142 substituted with Glycine; (6) amino acid corresponding to Lysine-135 substituted with Glycine and amino acid corresponding to Asparagine 142 substituted with Threonine; (7) amino acid corresponding to Lysine-135 substituted with Threonine and amino acid corresponding to Asparagine 142 substituted with Glycine; and (8) amino acid corresponding to Lysine-135 substituted with Threonine and amino acid corresponding to Asparagine 142. substituted with Threonine. The amino acid residue position numbering

in the listed substitutions corresponds with the position numbering found in the wild-type AAV5 Rep protein (SEQ ID NO: 269). Corresponding amino acid residues at these positions for other AAV serotypes and species of Rep proteins can be found in the sequence alignments provided in FIG. 8. Corresponding amino acid residues at these positions for additional parvoviral serotypes and species of Rep proteins can be determined by sequence alignment analysis using the methods described herein or otherwise known in the art.

[0150] In any of the above-described embodiments, the altered or modified amino acid residue(s) of the engineered Rep protein has decreased contact with a wild-type Rep binding sequence region relative to the amount of contact a wild-type Rep protein has with the wild-type Rep binding sequence region. In another embodiment, the modified amino acid residue(s) of the engineered Rep protein has decreased contact with an engineered Rep binding sequence region relative to the amount of contact a wild-type Rep protein has with the engineered Rep binding sequence region. In another embodiment, the modified amino acid residue(s) of the engineered Rep protein has decreased contact with an engineered Rep binding sequence region relative to the amount of contact it has with a wild-type Rep binding sequence region. In any of these embodiments, the engineered Rep binding sequence region can be an engineered Rep binding sequence region provided in the present disclosure.

[0151] In another embodiment, the one or more altered or modified amino acid residue(s) of the engineered Rep protein is modified by non-conservative substitution with a different amino acid residue.

[0152] In any of these embodiments in which the altered or modified amino acid residue of the engineered Rep protein is an amino acid residue that in its wild-type form contacts a Rep binding sequence region, the Rep binding sequence region can be a wild-type Rep binding sequence region or an engineered Rep binding sequence region. Thus, in one embodiment in which the altered or modified amino acid residue of the engineered Rep protein is an amino acid residue that in its wild-type form contacts a Rep binding sequence region, the Rep binding sequence region is a wild-type Rep binding sequence region. The wild-type Rep binding sequence region can be an AAV Rep binding sequence region. In certain specific embodiments, the AAV is selected from any of the AAV serotypes and species provided herein. In one embodiment wherein the altered or modified amino acid residue of the engineered Rep protein is an amino acid residue that in its wild-type form contacts a Rep binding sequence region, the Rep binding sequence region is selected from any of SEQ ID NOS: 1-20, 21-66, and/or 79-197. Also, in any of these embodiments in which the altered or modified amino acid residue of the engineered Rep protein is an amino acid residue that in its wild-type form contacts a Rep binding sequence region, the Rep binding sequence region can be an engineered Rep binding sequence region. The engineered Rep binding sequence region can be derived from art AAV Rep binding sequence region. In certain specific embodiments, the AAV is selected from any of the AAV serotypes and species provided herein. In another embodiment wherein the altered or modified amino acid residue of the engineered Rep protein is an amino acid residue that in its wild-type form contacts an engineered. Rep binding sequence region, the engineered

Rep binding sequence region is selected from any of the engineered Rep binding sequence regions provided in the present disclosure. In some embodiments the engineered Rep binding sequence region is selected from any of SEQ ID NOs: 1-20, 21-66, and/or 79-197.

[0153] In one embodiment, the modified amino acid residue of the engineered Rep protein is modified by chemical, enzymatic, or other post-translational modification.

[0154] The engineered Rep protein has decreased binding affinity with a Rep binding sequence region as a result of the one or more alterations and/or modifications. In one embodiment, the engineered Rep protein has decreased binding affinity with a wild-type Rep binding sequence region relative to the binding affinity of a wild-type Rep protein with the wild-type Rep binding sequence region as a result of the one or more alterations and/or modifications. In another embodiment, the engineered Rep protein has decreased binding affinity with an engineered Rep binding sequence region relative to the binding affinity of a wild-type Rep protein with the engineered Rep binding sequence region. In another embodiment, the engineered Rep protein has decreased binding affinity with an engineered Rep binding sequence region relative to the binding affinity it has with a wild-type Rep binding sequence region. In any of these embodiments, the binding affinity is decreased about 2-fold to about 20-fold. In another embodiment, the binding affinity is decreased at least 2-fold, at least 5-fold, at least 10-fold, or at least 20-fold. In any of these embodiments, the Rep binding sequence region is a wild-type Rep binding sequence region, for example, an AAV Rep binding sequence region, such as an AAV Rep binding sequence region selected from any of the AAV serotypes and species provided herein. In any of these embodiments, the Rep binding sequence region is an engineered Rep binding sequence region, for example, an engineered Rep binding sequence region derived from a wild-type AAV Rep binding sequence region, such as an AAV Rep binding sequence region selected from any of the AAV serotypes and species provided herein.

V. PAYLOAD: TRANSGENES, POLYPEPTIDE-ENCODING POLYNUCLEOTIDES AND/OR MODULATORY NUCLEIC ACIDS

[0155] The payload construct vector of the present invention comprises a nucleic acid sequence encoding at least one “payload molecule.” As used herein, a “payload molecule” refers to a transgene, a polynucleotide encoding a polypeptide or a modulatory nucleic acid. The payload molecule may comprise any nucleic acid encoded in the viral genome produced in accordance with the present invention for expression in a target cell transduced or contacted with the viral particle.

[0156] According to the present invention, the payload construct vector encodes a “payload construct.” As used herein, a “payload construct” is a polynucleotide sequence encoding at least a payload molecule and sufficient ITR sequence to allow for expression of the payload molecule in a cell transduced with the viral particle.

[0157] The payload molecule may comprise a polypeptide, an RNA molecule, or any other gene product that is desired for expression in the target cell. The payload construct may comprise a combination of coding and non-coding nucleic acid sequences.

[0158] In one embodiment, the payload construct vector comprises more than one nucleic acid sequences encoding more than one payload molecule of interest. In such an embodiment, a payload construct vector encoding more than one payload molecule may be replicated and packaged into a viral particle. A target cell transduced with a viral particle comprising more than one payload molecule may express each of the payload molecules in a single cell.

[0159] In some embodiments, the payload construct vector sequence may encode a coding or non-coding RNA.

[0160] Where the payload construct vector sequence encodes a polypeptide, the polypeptide may be a peptide or protein. A protein encoded by the payload construct vector sequence may comprise a secreted protein, an intracellular protein, an extracellular protein, and/or a membrane protein. The encoded proteins may be structural or functional. Proteins encoded by the payload construct vector or payload construct include, but are not limited to, mammalian proteins. The viral vectors encoding polypeptides (e.g., mRNA) of the invention may be used in the fields of human disease, antibodies, viruses, veterinary applications and a variety of in vivo and in vitro settings.

[0161] In some embodiments, the viral particles are useful in the field of medicine for the treatment, palliation or amelioration of conditions or diseases such as, but not limited to, blood, cardiovascular, CNS, dermatology, endocrinology, genetic, genitourinary, gastrointestinal, musculoskeletal, oncology, and immunology, respiratory, sensory and anti-infective,

[0162] In some embodiments, viral particles in accordance with the present invention may be used for the treatment of disorders, and/or conditions, including but not limited to one or more of the following: autoimmune disorders (e.g. diabetes, lupus, multiple sclerosis, psoriasis, rheumatoid arthritis); inflammatory disorders (e.g. arthritis, pelvic inflammatory disease); neurological disorders (e.g. Alzheimer’s disease, Huntington’s disease; autism; Parkinson’s disease); cardiovascular disorders (e.g. atherosclerosis, hypercholesterolemia, thrombosis, clotting disorders, angiogenic disorders such as macular degeneration); proliferative disorders (e.g. cancer, benign neoplasms); respiratory disorders (e.g. chronic obstructive pulmonary disease); digestive disorders (e.g. inflammatory bowel disease, ulcers); musculoskeletal disorders (e.g. fibromyalgia, arthritis); endocrine, metabolic, and nutritional disorders (e.g. diabetes, osteoporosis); urological disorders (e.g. renal disease); psychological disorders (e.g. depression, schizophrenia); skin disorders (e.g. wounds, eczema); blood and lymphatic disorders (e.g. anemia, hemophilia); etc.

[0163] In some embodiments, the payload construct encodes a messenger RNA (mRNA). As used herein, the term “messenger RNA” (mRNA) refers to any polynucleotide which encodes a polypeptide of interest and which is capable of being translated to produce the encoded polypeptide of interest in vitro, in vivo, in situ or ex vivo.

[0164] Traditionally, the basic components of an mRNA molecule include at least a coding region, a 5'UTR, a 3'UTR, a 5' cap and a poly-A tail. According to the present invention, payload constructs encoding mRNA may comprise a coding region only. They may also comprise a coding region and at least one UTR. They may also comprise a coding region, 3'UTR and a poly-A tail.

[0165] In one embodiment the polypeptide encoded by the payload construct is between 50-5000 amino acids in length.

In some embodiments the protein encoded is between 50-2000 amino acids in length. In some embodiments the protein encoded is between 50-1500 amino acids in length. In some embodiments the protein encoded is between 50-1000 amino acids in length. In some embodiments the protein encoded is between 50-800 amino acids in length. In some embodiments the protein encoded is between 50-600 amino acids in length. In some embodiments the protein encoded is between 50-400 amino acids in length. In some embodiments the protein encoded is between 50-200 amino acids in length. In some embodiments the protein encoded is between 50-100 amino acids in length.

[0166] In some embodiments the peptide encoded by the payload construct is between 4-50 amino acids in length. In one embodiment, the shortest length of a region of the payload molecule of the present invention encoding a peptide can be the length that is sufficient to encode for a tetrapeptide, a pentapeptide, a hexapeptide, a heptapeptide, an octapeptide, a nonapeptide, or a decapeptide. In another embodiment, the length may be sufficient to encode a peptide of 2-30 amino acids, e.g. 5-30, 10-30, 2-25, 5-25, 10-25, or 10-20 amino acids. The length may be sufficient to encode for a peptide of at least 11, 12, 13, 14, 15, 17, 20, 25 or 30 amino acids, or a peptide that is no longer than 50 amino acids, e.g. no longer than 35, 30, 25, 20, 17, 15, 14, 13, 12, 11 or 10 amino acids.

[0167] An RNA encoded by the payload construct may comprise an mRNA, tRNA, rRNA, tmRNA, miRNA, RNAi, siRNA, piRNA, shRNA antisense RNA, double stranded RNA, snRNA, snoRNA, and long non-coding RNA (lncRNA). Examples of such lncRNA molecules and RNAi constructs designed to target such lncRNA any of which may be encoded in the payload constructs are taught in International Publication, WO2012/018881 A2, the contents of which are incorporated herein by reference in their entirety.

[0168] In one embodiment, the payload construct encodes a microRNA or miRNA as the payload molecule. These payload molecules are also referred to as modulatory nucleic acid payloads.

[0169] microRNAs (or miRNA) are 19-25 nucleotide long noncoding RNAs that bind to the 3'UTR of nucleic acid molecules and down-regulate gene expression either by reducing nucleic acid molecule stability or by inhibiting translation. The payload constructs of the invention may comprise one or more microRNA target sequences, microRNA sequences, or microRNA seeds. Such sequences may correspond to any known microRNA such as those taught in US Publication US200510261218 and US Publication US2005/0059005, the contents of which are incorporated herein by reference in their entirety.

[0170] A microRNA sequence comprises a "seed" region, i.e., a sequence in the region of positions 2-8 of the mature microRNA, which has perfect Watson-Crick complementarity to the miRNA target sequence. A microRNA seed may comprise positions 2-8 or 2-7 of the mature microRNA. In some embodiments, a microRNA seed may comprise 7 nucleotides (e.g., nucleotides 2-8 of the mature microRNA), wherein the seed-complementary site in the corresponding miRNA target is flanked by an adenine (A) opposed to microRNA position 1. In some embodiments, a microRNA seed may comprise 6 nucleotides (e.g., nucleotides 2-7 of the mature microRNA), wherein the seed-complementary site in the corresponding miRNA target is flanked by an adenine (A) opposed to microRNA position 1. See for example,

Grimson A, Farh K K, Johnston W K, Garrett-Engle P, Lim L P, Bartel D P; *Mol Cell*. 2007 Jul. 6; 27(1):91-105; each of which is herein incorporated by reference in their entirety. The bases of the microRNA seed have complete complementarity with the target sequence:

[0171] A payload molecule may comprise polypeptides that serve as marker proteins to assess cell transformation and expression, fusion proteins, polypeptides having a desired biological activity, gene products that can complement a genetic defect, RNA molecules, transcription factors, and other gene products that are of interest in regulation and/or expression. A payload molecule may comprise nucleotide sequences that provide a desired effect or regulatory function (e.g., transposons, transcription factors). A payload molecule may comprise, but is not limited to: hormone receptors (e.g., mineral corticosteroid, glucocorticoid, and thyroid hormone receptors); intramembrane proteins (e.g., TM-1 and TM-7); intracellular receptors (e.g., orphans, retinoids, vitamin D3 and vitamin A receptors); signaling molecules (e.g., kinases, transcription factors, or molecules such as signal transducers and activators of transcription receptors of the cytokine superfamily (e.g., erythropoietin, growth hormone, interferons, and interleukins, and colony-stimulating factors); G-protein coupled receptors, e.g., hormones, calcitonin, epinephrine, gastrin, and paracrine or autocrine mediators, such as somatostatin or prostaglandins); neurotransmitter receptors (norepinephrine, dopamine, serotonin or acetylcholine); pathogenic antigens, which can be of viral, bacterial, allergenic, or cancerous origin; and tyrosine kinase receptors (such as insulin growth factor, and nerve growth factor).

[0172] A payload molecule may comprise a gene therapy product. A gene therapy product may comprise a polypeptide, an RNA molecule, or other gene product that, when expressed in a target cell, provides a desired therapeutic effect. In some embodiments, a gene therapy product may comprise a substitute for a non-functional gene that is absent or mutated. In some embodiments, a gene therapy product may comprise a method for elimination of a gene that is over-active or dysregulated. Goldsmith et al., WO 90/07936, the contents of which are incorporated herein by reference in their entirety.

[0173] A payload construct vector encoding a payload molecule may comprise a selectable marker. A selectable marker may comprise a gene sequence or a protein encoded by that gene sequence expressed in a host cell that allows for the identification, selection, and/or purification of the host cell from a population of cells that may or may not express the selectable marker. In one embodiment the selectable marker provides resistance to survive a selection process that would otherwise kill the host cell, such as treatment with an antibiotic. In some embodiments an antibiotic selectable marker may comprise one or more antibiotic resistance factors, including but not limited to neomycin resistance (e.g., neo), hygromycin resistance, kanamycin resistance, and/or puromycin resistance.

[0174] In some embodiments a selectable marker may comprise a cell-surface marker, such as any protein expressed on the surface of the cell including, but not limited to receptors, CD markers, lectins, integrins, or truncated versions thereof. In some embodiments, cells that comprise a cell-surface marker may be selected using an antibody targeted to the cell-surface marker. In some embodiments an antibody targeted to the cell-surface marker may be directly

conjugated with a selection agent including, but not limited to a fluorophore, sepharose, or magnetic bead. In some embodiments an antibody targeted to the cell-surface marker may be detected using a secondary labeled antibody or substrate which binds to the antibody targeted to the cell-surface marker. In some embodiments, a selectable marker may comprise negative selection by using an enzyme, including but not limited to Herpes simplex virus thymidine kinase (HSVIX) that converts a pro-toxin (ganciclovir) into a toxin or bacterial Cytosine Deaminase (CD) which converts the pro-toxin 5'-fluorocytosine (5'-FC) into the toxin 5'-fluorouracil (5'-FU). In some embodiments, any nucleic acid sequence encoding a polypeptide can be used as a selectable marker comprising recognition by a specific antibody.

[0175] In some embodiments, a payload construct vector encoding a payload molecule may comprise a selectable marker including, but not limited to, β -lactamase, luciferase, β -galactosidase, or any other reporter gene as that term is understood in the art, including cell-surface markers, such as CD4 or the truncated nerve growth factor (NGFR) (for GFP, see WO 96/23810; Heim et al., *Current Biology* 2:178-182 (1996); Heim et al., *Proc. Natl. Acad. Sci. USA* (1995); or Heim et al., *Science* 373:663-664 (1995); for β -lactamase, see WO 96/30540). In some embodiments, a nucleic acid encoding a selectable marker may comprise a fluorescent protein. A fluorescent protein as herein described may comprise any fluorescent marker including but not limited to green, yellow, and/or red fluorescent protein (GFP, YFP, and RFP).

[0176] In accordance with the invention, a payload molecule comprising a nucleic acid for expression in a target cell will be incorporated into the viral particle produced in the viral replication cell if the payload molecule is located between two ITR sequences, or is located on either side of an asymmetrical ITR engineered with two D regions.

[0177] A payload construct vector sequence encoding one or more payload molecules for expression in a target cell may comprise one or more nucleotide sequences operably linked to at least one target cell-compatible promoter. A person skilled in the art may recognize that a target cell may require a specific promoter including but not limited to a promoter that is species specific, inducible, tissue-specific, or cell cycle-specific Parr et al., *Nat Med.*3:1145-9 (1997).

VI. VIRAL PRODUCTION

[0178] The present disclosure provides a method for increasing the generation of self-complementary (SC) AAV during AAV genome replication in a viral replication cell comprising contacting the viral replication cell or tissue with an AAV polynucleotide or AAV genome or AAV particle comprising a chimeric Rep binding sequence region, or contacting the cell or tissue with a vector comprising the AAV polynucleotide or AAV genome, or contacting the cell or tissue with any of the described compositions, including pharmaceutical compositions.

[0179] The present disclosure provides a method for producing an AAV particle having enhanced (increased, improved) transduction efficiency comprising the steps of 1) co-transfecting competent bacterial cells with a bacmid vector and either a viral construct vector or payload construct vector, 2) isolating the resultant viral construct expression vector and payload construct expression vector and separately transfecting viral replication cells, 3) isolating

and purifying resultant payload and viral construct particles comprising viral construct expression vector or payload construct expression vector, 4) co-infecting a viral replication cell with both the payload and viral construct particles comprising viral construct expression vector or payload construct expression vector, 5) harvesting and purifying the viral particle comprising a self-complementary parvoviral genome (FIG. 9 and FIG. 10).

Vectors

[0180] The invention also provides nucleic acids encoding the mutated or modified virus capsids and capsid proteins of the invention. Further provided are vectors comprising the nucleic acids, and cells (in vivo or in culture) comprising the nucleic acids and/or vectors of the invention. Suitable vectors include without limitation viral vectors (e.g., adenovirus, AAV, herpesvirus, vaccinia, poxviruses, baculoviruses, and the like), plasmids, phage, YACs, BACs, and the like as are well known in the art. Such nucleic acids, vectors and cells can be used, for example, as reagents (e.g., helper packaging constructs or packaging cells) for the production of modified virus capsids or virus vectors as described herein.

[0181] The molecules of the invention which contain AAV sequences include any genetic element (vector) which may be delivered to a host cell, e.g., naked DNA, a plasmid, phage, transposon, cosmid, episome, a protein in a non-viral delivery vehicle (e.g., a lipid-based carrier), virus, etc., which transfers the sequences carried thereon. The selected vector may be delivered by any suitable method, including transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion. The methods used to construct any embodiment of this invention are known to those with skill in nucleic acid manipulation and include genetic engineering, recombinant engineering, and synthetic techniques. See, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y.

[0182] The transgene can be carried on any suitable vector, e.g., a plasmid, which is delivered to a host cell. The plasmids useful in this invention may be engineered such that they are suitable for replication and, optionally, integration in prokaryotic cells, mammalian cells, or both. These plasmids contain sequences permitting replication of the transgene in eukaryotes and/or prokaryotes and selection markers for these systems. Selectable markers or reporter genes may include sequences encoding geneticin, hygromycin or purimycin resistance, among others. The plasmids may also contain certain selectable reporters or marker genes that can be used to signal the presence of the vector in bacterial cells, such as ampicillin resistance. Other components of the plasmid may include an origin of replication and an amplicon, such as the amplicon system employing the Epstein Barr virus nuclear antigen. This amplicon system, or other similar amplicon components permit high copy episomal replication in the cells. Preferably, the molecule carrying the transgene is transfected into the cell, where it may exist transiently. Alternatively, the transgene may be stably integrated into the genome of the host cell, either chromosomally or as an episome. In certain embodiments, the transgene may be present in multiple copies, optionally in head-to-head, head-to-tail, or tail-to-tail concatamers.

Suitable transfection techniques are known and may readily be utilized to deliver the transgene to the host cell. Cells

[0183] The present disclosure provides a cell comprising an engineered Rep protein or an AAV polynucleotide, AAV genome, or AAV comprising an engineered Rep protein, a cell comprising a viral vector comprising the AAV polynucleotide or AAV genome, and a cell comprising a composition comprising the AAV polynucleotide or AAV genome or the AAV or a composition comprising the viral vector.

[0184] Viral production of the invention disclosed herein describes processes and methods for producing viral vector that contacts a target cell to deliver a payload construct, e.g., a recombinant viral construct, which comprises a nucleotide encoding a payload molecule.

[0185] In one embodiment, the viral vector of the invention may be produced in a viral replication cell that comprises an insect cell.

[0186] Growing conditions for insect cells in culture, and production of heterologous products in insect cells in culture are well-known in the art, see U.S. Pat. No. 6,204,059, the contents of which are herein incorporated by reference in their entirety.

[0187] Any insect cell which allows for replication of parvovirus and which can be maintained in culture can be used in accordance with the present invention. Cell lines may be used from *Spodoptera frugiperda*, including, but not limited to the Sf9 or Sf21 cell lines, *drosophila* cell lines, or mosquito cell lines, such as, *Aedes albopictus* derived cell lines. Use of insect cells for expression of heterologous proteins is well documented, as are methods of introducing nucleic acids, such as vectors, e.g., insect-cell compatible vectors, into such cells and methods of maintaining such cells in culture. See, for example, METHODS IN MOLECULAR BIOLOGY, ed. Richard, Humana Press, NJ (1995); O'Reilly et al., BACULOVIRUS EXPRESSION VECTORS, A LABORATORY MANUAL, Oxford Univ. Press (1994); Samulski et al., *J. Vir.* 63:3822-8 (1989); Kajigaya et al., *Proc. Nat'l Acad. Sci. USA* 88: 4646-50 (1991); Ruffing et al., *J. Vir.* 66:6922-30 (1992); Kinibauer et al., *Vir.* 219:37-44 (1996); Zhao et al., *Vir.* 272:382-93 (2000); and Samulski et al., U.S. Pat. No. 6,204,059, the contents of each of which are herein incorporated by reference in their entirety.

[0188] The viral replication cell may be selected from any biological organism, including prokaryotic (e.g., bacterial) cells, and eukaryotic cells, including, insect cells, yeast cells and mammalian cells. Viral replication cells may comprise mammalian cells such as A549, WEHI, 3T3, 10T1/2, BHK, MDCK, COS 1, COS 7, BSC 1, BSC 40, BMT 10, VERO, W138, HeLa, 293, Saos, C2C12, L cells, HT1080, HepG2 and primary fibroblast, hepatocyte and myoblast cells derived from mammals. Viral replication cells of the invention comprise cells derived from mammalian species including, but not limited to, human, monkey, mouse, rat, rabbit, and hamster or cell types, including but not limited to fibroblast, hepatocyte, tumor cell, cell line transformed cell, etc.

Small Scale Production of the Self-Complementary Viral Particle

[0189] Viral production of the invention disclosed herein describes processes and methods for producing viral vector

that contacts a target cell to deliver a payload construct, e.g., a recombinant viral construct, which comprises a nucleotide encoding a payload molecule.

[0190] In one embodiment, the viral vector of the invention may be produced in a viral replication cell that comprises a mammalian cell.

[0191] Viral replication cells commonly used for production of recombinant AAV viral vector include, but are not limited to 293 cells, COS cells, Heta cells, KB cells, and other mammalian cell lines as described in U.S. Pat. Nos. 6,156,303, 5,387,484, 5,741,683, 5,691,176, and 5,688,676; U.S. patent application 2002/0081721, and International Patent Applications WO 00/47757, WO 00/24916, and WO 96/17947, the contents of each of which are herein incorporated by reference in their entirety.

[0192] In one embodiment, viral vector produced in mammalian-cells wherein all three VP proteins are expressed at a stoichiometry approaching 1:1:10 (VP1:VP2:VP3). The regulatory mechanisms that allow this controlled level of expression include the production of two mRNAs, one for VP1, and the other for VP2 and VP3, produced by differential splicing.

[0193] In another embodiment, viral vector is produced in mammalian cells using a triple transfection method wherein a payload construct, parvoviral Rep, and Cap and helper virus are comprised within three different constructs. The triple transfection method of the three components of AAV viral vector production may be utilized to produce small lots of virus for assays including transduction efficiency, target tissue (tropism) evaluation, and stability.

Large Scale Production of the Self Complementary Viral Particle Baculovirus

[0194] Self-complementary viral particle production of the invention disclosed herein describes processes and methods for producing scAAV that contacts a target cell to deliver a payload construct which comprises a nucleotide encoding a payload molecule.

[0195] In one embodiment, the process comprises production of viral particles in a baculoviral system using a viral construct vector and a payload construct vector as depicted in FIG. 9 and FIG. 10. Briefly, the viral construct vector and the payload construct vector of the invention are each incorporated by a transposon donor/acceptor system into a bacmid, also known as a baculovirus plasmid, by standard molecular biology techniques known and performed by a person skilled in the art. Transfection of separate viral replication cell populations produces two baculoviruses, one that comprises the viral construct expression vector, and another that comprises the payload construct expression vector. The two baculoviruses depicted in FIG. 9 and FIG. 10 may be used to infect a single viral replication cell population for production of scAAV.

[0196] Baculovirus expression vectors for producing viral particles in insect cells, including but not limited to *Spodoptera frugiperda* (SP) cells, provide high titers of viral particle product. Recombinant baculovirus encoding the viral construct expression vector and payload construct expression vector initiates a productive infection of viral replicating cells. Infectious baculovirus particles released from the primary infection secondarily infect additional cells in the culture, exponentially infecting the entire cell culture population in a number of infection cycles that is a function of the initial multiplicity of infection, see Urabe, M. et al. J

Virol. 2006 February; 80(4):1874-85, the contents of which are herein incorporated by reference in their entirety.

[0197] Production of viral particles with baculovirus in an insect cell system may address known baculovirus genetic and physical instability. In one embodiment, the production system of the invention addresses baculovirus instability over multiple passages by utilizing a titerless infected-cells preservation and scale-up system. Small scale seed cultures of viral producing cells are transfected with viral expression constructs encoding the structural, non-structural, components of the viral particle. Baculovirus-infected viral producing cells are harvested into aliquots that may be cryopreserved in liquid nitrogen; the aliquots retain viability and infectivity for infection of large scale viral producing cell culture Wasilko D J et al. Protein Expr Purif. 2009 June; 65(2):122-32, the contents of which are herein incorporated by reference in their entirety.

[0198] A genetically stable baculovirus may be used to produce the source of one or more of the components for producing viral particles in invertebrate cells. In one embodiment, defective baculovirus expression vectors may be maintained episomally in insect cells. In such an embodiment the hacmid vector is engineered with replication control elements, including but not limited to promoters, enhancers, and/or cell-cycle regulated replication elements.

[0199] In one embodiment, baculoviruses may be engineered with a (non-) selectable marker for recombination into the chitinase/cathepsin locus. The chia/v-cath locus is non-essential for propagating baculovirus in tissue culture, and the V-cath (EC 3.4.22.50) is a cysteine endoprotease that is most active on Arg-Arg dipeptide containing substrates. The Arg-Arg dipeptide is present in densovirus and parvovirus capsid structural proteins but infrequently occurs in dependovirus VP1.

[0200] In one embodiment, stable viral replication cells permissive for baculovirus infection are engineered with at least one stable integrated copy of any of the elements necessary for AAV replication and viral particle production including, but not limited to, the entire AAV genome, Rep and Cap genes, Rep genes, Cap genes, each Rep protein as a separate transcription cassette, each VP protein as a separate transcription cassette, the AAP (assembly activation protein), or at least one of the baculovirus helper genes with native or non-native promoters.

Large-Scale Production Methods

[0201] In some embodiments, viral particle production may be modified to increase the scale of production. Large scale viral production methods according to the present invention may include any of those taught in U.S. Pat. Nos. 5,756,283, 6,258,595, 6,261,551, 6,270,996, 6,281,010, 6,365,394, 6,475,769, 6,482,634, 6,485,966, 6,943,019, 6,953,690, 7,022,519, 7,238,526, 7,291,498 and 7,491,508 or International Publication Nos. WO1996039530, WO1998010088, WO1999014354, WO1999015685, WO1999047691, WO2000055342, WO2000075353 and WO2001023597, the contents of each of Which are herein incorporated by reference by reference in their entirety. Methods of increasing viral particle production scale typically comprise increasing the number of viral replication cells. In some embodiments, viral replication cells comprise adherent cells. To increase the scale of viral particle production by adherent viral replication cells, larger cell culture surfaces are required. In some cases, large-scale production

methods comprise the use of roller bottles to increase cell culture surfaces. Other cell culture substrates with increased surface areas are known in the art. Examples of additional adherent cell culture products with increased surface areas include, but are not limited to CELLSTACK® (Corning Corp., Corning, N.Y.) and NUNC™ CELL FACTORY™ (Thermo Scientific, Waltham, Mass.) In some cases, large-scale adherent cell surfaces may comprise from about 1,000 cm² to about 100,000 cm². In some cases, large-scale adherent cell cultures may comprise from about 10⁷ to about 10⁹ cells, from about 10⁸ to about 10¹⁰ cells, from about 10⁹ to about 10¹² cells or at least 10¹² cells. In some cases, large-scale adherent cultures may produce from about 10⁹ to about 10¹², from about 10¹⁰ to about 10¹³, from about 10¹¹ to about 10¹⁴, from about 10¹² to about 10¹⁵ or at least 10¹⁵ viral particles.

[0202] In some embodiments, large-scale viral production methods of the present invention may comprise the use of suspension cell cultures. Suspension cell culture allows for significantly increased numbers of cells. Typically, the number of adherent cells that can be grown on about 10-50 cm² of surface area can be grown in about 1 cm³ volume in suspension.

[0203] Transfection of replication cells in large-scale culture formats may be carried out according to any methods known in the art. For large-scale adherent cell cultures, transfection methods may include, but are not limited to the use of inorganic compounds (e.g. calcium phosphate,) organic compounds [e.g. polyethyleneimine (PEI)] or the use of non-chemical methods (e.g. electroporation). With cells grown in suspension, transfection methods may include, but are not limited to the use of calcium phosphate and the use of PEI. In some cases, transfection of large scale suspension cultures may be carried out according to the section entitled "Transfection Procedure" described in Feng, L. et al., 2008. Biotechnol Appl Biochem. 50:121-32, the contents of which are herein incorporated by reference in their entirety. According to such embodiments, PEI-DNA complexes may be formed for introduction of plasmids to be transfected. In some cases, cells being transfected with PEI-DNA complexes may be 'shocked' prior to transfection. 'This comprises lowering cell culture temperatures to 4° C. for a period of about 1 hour. In some cases, cell cultures may be shocked for a period of from about 10 minutes to about 5 hours. In some cases, cell cultures may be shocked at a temperature of from about 0° C. to about 20° C.,

[0204] In some cases, transfections may include one or more vectors for expression of an RNA effector molecule to reduce expression of nucleic acids from one or more payload construct. Such methods may enhance the production of viral particles by reducing cellular resources wasted on expressing payload constructs. In some cases, such methods may be carried according to those taught in US Publication No. US2014/0099666, the contents of which are herein incorporated by reference in their entirety.

Bioreactors

[0205] In some embodiments, cell culture bioreactors may be used for large scale viral production. In some cases, bioreactors comprise stirred tank reactors. Such reactors generally comprise a vessel, typically cylindrical in shape, with a stirrer (e.g. impeller). In some embodiments, such bioreactor vessels may be placed within a water jacket to control vessel temperature and/or to minimize effects from

ambient temperature changes. Bioreactor vessel volume may range in size from about 500 ml to about 2 L, from about 1 L to about 5 L, from about 2.5 L to about 20 L, from about 10 L to about 50 L, from about 25 L to about 100 L, from about 75 L to about 500 L, from about 250 L to about 2,000 L, from about 1,000 L to about 10,000 L, from about 5,000 L to about 50,000 L or at least 50,000 L. Vessel bottoms may be rounded or flat. In some cases, animal cell cultures may be maintained in bioreactors with rounded vessel bottoms,

[0206] In some cases, bioreactor vessels may be warmed through the use of a thermocirculator. Thermocirculators pump heated water around water jackets. In some cases, heated water may be pumped through pipes (e.g. coiled pipes) that are present within bioreactor vessels. In some cases, warm air may be circulated around bioreactors, including, but not limited to air space directly above culture medium. Additionally, pH and CO₂ levels may be maintained to optimize cell viability.

[0207] In some cases, bioreactors may comprise hollow-fiber reactors. Hollow-fiber bioreactors may support the culture of both anchorage dependent and anchorage independent cells. Further bioreactors may include, but are not limited to packed-bed or fixed-bed bioreactors. Such bioreactors may comprise vessels with glass beads for adherent cell attachment. Further packed-bed reactors may comprise ceramic beads.

[0208] In sonic cases, viral particles are produced through the use of a disposable bioreactor. In some embodiments, such bioreactors may include WAVE™ disposable bioreactors,

[0209] In some embodiments, viral particle production in animal cell bioreactor cultures may be carried out according to the methods taught in U.S. Pat. Nos. 5,064,764, 6,194,191, 6,566,118, 8,137,948 or US Patent Application No. US2011/0229971, the contents of each of which are herein incorporated by reference in their entirety.

Cell Lysis

[0210] Cells of the invention, including, but not limited to viral production cells, may be subjected to cell lysis according to any methods known in the art. Cell lysis may be carried out to obtain one or more agents (e.g. viral particles) present within any cells of the invention. In some embodiments, cell lysis may be carried out according to any of the methods listed in U.S. Pat. Nos. 7,326,555, 7,579,181, 7,048,920, 6,410,300, 6,436,394, 7,732,129, 7,510,875, 7,445,930, 6,726,907, 6,194,191, 7,125,706, 6,995,006, 6,676,935, 7,968,333, 5,756,283, 6,258,595, 6,261,551, 6,270,996, 6,281,010, 6,365,394, 6,475,769, 6,482,634, 6,485,966, 6,943,019, 6,953,690, 7,022,519, 7,238,526, 7,291,498 and 7,491,508 or International Publication Nos. WO1996039530, WO1998010088, 101999014354, WO1999015685, WO1999047691, WO2000055342, WO2000075353 and WO2001023597, the contents of each of which are herein incorporated by reference in their entirety. Cell lysis methods may be chemical or mechanical. Chemical cell lysis typically comprises contacting one or more cells with one or more lysis agents. Mechanical lysis typically comprises subjecting one or more cells to one or more lysis conditions and/or one or more lysis forces.

[0211] In some embodiments, chemical lysis may be used to lyse cells. As used herein, the term “lysis agent” refers to any agent that may aid in the disruption of a cell. In some

cases, lysis agents are introduced in solutions, termed lysis solutions or lysis buffers. As used herein, the term “lysis solution” refers to a solution (typically aqueous) comprising one or more lysis agents. In addition to lysis agents, lysis solutions may include one or more buffering agents, solubilizing agents, surfactants, preservatives, cryoprotectants, enzymes, enzyme inhibitors and/or chelators. Lysis buffers are lysis solutions comprising one or more buffering agents. Additional components of lysis solutions may include one or more solubilizing agents. As used herein, the term “solubilizing agent” refers to a compound that enhances the solubility of one or more components of a solution and/or the solubility of one or more entities to which solutions are applied. In some cases, solubilizing agents enhance protein solubility. In some cases, solubilizing agents are selected based on their ability to enhance protein solubility while maintaining protein conformation and/or activity.

[0212] Exemplary lysis agents may include any of those described in U.S. Pat. Nos. 8,685,734, 7,901,921, 7,732,129, 7,223,585, 7,125,706, 8,236,495, 8,110,351, 7,419,956, 7,300,797, 6,699,706 and 6,143,567, the contents of each of which are herein incorporated by reference in their entirety. In some cases, lysis agents may be selected from lysis salts, amphoteric agents, cationic agents, ionic detergents and non-ionic detergents. Lysis salts may include, but are not limited to sodium chloride (NaCl) and potassium chloride (KCl). Further lysis salts may include any of those described in U.S. Pat. Nos. 8,614,101, 7,326,555, 7,579,181, 7,048,920, 6,410,300, 6,436,394, 7,732,129, 7,510,875, 7,445,930, 6,726,907, 6,194,191, 7,125,706, 6,995,006, 6,676,935 and 7,968,333, the contents of each of which are herein incorporated by reference in their entirety. Concentrations of salts may be increased or decreased to obtain an effective concentration for rupture of cell membranes. Amphoteric agents, as referred to herein, are compounds capable of reacting as an acid or a base. Amphoteric agents may include, but are not limited to lysophosphatidylcholine, 3-((3-Cholamidopropyl)dimethylammonium)-1-propane-sulfonate (CHAPS), ZWITTERGENT® and the like. Cationic agents may include, but are not limited to cetyltrimethylammonium bromide (C(16)TAB) and Benzalkonium chloride. Lysis agents comprising detergents may include ionic detergents or non-ionic detergents. Detergents may function to break apart or dissolve cell structures including, but not limited to cell membranes, cell walls, lipids, carbohydrates, lipoproteins and glycoproteins. Exemplary ionic detergents include any of those taught in U.S. Pat. Nos. 7,625,570 and 6,593,123 or US Publication No. US2014/0087361, the contents of each of which are herein incorporated by reference in their entirety. Some ionic detergents may include, but are not limited to sodium dodecyl sulfate (SDS), cholate and deoxycholate. In some cases, ionic detergents may be included in lysis solutions as a solubilizing agent. Non-ionic detergents may include, but are not limited to octylglucoside, digitonin, lubrol, C12E8, TWEEN®-20, TWEEN®-80, Triton X-100 and Nonidet P-40. Non-ionic detergents are typically weaker lysis agents, but may be included as solubilizing agents for solubilizing cellular and/or viral proteins. Further lysis agents may include enzymes and urea. In some cases, one or more lysis agents may be combined in a lysis solution in order to enhance one or more of cell lysis and protein solubility. In some cases, enzyme inhibitors may be included in lysis

solutions in order to prevent proteolysis that may be triggered by cell membrane disruption.

[0213] In some embodiments, mechanical cell lysis is carried out. Mechanical cell lysis methods may include the use of one or more lysis conditions and/or one or more lysis forces. As used herein, the term “lysis condition” refers to a state or circumstance that promotes cellular disruption. Lysis conditions may comprise certain temperatures, pressures, osmotic purity, salinity and the like. In some cases, lysis conditions comprise increased or decreased temperatures. According to some embodiments, lysis conditions comprise changes in temperature to promote cellular disruption. Cell lysis carried out according to such embodiments may include freeze-thaw lysis. As used herein, the term “freeze-thaw lysis” refers to cellular lysis in which a cell solution is subjected to one or more freeze-thaw cycles. According to freeze-thaw lysis methods, cells in solution are frozen to induce a mechanical disruption of cellular membranes caused by the formation and expansion of ice crystals. Cell solutions used according to freeze-thaw lysis methods, may further comprise one or more lysis agents, solubilizing agents, buffering agents, cryoprotectants, surfactants, preservatives, enzymes, enzyme inhibitors and/or chelators. Once cell solutions subjected to freezing are thawed, such components may enhance the recovery of desired cellular products. In some cases, one or more cryoprotectants are included in cell solutions undergoing freeze-thaw lysis. As used herein, the term “cryoprotectant” refers to an agent used to protect one or more substances from damage due to freezing. Cryoprotectants of the invention may include any of those taught in US Publication No. US2013/0323302 or U.S. Pat. Nos. 6,503,888, 6,180,613, 7,888,096, 7,091,030, the contents of each of which are herein incorporated by reference in their entirety. In some cases, cryoprotectants may include, but are not limited to dimethyl sulfoxide, 1,2-propanediol, 2,3-butanediol, formamide, glycerol, ethylene glycol, 1,3-propanediol and n-dimethyl formamide, polyvinylpyrrolidone, hydroxy ethyl starch, agarose, dextrans, inositol, glucose, hydroxyethylstarch, lactose, sorbitol, methyl glucose, sucrose and urea. In some embodiments, freeze-thaw lysis may be carried out according to any of the methods described in U.S. Pat. No. 7,704,721, the contents of which are herein incorporated by reference in their entirety.

[0214] As used herein, the term “lysis force” refers to a physical activity used to disrupt a cell. Lysis forces may include, but are not limited to mechanical forces, sonic forces, gravitational forces, optical forces, electrical forces and the like. Cell lysis carried out by mechanical force is referred to herein as “mechanical lysis.” Mechanical forces that may be used according to mechanical lysis may include high shear fluid forces. According to such methods of mechanical lysis, a microfluidizer may be used. Microfluidizers typically comprise an inlet reservoirs where cell solutions may be applied. Cell solutions may then be pumped into an interaction chamber via a pump (e.g. high-pressure pump) at high speed and/or pressure to produce shear fluid forces. Resulting lysates may then be collected in one or more output reservoir. Pump speed and/or pressure may be adjusted to modulate cell lysis and enhance recovery of products (e.g. viral particles). Other mechanical lysis methods may include physical disruption of cells by scraping.

[0215] Cell lysis methods may be selected based on the cell culture format of cells to be lysed. For example, with adherent cell cultures, some chemical and mechanical lysis methods may be used. Such mechanical lysis methods may include freeze-thaw lysis or scraping. In another example, chemical lysis of adherent cell cultures may be carried out through incubation with lysis solutions comprising surfactant, such as Triton-X-100. In some cases, cell lysates generated from adherent cell cultures may be treated with one more nucleases to lower the viscosity of the lysates caused by liberated DNA.

[0216] In one embodiment, a method for harvesting AAV without lysis may be used for efficient and scalable AAV production. In a non-limiting example, viral particles may be produced by culturing an AAV lacking a heparin binding site, thereby allowing the AAV to pass into the supernatant, in a cell culture, collecting supernatant from the culture; and isolating the AAV from the supernatant, as described in US Patent Application 20090275107, the contents of which are incorporated herein by reference in their entirety.

Clarification

[0217] Cell lysates comprising viral particles may be subjected to clarification. Clarification refers to initial steps taken in purification of viral particles from cell lysates. Clarification serves to prepare lysates for further purification by removing larger, insoluble debris. Clarification steps may include, but are not limited to centrifugation and filtration. During clarification, centrifugation may be carried out at low speeds to remove larger debris, only. Similarly, filtration may be carried out using filters with larger pore sizes so that only larger debris is removed. In some cases, tangential flow filtration may be used during clarification. Objectives of viral clarification include high throughput processing of cell lysates and to optimize ultimate viral recovery. Advantages of including a clarification step include scalability for processing of larger volumes of lysate. In some embodiments, clarification may be carried out according to any of the methods presented in U.S. Pat. Nos. 8,524,446, 5,756,283, 6,258,595, 6,261,551, 6,270,996, 6,281,010, 6,365,394, 6,475,769, 6,482,634, 6,485,966, 6,943,019, 6,953,690, 7,022,519; 7,238,526, 7,291,498; 7,491,508, US Publication Nos. US2013/0045186, US2011/0263027, US2011/0151434, US2003/0138772, and International Publication Nos. WO2002012455, WO1996039530, WO1998010088, WO1999014354, WO1999015685, WO1999047691, WO2000055342, WO2000075353 and WO2001023597; the contents of each of which are herein incorporated by reference in their entirety.

[0218] Methods of cell lysate clarification by filtration are well understood in the art and may be carried out according to a variety of available methods including, but not limited to passive filtration and flow filtration. Filters used may comprise a variety of materials and pore sizes. For example, cell lysate filters may comprise pore sizes of from about 1 μ M to about 5 μ M, from about 0.5 μ M to about 2 μ M, from about 0.1 μ M to about 1 μ M, from about 0.05 μ M to about 0.5 μ M and from about 0.001 μ M to about 0.1 μ M. Exemplary pore sizes for cell lysate filters may include, but are not limited to, 2.0, 1.9, 1.8, 1.7, 1.6, 1.5, 1.4, 1.3, 1.2, 1.1, 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.95, 0.9, 0.85, 0.8, 0.75, 0.7, 0.65, 0.6, 0.55, 0.5, 0.45, 0.4, 0.35, 0.3, 0.25, 0.15, 0.1, 0.05, 0.22, 0.21, 0.20, 0.19, 0.18, 0.17, 0.16, 0.15, 0.14, 0.13, 0.12, 0.11, 0.1, 0.09, 0.08, 0.07, 0.06, 0.05, 0.04, 0.03, 0.02,

0.01, 0.02, 0.019, 0.018, 0.017, 0.016, 0.015, 0.014, 0.013, 0.012, 0.011, 0.01, 0.009, 0.008, 0.007, 0.006, 0.005, 0.004, 0.003, 0.002, 0.001 and 0.001 μM . In one embodiment, clarification may comprise filtration through a filter with 2.0 μM pore size to remove large debris, followed by passage through a filter with 0.45 μM pore size to remove intact cells.

[0219] Filter materials may be composed of a variety of materials. Such materials may include, but are not limited to polymeric materials and metal materials (e.g. sintered metal and pored aluminum). Exemplary materials may include, but are not limited to nylon, cellulose materials (e.g. cellulose acetate), polyvinylidene fluoride (PVDF), polyethersulfone, polyamide, polysulfone, polypropylene and polyethylene terephthalate. In some cases, filters useful for clarification of cell lysates may include, but are not limited to ULTIPLEAT PROFILE™ filters (Pall Corporation, Port Washington N.Y.), SUPOR™ membrane filters (Pall Corporation Port Washington, N.Y.)

[0220] In some cases, flow filtration may be carried out to increase filtration speed and/or effectiveness. In some cases, flow filtration may comprise vacuum filtration. According to such methods, a vacuum is created on the side of the filter opposite that of cell lysate to be filtered. In some cases, cell lysates may be passed through filters by centrifugal forces. In some cases, a pump is used to force cell lysate through clarification filters. Flow rate of cell lysate through one or more filters may be modulated by adjusting one of channel size and/or fluid pressure.

[0221] According to some embodiments, cell lysates may be clarified by centrifugation. Centrifugation may be used to pellet insoluble particles in the lysate. During clarification, centrifugation strength [expressed in terms of gravitational units (g), which represents multiples of standard gravitational force] may be lower than in subsequent purification steps. In some cases, centrifugation may be carried out on cell lysates at from about 200 g to about 800 g, from about 500 g to about 1500 g, from about 1000 g to about 5000 g, from about 1200 g to about 10000 g or from about 8000 g to about 15000 g. In sonic embodiments, cell lysate centrifugation is carried out at 8000 g for 15 minutes. In some cases, density gradient centrifugation may be carried out in order to partition particulates in the cell lysate by sedimentation rate. Gradients used according to methods of the present invention may include, but are not limited to cesium chloride gradients and iodixanol step gradients.

Purification Chromatography

[0222] In some cases, viral particles may be purified from clarified cell lysates by one or more methods of chromatography. Chromatography refers to any number of methods known in the art for separating out one or more elements from a mixture. Such methods may include, but are not limited to ion exchange chromatography (e.g. cation exchange chromatography and anion exchange chromatography), immunoaffinity chromatography and size-exclusion chromatography. In some embodiments, methods of viral chromatography may include any of those taught in U.S. Pat. Nos. 5,756,283, 6,258,595, 6,261,551, 6,270,996, 6,281,010, 6,365,394, 6,475,769, 6,482,634, 6,485,966, 6,943,019, 6,953,690, 7,022,519, 7,238,526, 7,291,498 and 7,491,508 or international Publication Nos. WO1996039530, WO199801.0088, WO1999014354, WO1999015685, WO1999047691, WO2000055342, WO2000075353 and

WO2001023597, the contents of each of which are herein incorporated by reference in their entirety.

[0223] In some embodiments, ion exchange chromatography may be used to isolate viral particles. Ion exchange chromatography is used to bind viral particles based on charge-charge interactions between capsid proteins and charged sites present on a stationary phase, typically a column through which viral preparations (e.g. clarified lysates) are passed. After application of viral preparations, bound viral particles may then be eluted by applying an elution solution to disrupt the charge-charge interactions. Elution solutions may be optimized by adjusting salt concentration and/or to enhance recovery of bound viral particles. Depending on the charge of viral capsids being isolated, cation or anion exchange chromatography methods may be selected. Methods of ion exchange chromatography may include, but are not limited to any of those taught in U.S. Pat. Nos. 7,419,817, 6,143,548, 7,094,604, 6,593,123, 7,015,026 and 8,137,948, the contents of each of which are herein incorporated by reference in their entirety.

[0224] In some embodiments, immunoaffinity chromatography may be used. Immunoaffinity chromatography is a form of chromatography that utilizes one or more immune compounds (e.g. antibodies or antibody-related structures) to retain viral particles. Immune compounds may bind specifically to one or more structures on viral particle surfaces, including, but not limited to one or more viral coat proteins. In some cases, immune compounds may be specific for a particular viral variant. In some cases, immune compounds may bind to multiple viral variants. In some embodiments, immune compounds may include recombinant single-chain antibodies. Such recombinant single chain antibodies may include those described in Smith, R. H. et al., 2009, Mol Ther. 17(11):1888-96, the contents of which are herein incorporated by reference in their entirety. Such immune compounds are capable of binding to several AAV capsid variants, including, but not limited to AAV1, AAV2, AAV6 and AAV8.

[0225] In some embodiments, size-exclusion chromatography (SEC) may be used. SEC may comprise the use of a gel to separate particles according to size. In viral particle purification, SEC filtration is sometimes referred to as "polishing." In some cases, SEC may be carried out to generate a final product that is near-homogenous. Such final products may in some cases be used in pre-clinical studies and/or clinical studies (Kotin, R. M. 2011. Human Molecular Genetics. 20(1):R2-R6, the contents of which are herein incorporated by reference in their entirety). In some cases, SEC may be carried out according to any of the methods taught in U.S. Pat. Nos. 6,143,548, 7,015,026, 8,476,418, 6,410,300, 8,476,418, 7,419,817, 7,094,604, 6,593,123, and 8,137,948, the contents of each of which are herein incorporated by reference in their entirety.

[0226] In one embodiment, the compositions comprising at least one viral particle may be isolated or purified using the methods described in U.S. Pat. No. 6,146,874, the contents of which are herein incorporated by reference in their entirety.

[0227] In one embodiment, the compositions comprising at least one viral particle may be isolated or purified using the methods described in U.S. Pat. No. 6,660,514, the contents of which are herein incorporated by reference in their entirety.

[0228] In one embodiment, the compositions comprising at least one viral particle may be isolated or purified using the methods described in U.S. Pat. No. 8,283,151, the contents of which are herein incorporated by reference in their entirety.

[0229] In one embodiment, the compositions comprising at least one viral particle may be isolated or purified using the methods described in U.S. Pat. No. 8,524,446, the contents of which are herein incorporated by reference in their entirety.

VII. TREATMENT AND PHARMACEUTICAL COMPOSITIONS

[0230] The present disclosure additionally provides a method for treating a disease, disorder and/or condition in a mammalian subject, including a human subject, comprising administering to the subject any of the above-described AAV polynucleotides or AAV genomes or AAV comprising a chimeric Rep binding sequence region or administering to the subject a vector comprising the AAV polynucleotide or AAV genome, or administering to the subject any of the described compositions, including pharmaceutical compositions. In one embodiment, the disease, disorder and/or condition is a neurological disease, disorder and/or condition. In one embodiment, the neurological disease, disorder and/or condition is Parkinson's disease. In another embodiment, the neurological disease, disorder and/or condition is Friedreich's Ataxia. In another embodiment, the neurological disease, disorder and/or condition is Amyotrophic lateral sclerosis (ALS). In another embodiment, the disease, disorder and/or condition is a muscular or cardiac disease, disorder and/or condition. In another embodiment, the disease, disorder and/or condition is an immune system disease, disorder and/or condition.

[0231] The present disclosure provides a composition comprising an AAV polynucleotide or AAV genome or AAV particle comprising an engineered Rep binding sequence region and at least one excipient.

[0232] The present disclosure also provides a pharmaceutical composition comprising an adeno-associated virus (AAV) comprising any of the above-described AAV polynucleotides or AAV genomes and one or more pharmaceutically acceptable excipients.

[0233] Although the descriptions of pharmaceutical compositions, e.g., those viral vectors comprising a payload to be delivered, provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to any other animal, e.g., to non-human animals, e.g. non-human mammals. Modification of pharmaceutical compositions suitable for administration to humans in order to render the compositions suitable for administration to various animals is well understood, and the ordinarily skilled veterinary pharmacologist can design and/or perform such modification with merely ordinary, if any, experimentation. Subjects to which administration of the pharmaceutical compositions is contemplated include, but are not limited to, humans and/or other primates; mammals, including commercially relevant mammals such as cattle, pigs, horses, sheep, cats, dogs, mice, and/or rats; and/or birds, including commercially relevant birds such as poultry, chickens, ducks, geese, and/or turkeys.

[0234] In some embodiments, compositions are administered to humans, human patients or subjects. For the purposes of the present disclosure, the phrase "active ingredient" generally refers either to the viral particle carrying the payload or to the payload molecule delivered by the viral particle as described herein.

[0235] Formulations of the pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient into association with an excipient and/or one or more other accessory ingredients, and then, if necessary and/or desirable, dividing, shaping and/or packaging the product into a desired single- or multi-dose unit.

[0236] Relative amounts of the active ingredient, the pharmaceutically acceptable excipient, and/or any additional ingredients in a pharmaceutical composition in accordance with the invention will vary, depending upon the identity, size, and/or condition of the subject treated and further depending upon the route by which the composition is to be administered.

Formulation

[0237] The viral particles of the invention can be formulated using one or more excipients to:

[0238] (1) increase stability; (2) increase cell transfection or transduction; (3) permit the sustained or delayed release; (4) alter the biodistribution (e.g., target the viral particle to specific tissues or cell types); (5) increase the translation of encoded protein in vivo; and/or (6) alter the release profile of encoded protein in vivo.

[0239] Formulations of the present invention can include, without limitation, saline, lipidoids, liposomes, lipid nanoparticles, polymers, lipoplexes, core-shell nanoparticles, peptides, proteins, cells transfected with viral vectors (e.g., for transplantation into a subject), nanoparticle mimics and combinations thereof. Further, the viral vectors of the present invention may be formulated using self-assembled nucleic acid nanoparticles.

[0240] Formulations of the pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of associating the active ingredient with an excipient and/or one or more other accessory ingredients.

[0241] A pharmaceutical composition in accordance with the present disclosure may be prepared, packaged, and/or sold in bulk, as a single unit dose, and/or as a plurality of single unit doses. As used herein, a "unit dose" refers to a discrete amount of the pharmaceutical composition comprising a predetermined amount of the active ingredient. The amount of the active ingredient is generally equal to the dosage of the active ingredient which would be administered to a subject and/or a convenient fraction of such a dosage such as, for example, one-half or one-third of such a dosage.

[0242] Relative amounts of the active ingredient (e.g., scAAV), the pharmaceutically acceptable excipient, and/or any additional ingredients in a pharmaceutical composition in accordance with the present disclosure may vary, depending upon the identity, size, and/or condition of the subject being treated and further depending upon the route by which the composition is to be administered. For example, the composition may comprise between 0.1% and 99% (w/w) of the active ingredient. By way of example, the composition

may comprise between 0.1% and 100%, e.g., between 0.5 and 50%, between 1-30%, between 5-80%, at least 80% (w/w) active ingredient.

[0243] In some embodiments, the formulations described herein may contain at least one payload molecule. As a non-limiting example, the formulations may contain 1, 2, 3, 4 or .5 payload molecules. In one embodiment the formulation may contain a payload construct encoding proteins selected from categories such as, but not limited to, human proteins, veterinary proteins, bacterial proteins, biological proteins, antibodies, immunogenic proteins, therapeutic peptides and proteins, secreted proteins, plasma membrane proteins, cytoplasmic and cytoskeletal proteins, intracellular membrane bound proteins, nuclear proteins, proteins associated with human disease and/or proteins associated with non-human diseases. In one embodiment, the formulation contains at least three payload construct encoding proteins.

[0244] The formulations of the invention can include one or more excipients, each in an amount that together increases the stability of the viral particle, increases cell transfection or transduction by the viral particle, increases the expression of viral particle encoded protein, and/or alters the release profile of viral particle encoded proteins. In some embodiments, a pharmaceutically acceptable excipient may be at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% pure. In some embodiments, an excipient is approved for use for humans and for veterinary use. In some embodiments, an excipient may be approved by United States Food and Drug Administration. In some embodiments, an excipient may be of pharmaceutical grade. In some embodiments, an excipient may meet the standards of the United States Pharmacopoeia (USP), the European Pharmacopoeia (EP), the British Pharmacopoeia, and/or the International Pharmacopoeia.

[0245] Excipients, which, as used herein, includes, but is not limited to, any and all solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, and the like, as suited to the particular dosage form desired. Various excipients for formulating pharmaceutical compositions and techniques for preparing the composition are known in the art (see Remington: The Science and Practice of Pharmacy, 21st Edition, A. R., Gennaro, Lippincott, Williams & Wilkins, Baltimore, Md., 2006; incorporated herein by reference in its entirety). The use of a conventional excipient medium may be contemplated within the scope of the present disclosure, except insofar as any conventional excipient medium may be incompatible with a substance or its derivatives, such as by producing any undesirable biological effect or otherwise interacting in a deleterious manner with any other component(s) of the pharmaceutical composition.

[0246] Diluents may include, but are not limited to, calcium carbonate, sodium carbonate, calcium phosphate, dicalcium phosphate, calcium sulfate, calcium hydrogen phosphate, sodium phosphate lactose, sucrose, cellulose, microcrystalline cellulose, kaolin, mannitol, sorbitol, inositol, sodium chloride, dry starch, cornstarch, powdered sugar, etc., and/or combinations thereof.

[0247] Additives include physiologically biocompatible buffers (e.g., trimethylamine hydrochloride), addition of chelants (such as, for example, DTPA or DTPA-bisamide) or calcium chelate complexes (as for example calcium DTPA, CaNaDTPA-bisamide), or, optionally, additions of calcium

or sodium salts (for example, calcium chloride, calcium ascorbate, calcium gluconate or calcium lactate). In addition, antioxidants and suspending agents can be used.

Cryoprotectants

[0248] In some embodiments, viral particle formulations may comprise cryoprotectants. As used herein, the term “cryoprotectant” refers to one or more agents that when combined with a given substance, helps to reduce or eliminate damage to that substance that occurs upon freezing. In some embodiments, cryoprotectants are combined with viral particles in order to stabilize them during freezing. Frozen storage between -20° C. and -80° C. may be advantageous for long term (e.g. 36 months) stability of viral particles. In some embodiments, cryoprotectants are included in viral particle formulations to stabilize them through freeze/thaw cycles and under frozen storage conditions. Cryoprotectants of the present invention may include, but are not limited to sucrose, trehalose, lactose, glycerol, dextrose, raffinose and/or mannitol. Trehalose is listed by the Food and Drug Administration as being generally regarded as safe (GRAS) and is commonly used in commercial pharmaceutical formulations.

Bulking Agents

[0249] In some embodiments, viral particle formulations may comprise bulking agents. As used herein, the term “bulking agent” refers to one or more agents included in formulations to impart a desired consistency to the formulation and/or stabilization of formulation components. In some embodiments, bulking agents are included in lyophilized viral particle formulations to yield a “pharmaceutically elegant” cake, stabilizing the lyophilized viral particles during long term (e.g. 36 month) storage. Bulking agents of the present invention may include, but are not limited to sucrose, trehalose, mannitol, glycine, lactose and/or raffinose. In some embodiments, combinations of cryoprotectants and bulking agents (for example, sucrose/glycine or trehalose/mannitol) may be included to both stabilize viral particles during freezing and provide a bulking agent for lyophilization.

Inactive Ingredients

[0250] In some embodiments, formulations may comprise at least one excipient which is an inactive ingredient. As used herein, the term “inactive ingredient” refers to one or more inactive agents included in formulations. In some embodiments, all, none or some of the inactive ingredients which may be used in the formulations of the present invention may be approved by the US Food and Drug Administration (FDA).

[0251] Formulations of viral particles disclosed herein may include cations or anions. In one embodiment, the formulations include metal cations such as, but not limited to, Zn²⁺, Ca²⁺, Cu²⁺, Mg⁺ and combinations thereof.

Administration

[0252] The viral particles of the present invention may be administered by any route which results in a therapeutically effective outcome. These include, but are not limited to enteral (into the intestine), gastroenteral, epidural (into the dura matter), oral (by way of the mouth), transdermal, peridural, intracerebral (into the cerebrum), intracerebroven-

tricular (into the cerebral ventricles), epicutaneous (application onto the skin), intradermal, (into the skin itself), subcutaneous (under the skin), nasal administration (through the nose), intravenous (into a vein), intravenous bolus, intravenous drip, intraarterial (into an artery), intramuscular (into a muscle), intracardiac (into the heart), intraosseous infusion (into the bone marrow), intrathecal (into the spinal canal), intraperitoneal, (infusion or injection into the peritoneum), intravesical infusion, intravitreal, (through the eye), intracavernous injection (into a pathologic cavity) intracavitary (into the base of the penis), intravaginal administration, intrauterine, extra-amniotic administration, transdermal (diffusion through the intact skin for systemic distribution), transmucosal (diffusion through a mucous membrane), transvaginal, insufflation (snorting), sublingual, sublabial, enema, eye drops (onto the conjunctiva), in ear drops, auricular (in or by way of the ear), buccal (directed toward the cheek), conjunctival, cutaneous, dental (to a tooth or teeth), electro-osmosis, endocervical, endosinusial, endotracheal, extracorporeal, hemodialysis, infiltration, interstitial, intra-abdominal, intra-amniotic, intra-articular, intrabiliary, intrabronchial, intrabursal, intracartilaginous (within a cartilage), intracaudal (within the cauda equine), intracisternal (within the cisterna magna cerebellomedullaris), intracorneal (within the cornea), dental intraconal, intracoronary (within the coronary arteries), intracorporus cavernosum (within the dilatable spaces of the corpus cavernosa of the penis), intradiscal (within a disc), intraductal (within a duct of a gland), intraduodenal (within the duodenum), intradural (within or beneath the dura), intraepidermal (to the epidermis), intraesophageal (to the esophagus), intragastric (within the stomach), intragingival (within the gingivae), intraileal (within the distal portion of the small intestine), intralesional (within or introduced directly to a localized lesion), intraluminal (within a lumen of a tube), intralymphatic (within the lymph), intramedullary (within the marrow cavity of a bone), intrameningeal (within the meninges), intraocular (within the eye), intraovarian (within the ovary), intrapericardial (within the pericardium), intrapleural (within the pleura), intraprostatic (within the prostate gland), intrapulmonary (within the lungs or its bronchi), intrasinal (within the nasal or periorbital sinuses), intraspinal (within the vertebral column), intrasynovial (within the synovial cavity of a joint), intratendinous (within a tendon), intratesticular (within the testicle), intrathecal (within the cerebrospinal fluid at any level of the cerebrospinal axis), intrathoracic (within the thorax), intratubular (within the tubules of an organ), intratumor (within a tumor), intratympanic (within the auras media), intravascular (within a vessel or vessels), intraventricular (within a ventricle), iontophoresis (by means of electric current where ions of soluble salts migrate into the tissues of the body), irrigation (to bathe or flush open wounds or body cavities), laryngeal (directly upon the larynx), nasogastric (through the nose and into the stomach), occlusive dressing technique (topical route administration which is then covered by a dressing which occludes the area), ophthalmic (to the external eye), oropharyngeal (directly to the mouth and pharynx), parenteral, percutaneous, periarticular, peridural, perineural, periodontal, rectal, respiratory (within the respiratory tract by inhaling orally or nasally for local or systemic effect), retrobulbar (behind the pons or behind the eyeball), soft tissue, subarachnoid., subconjunctival, submucosal, topical, transplacental (through or across the placenta), transtracheal

(through the wall of the trachea), transtympanic (across or through the tympanic cavity), ureteral (to the ureter), urethral (to the urethra), vaginal, caudal block, diagnostic, nerve block, biliary perfusion, cardiac perfusion, photopheresis or spinal. In specific embodiments, compositions may be administered in a way which allows them cross the blood-brain barrier, vascular barrier, or other epithelial barrier. In one embodiment, a formulation for a route of administration may include at least one inactive ingredient.

Dosing

[0253] The present invention provides methods comprising administering viral particles and their payload or complexes in accordance with the invention to a subject in need thereof. Viral particle pharmaceutical, imaging, diagnostic, or prophylactic compositions thereof, may be administered to a subject using any amount and any route of administration effective for preventing, treating, diagnosing, or imaging a disease, disorder, and/or condition (e.g., a disease, disorder, and/or condition relating to working memory deficits). The exact amount required will vary from subject to subject, depending on the species, age, and general condition of the subject, the severity of the disease, the particular composition, its mode of administration, its mode of activity, and the like. Compositions in accordance with the invention are typically formulated in unit dosage form for ease of administration and uniformity of dosage. It will be understood, however, that the total daily usage of the compositions of the present invention may be decided by the attending physician within the scope of sound medical judgment. The specific therapeutically effective, prophylactically effective, or appropriate imaging dose level for any particular patient will depend upon a variety of factors including the disorder being treated and the severity of the disorder; the activity of the specific compound employed; the specific composition employed; the age, body weight, general health, sex and diet of the patient; the time of administration, route of administration, and rate of excretion of the specific payload employed; the duration of the treatment; drugs used in combination or coincidental with the specific compound employed; and like factors well known in the medical arts.

[0254] In certain embodiments, viral particle pharmaceutical compositions in accordance with the present invention may be administered at dosage levels sufficient to deliver from about 00001 mg/kg to about 100 mg/kg, from about 0.001 mg/kg to about 0.05 mg/kg, from about 0.005 mg/kg to about 0.05 mg/kg, from about 0.001 mg/kg to about 0.005 mg/kg, from about 0.05 mg/kg to about 0.5 mg, from about 0.01 mg/kg to about 50 mg/kg, from about 0.1 mg/kg to about 40 mg/kg, from about 0.5 mg/kg to about 30 mg/kg, from about 0.01 mg/kg to about 10 mg/kg, from about 0.1 mg/kg to about 10 mg/kg, or from about 1 mg/kg to about 25 mg/kg, of subject body weight per day, one or more times a day, to obtain the desired therapeutic, diagnostic, prophylactic, or imaging effect.

[0255] The desired dosage may be delivered three times a day, two times a day, once a day, every other day, every third day, every week, every two weeks, every three weeks, or every four weeks. In certain embodiments, the desired dosage may be delivered using multiple administrations (e.g., two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, thirteen, fourteen, or more administrations). When multiple administrations are employed, split dosing regi-

mens such as those described herein may be used. As used herein, a “split dose” is the division of single unit dose or total daily dose into two or more doses, e.g., two or more administrations of the single unit dose. As used herein, a “single unit dose” is a dose of any therapeutic administered in one dose/at one time/single route/single point of contact, i.e., single administration event. As used herein, a “total daily dose” is an amount given or prescribed in 24 hour period. It may be administered as a single unit dose. In one embodiment, the viral particles of the present invention are administered to a subject in split doses. The viral particles may be formulated in buffer only or in a formulation described herein.

[0256] A pharmaceutical composition described herein can be formulated into a dosage form described herein, such as a topical, intranasal, pulmonary, intratracheal, or injectable (e.g., intravenous, intraocular, intravitreal, intramuscular, intracardiac, intraperitoneal, and subcutaneous).

Combinations

[0257] The viral particles may be used in combination with one or more other therapeutic, prophylactic, diagnostic, or imaging agents. By “in combination with,” it is not intended to imply that the agents must be administered at the same time and/or formulated for delivery together, although these methods of delivery are within the scope of the present disclosure. Compositions can be administered concurrently with, prior to, or subsequent to, one or more other desired therapeutics or medical procedures. In general, each agent will be administered at a dose and/or on a time schedule determined for that agent. In some embodiments, the present disclosure encompasses the delivery of pharmaceutical, prophylactic, diagnostic, or imaging compositions in combination with agents that may improve their bioavailability, reduce and/or modify their metabolism, inhibit their excretion, and/or modify their distribution within the body.

Delivery

Delivery to Cells

[0258] The present disclosure provides a method of delivering to a cell or tissue any of the above-described engineered Rep proteins or AAV polynucleotides, AAV genomes, or AAV comprising an engineered Rep protein comprising contacting the cell or tissue with the Rep protein or AAV polynucleotide or AAV genome or AAV, or contacting the cell or tissue with a particle comprising the AAV polynucleotide or AAV genome, or contacting the cell or tissue with any of the described compositions, including pharmaceutical compositions. The method of delivering the AAV polynucleotide or AAV genome or AAV to a cell or tissue can be accomplished in vitro, ex vivo, or in vivo.

Delivery to Subjects

[0259] The present disclosure additionally provides a method of delivering to a subject, including a mammalian subject, any of the above-described AAV polynucleotides or AAV genomes or AAV comprising an engineered Rep binding sequence region comprising administering to the subject the AAV polynucleotide or AAV genome or AAV, or administering to the subject a particle comprising the AAV

polynucleotide or AAV genome, or administering to the subject any of the described compositions, including pharmaceutical compositions.

[0260] The pharmaceutical compositions of viral particles described herein may be characterized by one or more of bioavailability, therapeutic window and/or volume of distribution.

Bioavailability

[0261] Viral particles of the present invention, when formulated into compositions with delivery/formulation agents or vehicles as described herein, may exhibit increased bioavailability as compared to compositions lacking delivery agents as described herein. As used herein, the term “bioavailability” refers to the systemic availability of a given amount of a particular agent administered to a subject. Bioavailability may be assessed by measuring the area under the curve (AUC) or the maximum serum or plasma concentration (C_{max}) of the unchanged form of a compound following administration of the compound to a mammal. AUC is a determination of the area under the curve plotting the serum or plasma concentration of a compound along the ordinate (Y-axis) against time along the abscissa (X-axis). Generally, the AUC for a particular compound (e.g., scAAV) may be calculated using methods known to those of ordinary skill in the art and as described in G. S. Banker, *Modern Pharmaceutics, Drugs and the Pharmaceutical Sciences*, v. 72, Marcel Dekker, New York, Inc., 1996, the contents of which are herein incorporated by reference in their entirety.

[0262] C_{max} values are maximum concentrations of compounds achieved in serum or plasma of a subject following administration of compounds to the subject. C_{max} values of particular compounds may be measured using methods known to those of ordinary skill in the art. As used herein, the phrases “increasing bioavailability” or “improving the pharmacokinetics,” refer to actions that may increase the systemic availability of a viral particle of the present invention (as measured by AUC, C_{max} , or C_{min}) in a subject. In some embodiments, such actions may comprise co-administration with one or more delivery agents as described herein. In some embodiments, the bioavailability of viral particles (e.g., scAAV) may increase by at least about 2%, at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95% or about 100%.

Therapeutic Window

[0263] Viral particles of the present invention, when formulated with one or more delivery agents as described herein, may exhibit increases in the therapeutic window of compound and/or composition administration as compared to the therapeutic window of viral particles administered without one or more delivery agents as described herein. As used herein, the term “therapeutic window” refers to the range of plasma concentrations, or the range of levels of therapeutically active substance at the site of action, with a high probability of eliciting a therapeutic effect. In some embodiments, therapeutic windows of viral particles when administered in a formulation may increase by at least about

2%, at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95% or about 100%.

Volume of Distribution

[0264] Viral particles e.g., scAAV) of the present invention, when formulated with one or more delivery agents as described herein, may exhibit an improved volume of distribution (V_{dist}), e.g., reduced or targeted, relative to formulations lacking one or more delivery agents as described herein. V_{dist} relates the amount of an agent in the body to the concentration of the same agent in the blood or plasma. As used herein, the term “volume of distribution” refers to the fluid volume that would be required to contain the total amount of an agent in the body at the same concentration as in the blood or plasma: V_{dist} equals the amount of an agent in the body/concentration of the agent in blood or plasma. For example, for a 10 mg dose of a given agent and a plasma concentration of 10 mg/L, the volume of distribution would be 1 liter. The volume of distribution reflects the extent to which an agent is present in the extravascular tissue. Large volumes of distribution reflect the tendency of agents to bind to the tissue components as compared with plasma proteins. In clinical settings, V_{dist} may be used to determine loading doses to achieve steady state concentrations. In some embodiments, volumes of distribution of viral particle compositions of the present invention when co-administered with one or more delivery agents as described herein may decrease at least about 2%, at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%.

Kits and Devices

[0265] The invention provides a variety of kits for conveniently and/or effectively carrying out methods of the present invention. Typically kits will comprise sufficient amounts and/or numbers of components to allow a user to perform multiple treatments of a subject(s) and/or to perform multiple experiments.

[0266] In one aspect, the present invention provides kits comprising the molecules (viral particles) of the invention. In one embodiment, the kit comprises one or more functional antibodies or function fragments thereof.

[0267] Said kits can be for viral particle production. The kit may further comprise packaging and instructions and/or a delivery agent to form a formulation composition. The delivery agent may comprise a saline, a buffered solution, or any delivery agent disclosed herein.

[0268] In one embodiment, the buffer solution may include sodium chloride, calcium chloride, phosphate and/or EDTA. In another embodiment, the buffer solution may include, but is not limited to, saline, saline with 2 mM calcium, 50 sucrose, 5% sucrose with 2 mM calcium, 5% Mannitol, 5% Mannitol with 2 mM calcium, Ringer's lactate, sodium chloride, sodium chloride with 2 mM calcium and mannose (See e.g., U.S. Pub. No. 20120258046; herein

incorporated by reference in its entirety). In a further embodiment, the buffer solutions may be precipitated or may be lyophilized. The amount of each component may be varied to enable consistent, reproducible higher concentration saline or simple buffer formulations. The components may also be varied in order to increase the stability of the viral particle or any expression construct taught herein in the buffer solution over a period of time and/or under a variety of conditions.

[0269] In one aspect, the present invention provides kits for viral particle production, comprising: an expression particle and a payload construct particle provided in an amount effective to produce a desired amount of a viral particle when introduced into a target cell and packaging and instructions.

[0270] Any of the particles, constructs, polynucleotides or polypeptides of the present invention may be comprised in a kit. In some embodiments, kits may further include reagents and/or instructions for creating and/or synthesizing compounds and/or compositions of the present invention. In some embodiments, kits may also include one or more buffers. In some embodiments, kits of the invention may include components for making protein or nucleic acid arrays or libraries and thus, may include, for example, solid supports.

[0271] In some embodiments, kit components may be packaged either in aqueous media or in lyophilized form. The container means of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other container means, into which a component may be placed, and preferably, suitably aliquoted. Where there are more than one kit component, (labeling reagent and label may be packaged together), kits may also generally contain second, third or other additional containers into which additional components may be separately placed. In some embodiments, kits may also comprise second container means for containing sterile, pharmaceutically acceptable buffers and/or other diluents. In some embodiments, various combinations of components may be comprised in one or more vial. Kits of the present invention may also typically include means for containing compounds and/or compositions of the present invention, e.g., proteins, nucleic acids, and any other reagent containers in close confinement for commercial sale. Such containers may include injection or blow-molded plastic containers into which desired vials are retained,

[0272] In some embodiments, kit components are provided in one and/or more liquid solutions. In some embodiments, liquid solutions are aqueous solutions, with sterile aqueous solutions being particularly preferred. In some embodiments, kit components may be provided as dried powder(s). When reagents and/or components are provided as dry powders, such powders may be reconstituted by the addition of suitable volumes of solvent. In some embodiments, it is envisioned that solvents may also be provided in another container means. In some embodiments, labeling dyes are provided as dried powders. In some embodiments, it is contemplated that 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400, 500, 600, 700, 800, 900, 1000 micrograms or at least or at most those amounts of dried dye are provided in kits of the invention. In such embodiments, dye may then be resuspended in any suitable solvent, such as DMSO,

[0273] In some embodiments, kits may include instructions for employing kit components as well the use of any

other reagent not included in the kit. Instructions may include variations that may be implemented.

Devices

[0274] In some embodiments, compounds and/or compositions of the present invention may be combined with, coated onto or embedded in a device. Devices may include, but are not limited to, dental implants, stents, bone replacements, artificial joints, valves, pacemakers and/or other implantable therapeutic devices.

[0275] The present invention provides for devices which may incorporate viral particles that encode one or more payload molecules. These devices contain in a stable formulation the viral particles which may be immediately delivered to a subject in need thereof, such as a human patient.

[0276] Devices for administration may be employed to deliver the viral particles of the present invention according to single, multi- or split-dosing regimens taught herein.

[0277] Method and devices known in the art for multi-administration to cells, organs and tissues are contemplated for use in conjunction with the methods and compositions disclosed herein as embodiments of the present invention. These include, for example, those methods and devices having multiple needles, hybrid devices employing for example lumens or catheters as well as devices utilizing heat, electric current or radiation driven mechanisms.

[0278] In one embodiment, the chimeric polynucleotide is administered subcutaneously or intramuscularly via at least 3 needles to three different, optionally adjacent, sites simultaneously, or within a 60 minutes period (e.g., administration to 4, 5, 6, 7, 8, 9, or 10 sites simultaneously or within a 60 minute period).

VIII. DEFINITIONS

[0279] At various places in the present specification, substituents of compounds of the present disclosure are disclosed in groups or in ranges. It is specifically intended that the present disclosure include each and every individual sub-combination of the members of such groups and ranges. The following is a non-limiting list of term definitions.

[0280] Activity: As used herein, the term “activity” refers to the condition in which things are happening or being done. Compositions of the invention may have activity and this activity may involve one or more biological events.

[0281] Administered in combination: As used herein, the term “administered in combination” or “combined administration” refers to simultaneous exposure to two or more agents (e.g., scAAV) administered at the same time or within an interval such that the subject is at some point in time simultaneously exposed to both and/or such that there may be an overlap in the effect of each agent on the patient. In some embodiments, at least one dose of one or more agents is administered within about 24 hours, 12 hours, 6 hours, 3 hours, 1 hour, 30 minutes, 15 minutes, 10 minutes, 5 minutes, or 1 minute of at least one dose of one or more other agents. In some embodiments, administration occurs in overlapping dosage regimens. As used herein, the term “dosage regimen” refers to a plurality of doses spaced apart in time. Such doses may occur at regular intervals or may include one or more hiatus in administration. In some embodiments, the administration of individual doses of one or more compounds and/or compositions of the present

invention, as described herein, are spaced sufficiently closely together such that a combinatorial (e.g., a synergistic) effect is achieved.

[0282] Animal: As used herein, the term “animal” refers to any member of the animal kingdom. In some embodiments, “animal” refers to humans at any stage of development. In some embodiments, “animal” refers to non-human animals at any stage of development. In certain embodiments, the non-human animal is a mammal (e.g., a rodent, a mouse, a rat, a rabbit, a monkey, a dog, a cat, a sheep, cattle, a primate, or a pig). In some embodiments, animals include, but are not limited to, mammals, birds, reptiles, amphibians, fish, and worms. In some embodiments, the animal is a transgenic animal, genetically-engineered animal, or a clone.

[0283] Approximately: As used herein, the term “approximately” or “about,” as applied to one or more values of interest, refers to a value that is similar to a stated reference value. In certain embodiments, the term “approximately” or “about” refers to a range of values that fall within 25%, 20%, 19%, 18%, 17%, 16%, 15%, 14%, 13%, 12%, 11%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, or less in either direction (greater than or less than) of the stated reference value unless otherwise stated or otherwise evident from the context (except where such number would exceed 100% of a possible value).

[0284] Associated with: As used herein, the terms “associated with,” “conjugated,” “linked,” “attached,” and “tethered,” when used with respect to two or more moieties, mean that the moieties are physically associated or connected with one another, either directly or via one or more additional moieties that serve as linking agents, to form a structure that is sufficiently stable so that the moieties remain physically associated under the conditions in which the structure is used, e.g., physiological conditions. An “association” need not be strictly through direct covalent chemical bonding. It may also suggest ionic or hydrogen bonding or a hybridization based connectivity sufficiently stable such that the “associated” entities remain physically associated.

[0285] Biomolecule: As used herein, the term “biomolecule” is any natural molecule which is amino acid-based, nucleic acid-based, carbohydrate-based or lipid-based, and the like.

[0286] Biologically active: As used herein, the phrase “biologically active” refers to a characteristic of any substance (e.g., an scAAV) that has activity in or on a biological system and/or organism. For instance, a substance that, when administered to an organism, has a biological effect on that organism, is considered to be biologically active. In particular embodiments, a compounds and/or compositions of the present invention may be considered biologically active if even a portion of is biologically active or mimics an activity considered to be biologically relevant.

[0287] Biological system: As used herein, the term “biological system” refers to a group of organs, tissues, cells, intracellular components, proteins, nucleic acids, molecules (including, but not limited to biomolecules) that function together to perform a certain biological task within cellular membranes, cellular compartments, cells, tissues, organs, organ systems, multicellular organisms, or any biological entity. In some embodiments, biological systems are cell signaling pathways comprising intracellular and/or extracellular cell signaling biomolecules. In some embodiments,

biological systems comprise growth factor signaling events within the extracellular/cellular matrix and/or cellular niches.

[0288] Compound: As used herein, the term “compound,” refers to a distinct chemical entity. In some embodiments, a particular compound may exist in one or more isomeric or isotopic forms (including, but not limited to stereoisomers, geometric isomers and isotopes). In some embodiments, a compound is provided or utilized in only a single such form. In some embodiments, a compound is provided or utilized as a mixture of two or more such forms (including, but not limited to a racemic mixture of stereoisomers). Those of skill in the art appreciate that some compounds exist in different such forms, show different properties and/or activities (including, but not limited to biological activities). In such cases it is within the ordinary skill of those in the art to select or avoid particular forms of the compound for use in accordance with the present invention. For example, compounds that contain asymmetrically substituted carbon atoms can be isolated in optically active or racemic forms. Methods on how to prepare optically active forms from optically active starting materials are known in the art, such as by resolution of racemic mixtures or by stereoselective synthesis.

[0289] Conserved: As used herein, the term “conserved” refers to nucleotides or amino acid residues of polynucleotide or polypeptide sequences, respectively, that are those that occur unaltered in the same position of two or more sequences being compared. Nucleotides or amino acids that are relatively conserved are those that are conserved among more related sequences than nucleotides or amino acids appearing elsewhere in the sequences.

[0290] In some embodiments, two or more sequences are said to be “completely conserved” if they are 100% identical to one another. In some embodiments, two or more sequences are said to be “highly conserved” if they are at least 70% identical, at least 80% identical, at least 90% identical, or at least 95% identical to one another. In some embodiments, two or more sequences are said to be “highly conserved” if they are about 70% identical, about 80% identical, about 90% identical, about 95%, about 98%, or about 99% identical to one another. In some embodiments, two or more sequences are said to be “conserved” if they are at least 30% identical, at least 40% identical, at least 50% identical, at least 60% identical, at least 70% identical, at least 80% identical, at least 90% identical, or at least 95% identical to one another. In some embodiments, two or more sequences are said to be “conserved” if they are about 30% c identical, about 40% identical, about 50% identical, about 60% identical, about 70% identical, about 80% c identical, about 90% identical, about 95% identical, about 98% identical, or about 99% identical to one another. Conservation of sequence may apply to the entire length of an oligonucleotide or polypeptide or may apply to a portion, region or feature thereof.

[0291] In one embodiment, conserved sequences are not contiguous. Those skilled in the art are able to appreciate how to achieve alignment when gaps in contiguous alignment are present between sequences, and to align corresponding residues notwithstanding insertions or deletions present.

[0292] Delivery: As used herein, “delivery” refers to the act or manner of delivering a compound, substance, entity,

moiety, cargo or payload to a target. Such target may be a cell, tissue, organ, organism, or system (whether biological or production).

[0293] Delivery Agent: As used herein, “delivery agent” refers to any agent which facilitates, at least in part, the *in vivo* delivery of one or more substances (including, but not limited to a compounds and/or compositions of the present invention, e.g., viral particles or expression vectors) to targeted cells.

[0294] Destabilized As used herein, the term “destable,” “destabilize,” or “destabilizing region” means a region or molecule that is less stable than a starting, reference, wild-type or native form of the same region or molecule.

[0295] Detectable label: As used herein, “detectable label” refers to one or more markers, signals, or moieties which are attached, incorporated or associated with another entity, which markers, signals or moieties are readily detected by methods known in the art including radiography, fluorescence, chemiluminescence, enzymatic activity, absorbance, immunological detection and the like. Detectable labels may include radioisotopes, fluorophores, chromophores, enzymes, dyes, metal ions, ligands, biotin, avidin, streptavidin and haptens, quantum dots, polyhistidine tags, myc tags, flag tags, human influenza hemagglutinin (HA) tags and the like. Detectable labels may be located at any position in the entity with which they are attached, incorporated or associated. For example, when attached, incorporated in or associated with a peptide or protein, they may be within the amino acids, the peptides, or proteins, or located at the N- or C-termini.

[0296] Engineered: used herein, embodiments of the invention are engineered” when they are designed to have a feature or property, whether structural or chemical, that varies from a starting point, wild-type or native molecule. Thus, engineered agents or entities are those whose design and/or production include an act of the hand of man.

[0297] Epitope: As used herein, an “epitope” refers to a surface or region on a molecule that is capable of interacting with a biomolecule. For example a protein may contain one or more amino acids, e.g., an epitope, which interacts with an antibody, e.g., a biomolecule. In some embodiments, when referring to a protein or protein module, an epitope may comprise a linear stretch of amino acids or a three dimensional structure formed by folded amino acid chains.

[0298] eRBSR: As used herein, an “eRBSR” is an engineered Rep binding sequence region. eRBSRs may range in size from 4-160 nucleotides. When making reference to any of Rep binding sequence, Rep binding sequence region, or eRBSR, it is understood that these sequences and/or regions may be substantially double stranded and therefore the recited lengths or sizes may be greater.

[0299] Expression: As used herein, “expression” of a nucleic acid sequence refers to one or more of the following events: (1) production of an RNA template from a DNA sequence (e.g., by transcription); (2) processing of an RNA transcript (e.g., by splicing, editing, 5' cap formation, and/or 3' end processing); (3) translation of an RNA into a polypeptide or protein; (4) folding of a polypeptide or protein; and (5) post-translational modification of a polypeptide or protein.

[0300] Feature: As used herein, a “feature” refers to a characteristic, a property, or a distinctive element.

[0301] Formulation: As used herein, a “formulation” includes at least a compound and/or composition of the present invention (e.g., a vector, scAAV particle, etc.) and a delivery agent.

[0302] Fragment: A “fragment,” as used herein, refers to a contiguous portion of a whole. For example, fragments of proteins may comprise polypeptides obtained by digesting full-length protein isolated from cultured cells. In some embodiments, a fragment of a protein includes at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250 or more amino acids. In some embodiments, fragments of an antibody include portions of an antibody subjected to enzymatic digestion or synthesized as such.

[0303] Functional: As used herein, a “functional” biological molecule is a biological entity with a structure and in a form in which it exhibits a property and/or activity by which it is characterized.

[0304] Homology: As used herein, the term “homology” refers to the overall relatedness between polymeric molecules, e.g. between nucleic acid molecules (e.g. DNA molecules and/or RNA molecules) and/or between polypeptide molecules. In some embodiments, polymeric molecules are considered to be “homologous” to one another if their sequences are at least 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99% identical or similar. The term “homologous” necessarily refers to a comparison between at least two sequences (polynucleotide or polypeptide sequences). In accordance with the invention, two polynucleotide sequences are considered to be homologous if the polypeptides they encode are at least about 50%, 60%, 70%, 80%, 90%, 95%, or even 99% for at least one stretch of at least about 20 amino acids. In some embodiments, homologous polynucleotide sequences are characterized by the ability to encode a stretch of at least 4-5 uniquely specified amino acids. For polynucleotide sequences less than 60 nucleotides in length, homology is typically determined by the ability to encode a stretch of at least 4-5 uniquely specified amino acids. In accordance with the invention, two protein sequences are considered to be homologous if the proteins are at least about 50%, 60%, 70%, 80%, or 90% identical for at least one stretch of at least about 20 amino acids. In many embodiments, homologous protein may show a large overall degree of homology and a high degree of homology over at least one short stretch of at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50 or more amino acids. In many embodiments, homologous proteins share one or more characteristic sequence elements. As used herein, the term “characteristic sequence element” refers to a motif present in related proteins. In some embodiments, the presence of such motifs correlates with a particular activity (such as biological activity).

[0305] Identity: As used herein, the term “identity” refers to the overall relatedness between polymeric molecules, e.g., between oligonucleotide molecules (e.g. DNA molecules and/or RNA molecules) and/or between polypeptide molecules. Calculation of the percent identity of two polynucleotide sequences, for example, may be performed by aligning the two sequences for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second nucleic acid sequences for optimal alignment and non-identical sequences can be disregarded for comparison purposes). In certain embodiments, the length of a sequence

aligned for comparison purposes is at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or 100% of the length of the reference sequence. The nucleotides at corresponding nucleotide positions are then compared. When a position in the first sequence is occupied by the same nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which needs to be introduced for optimal alignment of the two sequences. The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. For example, the percent identity between two nucleotide sequences can be determined using methods such as those described in Computational Molecular Biology, Lesk, A. M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D. W., ed., Academic Press, New York, 1993; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; Computer Analysis of Sequence Data, Part I. Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; each of which is incorporated herein by reference. For example, the percent identity between two nucleotide sequences can be determined, for example using the algorithm of Meyers and Miller (CABIOS, 1989, 4:11-17), which has been incorporated into the ALIGN program (version 2.0) using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4. The percent identity between two nucleotide sequences can, alternatively, be determined using the GAP program in the GCG software package using an NWSgapdna.CMP matrix. Methods commonly employed to determine percent identity between sequences include, but are not limited to those disclosed in Carillo, H., and Lipman, D., SIAM J Applied Math., 48:1073 (1988); incorporated herein by reference. Techniques for determining identity are codified in publicly available computer programs. Computer software to determine homology between two sequences include, but are not limited to, GCG program package, Devereux, J., et al., *Nucleic Acids Research*, 12(1), 387 (1984)), BLASTP, BLASTN, and FASTA Altschul, S. F. et al. *J. Molec. Biol.*, 215, 403 (1990)).

[0306] Inhibit expression of a gene: As used herein, the phrase “inhibit expression of a gene” means to cause a reduction in the amount of an expression product of the gene. The expression product may be RNA transcribed from the gene (e.g. mRNA) or a polypeptide translated from snRNA transcribed from the gene. Typically a reduction in the level of mRNA results in a reduction in the level of a polypeptide translated therefrom. The level of expression may be determined using standard techniques for measuring mRNA or protein.

[0307] In vitro: As used herein, the term “in vitro” refers to events that occur in an artificial environment, e.g., in a test tube or reaction vessel, in cell culture, in a Petri dish, etc., rather than within an organism (e.g., animal, plant, or microbe).

[0308] In vivo: As used herein, the term “in vivo” refers to events that occur within an organism (e.g., animal, plant, or microbe or cell or tissue thereof).

[0309] Isolated: As used herein, the term “isolated” is synonymous with “separated,” but carries with it the inference separation was carried out by the hand of man. In one embodiment, an isolated substance or entity is one that has been separated from at least some of the components with which it was previously associated (whether in nature or in an experimental setting). Isolated substances may have varying levels of purity in reference to the substances from which they have been associated. Isolated substances and/or entities may be separated from at least about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, or more of the other components with which they were initially associated. In some embodiments, isolated agents are more than about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 950, about 96%, about 97%, about 98%, about 99%, or more than about 99% pure. As used herein, a substance is “pure” if it is substantially free of other components.

[0310] Substantially isolated: By “substantially isolated” is meant that the compound is substantially separated from the environment in which it was formed or detected. Partial separation can include, for example, a composition enriched in the compound of the present disclosure. Substantial separation can include compositions containing at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 97%, or at least about 99% by weight of the compound of the present disclosure, or salt thereof. Methods for isolating compounds and their salts are routine in the art. In some embodiments, isolation of a substance or entity includes disruption of chemical associations and/or bonds. In some embodiments, isolation includes only the separation from components with which the isolated substance or entity was previously combined and does not include such disruption.

[0311] Linker: As used herein, a linker refers to a moiety that connects two or more domains, moieties or entities. In one embodiment, a linker may comprise 10 or more atoms. In a further embodiment, a linker may comprise a group of atoms, e.g., 10-1,000 atoms, and can be comprised of the atoms or groups such as, but not limited to, carbon, amino, alkylamino, oxygen, sulfur, sulfoxide, sulfonyl, carbonyl, and imine. In some embodiments, a linker may comprise one or more nucleic acids comprising one or more nucleotides. In some embodiments, the linker may comprise an amino acid, peptide, polypeptide or protein. In some embodiments, a moiety bound by a linker may include, but is not limited to an atom, a chemical group, a nucleoside, a nucleotide, a nucleobase, a sugar, a nucleic acid, an amino acid, a peptide, a polypeptide, a protein, a protein complex, a payload (e.g., a therapeutic agent marker (including, but not limited to a chemical, fluorescent, radioactive or bioluminescent marker). The linker can be used for any useful purpose, such as to form multimers or conjugates, as well as to administer a payload, as described herein. Examples of chemical groups that can be incorporated into the linker include, but are not limited to, alkyl, alkenyl, alkynyl, amino, imino, ether, thioether, ester, alkylene, heteroalkylene, aryl, or heterocyclyl, each of which can be optionally substituted, as described herein. Examples of linkers include, but are not limited to, unsaturated alkanes, polyethylene glycols (e.g., ethylene or propylene glycol monomeric units, e.g., diethylene glycol, dipropylene glycol, triethylene glycol, tripropylene glycol,

tetraethylene glycol, or tetraethylene glycol), and dextran polymers. Other examples include, but are not limited to, cleavable moieties within the linker, such as, for example, a disulfide bond (—S—S—) or an azo bond (—N=N—), which can be cleaved using a reducing agent or photolysis. Non-limiting examples of a selectively cleavable bonds include an amido bond which may be cleaved for example by the use of tris(2-carboxyethyl)phosphine (TCEP), or other reducing agents, and/or photolysis, as well as an ester bond which may be cleaved for example by acidic or basic hydrolysis.

[0312] Modified: As used herein, the term “modified” refers to a changed state or structure of a molecule or entity as compared with a parent or reference molecule or entity. Molecules may be modified in many ways including chemically, structurally, and functionally. In some embodiments, compounds and/or compositions of the present invention are modified by the introduction of non-natural amino acids, or non-natural nucleotides.

[0313] Mutation: As used herein, the term “mutation” refers to a change and/or alteration. In some embodiments, mutations may be changes and/or alterations to proteins (including peptides and polypeptides) and/or nucleic acids (including polynucleic acids). In some embodiments, mutations comprise changes and/or alterations to a protein and/or nucleic acid sequence. Such changes and/or alterations may comprise the addition, substitution and or deletion of one or more amino acids (in the case of proteins and/or peptides) and/or nucleotides (in the case of nucleic acids and or polynucleic acids). In embodiments wherein mutations comprise the addition and/or substitution of amino acids and/or nucleotides, such additions and/or substitutions may comprise 1 or more amino acid and/or nucleotide residues and may include modified amino acids and/or nucleotides.

[0314] Naturally occurring: As used herein, “naturally occurring” means existing in nature without artificial aid, or involvement of the hand of man.

[0315] Non-human vertebrate: As used herein, a “non-human vertebrate” includes all vertebrates except Homo sapiens, including wild and domesticated species. Examples of non-human vertebrates include, but are not limited to, mammals, such as alpaca, banteng, bison, camel, cat, cattle, deer, dog, donkey, gayal, goat, guinea pig, horse, llama, mule, pig, rabbit, reindeer, sheep water buffalo, and yak.

[0316] Off-target: As used herein, “off target” refers to any unintended effect on any one or more target, gene and/or cellular transcript.

[0317] Operably linked: As used herein, the phrase “operably linked” refers to a functional connection between two or more molecules, constructs, transcripts, entities, moieties or the like.

[0318] Particle: As used herein, a “particle” is a virus comprised of at least two components, a protein capsid and a polynucleotide sequence enclosed within the capsid.

[0319] Patient: As used herein, “patient” refers to a subject who may seek or be in need of treatment, requires treatment, is receiving treatment, will receive treatment, or a subject who is under care by a trained (e.g., licensed) professional for a particular disease or condition.

[0320] Payload construct: As used herein, “payload construct” is one or more polynucleotide regions encoding or comprising a payload molecule of interest, e.g., a transgene, a polynucleotide encoding a polypeptide or multi-polypeptide or a modulatory or regulatory nucleic acid, that that is

flanked on one or both sides by an inverted terminal repeat (ITR) sequence. A vector which comprises a payload construct is a “payload construct vector.”

[0321] Payload construct expression vector: As used herein, a “payload construct expression vector” is a vector which comprises a payload construct and which further comprises one or more polynucleotide regions encoding or comprising components for viral expression in a viral replication cell.

[0322] Peptide: As used herein, the term “peptide” refers to a chain of amino acids that is less than or equal to about 50 amino acids long, e.g., about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 amino acids long.

[0323] Pharmaceutically acceptable: The phrase “pharmaceutically acceptable” is employed herein to refer to those compounds, materials, compositions, and/or dosage forms which are, within the scope of sound medical judgment, suitable for use in contact with the tissues of human beings and animals without excessive toxicity, irritation, allergic response, or other problem or complication, commensurate with a reasonable benefit/risk ratio.

[0324] Pharmaceutically acceptable excipients: As used herein, the term “pharmaceutically acceptable excipient,” as used herein, refers to any ingredient other than active agents (e.g., as described herein) present in pharmaceutical compositions and having the properties of being substantially nontoxic and non-inflammatory in subjects. In some embodiments, pharmaceutically acceptable excipients are vehicles capable of suspending and/or dissolving active agents. Excipients may include, for example: antiadherents, antioxidants, binders, coatings, compression aids, disintegrants, dyes (colors), emollients, emulsifiers, fillers (diluent), film formers or coatings, flavors, fragrances, glidants (flow enhancers), lubricants, preservatives, printing inks, sorbents, suspending or dispersing agents, sweeteners, and waters of hydration. Excipients include, but are not limited to: butylated hydroxytoluene (BHT), calcium carbonate, calcium phosphate (dibasic), calcium stearate, croscarmellose, crosslinked polyvinyl pyrrolidone, citric acid, crospovidone, cysteine, ethylcellulose, gelatin, hydroxypropyl cellulose, hydroxypropyl methylcellulose, lactose, magnesium stearate, maltitol, mannitol, methionine, methylcellulose, methyl paraben, microcrystalline cellulose, polyethylene glycol, polyvinyl pyrrolidone, povidone, pregelatinized starch, propyl paraben, retinyl palmitate, shellac, silicon dioxide, sodium carboxymethyl cellulose, sodium citrate, sodium starch glycolate, sorbitol, starch (corn), stearic acid, sucrose, talc, titanium dioxide, vitamin A, vitamin E, vitamin C, and.

[0325] Pharmaceutically acceptable salts: Pharmaceutically acceptable salts of the compounds described herein are forms of the disclosed compounds wherein the acid or base moiety is in its salt form (e.g., as generated by reacting a free base group with a suitable organic acid). Examples of pharmaceutically acceptable salts include, but are not limited to, mineral or organic acid salts of basic residues such as amines; alkali or organic salts of acidic residues such as carboxylic acids; and the like. Representative acid addition salts include acetate, adipate, alginate, ascorbate, aspartate, benzenesulfonate, benzoate, bisulfate, borate, butyrate, camphorate, camphorsulfonate, citrate, cyclopentanepropionate, digluconate, dodecylsulfate, ethanesulfonate, fumarate, glucoheptonate, glycerophosphate, hemisulfate, heptonate, hexanoate, hydrobromide, hydrochloride, hydroiodide,

2-hydroxy-ethanesulfonate, lactobionate, lactate, laurate, lauryl sulfate, malate, maleate, malonate, methanesulfonate, 2-naphthalenesulfonate, nicotinate, nitrate, oleate, oxalate, palmitate, pamoate, pectinate, persulfate, 3-phenylpropionate, phosphate, picrate, pivalate, propionate, stearate, succinate, sulfate, tartrate, thiocyanate, toluenesulfonate, undecanoate, valerate salts, and the like. Representative alkali or alkaline earth metal salts include sodium, lithium, potassium, calcium, magnesium, and the like, as well as nontoxic ammonium, quaternary ammonium, and amine cations, including, but not limited to ammonium, tetramethylammonium, tetraethylammonium, methylamine, dimethylamine, trimethylamine, triethylamine, ethylamine, and the like. Pharmaceutically acceptable salts include the conventional non-toxic salts, for example, from non-toxic inorganic or organic acids. In some embodiments a pharmaceutically acceptable salt is prepared from a parent compound which contains a basic or acidic moiety by conventional chemical methods. Generally, such salts can be prepared by reacting the free acid or base forms of these compounds with a stoichiometric amount of the appropriate base or acid in water or in an organic solvent, or in a mixture of the two; generally, nonaqueous media like ether, ethyl acetate, ethanol, isopropanol, or acetonitrile are preferred. Lists of suitable salts are found in *Remington's Pharmaceutical Sciences*, 17th ed., Mack Publishing Company, Easton, Pa., 1985, p. 1418, *Pharmaceutical Salts: Properties, Selection, and Use*, P. H. Stahl and C. G. Wermuth (eds.), Wiley-VCH, 2008, and Berge et al., *Journal of Pharmaceutical Science*, 66, 1-19 (1977), each of which is incorporated herein by reference in its entirety. Pharmaceutically acceptable solvate: The term “pharmaceutically acceptable solvate,” as used herein, refers to a crystalline form of a compound wherein molecules of a suitable solvent are incorporated in the crystal lattice. For example, solvates may be prepared by crystallization, recrystallization, or precipitation from a solution that includes organic solvents, water, or a mixture thereof. Examples of suitable solvents are ethanol, water (for example, mono-, di-, and tri-hydrates), N-methylpyrrolidone (NMP), dimethyl sulfoxide (DMSO), N,N'-dimethylformamide (DMF), N,N'-dimethylacetamide (DMAC), 1,3-dimethyl-2-imidazolidinone (DMEU), 1,3-dimethyl-3,4,5,6-tetrahydro-2-(1H)-pyrimidinone (DMPU), acetonitrile (ACN), propylene glycol, ethyl acetate, benzyl alcohol, 2-pyrrolidone, benzyl benzoate, and the like. When water is the solvent, the solvate is referred to as a “hydrate.” In some embodiments, the solvent incorporated into a solvate is of a type or at a level that is physiologically tolerable to an organism to which the solvate is administered (e.g., in a unit dosage form of a pharmaceutical composition).

[0326] Pharmacokinetic: As used herein, “pharmacokinetic” refers to any one or more properties of a molecule or compound as it relates to the determination of the fate of substances administered to living organisms. Pharmacokinetics are divided into several areas including the extent and rate of absorption, distribution, metabolism and excretion. This is commonly referred to as ADME where: (A) Absorption is the process of a substance entering the blood circulation; (D) Distribution is the dispersion or dissemination of substances throughout the fluids and tissues of the body; (M) Metabolism (or Biotransformation) is the irreversible transformation of parent compounds into daughter metabolites; and (E) Excretion (or Elimination) refers to the elimination

of the substances from the body. In rare cases, some drugs irreversibly accumulate in body tissue.

[0327] Physicochemical: As used herein, “physicochemical” means of or relating to a physical and/or chemical property.

[0328] Preventing: As used herein, the term “preventing” refers to partially or completely delaying onset of an infection, disease, disorder and/or condition; partially or completely delaying onset of one or more symptoms, features, or clinical manifestations of a particular infection, disease, disorder, and/or condition; partially or completely delaying onset of one or more symptoms, features, or manifestations of a particular infection, disease, disorder, and/or condition; partially or completely delaying progression from an infection, a particular disease, disorder and/or condition; and/or decreasing the risk of developing pathology associated with the infection, the disease, disorder, and/or condition.

[0329] Prolitrates As used herein, the term “proliferate” means to grow, expand, replicate or increase or cause to grow, expand, replicate or increase. “Proliferative” means having the ability to proliferate. “Anti-proliferative” means having properties counter to or in opposition to proliferative properties.

[0330] Protein of interest: As used herein, the terms “proteins of interest” or “desired proteins” include those provided herein and fragments, mutants, variants, and alterations thereof.

[0331] Purified: As used herein, the term “purify” means to make substantially pure or clear from unwanted components, material defilement, admixture or imperfection. “Purified” refers to the state of being pure. “Purification” refers to the process of making pure.

[0332] Region: As used herein, the term “region” refers to a zone or general area. In some embodiments, when referring to a protein or protein module, a region may comprise a linear sequence of amino acids along the protein or protein module or may comprise a three dimensional area, an epitope and/or a cluster of eptiopes. In some embodiments, regions comprise terminal regions. As used herein, the term “terminal region” refers to regions located at the ends or termini of a given agent. When referring to proteins, terminal regions may comprise N- and/or C-termini. N-termini refer to the end of a protein comprising an amino acid with a free amino group. C-termini refer to the end of a protein comprising an amino acid with a free carboxyl group. N- and/or C-terminal regions may there for comprise the N- and/or C-termini as well as surrounding amino acids. In some embodiments, N- and/or C-terminal regions comprise from about 3 amino acid to about 30 amino acids, from about 5 amino acids to about 40 amino acids, from about 10 amino acids to about 50 amino acids, from about 20 amino acids to about 100 amino acids and/or at least 100 amino acids. In some embodiments, N-terminal regions may comprise any length of amino acids that includes the N-terminus, but does not include the C-terminus. In some embodiments, C-terminal regions may comprise any length of amino acids, which include the C-terminus, but do not comprise the N-terminus.

[0333] In some embodiments, when referring to a polynucleotide, a region may comprise a linear sequence of nucleic acids along the polynucleotide or may comprise a three dimensional area, secondary structure, or tertiary structure. In some embodiments, regions comprise terminal regions. As used herein, the term “terminal region” refers to

regions located, at the ends or termini of a given agent. When referring to polynucleotides, terminal regions may comprise 5' and 3' termini. 5' termini refer to the end of a polynucleotide comprising a nucleic acid with a free phosphate group. 3' termini refer to the end of a polynucleotide comprising a nucleic acid with a free hydroxyl group, 5' and 3' regions may there for comprise the 5' and 3' termini as well as surrounding nucleic acids. In some embodiments, 5' and 3' terminal regions comprise from about 9 nucleic acids to about 90 nucleic acids, from about 15 nucleic acids to about 120 nucleic acids, from about 30 nucleic acids to about 150 nucleic acids, from about 60 nucleic acids to about 300 nucleic acids and/or at least 300 nucleic acids. In some embodiments, 5' regions may comprise any length of nucleic acids that includes the 5' terminus, but does not include the 3' terminus. In some embodiments, 3' regions may comprise any length of nucleic acids, which include the 3' terminus, but does not comprise the 5' terminus,

[0334] Rep binding sequence: As used herein, a “Rep binding sequence” is a series of linked nucleosides which interact with one or more Rep proteins. Rep binding sequences may range from 4-160 nucleotides.

[0335] Rep binding sequence region: As used herein, a “Rep binding sequence region” is a nucleotide sequence which interacts with one or more Rep proteins. A Rep binding sequence region may range from 4-160 nucleotides. In one embodiment, the Rep binding region is a nucleotide sequence located within an inverted terminal repeat (ITR).

[0336] Sample: As used herein, the term “sample” refers to an aliquot or portion taken from a source and/or provided for analysis or processing. In some embodiments, a sample is from a biological source such as a tissue, cell or component part (e.g. a body fluid, including but not limited to blood, mucus, lymphatic fluid, synovial fluid, cerebrospinal fluid, saliva, amniotic fluid, amniotic cord blood, urine, vaginal fluid and semen). In some embodiments, a sample may be or comprise a homogenate, lysate or extract prepared from a whole organism or a subset of its tissues, cells or component parts, or a fraction or portion thereof, including but not limited to, for example, plasma, serum, spinal fluid, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, blood cells, tumors, organs. In some embodiments, a sample is or comprises a medium, such as a nutrient broth or gel, which may contain cellular components, such as proteins or nucleic acid molecule. In some embodiments, a “primary” sample is an aliquot of the source. In some embodiments, a primary sample is subjected to one or more processing (e.g., separation, purification, etc) steps to prepare a sample for analysis or other use.

[0337] Self-complementary genome: As used herein, a “self-complementary genome” is a polynucleotide comprising, in the 5' to 3' direction, a first parvoviral inverted terminal repeat sequence, a first heterologous sequence, a parvoviral inverted terminal repeat (ITR) nucleotide sequence comprising an engineered Rep binding sequence region, a second heterologous sequence, wherein the second heterologous sequence is complementary to the first heterologous sequence, and a second parvoviral inverted terminal repeat sequence.

[0338] Self-complementary viral particle: As used herein, a “self-complementary viral particle” is a particle comprised of at least two components, a protein capsid and a poly-

nucleotide sequence encoding a self-complementary genome enclosed within the capsid.

[0339] Signal Sequences: As used herein, the phrase “signal sequences” refers to a sequence which can direct the transport or localization.

[0340] Single unit dose: As used herein, a “single unit dose” is a dose of any therapeutic administered in one dose/at one time/single route/single point of contact, i.e., single administration event. In some embodiments, a single unit dose is provided as a discrete dosage form (e.g., a tablet, capsule, patch, loaded syringe, vial, etc.).

[0341] Similarity: As used herein, the term “similarity” refers to the overall relatedness between polymeric molecules, e.g. between polynucleotide molecules (e.g. DNA molecules and/or RNA molecules) and/or between polypeptide molecules. Calculation of percent similarity of polymeric molecules to one another can be performed in the same manner as a calculation of percent identity, except that calculation of percent similarity takes into account conservative substitutions as is understood in the art.

[0342] Split dose: As used herein, a “split dose” is the division of single unit dose or total daily dose into two or more doses.

[0343] Stable: As used herein “stable” refers to a compound or entity that is sufficiently robust to survive isolation to a useful degree of purity from a reaction mixture, and preferably capable of formulation into an efficacious therapeutic agent.

[0344] Stabilized: As used herein, the term “stabilize,” “stabilized,” “stabilized region” means to make or become stable. In some embodiments, stability is measured relative to an absolute value. In some embodiments, stability is measured relative to a reference compound or entity.

[0345] Subject: As used herein, the term “subject” or “patient” refers to any organism to which a composition in accordance with the invention may be administered, e.g., for experimental, diagnostic, prophylactic, and/or therapeutic purposes. Typical subjects include animals (e.g., mammals such as mice, rats, rabbits, non-human primates, and humans) and/or plants.

[0346] Substantially As used herein, the term “substantially” refers to the qualitative condition of exhibiting total or near-total extent or degree of a characteristic or property of interest. One of ordinary skill in the biological arts will understand that biological and chemical phenomena rarely, if ever, go to completion and/or proceed to completeness or achieve or avoid an absolute result. The term “substantially” is therefore used herein to capture the potential lack of completeness inherent in many biological and chemical phenomena.

[0347] Substantially equal: As used herein as it relates to time differences between doses, the term means plus/minus 2%.

[0348] Substantially simultaneously: As used herein and as it relates to plurality of doses, the term typically means within about 2 seconds.

[0349] Suffering from: An individual who is “suffering from” a disease, disorder, and/or condition has been diagnosed with or displays one or more symptoms of a disease, disorder, and/or condition.

[0350] Susceptible to: An individual who is “susceptible to” a disease, disorder, and/or condition has not been diagnosed with and/or may not exhibit symptoms of the disease, disorder, and/or condition but harbors a propensity to

develop a disease or its symptoms. In some embodiments, an individual who is susceptible to a disease, disorder, and/or condition example, cancer) may be characterized by one or more of the following: (1) a genetic mutation associated with development of the disease, disorder, and/or condition; (2) a genetic polymorphism associated with development of the disease, disorder, and/or condition; (3) increased and/or decreased expression and/or activity of a protein and/or nucleic acid associated with the disease, disorder, and/or condition; (4) habits and/or lifestyles associated with development of the disease, disorder, and/or condition; (5) a family history of the disease, disorder, and/or condition; and (6) exposure to and/or infection with a microbe associated with development of the disease, disorder, and/or condition. In some embodiments, an individual who is susceptible to a disease, disorder, and/or condition will develop the disease, disorder, and/or condition. In some embodiments, an individual who is susceptible to a disease, disorder, and/or condition will not develop the disease, disorder, and/or condition.

[0351] Synthetic: The term “synthetic” means produced, prepared, and/or manufactured by the hand of man. Synthesis of polynucleotides or polypeptides or other molecules of the present invention may be chemical or enzymatic.

[0352] Targeted Cells: As used herein, “targeted cells” refers to any one or more cells of interest. The cells may be found in vitro, in vivo, in situ or in the tissue or organ of an organism. The organism may be an animal, preferably a mammal, more preferably a human and most preferably a patient.

[0353] Therapeutic Agent: The term “therapeutic agent” refers to any agent that, when administered to a subject, has a therapeutic, diagnostic, and/or prophylactic effect and/or elicits a desired biological and/or pharmacological effect.

[0354] Therapeutically effective amount: As used herein, the term “therapeutically effective amount” means an amount of an agent to be delivered (e.g., nucleic acid, drug, therapeutic agent, diagnostic agent, prophylactic agent, etc.) that is sufficient, when administered to a subject suffering from or susceptible to an infection, disease, disorder, and/or condition, to treat, improve symptoms of, diagnose, prevent, and/or delay the onset of the infection, disease, disorder, and/or condition. In some embodiments, a therapeutically effective amount is provided in a single dose. In some embodiments, a therapeutically effective amount is administered in a dosage regimen comprising a plurality of doses. Those skilled in the art will appreciate that in some embodiments, a unit dosage form may be considered to comprise a therapeutically effective amount of a particular agent or entity if it comprises an amount that is effective when administered as part of such a dosage regimen.

[0355] Therapeutically effective outcome: As used herein, the term “therapeutically effective outcome” means an outcome that is sufficient in a subject suffering from or susceptible to an infection, disease, disorder, and/or condition, to treat, improve symptoms of, diagnose, prevent, and/or delay the onset of the infection, disease, disorder, and/or condition.

[0356] Total daily dose: As used herein, a “total daily dose” is an amount given or prescribed in a 24 hour period. It may be administered as a single unit dose.

[0357] Treating: As used herein, the term “treating” refers to partially or completely alleviating, ameliorating, improving, relieving, delaying onset of, inhibiting progression of, reducing severity of, and/or reducing incidence of one or

more symptoms or features of a particular infection, disease, disorder, and/or condition. For example, “treating” cancer may refer to inhibiting survival, growth, and/or spread of a tumor. Treatment may be administered to a subject who does not exhibit signs of a disease, disorder, and/or condition anchor to a subject who exhibits only early signs of a disease, disorder, and/or condition for the purpose of decreasing the risk of developing pathology associated with the disease, disorder, and/or condition.

[0358] Unmodified: As used herein, “unmodified” refers to any substance, compound or molecule prior to being changed in any way. Unmodified may, but does not always, refer to the wild-type or native form of a biomolecule or entity. Molecules or entities may undergo a series of modifications whereby each modified product may serve as the “unmodified” starting molecule or entity for a subsequent modification.

[0359] Vector: As used herein, a “vector” is any molecule or moiety which transports, transduces or otherwise acts as a carrier of a heterologous molecule.

[0360] Viral construct expression vector: As used herein, a “viral construct expression vector” is a vector which comprises one or more polynucleotide regions encoding or comprising Rep and/or Cap that further comprises one or more polynucleotide regions encoding or comprising components for viral expression in a viral replication cell. Viral vectors of the present invention may be produced recombinantly and may be based on and/or may comprise adeno-associated virus (AAV) parent or reference sequence. Such parent or reference AAV sequences may serve as an original, second, third or subsequent sequence for engineering viral vectors. In non-limiting examples, such parent or reference AAV sequences may comprise any one or more of the following sequences: a polynucleotide sequence encoding a polypeptide or multi-polypeptide, which may be wild-type or modified from wild-type and which may encode full-length or partial sequence of a protein, protein domain, or one or more subunits of a protein; a polynucleotide comprising a modulatory or regulatory nucleic acid which may be wild-type or modified from wild-type; and a transgene that may or may not be modified from wild-type sequence. These AAV sequences may serve as either the “donor” sequence of one or more codons (at the nucleic acid level) or amino acids (at the polypeptide level) or “acceptor” sequences of one or more codons (at the nucleic acid level) or amino acids (at the polypeptide level).

[0361] Viral construct vector: As used herein, a “viral construct vector” is a vector which comprises one or more polynucleotide regions encoding or comprising Rep and/or Cap.

[0362] Viral genome: As used herein, a “viral genome” is a payload construct that has been rescued from a payload construct expression vector and is packaged within a viral particle.

Equivalents and Scope

[0363] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments in accordance with the invention described herein. The scope of the present invention is not intended to be limited to the above Description, but rather is as set forth in the appended claims.

[0364] In the claims, articles such as “a,” “an” and “the” may mean one or more than one unless indicated to the contrary

or otherwise evident from the context. Claims or descriptions that include “or” between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention includes embodiments in which more than one, or the entire group members are present in, employed in, or otherwise relevant to a given product or process.

[0365] It is also noted that the term “comprising” is intended to be open and permits but does not require the inclusion of additional elements or steps. When the term “comprising” is used herein, the term “consisting of” is thus also encompassed and disclosed.

[0366] Where ranges are given, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value or subrange within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise.

[0367] In addition, it is to be understood that any particular embodiment of the present invention that falls within the prior art may be explicitly excluded from any one or more of the claims. Since such embodiments are deemed to be known to one of ordinary skill in the art, they may be excluded even if the exclusion is not set forth explicitly herein. Any particular embodiment of the compositions of the invention (e.g., any nucleic acid or protein encoded thereby; any method of production; any method of use; etc.) can be excluded from any one or more claims, for any reason, whether or not related to the existence of prior art.

[0368] All cited sources, for example, references, publications, databases, database entries, and art cited herein, are incorporated into this application by reference, even if not expressly stated in the citation. In case of conflicting statements of a cited source and the instant application, the statement in the instant application shall control.

[0369] Section and table headings are not intended to be limiting.

IX. EXAMPLES

Example 1

eRBSR Design

[0370] To identify the Rep binding sequence regions at the genomic sequence level, the genome sequences of several different AAV serotypes and species including AAV1, AAV2, AAV4, AAV5, AAV7, AAV8, Bat AAV, Bovine AAV, Snake AAV, Avian AAV DA-1, and Avian AAV VR865 (SEQ ID NO: 280-291) were obtained from the NCBI Genome database. The 5' and 3' inverted terminal repeat sequences of selected genomes were aligned with the Clustal Omega multiple sequence alignment program using either AAV2 or AAV5 as the parent sequence, as shown in FIG. 3 and FIG. 4, respectively. Additional Rep binding sequence regions found in the ITR sequences of other parvoviral serotypes and species, including AAV8 5' ITR (SEQ ID NO: 253), AAV8 3' ITR (SEQ ID NO: 254), Bat AAV 5' ITR (SEQ ID

NO: 261). Bat AAV 3' ITR (SEQ ID NO: 262) are obtained in the same manner using this alignment program.

[0371] Based on the genomic sequence alignments, two groups of Rep binding sequence regions, "AAV2-like" and "AAV5-like," were identified as regions of interest with FIG. 3 showing the sequences having the AAV2-like Rep binding sequence region and FIG. 4 showing the sequences having the AAV5-like Rep binding sequence region, in which the underlined regions represent the complementary Rep binding sequence region in the 5' and 3' ITR regions. The Rep binding sequence regions share a consensus GCTC motif that is repeated four to five times. In total, the identified consensus Rep binding sequence regions are about 16-20 nucleotides in length.

[0372] In some embodiments, eRBSRs comprise one, two, or three GCTC consensus motif(s). In other embodiments, eRBSRs comprise four or five GCTC consensus motifs in which at least one nucleotide in one or more of the consensus motif(s) is altered to a motif sequence selected from the group consisting of (1) NCTC, (2) GNTC, (3) GCNC, (4) GCTN, and (5) in the case of two or more altered motif sequences, any combination of these modified altered motifs, wherein N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine. In a further embodiment, eRBSRs comprise four or five GCTC: consensus motifs in which at least one nucleotide in one or more of the GCTC consensus motif(s) contains at least one nucleotide which is chemically modified.

[0373] Table 1 provides the Rep binding sequence regions of these different AAV serotypes and species (SEQ ID NOs: 1-20). Table 2 provides engineered Rep binding sequence regions which can function as an eRBSR (SEQ ID NOs: 21-66). Table 3 provides oligonucleotide sequences that comprise Rep protein binding properties which may therefore function as an eRBSR (SEQ ID NOs: 79-197) (Chiorini et al. 1995 Journal of Virology- 69(11) 7334-7338). Any of SEQ ID NOs: 1-20, 21-66, and/or 79-197 can be used as an eRBSR, for example, by replacing a native or wild-type Rep binding sequence region with any heterologous sequence of SEQ ID NOs: 1-20, 21-66, and/or 79-197. For example, any of the native Rep binding sequence regions found in AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-a18 can be replaced with any of SEQ ID NOs: 1-20, 21-66, and/or 79-197 as long as the eRBSR is derived from a different AAV serotype or species than that of the AAV ITR. Alternatively, any of SEQ ID NOs: 1-20, 21-66, and/or 79-197 can be used as an eRBSR by altering or modifying one or more nucleotides in any of the sequences.

Example 2

eRBSR with Reduced Binding Affinity for Rep Protein

[0374] Any of the eRBSRs provided in Example 1 or elsewhere herein can be tested to confirm its reduced binding to Rep protein by determining the dissociation constant or K_d. The K_d of wild-type Rep78 protein for its cognate Rep binding sequence region in AAV2 is approximately 1 nM as measured by electro-mobility shift assay (EMSA) (Chiorini et al. 1994, Sequence Requirements for Stable Binding and Function of Rep68 on the Adeno-Associated Virus Type 2 Inverted Terminal Repeats, Journal of Virology

68:11; 7448-7457). Decreased binding is measured as an increase in the dissociation constant, K_d, over wild-type, i.e., a K_d greater than 1 nM.

[0375] In previous experiments, the 5-5/EGR oligonucleotide (SEQ ID NO:71) shown in FIG. 11C exhibits stable, but weaker binding affinity for Rep78, the wild-type Rep found in AAV2, (5-10-fold less binding) than wild-type AAV2 Rep binding sequence region (Chiorini et al. Journal of Virology 1995, 7334-7338). In the same experiments it was found that alpha-PAL (SEQ ID NO: 68) and JcDNV NS1 (SEQ ID NO: 74) sequences, shown in FIGS. 11A and 11E respectively, were determined to exhibit weak, non-specific binding to Rep78 in EMSA experiments

[0376] The binding affinity of three different eRBSRs are tested in comparison with the binding of the known AAV2 Rep binding sequence region, JcDNV (SEQ ID NO: 74), α PAL (SEQ ID NO: 68), and 5-5/EGR (SEQ ID NO: 71) oligonucleotides are individually inserted into AAV2 ITR sequence by swapping them into (replacing) the native Rep binding sequence region of the AAV2 ITR, which has the formula [GCGC]-[GCTC]₃. (SEQ ID NO: 311), FIG. 11F, FIG. 11B, and FIG. 11D, respectively. The wild-type AAV2 ITR having the native AAV2 Rep binding sequence region is used as a positive control. All of the sequences are first designed in silico and then synthesized using standard molecular biology techniques well known to those of skill in the art See, e.g., Sambrook et al, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press (Cold Spring Harbor, N.Y.). Rep protein utilized in binding assays as described herein are expressed from polynucleotides transfected into *E. coli* cells and purified by methods known to those of skill in the art. Polypeptides can also be synthesized by well-known solid phase peptide synthesis methods (Merrifield, J. Am. Chem. Soc., 85:2149 (1962); Stewart and Young, Solid Phase Peptide Synthesis (Freeman, San Francisco, 1969) pp.27-62). These and other suitable production methods are within the knowledge of those of skill in the art and are not a limitation of the present invention.

[0377] To determine initial binding, the modified AAV2 ITRs having the various described sequences are chemically synthesized and purified according to standard procedures. Double-stranded DNA probes are individually generated by annealing primers to the modified AAV2 ITRs comprising the α PAL eRBSR, 5-5/EGR eRBSR, and JcDNV eRBSR sequences, as well as the wild-type AAV2 ITR, and extending the sequence with the Klenow fragment of DNA polymerase and radiolabeled nucleotides.

[0378] An in vitro binding assay is performed in which the radio-labeled double-stranded AAV2 ITRs comprising α PAL eRBSR, 5-5/EGR eRBSR, and JcDNV eRBSR sequences, as well as wild-type AAV2 ITR having the native AAV2 Rep binding sequence region, are combined with varying concentrations of wild-type Rep protein (Rep78) under conditions which favor Rep protein binding to the AAV2 Rep binding sequence region substrate. Chiorini et al. Journal of Virology 1995, 7334-7338 describes a Rep78 protein available for use in various experiments (see also, Chiorini et al., 1994, J. Virol., 68:797-804; Chiorini et al., J. Virol., 68: 7448-7457). The resultant protein-DNA complexes formed in the binding experiments are separated on a non-denaturing polyacrylamide gel under conditions known for electro-mobility shift assay. The concentrations of bound and free radiolabeled eRBSR are determined using a PhosphorImager (Molecular Dynamics). A plot of the ratio

of bound eRBSR versus Rep protein concentration is used to determine the K_d of each eRBSR.

[0379] To determine the binding affinities (K_d) of ∞ PAL eRBSR, 5-5/EGR eRBSR, and JcDNV eRBSR binding, competition binding EMSA experiments are performed, in which radio-labeled double-stranded wild-type AAV2 ITR is incubated with Rep78 in the presence of 5-, 10-, or 15-fold excess of the following unlabeled competitor sequences: (1) AAV2 ITR having the native AAV2 Rep binding sequence region (positive control), (2) modified AAV2 ITR having the ∞ PAL sequence, (3) modified ∞ AAV2 FIR having the 5-5/EGR sequence (eRBSR), and (4) modified AAV2 ITR having the JcDNV eRBSR sequence. The resultant protein-DNA complexes formed in the competition binding experiments are determined by EMSA. The concentrations of bound radiolabeled wild-type AAV2 ITR in the presence of various concentrations of unlabeled competitor DNA (eRBSRs) is determined using a PhosphorImager (Molecular Dynamics). A plot of the ratio of bound AAV2 ITR probe in the presence or absence of various concentrations of unlabeled competitor eRBSR DNA is used to determine the K_d of each eRBSR. eRBSR sequences having a K_d greater than that of the wild-type AAV2 ITR (with native Rep binding sequence region) bind with less affinity to Rep protein than the native Rep binding sequence region and are candidates for use in generating self-complementary viral particles. EMSA methods are provided in Chiorini et al. 1995 Determination of Adeno-Associated Virus Rep68 and Rep78 Binding Sites by Random Sequence Oligonucleotide Selection, *Journal of Virology* 69:11, 7334-7338, the contents of which are incorporated herein in their entirety.

Example 3

eRBSR with Reduced Banding Affinity for Rep Protein

[0380] Another technique used to test the binding of an eRBSR is surface plasmon resonance, which is a label free technique used to determine the binding affinity (K_d) between molecules. Briefly, a first molecule is immobilized on a biosensor adjacent to an internal reference surface. White light reflected from the two surfaces creates an interference pattern that is shifted when a second molecule is bound to the first. The second molecule in a first solution is passed over the biosensor in a constant flow to determine the rate of association with the first molecule until the signal is saturated to determine the K_d value of the binding interaction. A second solution devoid of the second molecule is then passed over the biosensor to determine the K_d value of the binding interaction.

[0381] The binding affinities of the following polypeptide-polynucleotide pairs are determined using an Octet surface plasmon resonance instrument (ForteBio, Menlo Park, Calif.): (1) AAV2 wild-type Rep protein (Rep78) and AAV2 wild-type ITR containing the native Rep binding sequence region; (2) AAV2 wild-type Rep protein (Rep78) and modified AAV2 ITR having the ∞ PAL eRBSR sequence, (3) AAV2 wild-type Rep protein (Rep78) and modified AAV2 ITR having the 5-5/EGR eRBSR sequence, and (4) AAV2 wild-type Rep protein (Rep78) and modified AAV2 ITR having the JcDNV eRBSR sequence. The wild-type Rep protein (Rep78) is immobilized on the biosensor layer. A solution containing either (1) AAV2 wild-type ITR containing the native Rep binding sequence region; (2) modified

AAV2 ITR having the ∞ PAL eRBSR sequence; (3) modified AAV2 ITR having the 5-5/EGR eRBSR sequence; or (4) modified AAV2 ITR having the JcDNV eRBSR sequence is passed over the biosensor having immobilized Rep78 protein to determine the rate of association for each of the polynucleotides. A second solution devoid of the polynucleotide is passed over the saturated biosensor layer to determine the rate of dissociation.

Example 4

Chimeric Nicking Stem Loop Design

[0382] To identify the Rep nicking stem loop at the genomic sequence level, the genome sequences of several different AAV serotypes and species including AAV1, AAV2, AAV3, AAV4, AAV5, AAV7, AAV8, Bat AAV, Bovine AAV, Snake AAV, Avian AAV DA-1, and Avian AAV VR865 (SEQ ID NO: 280-291) were obtained from the NCBI Genome database. The 5' and 3' inverted terminal repeat sequences of selected genomes were aligned with the Clustal Omega multiple sequence alignment program using either AAV2 or AAV5 as the parent sequence, as shown in FIG. 3 and FIG. 4, respectively. Additional Rep nicking stem loops found in the ITR sequences of other parvoviral serotypes and species, including AAV8 5' ITR (SEQ ID NO: 253), AAV8 3' ITR (SEQ ID NO: 254), Bat AAV 5' ITR (SEQ ID NO: 261), Bat AAV 3' ITR (SEQ ID NO: 262) are obtained in the same manner using this alignment program.

[0383] Based on the genomic sequence alignments, several different Rep nicking stem loops were identified (see boxed sequence in FIG. 3 and FIG. 4). The Rep nicking stem loops for AAV2, AAV1, AAV3, and AAV7 share the same nicking stem loop: AGAGGGAGTGGCCAACTCCATCA (SEQ ID NO:198) on the 5' ITR and its reverse complement TGATGGAGTTGGCCACTCCCTCT (SEQ ID NO:199) on the 3' ITR. AAV-4 has the same nicking sequence on the 5' ITR and a slightly different sequence on the 3' ITR, (TGATGGAGTTGGCCACATTAGCT; SEQ ID NO: 207). The Rep nicking stem loop for AAV5 is GGGGGAGAGT-GCCCACTCTCA (SEQ ID NO: 200) on the 5' ITR and its reverse complement is TGAGAGTGTGGCACTCTCCCCC (SEQ ID NO: 201) on the 3' ITR. The Rep nicking stem loop for bovine AAV is GGGGGGGAGTGGCCAACTCTCT (SEQ ID NO: 214) on the 5' ITR and its reverse complement is AGAGAGTGTGGCACTCCCCCCC (SEQ ID NO: 215) on the 3' ITR. The Rep nicking stem loop for snake AAV is TGGGGCGAGTGGCCCTGCTC (SEQ ID NO: 216) on the 5' ITR and its reverse complement is GAGCA-GGGCACTCGCCCCA (SEQ ID NO: 217) on the 3' ITR. The Rep nicking stem loop for avian AAV (AAVDA1 and VR865) is ACTGGCCAGCACTCCGGTGA (SEQ ID NO: 210) on the 5' ITR and its reverse complement is TCAC-CGGAGTGCTGGCCAGT (SEQ ID NO: 211) on the 3' ITR.

[0384] A chimeric nicking stem loop is achieved by swapping the nicking stem loop sequence of one parvoviral serotype or species with the native nicking stem loop sequence of a different parvoviral serotype or species. For example, the chimeric nicking stem loop for ∞ AAV5 (SEQ ID NO: 201) is swapped into (replaces) the native nicking stem loop sequence in any of the following ITR sequences: AAV4, AAV5, AAV6, AAV8, AAV9, AAV 10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 ITR. In one embodiment, the chimeric nicking stem loop for AAV5

(SEQ ID NO: 201) is swapped into (replaces) the native nicking stem loop sequence in AAV2 (SEQ ID NO: 199). Additional nicking stem loop sequences are shown in Table 4. The sequences for certain chimeric nicking stem loop ITR sequences are provided in Table 5.

Example 5

Design of Engineered Rep Protein

[0385] The Rep protein is altered so as to decrease its binding with its cognate Rep binding sequence region. Residues in the Rep protein that contact the phosphate backbone, one or more nucleosides, or both the phosphate backbone and one or more nucleosides are altered.

[0386] Modified Rep protein design was performed in silico utilizing structural and sequence analysis of the Rep protein interaction with the AAV Rep binding sequence region. FIG. 7 is a schematic depicting a 3-dimensional model showing Rep protein binding to the Rep binding domain of the AAV5 ITR. A crystal structure showing the binding interaction between AAV5 Rep protein residues 1-197 and a double stranded oligonucleotide sequence encoding the AAV5 Rep binding sequence region (SEQ ID NOs: 3-4) was previously analyzed (Hickman et al 2004 The Nuclease Domain of Adeno-Associated Virus Rep Coordinates Replication Initiation Using Two Distinct DNA Recognition interfaces, *Molecular Cell* 13:403-414, the contents of which are referenced herein in their entirety). According to these studies, five Rep protein monomers bind independently to the Rep binding sequence region on the ITR and spiral around the DNA axis, off-set by four base pairs. As discussed elsewhere herein a repeated GCTC tetranucleotide sequence is necessary for recognition by Rep protein. The two structural elements of Rep protein that are important for binding are the surface loop between the β 4 and β 5 (referred to as the β 4/ β 5 loop; residues 135-144) and the α -helix C (residues 101-118), which are located along one edge of the central β sheet. The β 4/ β 5 loop interacts with four bases from the major groove side. Of the eight residues that form direct side chain interactions with the Rep binding sequence region, seven are strictly conserved among serotypes AAV2-6. There are three consecutive amino acids in the β 4/ β 5 loop that are highly conserved- Gly-139, Gly-140, and Ala-141 (in AAV5), which correlate with three sequential Glycines in AAV2-6. Other conserved residues in the β 4/ β 5 loop are Asp-142, Lys-135, Lys-137, and Lys-138. Conserved residues in the α -helix C are Met-102, Arg-106, Ser-109, and Glu-110. Arg-106 and Lys-137 provide important base contacts and Lys-135 forms a salt bridge to the phosphate backbone. Mutations of three residues: Arg-106, Lys-135, and Lys-137 results in complete loss of Rep binding (Hickman et al, 2004 *Molecular Cell* 13:403-414).

[0387] An alignment of the polynucleotide sequences encoding Rep protein from different AAV serotypes (AAV5, AAV2Rep68, AAV2Rep78, AAV2Rep40, AAV2Rep52, AAV1, AAV3, AAV4, AAV7, and AAV8) and AAV species (Avian AAVDA-1, Avian AAVVR865, Bat, and Bovine) was constructed to identify those amino acids having conserved specificity and conserved affinity in areas of Rep protein secondary structure (α -helix and β sheet) at the sequence level. FIG. 8 shows the sequence alignment. Amino acids having conserved specificity and conserved affinity in Rep proteins found in the genomes of other parvoviral serotypes and species, including snake AAV (SEQ ID NO: 279), are

obtained in the same manner using this alignment program. An analysis of the correlation between structural features of the Rep binding sequence region in the ITR and the degree of conservation of the Rep protein sequence was performed in silico. The combination of sequence and structural analysis of Rep proteins identified nine conserved residues of the α -helix C and β -sheet 4/ β -sheet 5 loop, which were identified as targets for modification (see Table 6 and Table 7).

[0388] Residues in the Rep protein identified for alteration or modification are listed here. The residue numbering is that of AAV5, and which corresponding residues in other AAV serotypes and species can be obtained using the sequence alignment in FIG. 8. (1) the amino acid corresponding to Gly-139 can be substituted with a Proline, Alanine, or Serine residue; (2) the amino acid corresponding to Gly-140 can be substituted with a Proline, Alanine, or Serine residue; (3) the amino acid corresponding to Ala-141 can be substituted with a Serine, Glycine, Threonine, Cysteine, or Valine residue; (4) the amino acid corresponding to Lys-138 can be substituted with Arginine, Glutamine, Glutamic acid, Asparagine, and Serine; (5) the amino acid corresponding to Met-102 can be substituted with Leucine, Isoleucine, Glutamine, Valine, or Phenylalanine; (6) the amino acid corresponding to Ser-109 can be substituted with Threonine, Alanine, Asparagine, Aspartic acid, Glutamine, Glutamic acid, Glycine, Lysine, or Threonine; (7) the amino acid corresponding to Glu-110 can be substituted with Arginine, Asparagine, Aspartic acid, Histidine, Lysine, or Serine; (8) the amino acid corresponding with Asp 142 can be substituted with Asparagine, Aspartic acid, Serine, Glutamine, or Glutamic acid.

Example 6

Engineered Rep Protein with Reduced Binding Affinity for Native Rep Binding Sequence Region

[0389] Rep protein affinity for a Rep binding sequence region is reduced by modification of one or more highly conserved residues that have been shown to contact the DNA and are highly conserved across all Rep proteins. The identified residues contact either the phosphodiester backbone and/or the nucleosides, contributing to stabilized protein-DNA interaction of a Rep protein binding site.

[0390] Residues K135 and N142 of the AAV5 Rep 22 protein β 4/ β 5 loop, are identified as contributing to DNA binding affinity. K135 and N142 are altered to either Glycine or Threonine. The following modifications are tested to determine the effect on DNA binding: (1) K135G, (2) K135T, (3) N142G, (4) N142T, (5) K135G-N142G, (6) K135G-N142T, (7) K135T-N142G, and/or (8) K135T-N142T. Engineered Rep protein polynucleotide sequences comprising modified residues are designed in vitro and synthesized by standard techniques as described herein. The modified Rep proteins are expressed in *E. coli* and purified according to standard procedures. An in vitro binding assay is performed with one or more of the listed engineered Rep proteins and a radiolabeled AAV5 Rep binding sequence region oligonucleotide probe (SEQ ID NO: 3-4) at varying concentrations. The resultant protein-DNA complexes are separated on a non-denaturing polyacrylamide gel. The concentrations of bound and free probe is determined using a PhosphorImager (Molecular Dynamics). A plot of the ratio of bound-free probe versus protein concentration is used to determine the K_d of each modified Rep protein.

[0391] The DNA binding affinity of the engineered Rep proteins can also be measured using surface plasmon resonance (ForteBio, Menlo Park, Calif.) using the following engineered Rep protein and DNA species: (1) AAV5 wild-type Rep protein and AAV5 wild-type ITR containing the native Rep binding sequence region (SEQ ID NO: 3-4) and (2) any of the engineered Rep proteins listed above and AAV5 wild-type ITR containing the native Rep binding sequence region (SEQ ID NO: 3-4). The polynucleotide encoding the wild-type AAV5 Rep binding sequence region is immobilized on the biosensor layer. A solution is prepared comprising AAV5 wild-type Rep protein or any of the engineered Rep proteins listed above, which is then passed over the biosensor to determine the rate of association. A second solution devoid of the engineered Rep protein or wild-type Rep protein is passed over the saturated biosensor layer to determine the rate of dissociation.

Example 7

Engineered Rep Protein with Reduced Binding Affinity for eRBSR

[0392] Rep protein affinity for a Rep binding sequence region is reduced by modification of one or more highly conserved residues that have been shown to contact the DNA and are highly conserved across all Rep proteins. The identified residues contact either the phosphodiester backbone and/or the nucleosides, contributing to stabilized protein-DNA interaction of a Rep protein binding site.

[0393] Residues K135 and N142 of the AAV5 Rep 22 protein $\beta 4/\beta 5$ loop, are identified as contributing to DNA binding affinity. K135 and N142 are altered to either Glycine or Threonine. The following modifications are tested to determine the effect on DNA binding: (1) K135G, (2) K135T, (3) N142G, (4) N142T, (5) K135G-N142G, (6) K135G-N142T, (7) K135T-N142G, and/or (8) K135T-N142T. Engineered Rep protein polynucleotide sequences comprising modified residues are designed in vitro and synthesized by standard techniques as described herein. The modified Rep proteins are expressed in *E. coli* and purified according to standard procedures. An in vitro binding assay is performed with one or more of the listed engineered Rep proteins and a radiolabeled oligonucleotide probe comprising an eRBSR at varying concentrations. The resultant protein-DNA complexes are separated on a non-denaturing polyacrylamide gel. The concentrations of bound and free probe is determined using a PhosphorImager (Molecular Dynamics). A plot of the ratio of bound-free probe versus protein concentration is used to determine the K_d of each modified Rep protein.

[0394] The DNA binding affinity of the engineered Rep proteins can also be measured using surface plasmon resonance (ForteBio, Menlo Park, Calif.) using the following engineered Rep DNA species: (1) AAV2 ITR having the native AAV2 Rep binding sequence region (SEQ ID NOs: 1-2), (2) modified AAV2 ITR having the α PAL sequence (eRBSR) (SEQ ID NO: 68), (3) modified AAV2 ITR having the 5-5/EGR sequence (eRBSR) (SEQ ID NO: 71), (4) modified AAV2 ITR having the JcDENV NS1 sequence (eRBSR) (SEQ ID NO: 74). The polynucleotide encoding either the wild-type or engineered Rep binding sequence region is immobilized on the biosensor layer. A solution is prepared comprising AAV5 wild-type Rep protein or any of the engineered Rep proteins listed above, which is then

passed over the biosensor to determine the rate of association. A second solution devoid of the engineered Rep protein or wild-type Rep protein is passed over the saturated biosensor layer to determine the rate of dissociation.

Example 8

Determination of Production of Recombinant scAAV

[0395] scAAV is produced utilizing vectors comprising eRBSRs, engineered Rep protein, or chimeric nicking stem loops and combinations thereof. The relative production of scAAV using any of the disclosed sequences of the invention alone or in combination is quantified in the following manner.

[0396] A culture of 293 cells engineered to produce helper components required for AAV production is co-transfected with the viral construct expression vector, comprising any engineered Rep protein, and payload construct expression vector, comprising any eRBSR, chimeric nicking stem loop, or combination thereof. The culture is maintained for 48 hours while scAAV is produced and released into the medium.

[0397] The viral replication cells are lysed using the Microfluidizer™ (Microfluidics International Corp., Newton, Mass.), high shear force fluid processor. The resultant cell lysate is clarified by low speed centrifugation followed by tangential flow filtration. The resultant clarified lysate is then processed by ethanol precipitation to isolate the scAAV genomes.

[0398] The titer of AAV particles produced and purified by the methods described herein is determined by real-time quantitative polymerase chain reaction (qPCR) on a thermal cycler equipped with an excitation source filters, and detector for quantification of the reaction such as, but not limited to, the 7500 FAST Real-Time PCR system (Applied Biosystems, Foster City Calif.). AAV particles produced and purified by the methods described herein is treated with proteinase K, serially diluted, and PCR-amplified using a fluor such as, but not limited to, SYBR green (Applied Biosystems, Foster City, Calif.) with primers specific to the NAV genome ITR sequences. Linearized AAV vector genome is used as a copy number standard. The cycling conditions are: 95° C. for 3 min, followed by 35 cycles of 95° C. for 30 sec, 60° C. for 30 sec, and 72° C. for 30 sec.

[0399] The production of scAAV genomes is assessed by gel electrophoresis. The genomes isolated as described above are separated by denaturing (alkaline) gel electrophoresis. Analysis of the DNA content banding pattern has shown multiple sizes of viral genomes including scAAV, scAAV intermediate species, ssAAV monomer, and an ssAAV repaired from scAAV. Quantitation of the relative amounts of intermediate species is determined by densitometry and the efficiency of scAAV production quantified as a ratio of scAAV produced by wild-type versus eRBSR, engineered Rep protein, chimeric nicking stem loop constructs, and combinations thereof.

Example 9

Gene Expression

[0400] The level of transgene expression by AAV particles produced and purified by the methods described herein is determined by real-time quantitative polymerase chain reac-

tion (qPCR). A culture of 293 cells engineered to produce helper components required for AAV production is infected by scAAV particles produced as described herein.

[0401] The target 293 cells are harvested at a series of time points, lysed and the mRNA is purified. The level of transgene expressed is determined by reverse transcription (qPCR) on a thermal cycler equipped with an excitation source filters, and detector for quantification of the reaction such as, but not limited to, the 7500 FAST Real-Time PCR system (Applied Biosystems, Foster City Calif.). AAV particles produced and purified by the methods described herein is treated with proteinase K, serially diluted, and PCR-amplified using a fluor such as, but not limited to, SYBR green (Applied Biosystems, Foster City, Calif.) with primers specific to the transgene sequence. A reference transgene oligonucleotide is used as a copy number standard. The cycling conditions are: 95° C. for 3 min, followed by 35 cycles of 95° C. for 30 sec, 60° C. for 30 sec, and 72° C. for 30 sec.

Example 10

Recombinant scAAV Production in Invertebrate Cells

[0402] The AAV viral construct vector encodes the three structural cap proteins, VP1, VP2, and VP3, in a single open reading frame regulated by utilization of both alternative splice acceptor and non-canonical translational initiation codon(s). In-frame and out-of-frame ATG triplets preventing translation initiation at a position between the VP1 and VP2 start codons are eliminated. Both Rep78 and Rep52 are translated from a single transcript: Rep78 translation initiates at a non-AUG codon and Rep52 translation initiates at the first AUG in the transcript.

[0403] The nucleotides that encode the structural VP 1, VP2, and VP3 capsid proteins and non-structural Rep78 and Rep52 proteins are contained on one viral expression construct under control of the baculovirus major late promoter.

[0404] The payload construct vector encodes two ITR sequences flanking a transgene polynucleotide encoding a polypeptide or modulatory nucleic acid. The ITR sequences allow for replication of a polynucleotide encoding the transgene and ITR sequences alone that will be packaged within the capsid of the viral vector. The replicated polynucleotide encodes ITR sequences on the 5' and 3' ends of the molecule. The 3' [FR encodes an eRBSR that exhibits a ten-fold reduction in binding affinity as compared to a wild-type Rep binding sequence region. The reduced binding affinity of Rep protein for the eRBSR results in reduced nicking of the terminal resolution site located in the nicking stem loop. A failure to resolve the intermediate replication structure by nicking of the stem loop results in an alternative replication pathway and generation of the self-complementary AAV genome species of the polynucleotide encoding the payload and ITR sequences.

[0405] The payload construct vector and viral construct vector each comprise a Tn7 transposon element that transposes the [FR-payload sequences or the Rep and Cap sequences respectively to a bacmid that comprises the attTn7 attachment site. Competent bacterial DH10 cells are transfected with either the payload construct vector or viral construct vector. The resultant viral construct expression vector and payload construct expression vector produced in

the competent cell are then purified by detergent lysis and purification on DNA columns.

[0406] Separate seed cultures of Sf9 cells in serum free suspension culture are transfected with the viral construct expression vector or payload construct expression vector. The cultures are maintained for 48 hours while baculovirus is produced and released into the medium. The baculovirus released into the media continue to infect Sf9 cells in an exponential manner until all of the Sf9 cells in the culture are infected at least once. The baculoviral infected insect cells (BIIC) and media of the seed culture is harvested and divided into aliquots before being frozen in liquid nitrogen.

[0407] A naïve population of un-transfected Sf9 cells is expanded in serum free suspension cell culture conditions. Once the culture growth has reached peak log phase in 1 L of media as measured by optical density the culture is added to a large volume 20 L bioreactor. The bioreactor culture is co-inoculated with a frozen viral construct expression vector and payload construct expression vector BIIC aliquot. The conditions of the Sf9 cell suspension culture is monitored by instruments that measure and/or control external variables that support the growth and activity of viral replication cells such as mass, temperature, CO₂, O₂, pH, and/or optical density (OD) The Sf9 culture is maintained at optimal conditions until cell population growth has reached peak log phase and before cell growth has plateaued, as measured by optical density.

[0408] In each viral replication cell that has been infected with both baculoviruses the payload flanked on one end with an ITR sequence containing an eRBSR is replicated by an alternative pathway producing a self-complementary AAV genome and packaged in a capsid assembled from the proteins VP1, VP2, and VP3.

[0409] The viral replication cells are lysed using the Microfluidizer™ (Microfluidics International Corp., Newton, Mass.), high shear force fluid processor. The resultant cell lysate is clarified by low speed centrifugation followed by tangential flow filtration. The resultant clarified lysate is filtered by a size exclusion column to remove any remaining baculoviral particles from solution. The final steps utilize ultracentrifugation and sterile filtration to produce viral particles suitable for use as described herein.

[0410] The titer of AAV particles produced and purified by the methods described herein is determined by real-time quantitative polymerase chain reaction (qPCR) on a thermal cycler equipped with an excitation source filters, and detector for quantification of the reaction such as, but not limited to, the 7500 FAST Real-Time PCR system (Applied Biosystems, Foster City Calif.). AAV particles produced and purified by the methods described herein is treated with proteinase K, serially diluted, and PCR-amplified using a fluor such as, but not limited to, SYBR green (Applied Biosystems, Foster City, Calif.) with primers specific to the AAV genome ITR sequences. Linearized AAV vector genome is used as a copy number standard. The cycling conditions are: 95° C. for 3 min, followed by 35 cycles of 95° C. for 30 sec, 60° C. for 30 sec, and 72° C. for 30 sec.

Example 11

Recombinant AAV Production in Mammalian Cells

[0411] The AAV viral construct vector encodes the three structural cap proteins, VP1, VP2, and VP3, in a single open reading frame regulated by utilization of both alternative

splice acceptor and non-canonical translational initiation codon(s). In-frame and out-of-frame ATG triplets preventing translation initiation at a position between the VP1 and VP2 start codons are eliminated. Both Rep78 and Rep52 are translated from a single transcript: Rep78 translation initiates at a non-AUG codon and Rep52 translation initiates at the first AUG in the transcript.

[0412] The nucleotides that encode the structural VP1, VP2, and VP3 capsid proteins and non-structural Rep78 and Rep52 proteins are contained on one viral expression construct under control of the CMV promoter.

[0413] The payload construct vector encodes two ITR sequences flanking a transgene polynucleotide encoding a polypeptide or modulatory nucleic acid. The ITR sequences allow for replication of a polynucleotide encoding the transgene and ITR sequences alone that will be packaged within the capsid of the viral vector. The replicated polynucleotide encodes ITR sequences on the 5' and 3' ends of the molecule. The 3' ITR encodes an eRBSR that exhibits a ten-fold reduction in binding affinity as compared to a wild-type Rep binding sequence region. The reduced binding affinity of Rep protein for the eRBSR results in reduced nicking of the terminal resolution site located in the nicking stem loop. A failure to resolve the intermediate replication structure by nicking of the stem loop results in an alternative replication pathway and generation of the self-complementary AAV genome species of the polynucleotide encoding the payload and ITR sequences.

[0414] The payload construct vector and viral construct vector each comprise a Tn7 transposon element that transposes the ITR-payload sequences or the Rep and Cap sequences respectively to a bacmid that comprises the attTn7 attachment site. Competent bacterial DH10 cells are transfected with either the payload construct vector or viral construct vector. The resultant viral construct expression vector and payload construct expression vector produced in the competent cell are then purified by detergent lysis and purification on DNA columns.

[0415] A seed culture of Chinese Hamster Ovary (CHO) cells adapted for growth in serum free suspension culture is co-transfected with the viral construct expression vector and payload construct expression vector. The culture is maintained for 48 hours while two baculoviruses are produced and released into the medium, one containing the payload construct vector and a second containing the viral expression construct. The baculovirus released into the media continue to infect CHO cells in an exponential manner until all of the CHO cells in the culture are infected at least once with both baculoviruses. In each viral replication cell that has been infected with both baculoviruses the payload flanked on either end with an ITR sequence is replicated and packaged in a capsid assembled from the proteins VP1, VP2, and VP3. The cells and media of the seed culture is harvested and divided into aliquots before being frozen in, for example, liquid nitrogen.

[0416] A naïve population of un-transfected CHO cells is expanded in serum free suspension cell culture conditions.

Once the culture growth has reached peak log phase in 1 L of media as measured by optical density the culture is added to a large volume 20 L bioreactor. The bioreactor culture is inoculated with a frozen aliquot from the baculovirus seed culture. The conditions of the CHO cell suspension culture is monitored by instruments that measure and/or control external variables that support the growth and activity of viral replication cells such as mass, temperature, CO₂, O₂, pH, and/or optical density (OD). The CHO culture is maintained at optimal conditions until cell population growth has reached peak log phase and before cell growth has plateaued, as measured by optical density.

[0417] The viral replication cells are lysed using the Microfluidizer™ (Microfluidics International Corp., Newton, Mass.), high shear force fluid processor. The resultant cell lysate is clarified by low speed centrifugation followed by tangential flow filtration. The resultant clarified lysate is filtered by a size exclusion column to remove any remaining baculoviral particles from solution. The final steps utilize ultracentrifugation and sterile filtration to produce viral particles suitable for uses described herein.

Example 12

Large Scale PEI Transfection

[0418] Polyethyleneimine (PEI) is used to form PEI-DNA complexes. Plasmids being transfected are combined with PEI in PBS and allowed to incubate at room temperature for 10 minutes. HEK 293 cell cultures being transfected are 'shocked' at 4° C. for 1 hour before being returned to the 37° C. incubator for a period of 6-24 hours (to arrest cell cycle at the junction between G2 phase and M phase). PEI-DNA transfection complexes are then added to the cells under shaking conditions and allowed to incubate 6 hours. After incubation, an equal volume of fresh medium is added and cells are incubated for 24-96 hours.

Example 13

Central Nervous System AAV Delivery

[0419] Viral particles are produced as taught herein and prepared for delivery to the central nervous system. In one aspect, preparation for CNS delivery is according to the method of Foust et al (Foust, K. D, et al., 2009, Nat Biotechnol 27:59-65, the contents of which are herein incorporated by reference in their entirety).

[0420] According to the Foust method, AAV9 viral particles delivered by venous injection are transported across the blood brain barrier (BBB) and carry out astrocyte transduction. Viruses are produced and purified by cesium chloride gradient purification, followed by dialysis against phosphate buffered saline (PBS). Resulting preparations are formulated with 0.001% Pluronic-F68 to discourage viral aggregation. Viral preparations are titrated following quantitative-PCR analysis of viral levels. Purity of viral preparations is further assessed by gel electrophoresis and subsequent silver staining (Invitrogen, Carlsbad, Calif.) Viral preparations are then delivered to subjects by intravenous injection. Viral payloads are delivered to cells of the CNS.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 312

<210> SEQ ID NO 1

<211> LENGTH: 16

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 1

gcgcgctcgc tcgctc

16

<210> SEQ ID NO 2

<211> LENGTH: 16

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 2

gagcgagcga gcgcgc

16

<210> SEQ ID NO 3

<211> LENGTH: 16

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 3

gctcgctcgc tggctc

16

<210> SEQ ID NO 4

<211> LENGTH: 16

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 4

gagccagcga gcgagc

16

<210> SEQ ID NO 5

<211> LENGTH: 16

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 5

gcgcgctcgc tcgctc

16

<210> SEQ ID NO 6

<211> LENGTH: 16

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 6

gagcgagcga gcgcgc

16

<210> SEQ ID NO 7

<211> LENGTH: 16

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 7

gcgcactcgc tcgctc

16

<210> SEQ ID NO 8

<211> LENGTH: 16

<212> TYPE: DNA

-continued

```

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 8

gagcgagcga gtgcgc                                16

<210> SEQ ID NO 9
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 9

gcgcgctcgc tcactc                                16

<210> SEQ ID NO 10
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 10

gagtgagcga gcgcgc                                16

<210> SEQ ID NO 11
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 11

gcgcgctcgc tcgctc                                16

<210> SEQ ID NO 12
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 12

gagcgagcga gcgcgc                                16

<210> SEQ ID NO 13
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 13

gctcgctcgc tc                                    12

<210> SEQ ID NO 14
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 14

gagcgagcga gc                                    12

<210> SEQ ID NO 15
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 15

gctcgctcgc tc                                    12

```

-continued

```

<210> SEQ ID NO 16
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 16

gagcgagcga gc                                     12

<210> SEQ ID NO 17
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Bovine adeno-associated virus

<400> SEQUENCE: 17

gctcgttcgc tggctc                                 16

<210> SEQ ID NO 18
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Bovine adeno-associated virus

<400> SEQUENCE: 18

gagccagcga acgagc                                 16

<210> SEQ ID NO 19
<211> LENGTH: 10
<212> TYPE: DNA
<213> ORGANISM: Snake parvovirus

<400> SEQUENCE: 19

gcgcgcgcctc                                       10

<210> SEQ ID NO 20
<211> LENGTH: 10
<212> TYPE: DNA
<213> ORGANISM: Snake parvovirus

<400> SEQUENCE: 20

gagcgcgcgcg                                       10

<210> SEQ ID NO 21
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 21

gctcnnnnnnn nnnnnnnnnnnn                         20

<210> SEQ ID NO 22
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:

```

-continued

```

<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 22

gctcnnnnnnn nnnnnn                                16

<210> SEQ ID NO 23
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 23

gctcnnnnnnn nn                                    12

<210> SEQ ID NO 24
<211> LENGTH: 8
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 24

gctcnnnnn                                           8

<210> SEQ ID NO 25
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 25

nnnnnnnnnnn nnnnnngctc                            20

<210> SEQ ID NO 26
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 26

nnnnnnnnnnn nngctc                                16

```

-continued

```

<210> SEQ ID NO 27
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 27

nnnnnnnnngc tc                                     12

<210> SEQ ID NO 28
<211> LENGTH: 8
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(4)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 28

nnnnngctc                                           8

<210> SEQ ID NO 29
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 29

gctcnnnnnnn nnnnnngctc                             20

<210> SEQ ID NO 30
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 30

gctcnnnnnnn nngctc                                  16

<210> SEQ ID NO 31
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide

```

-continued

```

<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 31

gctcnnnnngc tc                                     12

<210> SEQ ID NO 32
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 32

gctcgctcnn nnnnnnnnnn                               20

<210> SEQ ID NO 33
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 33

gctcgctcnn nnnnnn                                   16

<210> SEQ ID NO 34
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 34

gctcgctcnn nn                                       12

<210> SEQ ID NO 35
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (13)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 35

gctcgctcgc tcnnnnnnnn                               20

```

-continued

<210> SEQ ID NO 36
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (13)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 36

gctcgctcgc tcnnnn 16

<210> SEQ ID NO 37
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 37

nnnnnnnnnn nngctcgc 20

<210> SEQ ID NO 38
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 38

nnnnnnnnngc tcgctc 16

<210> SEQ ID NO 39
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(4)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 39

nnnngctcgc tc 12

<210> SEQ ID NO 40
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic

-continued

oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 40

nnnnnnnnngc tcgctcgcgc 20

<210> SEQ ID NO 41
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(4)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 41

nnnngctcgc tcgctc 16

<210> SEQ ID NO 42
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 42

gctcgcctcnn nnnnnngctc 20

<210> SEQ ID NO 43
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 43

gctcgcctcnn nngctc 16

<210> SEQ ID NO 44
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 44

-continued

gctcnnnnnn nngctcgtc 20

<210> SEQ ID NO 45
 <211> LENGTH: 16
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 oligonucleotide
 <220> FEATURE:
 <221> NAME/KEY: modified_base
 <222> LOCATION: (5)..(8)
 <223> OTHER INFORMATION: a, c, t, g, unknown or other

 <400> SEQUENCE: 45

gctcnnnngc tcgtc 16

<210> SEQ ID NO 46
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 oligonucleotide
 <220> FEATURE:
 <221> NAME/KEY: modified_base
 <222> LOCATION: (1)..(4)
 <223> OTHER INFORMATION: a, c, t, g, unknown or other
 <220> FEATURE:
 <221> NAME/KEY: modified_base
 <222> LOCATION: (17)..(20)
 <223> OTHER INFORMATION: a, c, t, g, unknown or other

 <400> SEQUENCE: 46

nnnngctcgc tcgctcnnnn 20

<210> SEQ ID NO 47
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 oligonucleotide
 <220> FEATURE:
 <221> NAME/KEY: modified_base
 <222> LOCATION: (1)..(8)
 <223> OTHER INFORMATION: a, c, t, g, unknown or other
 <220> FEATURE:
 <221> NAME/KEY: modified_base
 <222> LOCATION: (17)..(20)
 <223> OTHER INFORMATION: a, c, t, g, unknown or other

 <400> SEQUENCE: 47

nnnnnnnngc tcgctcnnnn 20

<210> SEQ ID NO 48
 <211> LENGTH: 20
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
 oligonucleotide
 <220> FEATURE:
 <221> NAME/KEY: modified_base
 <222> LOCATION: (1)..(4)
 <223> OTHER INFORMATION: a, c, t, g, unknown or other
 <220> FEATURE:
 <221> NAME/KEY: modified_base
 <222> LOCATION: (13)..(20)

-continued

```

<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 48

nnnngctcgc tcnnnnnnnn                                20

<210> SEQ ID NO 49
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(4)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (13)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 49

nnnngctcgc tcnnnn                                16

<210> SEQ ID NO 50
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (17)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 50

nnnnnnnnnn nngctcnnnn                                20

<210> SEQ ID NO 51
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (13)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 51

nnnnnnnngc tcnnnnnnnn                                20

<210> SEQ ID NO 52
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic

```

-continued

```

    oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (13)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 52

nnnnnnnnngc tcnnnn                                     16

<210> SEQ ID NO 53
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
    oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(4)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 53

nnnngctcnn nn                                           12

<210> SEQ ID NO 54
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
    oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(4)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 54

nnnngctcnn nnnnnnnnnn                                   20

<210> SEQ ID NO 55
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
    oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(4)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 55

nnnngctcnn nnnnnn                                       16

```

-continued

<210> SEQ ID NO 56
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (13)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 56

gctcnnnnngc tcnnnn 16

<210> SEQ ID NO 57
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (17)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 57

gctcnnnnnnn nngctcnnnn 20

<210> SEQ ID NO 58
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (13)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 58

gctcnnnnngc tcnnnnnnnn 20

<210> SEQ ID NO 59
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(4)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

-continued

```

<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 59

nnnngctcnn nnctc                                     16

<210> SEQ ID NO 60
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(4)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 60

nnnngctcnn nnnnnngctc                                 20

<210> SEQ ID NO 61
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (13)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 61

nnnnnnnnngc tcnnnngctc                                20

<210> SEQ ID NO 62
<211> LENGTH: 8
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide

<400> SEQUENCE: 62

gctcgggg                                              8

<210> SEQ ID NO 63
<211> LENGTH: 9
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide

<400> SEQUENCE: 63

gctcggggg                                              9

```

-continued

```

<210> SEQ ID NO 64
<211> LENGTH: 10
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide

<400> SEQUENCE: 64

gctcgggggg                                     10

<210> SEQ ID NO 65
<211> LENGTH: 11
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide

<400> SEQUENCE: 65

gctcgggggg g                                   11

<210> SEQ ID NO 66
<211> LENGTH: 12
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      oligonucleotide

<400> SEQUENCE: 66

gctcgggggg gg                                 12

<210> SEQ ID NO 67
<211> LENGTH: 2303
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polynucleotide

<400> SEQUENCE: 67

cctgcaggca gctgcgcgc atcgggctca ctgaggccgc ccgggcaaag cccgggcgtc      60
gggcgacctt tggtcgcccg gcctcagtga gccgcatgcg gcgcagagag ggagtggcca      120
actccatcac taggggttcc ttgtagttaa tgattaaccc gccatgctac ttatctacgt      180
agccatgctc tagagcggcc gcacgcgtgt tactagttat taatagtaat caattacggg      240
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc      300
gcctggctga ccgcccacg acccccgcc attgacgtca ataagacgt atgttcccat      360
agtaacgtca atagggaact tccattgacg tcaatgggtg gagtatttac ggtaaaactgc      420
ccacttgca gtacatcaag tgtatcatat gccaaagtacg cccctattg acgtcaatga      480
cggtaaatgg cccgcctggc attatgccca gtacatgacc ttatgggact ttcctacttg      540
gcagtacatc tacgtattag tcacgctat taccatggtg atgcgggtttt ggcagtacat      600
caatgggcgt gtagatcggg ttgactcagc gggatttcca agtctccacc ccattgacgt      660
caatgggagt ttgttttgca ccaaaatcaa cgggacttcc caaatgtcg taacaactcc      720
gccccattga cgcaaatggg cggtaggcgt gtacggtggg aggtctatat aagcagagct      780

```

-continued

```

cgtttagtga accgtcagat cgcttgagaga cgccatccac gctgttttga cctccataga      840
agacaccggg accgatccag cctccaaccg gttcgaacag gtaagcgccc ctaaaatccc      900
tttggcacaa tgtgtcctga ggggagagagc agcgacctgt agatgggacg ggggcactaa      960
ccctcagggg ttggggttct gaatgtgagt atcgccatgt aagcccagta ttgggccaat     1020
ctcagaaagc tcctggctcc ctggaggatg gagagagaaa aacaaacagc tcctggagca     1080
gggagagtgc tggcctcttg ctctccggct ccctctgttg ccctctgggt tctccccagg     1140
ttcgaagcgc gcaaggagag cggccatggt gagcaaggcg gaggagctgt tcaccggggg     1200
ggtgcccata ctggtcgagc tggacggcga cgtaaacggc cacaagttca gcgtgtccgg     1260
cgaggggcag ggcgatgcca cctacggcaa gctgacctg aagttcatct gcaccaccgg     1320
caagtgtccc gtgccctggc ccacctctgt gaccacctg acctacggcg tgcaagtctt     1380
cagccgctac cccgaccaca tgaagcagca cgacttcttc aagtcgccca tgcccgaagg     1440
ctacgtccag gagcgacca tcttcttcaa ggacgacggc aactacaaga ccgcgcgcga     1500
ggtgaagtgc gagggcgaca cctggtgaa ccgcacgag ctgaagggca tcgacttcaa     1560
ggaggacggc aacatcctgg ggcacaagct ggagtacaac tacaacagcc acaacgtcta     1620
tatcatggcc gacaagcaga agaacggcat caaggtgaac ttcaagatcc gccacaacat     1680
cgaggacggc agcgtgcagc tcgcccacca ctaccagcag aacaccccca tcggcgacgg     1740
ccccgtgtcg ctgcccgaca accactacct gagcaccag tccgccctga gcaaagacc     1800
caacgagaag cgcgatcaca tggctctgct ggagttcgtg accgccgcgc ggatcactct     1860
cggcatggac gagctgtaca agtaaacggc cgggcgcgac agctgtatac acgtgcaagc     1920
cgcgcgacgc ggccgacct ggcccaactt gtttattgca gcttataatg gttacaaata     1980
aagcaatagc atcacaatt tcacaaataa agcatttttt tcaactgcatt ctagtgtggt     2040
tttgtccaaa ctcatcaatg tatcttatca tgtctggatc tccggaccac gtgcggaccg     2100
agcggccgct ctagagcatg gctacgtaga taagtagcat ggcgggttaa tcattaacta     2160
caaggaaacc ctagtgtagg agttggccac tccctctctg cgcgctcgct cgctcactga     2220
ggccggggcg ccaaaggctg ccgcagcgcc gggctttgcc cgggcggcct cagtgcgcga     2280
gcgagcgcgc agctgcctgc agg                                     2303

```

```

<210> SEQ ID NO 68
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        oligonucleotide

```

```

<400> SEQUENCE: 68

```

```

gcgcgcgatg cggctc                                     16

```

```

<210> SEQ ID NO 69
<211> LENGTH: 145
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide

```

```

<400> SEQUENCE: 69

```

-continued

```

aggaaccct agtgatggag ttggccactc cctctctgcg ccgcatgagg ctcactgagg      60
ccgcccgggc aaagcccggg cgtcggggga cctttggteg cccggcctca gtgagcgagc      120
gagcgcgagc agagggagtg gccaa                                     145

```

```

<210> SEQ ID NO 70
<211> LENGTH: 2303
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide

```

```

<400> SEQUENCE: 70

```

```

cctgcaggca gctgagcggg ggcgcgctca ctgaggccgc ccgggcaaag cccgggcgtc      60
gggcgacctt tggtcgcccg gcctcagtga gcgcgcccc gctcagagag ggagtggcca      120
actccatcac taggggttcc ttgtagttaa tgattaacct gccatgctac ttatctactg      180
agccatgctc tagagcggcc gcacgcgtgt tactagttaa taatagtaat caattacggg      240
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc      300
gcctggctga ccgcccacg acccccgcgc attgacgtca ataatacgt atgttcccat      360
agtaacgtca atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaaactgc      420
ccacttgcca gtacatcaag tgtatcatat gccaaagtac cccctattg acgtcaatga      480
cggtaaatgg ccgcctggc attatgccca gtacatgacc ttatgggact ttctacttg      540
gcagtacatc tacgtattag tcacgctat taccatggtg atgcggtttt ggcagtacat      600
caatgggcgt ggatagcggg ttgactcacc gggatttcca agtctccacc ccattgacgt      660
caatgggagt ttgttttgca ccaaaatcaa cgggacttcc caaatgtcg taacaactcc      720
gccccattga cgcaaatggg cggtaggcgt gtacgggtgg aggtctatat aagcagagct      780
cgtttagtga accgtcagat cgctcgaga cgccatccac gctgttttga cctccataga      840
agacaccggg accgatccag cctccaaccg gttcgaacag gtaagcgccc ctaaaatccc      900
tttggcacaa tgtgtcctga ggggagaggc agcgacctgt agatgggacg ggggcactaa      960
ccctcagggt ttgggggtct gaatgtgagt atcgccatgt aagcccagta ttggccaat      1020
ctcagaaaag tcttggtctc ctggaggatg gagagagaaa aacaaacagc tcttgagca      1080
gggagagtgc tggcctcttg ctctccggct ccctctgttg ccctctggtt tctcccagg      1140
ttcgaagcgc gcaaggagag ccgccatggt gagcaaggcg gaggagctgt tcaccggggt      1200
gggtgcccatc ctggtcgagc tggacggcga cgtaaacggc cacaagttca gcgtgtccgg      1260
cgagggcgag ggcgatgcca cctacggcaa gctgaccctg aagttcatct gcaccaaccg      1320
caagctgccc gtgccctggc ccccccctgt gaccaccctg acctacggcg tgcagtgctt      1380
cagccgctac ccgaccaca tgaagcagca cgacttcttc aagtcgccca tgcccgaagg      1440
ctacgtccag gagcgcacca tcttcttcaa ggacgacggc aactacaaga cccgcgccga      1500
gggtgaagtc gagggcgaca cctggtgaa ccgcatcgag ctgaagggca tcgacttcaa      1560
ggagacggc aacatcctgg ggcacaagct ggagtacaac tacaacagcc acaacgtcta      1620
tatcatggcc gacaagcaga agaacggcat caaggtgaac ttcaagatcc gccacaacat      1680
cgaggacggc agcgtgcagc tcgccgacca ctaccagcag aacaccccca tcggcgacgg      1740

```

-continued

```

ccccgtgctg ctgcccagaca accactacct gagcaccag tccgccctga gcaaagaccc 1800
caacgagaag cgcgatcaca tggctctgct ggagttcgtg accgccgcgc ggatcactct 1860
cggcattggac gagctgtaca agtaaagcgg ccggccgcac agctgtatac acgtgcaagc 1920
cgcgcgacgc ggccgacccat ggcccaactt gtttattgca gcttataatg gttacaaata 1980
aagcaatagc atcacaatt tcacaaataa agcatttttt tcaactgcatt ctagtgtgtg 2040
tttgtccaaa ctcattcaatg tatcttatca tgtctggatc tccggaccac gtgcggaccg 2100
agcggccgct ctagagcatg gctacgtaga taagtagcat ggccgggttaa tcattaacta 2160
caaggaaccc ctagtgtatg agttggccac tccctctctg cgcgctcgtc cgctcactga 2220
ggccgggcga ccaaggtcg ccgcagccc gggttttgc ccggcgccct cagtgcgcga 2280
gcgagcgcgc agctgcctgc agg 2303

```

```

<210> SEQ ID NO 71
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        oligonucleotide

```

```

<400> SEQUENCE: 71

```

```

gagcgggggc gcgtc 16

```

```

<210> SEQ ID NO 72
<211> LENGTH: 145
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide

```

```

<400> SEQUENCE: 72

```

```

aggaaccct agtgatggag ttggccactc cctctctgag cgggggcgcg ctactgagg 60
ccgccgggc aaagccggg cgctgggcga cctttggtcg ccggcctca gtgagcgagc 120
gagcgcgcag agagggagtg gccaa 145

```

```

<210> SEQ ID NO 73
<211> LENGTH: 2303
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide

```

```

<400> SEQUENCE: 73

```

```

cctgcaggca gctccgtcgt cgtggtcgga ctgaggccgc ccgggcaaag cccgggcgtc 60
gggcgacctt tggctgcccg gcctcagtc gaccacgacg acggagagag ggagtggcca 120
actccatcac taggggttcc ttgtagttaa tgattaacct gccatgctac ttatctactg 180
agccatgctc tagagcggcc gcacgcgtgt tactagttaa taatagtaat caattacggg 240
gtcattagtt catagcccat atatggagt ccgcgttaca taacttacgg taaatggccc 300
gcctggctga ccgcccaacg acccccgcgc attgacgtca ataagacgt atgttcccat 360
agtaacgtca atagggaact tccattgacg tcaatgggtg gagtatttac ggtaaactgc 420
ccacttgcca gtacatcaag tgtatcatat gccaaagtacg cccctattg acgtcaatga 480

```

-continued

```

cggtaaatgg cccgcctggc attatgccca gtacatgacc ttatgggact ttctacttg 540
gcagtacatc tacgtattag tcacgctat taccatgggtg atgcgggtttt ggcagtacat 600
caatgggcgt ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt 660
caatgggagt ttgttttgca ccaaaatcaa cgggactttc caaatgtcg taacaactcc 720
gccccattga cgcaaatggg cggtaggcgt gtacgggtggg aggtctatat aagcagagct 780
cgttttagtga accgtcagat cgctggaga cgccatccac gctgttttga cctccataga 840
agacaccggg accgatccag cctccaaccg gtctgaacag gtaagcgccc ctaaaatccc 900
tttgccacaa tgtgtcctga ggggagaggc agcgacctgt agatgggacg ggggcactaa 960
ccctcagggt ttgggggtct gaattgtgagt atcgccatgt aagcccagta ttggccaat 1020
ctcagaaagc tcttggtctc ctggaggatg gagagagaaa aacaaacagc tctggagca 1080
gggagagtgc tggcctcttg ctctcggct ccctctgttg ccctctggtt tctcccagg 1140
ttcgaagcgc gcaaggagag ccgccatggt gagcaaggcg gaggagctgt tcaccggggt 1200
ggtgcccacg ctggtcgagc tggacggcga cgtaaacggc cacaagttca gcgtgtccg 1260
cgagggcgag ggcgatgcca cctacggcaa gctgacctg aagttcatct gcaccaccg 1320
caagctgccc gtgccctggc ccacctcgt gaccacctg acctacggcg tgcagtgtt 1380
cagccgtac cccgaccaca tgaagcagca cgactttctc aagtcgccca tgcccgaag 1440
ctacgtccag gagcgacca tcttcttcaa ggacgacggc aactacaaga cccgcgccga 1500
ggtgaagtgc gagggcgaca cctggtgaa ccgcatcgag ctgaagggca tcgacttcaa 1560
ggaggacggc aacatcctgg ggcacaagct ggagtacaac tacaacagcc acaacgtcta 1620
tatcatggcc gacaagcaga agaacggcat caaggtgaac ttcaagatcc gccacaacat 1680
cgaggacggc agcgtgcagc tcgccgacca ctaccagcag aacaccccca tcggcgacgg 1740
ccccgtgtg ctgcccgaca accactacct gagcaccag tccgccctga gcaaagacc 1800
caacgagaag cgcgatcaca tggctctgct ggagtctgtg accgccgccg ggatcaactct 1860
cggcatggac gagctgtaca agtaaaggcg ccggccgcac agctgtatac acgtgcaagc 1920
cgcgcgcagc ggcgcgacct ggcccaactt gtttattgca gcttataatg gttacaaata 1980
aagcaatagc atcacaatt tcacaaataa agcatttttt tcaactgcatt ctagtgtgg 2040
tttgtccaaa ctcatcaatg tatcttatca tgtctggatc tcgggaccac gtgcggaccg 2100
agcggccgct ctagagcatg gctacgtaga taagtagcat ggcggtttaa tcattaacta 2160
caaggaaccc ctagtgtagg agttggccac tccctctctg cgcgctcgct cgctcaactga 2220
ggccggcgca ccaaggctcg ccgcacgccc gggctttgcc cgggcggcct cagtgcgca 2280
gcgagcgcgc agctgcctgc agg 2303

```

<210> SEQ ID NO 74

<211> LENGTH: 16

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 74

ccgaccacga cgacgg

16

-continued

```

<210> SEQ ID NO 75
<211> LENGTH: 145
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide

<400> SEQUENCE: 75
aggaaccct agtgatggag ttggccactc cctctctccg accacgacga cggactgagg      60
ccgccccggc aaagccccgg cgctggggcga cctttggtcg cccggcctca gtgagcgagc      120
gagcgcgcag agagggagtg gccaa                                     145

<210> SEQ ID NO 76
<211> LENGTH: 2304
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide

<400> SEQUENCE: 76
cctgcaggca gctgcgcgct cgctcgctca ctgaggccgc ccgggcaaaag cccgggcgctc      60
gggcgacctt tggtcgcccc gcctcagtga gcgagcgagc gcgcagagag ggagtggctc      120
gctcccatca ctagggggtc ctgtagtta atgattaacc cgccatgcta cttatctacg      180
tagccatgct ctgagcgggc cgcacgcgtg ttactagtta ttaatagtaa tcaattacgg      240
ggtcattagt tcatagccca tatatggagt tccgcgttac ataacttacg gtaaatggcc      300
cgccctggctg accgcccac gacccccgcc cattgacgtc aataatgacg tatgttccca      360
tagtaacgtc aatagggact ttccattgac gtcaatgggt ggagtattta cggtaaaactg      420
cccacttggc agtacatcaa gtgtatcata tgccaagtac gcccctatt gacgtcaatg      480
acggtaaatg gcccgctcgg cattatgccc agtacatgac cttatgggac ttctactctt      540
ggcagtacat ctacgtatta gtcacgtcta ttaccatggt gatgcggttt tggcagtaca      600
tcaatgggcg tggatagcgg ttgactcac ggggatttcc aagtctccac ccattgacg      660
tcaatgggag tttgttttgc accaaaatca acgggacttt ccaaaatgac gtaacaactc      720
cgccccattg acgcaaatgg gcggtaggcg tgtacgggtg gaggtctata taagcagagc      780
tcgtttagtg aaccgtcaga tcgcctggag acgccaacca cgctgttttg acctccatag      840
aagacaccgg gaccgatcca gcctccaacc gggtcgaaca ggtaagcgcc cctaaaaatcc      900
ctttggcaca atgtgtctcg aggggagagg cagcgacctg tagatgggac gggggcacta      960
accctcaggg tttgggggtc tgaatgtgag tatcgccatg taagcccagt atttgcccaa      1020
tctcagaaag ctcttggtc cctggaggat ggagagagaa aaacaaacag ctctggagc      1080
agggagagtg ctggcctctt gctctccggc tccctctggt gccctctggt ttctccccag      1140
gttcgaagcg cgcaaggaga gccgccatgg tgagcaaggg cgaggagctg ttcaccgggg      1200
tggtgcccat cctggctcag ctggacggcg acgtaaacgg ccacaagtgc agcgtgtccg      1260
gcgagggcga gggcgatgcc acctacggca agctgaccct gaagtctatc tgcaccaccg      1320
gcaagctgcc cgtgccctgg cccaccctcg tgaccacctc gacctacggc gtgcagtgtc      1380
tcagccgcta ccccgaccac atgaagcagc acgacttctt caagtccgcc atgccggaag      1440

```

-continued

```

gctacgtcca ggagcgcacc atcttcttca aggacgacgg caactacaag acccgcgccg 1500
aggtgaagtt cgagggcgac accctgggtga accgcatcga gctgaagggc atcgacttca 1560
aggaggacgg caacatcctg gggcacaagc tggagtacaa ctacaacagc cacaacgtct 1620
atatcatggc cgacaagcag aagaacggca tcaagggtgaa cttcaagatc cgccacaaca 1680
tcgaggacgg cagcgtgcag ctgcgcgacc actaccagca gaacaccccc atcggcgacg 1740
gccccgtgct gctgcccgcac aaccactacc tgagcaccca gtccgccctg agcaaagacc 1800
ccaacgagaa gcgcgatcac atggctcctgc tggagttcgt gaccgcccgc gggatcactc 1860
tcggcatgga cgagctgtac aagtaaagcg gccggccgca cagctgtata cagtgcaag 1920
ccgcgcgcag cggccgacca tggcccaact tgtttattgc agcttataat ggttacaaat 1980
aaagcaatag catcacaaat ttcacaaata aagcattttt ttactgcat tctagttgtg 2040
gtttgtccaa actcatcaat gtatcttata atgtctggat ctccggacca cgtgcggacc 2100
gagcggccgc tctagagcat ggctacgtag ataagtagca tggcgggtta atcattaact 2160
acaaggaacc cctagtgatg gagttggcca ctccctctct gcgcgctcgc tcgctcactg 2220
aggccgggcg accaaaggtc gcccgacgcc cgggctttgc ccgggcccgc tcagtgagcg 2280
agcgagcgcg cagctgcctg cagg 2304

```

```

<210> SEQ ID NO 77
<211> LENGTH: 13
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        oligonucleotide

```

```

<400> SEQUENCE: 77

```

```

gagtggtggca ctc 13

```

```

<210> SEQ ID NO 78
<211> LENGTH: 145
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        polynucleotide

```

```

<400> SEQUENCE: 78

```

```

aggaaccct agtgatggag tgtggcactc cctctctcgc cgctcgtcgc ctactgagg 60
ccgccgggc aaagccggg cgctcgggca ctttggctg ccgggcctca gtgagcgagc 120
gagcgcgcag agagggagtg gccaa 145

```

```

<210> SEQ ID NO 79
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        oligonucleotide

```

```

<400> SEQUENCE: 79

```

```

atacgccgcc tcgcgtcag 20

```

```

<210> SEQ ID NO 80
<211> LENGTH: 20

```

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 80

atctgtcgct cgcccgcta 20

<210> SEQ ID NO 81
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 81

tccgcgctgg ctcacgtcc 20

<210> SEQ ID NO 82
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 82

tccgcgctgg ctcacgtcc 20

<210> SEQ ID NO 83
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 83

tccccgcccc cgctcattct 20

<210> SEQ ID NO 84
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 84

gtgccccgc tcagagtcca 20

<210> SEQ ID NO 85
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 85

tttaccgccg ctcagataga 20

-continued

<210> SEQ ID NO 86
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 86

gacccccagg cgctcctatg 20

<210> SEQ ID NO 87
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 87

atctcgctca tgccccttag 20

<210> SEQ ID NO 88
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 88

tccccggtea ggggctcact 20

<210> SEQ ID NO 89
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 89

ccgcgcgtct atccactggt 20

<210> SEQ ID NO 90
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 90

ttgcctcgct gctactgttc 20

<210> SEQ ID NO 91
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 91

catatctccg cttagttgcc 20

-continued

<210> SEQ ID NO 92
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 92

tcgttaagaa ccttcctcat 20

<210> SEQ ID NO 93
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 93

cccccgcatc ctccgccttc 20

<210> SEQ ID NO 94
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 94

catatctccg cttagttgcc 20

<210> SEQ ID NO 95
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 95

cacagttctc gcctaccgt 20

<210> SEQ ID NO 96
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 96

ctctccttca gggcctcagc 20

<210> SEQ ID NO 97
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 97

-continued

gctgccccg tactcacccg 20

<210> SEQ ID NO 98
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 98

gctcgaagga agcggggaac 20

<210> SEQ ID NO 99
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 99

gtcagttcgc tgggtgattc 20

<210> SEQ ID NO 100
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 100

atcggaacgc ttcgttgtgc 20

<210> SEQ ID NO 101
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 101

tcgctgacca agccgcatgc 20

<210> SEQ ID NO 102
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 102

ccagtattct ggcagctgg 20

<210> SEQ ID NO 103
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

-continued

<400> SEQUENCE: 103

ggcgtcccttcttcttctg

20

<210> SEQ ID NO 104

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 104

taatcgatatg catcgctcgtg

20

<210> SEQ ID NO 105

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 105

tgaacgggtcg cgacgcagca

20

<210> SEQ ID NO 106

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 106

acgttcaacc cgccgcgtcg

20

<210> SEQ ID NO 107

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 107

atcggacggc ttcgttgtgc

20

<210> SEQ ID NO 108

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 108

ccatgtgttg cgcccgctgc

20

<210> SEQ ID NO 109

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 109

ggtcccccca tgcactgccc 20

<210> SEQ ID NO 110
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 110

cccgtatcca cgccccacgc 20

<210> SEQ ID NO 111
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 111

tgccccacgc gcgtcgtac 20

<210> SEQ ID NO 112
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 112

atgctttctc gctcagtc 19

<210> SEQ ID NO 113
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 113

acacttctca ctgcgtgcct 20

<210> SEQ ID NO 114
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 114

ccttcgcgct cgttcgaata 20

<210> SEQ ID NO 115
<211> LENGTH: 18
<212> TYPE: DNA

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 115

tgctttctcg ctcagtcc 18

<210> SEQ ID NO 116
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 116

gcgattgcct cggtaggctca 20

<210> SEQ ID NO 117
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 117

cttcacttac tcgcgacccc 20

<210> SEQ ID NO 118
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 118

tgcccttggc ttgctcagtg 20

<210> SEQ ID NO 119
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 119

gtctcgctcg gtcagctact 20

<210> SEQ ID NO 120
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 120

atccggtcac tcgttcgct 20

<210> SEQ ID NO 121

-continued

<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 121

acgcactcac tcgccggcgc 20

<210> SEQ ID NO 122
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 122

ctcttgctcg ctcaattgct 20

<210> SEQ ID NO 123
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 123

gggtgctcgc ctgggttgcg 20

<210> SEQ ID NO 124
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 124

ccagtgcctca ctcagctcgc 20

<210> SEQ ID NO 125
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 125

tcggcgctcg ctcggtcctc 20

<210> SEQ ID NO 126
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 126

cgcgggttgt tttcactcac 20

-continued

<210> SEQ ID NO 127
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 127

tgatcatctgc tcgctcactt 20

<210> SEQ ID NO 128
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 128

agatccttgt ggctcactcg 20

<210> SEQ ID NO 129
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 129

actggctggt tcgctcagac 20

<210> SEQ ID NO 130
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 130

ccacagatgt agctcactca 20

<210> SEQ ID NO 131
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 131

ctggtatcgc tcacttgacc 20

<210> SEQ ID NO 132
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 132

-continued

tcgcgctcgt tcgcctctgc 20

<210> SEQ ID NO 133
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 133

agateccttgt ggctcactcg 20

<210> SEQ ID NO 134
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 134

caattcagtc aaccactcaa 20

<210> SEQ ID NO 135
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 135

ctcgcttgca gtcactcact 20

<210> SEQ ID NO 136
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 136

gagtttatct catgttctgc 20

<210> SEQ ID NO 137
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 137

caattcagcc aaccactcaa 20

<210> SEQ ID NO 138
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

-continued

<400> SEQUENCE: 138

gctgactcgt tgactcatct 20

<210> SEQ ID NO 139

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 139

tctccgctgg ttcaccgtac 20

<210> SEQ ID NO 140

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 140

tgtgcggccc actgagacgt 20

<210> SEQ ID NO 141

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 141

ccacgtgctc gctcaacctt 20

<210> SEQ ID NO 142

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 142

tcctacagct tgctcactct 20

<210> SEQ ID NO 143

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 143

cacagttact ggctcactga 20

<210> SEQ ID NO 144

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic

-continued

oligonucleotide

<400> SEQUENCE: 144

ggggttggtt ggttcactcc 20

<210> SEQ ID NO 145
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 145

gggtgctcgc ctgggttgcg 20

<210> SEQ ID NO 146
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 146

gccacttagc gtactggttc 20

<210> SEQ ID NO 147
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 147

ccctgttcgc ttgctcgttc 20

<210> SEQ ID NO 148
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 148

ttaacatggc gcgttcactg 20

<210> SEQ ID NO 149
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 149

ccatcttatt ggttcactgg 20

<210> SEQ ID NO 150
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 150

ctcttcccg gcactgactc 20

<210> SEQ ID NO 151
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 151

ctaagatgcc ccgctcgctc 20

<210> SEQ ID NO 152
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 152

gtatggcact cgctcgctgg 20

<210> SEQ ID NO 153
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 153

tgcattctgcc cttgctcact 20

<210> SEQ ID NO 154
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 154

actcactcac tgagttgtcc 20

<210> SEQ ID NO 155
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 155

ttcggtggct cactcacccc 20

<210> SEQ ID NO 156
<211> LENGTH: 20

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 156

gtgcccgttt ctactcactg 20

<210> SEQ ID NO 157
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 157

tggtcgctcg cccagccgc 20

<210> SEQ ID NO 158
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 158

ttgcttgctc actccgcgct 20

<210> SEQ ID NO 159
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 159

cgccagctcc ctctcatgc 20

<210> SEQ ID NO 160
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 160

cctccactcg ctgggtgagc 20

<210> SEQ ID NO 161
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 161

tcgtcgctcg ctctgcatcc 20

-continued

<210> SEQ ID NO 162
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 162

taccgcgatac tcgctctctg 20

<210> SEQ ID NO 163
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 163

cccacgccct cgctcactgc 20

<210> SEQ ID NO 164
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 164

tctctgtcac tcgtctccgg 20

<210> SEQ ID NO 165
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 165

ctcgtgcccc ggtgcgctca 20

<210> SEQ ID NO 166
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 166

cgtgtacttc gtgctcagcc 20

<210> SEQ ID NO 167
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 167

cccgtgtagg tgcgctcgct 20

-continued

<210> SEQ ID NO 168
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 168

tccaccctac tccgtcggcc 20

<210> SEQ ID NO 169
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 169

ctcaccgcac tcaactggccc 20

<210> SEQ ID NO 170
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 170

gccccctcctt tcgctcgttt 20

<210> SEQ ID NO 171
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 171

cctctccgcg tgcccactcg 20

<210> SEQ ID NO 172
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 172

ccttcttttc cgctggetca 20

<210> SEQ ID NO 173
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 173

-continued

ctcgtgcccc ggtgcgctca 20

<210> SEQ ID NO 174
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 174

ggcgattcac tcattctgacc 20

<210> SEQ ID NO 175
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 175

actgcccggc cgtttcgctc 20

<210> SEQ ID NO 176
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 176

cctctccgcg tgcccactcg 20

<210> SEQ ID NO 177
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 177

gcttgctcga cccagccacg 20

<210> SEQ ID NO 178
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 178

cctggttcgg tccttctccc 20

<210> SEQ ID NO 179
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

-continued

<400> SEQUENCE: 179

atcctcctat ctactcgct

20

<210> SEQ ID NO 180

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 180

cccgggtact gactcgcgct

20

<210> SEQ ID NO 181

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 181

tccatgccac ttggttcact

20

<210> SEQ ID NO 182

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 182

ctgctatcgt taactcacct

20

<210> SEQ ID NO 183

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 183

tgcactcatt cgactgcctc

20

<210> SEQ ID NO 184

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 184

cgagcgagct tagtgaacgt

20

<210> SEQ ID NO 185

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

-continued

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 185

agctggtcgg tgttcactgc 20

<210> SEQ ID NO 186
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 186

attcagtcgc ggttgcgcac 20

<210> SEQ ID NO 187
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 187

tctcgctggt tcaagtcttc 20

<210> SEQ ID NO 188
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 188

gcatttctgg gtagctcgct 20

<210> SEQ ID NO 189
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 189

tagctgaatc gccaggcttg 20

<210> SEQ ID NO 190
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 190

tggctcattc attgagtcaca 20

<210> SEQ ID NO 191
<211> LENGTH: 18
<212> TYPE: DNA

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 191

cattggctgt tgctgact 18

<210> SEQ ID NO 192
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 192

tccatgccac ttggttcact 20

<210> SEQ ID NO 193
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 193

gcccctcctt tcgctcgttt 20

<210> SEQ ID NO 194
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 194

cgcccccttc ccaactggctc 20

<210> SEQ ID NO 195
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 195

cgccagctcc ctctcatgc 20

<210> SEQ ID NO 196
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 196

gtcgctccct ccttaccgcg 20

<210> SEQ ID NO 197

-continued

<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 197

ccagcgaacg ccctcccgcgca 20

<210> SEQ ID NO 198
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 198

agaggggagtg gccaaactcca tca 23

<210> SEQ ID NO 199
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 199

tgatggagtt ggccactccc tct 23

<210> SEQ ID NO 200
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 200

gggggagagt gccacactct ca 22

<210> SEQ ID NO 201
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 201

tgagagtgtg gcactctccc cc 22

<210> SEQ ID NO 202
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 202

agaggggagtg ggcaactcca tca 23

<210> SEQ ID NO 203
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 203

tgatggagtt gcccactccc tct 23

<210> SEQ ID NO 204
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

-continued

<400> SEQUENCE: 204

agagggagtg gccaaactcca tca 23

<210> SEQ ID NO 205

<211> LENGTH: 23

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 205

tgatggagtt ggccactccc tct 23

<210> SEQ ID NO 206

<211> LENGTH: 23

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 206

agagggagtg gccaaactcca tca 23

<210> SEQ ID NO 207

<211> LENGTH: 23

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 207

tgatggagtt ggccacatta gct 23

<210> SEQ ID NO 208

<211> LENGTH: 23

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 208

agagggagtg gccaaactcca tca 23

<210> SEQ ID NO 209

<211> LENGTH: 23

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 209

tgatggagtt ggccactccc tct 23

<210> SEQ ID NO 210

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 210

actggccagc actccggtga 20

<210> SEQ ID NO 211

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 211

tcaccggagt gctggccagt 20

-continued

```

<210> SEQ ID NO 212
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 212

actggccagc actccggtga                20

<210> SEQ ID NO 213
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 213

tcaccggagt gctggccagt                20

<210> SEQ ID NO 214
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Bovine adeno-associated virus

<400> SEQUENCE: 214

gggggggagt gccacactct ct            22

<210> SEQ ID NO 215
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Bovine adeno-associated virus

<400> SEQUENCE: 215

agagagtgtg gcactcccc cc            22

<210> SEQ ID NO 216
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Snake parvovirus

<400> SEQUENCE: 216

tggggcgagt gccctgctc                19

<210> SEQ ID NO 217
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Snake parvovirus

<400> SEQUENCE: 217

gagcagggca ctgccccca                19

<210> SEQ ID NO 218
<211> LENGTH: 168
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 218

tacaaaaacc ccttgcttga tggagttggc cactccctct ctgtcgcgtt cgctcgctcg    60
ctggctcgctt tggggggggcg acggccagag ggccgctgctc tggcagctct ttgagctgcc    120
acccccccaa acgagccagc gagcgagcga acgcgacagg ggggagag                168

<210> SEQ ID NO 219
<211> LENGTH: 168

```

-continued

```

<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 219

tacaaaaccc ccttgcttga tggagttggc cacattagct ctgtcgcgtt cgctcgtcgc      60
ctggctcgtt tgggggggag acggccagag ggccgtcgtc tggcagctct ttgagctgcc      120
acccccccaa acgagccagc gagcgagcga acgcgacagg ggggagag                      168

<210> SEQ ID NO 220
<211> LENGTH: 145
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 220

tggcaaacca gatgatggag ttggccactc cctctatgag cgctcgtcga ctactcggc      60
cctggagacc aaaggtctcc agactgccgg cctctggccg gcagggccga gtgagtgagc      120
gagcgcgcat agagggagtg gccaa                                              145

<210> SEQ ID NO 221
<211> LENGTH: 144
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 221

aggaaccctt agtgagagtg tggcactctc cccctgcgcg gctcgtcgcg tcactgaggc      60
cgcccgggca aagcccgggc gtcgggcgac ctttggtcgc ccggccctcag tgagcgagcg      120
agcgcgcaga gagggagtg ccaa                                              144

<210> SEQ ID NO 222
<211> LENGTH: 142
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 222

ttacccttag tgagagtgtg gcactctccc cctgcgcgcg tcgctcgtc ggtggggccg      60
gcagagcaga gctctgccgt ctgcggacct ttggtccgca ggccccaccg agcgagcgag      120
cgcgagaga gggagtgggc aa                                              142

<210> SEQ ID NO 223
<211> LENGTH: 145
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 223

gccatacttc tagtgagagt gtggcactct ccccatgcgc cactcgtcgc ctcggtgggg      60
ccggacgtgc aaagcacgtc cgtctggcga cctttggtcg ccagggccca ccgagcgagc      120
gagtgcgcat agagggagtg gccaa                                              145

<210> SEQ ID NO 224
<211> LENGTH: 144
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 224

tggcaaacca gatgagagtg tggcactctc ccccatgcgc gctcgtcac tcactcggcc      60

```

-continued

ctggagacca aaggtctoca gactgccggc ctctggccgg cagggccgag tgagtgagcg 120

agcgcgcata gagggagtgg ccaa 144

<210> SEQ ID NO 225

<211> LENGTH: 146

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 225

cgcggtaccc ctagttagag tgtggcactc tccccatgc gcctcgctc gctcggtggg 60

gccggcagag cagagctctg ccgtctgagg acctttgggc cgcaggcccc accgagcgag 120

cgagcgcgca tagagggagt ggccaa 146

<210> SEQ ID NO 226

<211> LENGTH: 144

<212> TYPE: DNA

<213> ORGANISM: Bovine adeno-associated virus

<400> SEQUENCE: 226

aggaaccctt agagagagtg tggcactccc cccctgctgc gctcgctgc tcaactgaggc 60

cgccccggca aagccccggc gtcgggagac ctttggtgc cgcgcctcag tgagcgagcg 120

agcgcgcaga gagggagtgg ccaa 144

<210> SEQ ID NO 227

<211> LENGTH: 142

<212> TYPE: DNA

<213> ORGANISM: Bovine adeno-associated virus

<400> SEQUENCE: 227

ttacccttag agagagagtg gcaactcccc cctgctgcgc tcgctcgctc ggtggggcgc 60

gcagagcaga gctctgacct ctgcggacct ttggtccgca ggccccaccg agcgagcgag 120

cgcgagaga gggagtgggc aa 142

<210> SEQ ID NO 228

<211> LENGTH: 145

<212> TYPE: DNA

<213> ORGANISM: Bovine adeno-associated virus

<400> SEQUENCE: 228

gccatactc tagagagagt gtggcactcc ccccatgcgc cactcgctc ctcggtgggg 60

ccggacgtgc aaagcacgtc cgtctggcga cctttggtcg ccaggcccca ccgagcgagc 120

gagtgcgcat agagggagtg gccaa 145

<210> SEQ ID NO 229

<211> LENGTH: 144

<212> TYPE: DNA

<213> ORGANISM: Bovine adeno-associated virus

<400> SEQUENCE: 229

tggcaaacca gaagagagtg tggcactccc ccccatgcgc gctcgctcac tcaactgggc 60

ctggagacca aaggtctoca gactgccggc ctctggccgg cagggccgag tgagtgagcg 120

agcgcgcata gagggagtgg ccaa 144

-continued

```

<210> SEQ ID NO 230
<211> LENGTH: 146
<212> TYPE: DNA
<213> ORGANISM: Bovine adeno-associated virus

<400> SEQUENCE: 230

cgcggtaccc ctagagagag tgtggcactc ccccccatgc gcgtcgctc gtcgggtggg    60
gccggcagag cagagctctg ccgtctgcgg acctttggtc cgcaggcccc accgagcgag    120
cgagcgcgca tagagggagt ggccaa                                         146

```

```

<210> SEQ ID NO 231
<211> LENGTH: 141
<212> TYPE: DNA
<213> ORGANISM: Snake parvovirus

<400> SEQUENCE: 231

aggaaccctt aggagcaggg cactcgcccc actgcgcgct cgctcgctc ctgaggccgc    60
ccgggcaaag cccgggcgtc gggcgacctt tggtcgccc gcctcagtga gcgagcgagc    120
gcgcagagag ggagtggcca a                                              141

```

```

<210> SEQ ID NO 232
<211> LENGTH: 142
<212> TYPE: DNA
<213> ORGANISM: Snake parvovirus

<400> SEQUENCE: 232

ttacccttag agagagtgtg gcactcccc cctgcgcgc tcgtcgctc ggtggggccg    60
gcagagcaga gctctgccgt ctgcggacct ttggtcgcca ggccccaccg agcgagcgag    120
cgcgcagaga gggagtgggc aa                                             142

```

```

<210> SEQ ID NO 233
<211> LENGTH: 142
<212> TYPE: DNA
<213> ORGANISM: Snake parvovirus

<400> SEQUENCE: 233

gccatactc taggagcagg gcactcgccc caatgcgcac tcgtcgctc ggtggggccg    60
gacgtgcaaa gcacgtccgt ctggcgacct ttggtcgcca ggccccaccg agcgagcgag    120
tgcgcataga gggagtgggc aa                                             142

```

```

<210> SEQ ID NO 234
<211> LENGTH: 141
<212> TYPE: DNA
<213> ORGANISM: Snake parvovirus

<400> SEQUENCE: 234

tggcaaacca gagagcaggg cactcgcccc aatgcgcgct cgctcactc ctcgccctg    60
gagaccaaag gtctccagac tgccggcctc tggccggcag ggccgagtga gtgagcgagc    120
gcgcatagag ggagtggcca a                                              141

```

```

<210> SEQ ID NO 235
<211> LENGTH: 143
<212> TYPE: DNA
<213> ORGANISM: Snake parvovirus

<400> SEQUENCE: 235

```

-continued

```

cgcggtaccc ctaggagcag ggcactcgcc ccaatgcgcg ctgctcgct cggtggggcc      60
ggcagagcag agctctgccc tctgcggacc ttggtccgc agggcccacc gagcgagcga      120
gcgcgcatag agggagtggc caa                                              143

```

```

<210> SEQ ID NO 236
<211> LENGTH: 142
<212> TYPE: DNA
<213> ORGANISM: Avian adeno-associated virus

```

```

<400> SEQUENCE: 236

```

```

aggaaccct agtcaccgga gtgctggcca gtctgcgcgc tcgctcgctc actgaggccg      60
cccgggcaaa gcccgggcgt cgggcgacct ttggtcgccc ggctcagtg agcgagcgag      120
cgcgcgagaga gggagtggcc aa                                              142

```

```

<210> SEQ ID NO 237
<211> LENGTH: 140
<212> TYPE: DNA
<213> ORGANISM: Avian adeno-associated virus

```

```

<400> SEQUENCE: 237

```

```

ttacccttag tcaccggagt gctggccagt ctgcgcgcgc gctcgctcgg tggggccggc      60
agagcagagc tctgccgtct gcggaccttt ggtccgcagg cccaccagag cgagcgagcg      120
cgcgagagagg gagtgggcaa                                              140

```

```

<210> SEQ ID NO 238
<211> LENGTH: 143
<212> TYPE: DNA
<213> ORGANISM: Avian adeno-associated virus

```

```

<400> SEQUENCE: 238

```

```

gccatacctc tagtcaccgg agtgcctggc agtatgcgca ctgctcgct cggtggggcc      60
ggacgtgcaa agcacgtccg tctggcgacc ttggtcgcc agggcccacc gagcgagcga      120
gtgcgcatag agggagtggc caa                                              143

```

```

<210> SEQ ID NO 239
<211> LENGTH: 142
<212> TYPE: DNA
<213> ORGANISM: Avian adeno-associated virus

```

```

<400> SEQUENCE: 239

```

```

tggcaaacca gatcaccgga gtgctggcca gtatgcgcgc tcgctcactc actcggccct      60
ggagacacaaa ggtctccaga ctgccggcct ctggccgcca gggccgagtg agtgagcgag      120
cgcgcataga gggagtggcc aa                                              142

```

```

<210> SEQ ID NO 240
<211> LENGTH: 144
<212> TYPE: DNA
<213> ORGANISM: Avian adeno-associated virus

```

```

<400> SEQUENCE: 240

```

```

cgcggtaccc ctagtaccg gagtgcctggc cagtatgcgc gctcgctcgc tcggtggggc      60
cggcagagca gagctctgcc gtctgcggac ctttgggtccg caggccccac cgagcgagcg      120
agcgcgcata gagggagtgg ccaa                                              144

```

-continued

```

<210> SEQ ID NO 241
<211> LENGTH: 145
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 241

ttggccactc cctctctgcg cgctcgctcg ctactgagg cggggcgacc aaaggtcgcc      60
cgacgcccgg gctttgcccg ggcgccctca gtgagcgagc gagcgcgag agagggagtg      120
gccaaactcca tctactagggg ttctt                                           145

<210> SEQ ID NO 242
<211> LENGTH: 145
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 242

aggaaccctc agtgatggag ttggccactc cctctctgcg cgctcgctcg ctactgagg      60
ccgcccgggc aaagcccggg cgtcgggcga cctttggtcg cccggcctca gtgagcgagc      120
gagcgcgagc agagggagtg gccaa                                           145

<210> SEQ ID NO 243
<211> LENGTH: 167
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 243

ctctcccccc tgtcgcttc gctcgctcg tggtcgttt ggggggggtg cagctcaaag      60
agctgccaga cgacggccct ctggccgctg ccccccaaa cgagccagcg agcgagcgaa      120
cgcgacaggg gggagagtgc cacactctca agcaaggggg ttttgta                    167

<210> SEQ ID NO 244
<211> LENGTH: 167
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 244

tacaaaacc ccttgcttga gagtgtggca ctctcccccc tgtcgcttc gctcgctcg      60
tggtcgcttt gggggggcga cggccagagg gccgctgctt ggcagctctt tgagtgcca      120
ccccccaaa cgagccagcg agcgagcgaa cgcgacaggg gggagag                    167

<210> SEQ ID NO 245
<211> LENGTH: 143
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 245

ttgcccactc cctctctgcg cgctcgctcg ctcggtgggg cctcgcgacc aaaggtcgc      60
agacggcaga gctctgctct gccggcccca ccgagcgagc gagcgcgag agagggagtg      120
ggcaactcca tctactagggg taa                                           143

<210> SEQ ID NO 246
<211> LENGTH: 143
<212> TYPE: DNA
<213> ORGANISM: Adeno-associated virus

```

-continued

<400> SEQUENCE: 246

```

ttacccttag tgatggagtt gccactccc tctctgcgcg ctctctcgt cggtggggcc      60
ggcagagcag agctctgcgc tctgcggacc ttggtccgc agggcccacc gagcgagcga      120
gcgcgcagag agggagtggg caa                                           143

```

<210> SEQ ID NO 247

<211> LENGTH: 146

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 247

```

ttggccactc cctctatgcg cactcgcctc ctcggtgggg cctggcgacc aaaggtcgcc      60
agacggacgt gctttgcacg tccggcccca ccgagcgagc gagtgcgcac agagggagtg      120
gccaaactcca tcactagagg tatggc                                           146

```

<210> SEQ ID NO 248

<211> LENGTH: 146

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 248

```

gccatacctc tagtgatgga gttggccact cctctatgcg gactcgcctc gctcggtggg      60
gccggacgtg caaagcacgt ccgtctggcg acctttggtc gccaggcccc accgagcgag      120
cgagtgcgca tagagggagt ggccaa                                           146

```

<210> SEQ ID NO 249

<211> LENGTH: 145

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 249

```

ttggccactc cctctatgcg cgctcgcctc ctcactcggc cctggagacc aaaggtctcc      60
agactgccgg cctctggcgc gcagggccga gtgagtgcgc gagcgcgcat agagggagtg      120
gccaaactcca tcacttaggt ttgcc                                           145

```

<210> SEQ ID NO 250

<211> LENGTH: 145

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 250

```

tggcaaacca gatgatggag ttggccacat tagctatgcg cgctcgcctc ctcactcggc      60
cctggagacc aaaggtctcc agactgccgg cctctggccc gcagggccga gtgagtgcgc      120
gagcgcgcat agagggagtg gccaa                                           145

```

<210> SEQ ID NO 251

<211> LENGTH: 147

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 251

```

ttggccactc cctctatgcg cgctcgcctc ctcggtgggg cctgcggacc aaaggtccgc      60
agacggcaga gctctgtctt gccggcccca ccgagcgagc gagcgcgcat agagggagtg      120

```

-continued

gccaactcca tcactagggg taccgcg 147

<210> SEQ ID NO 252
 <211> LENGTH: 147
 <212> TYPE: DNA
 <213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 252

cgcggtaccc ctagtgtatgg agttggccac tccctctatg cgcgctcgct cgctcggtgg 60

ggccggcaga gcagagctct gccgtctgcg gacctttggt ccgcaggccc caccgagcga 120

gcgagcgcgc atagagggag tggccaa 147

<210> SEQ ID NO 253
 <211> LENGTH: 150
 <212> TYPE: DNA
 <213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 253

cagagaggga gtggccaact ccatcactag gggtagcgcg aagcgctcc cgcgtgccg 60

cgtcagcgct gacgtaaatt acgtcatagg ggagtgggcc tgtattagct gtcacgtgag 120

tgcttttgcg gcattttgcg acaccacgtg 150

<210> SEQ ID NO 254
 <211> LENGTH: 150
 <212> TYPE: DNA
 <213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 254

cacgtggtgt cgaaaatgc cgaaaagca ctcacgtgac agctaataca ggaccactcc 60

cctatgacgt aatttacgtc agcgtgacg cggcagcggt ggaggcgctt cgcgtaccc 120

ctagtgtatgg agttggccac tccctctctg 150

<210> SEQ ID NO 255
 <211> LENGTH: 143
 <212> TYPE: DNA
 <213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 255

tggccagttt ccaaaacagg ctgcgtcgct cactcgggcc ccggcccaa gggggccggt 60

agcgacgcct ttggcgtgc ggcccagtg agcgagcgag cctgttttgg aaactggcca 120

gcactccggt gaggtaatgc cgt 143

<210> SEQ ID NO 256
 <211> LENGTH: 143
 <212> TYPE: DNA
 <213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 256

acggcattac ctcaccggag tgctggccag ttccaaaac aggctcgctc gctcactcgg 60

gccgcgacgc caaaggcgtc gctaccggcc cccttggggc cggggcccga gtgagcgagc 120

gagcctgttt tggaaactgg cca 143

<210> SEQ ID NO 257
 <211> LENGTH: 142

-continued

```

<212> TYPE: DNA
<213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 257

tggccagttt ccaagacagg ctcgctcgct cactcgggcc ggggcccacaa aggggcccct      60
agcgaccgct tcgcggtcgc ggcccagtg agcgagcgag cctgtcttgg aaactggcca      120
gcactccggt gaggtaatgc cg                                          142

<210> SEQ ID NO 258
<211> LENGTH: 142
<212> TYPE: DNA
<213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 258

cggcattacc tcaccggagt gctggccagt ttccaagaca ggctcgctcg ctcaactcggg      60
ccgcgaccgc gaagcggtcg ctatggggccc ctttggggcc ccggcccagag tgagcgagcg      120
agcctgtctt ggaaactggc ca                                          142

<210> SEQ ID NO 259
<211> LENGTH: 173
<212> TYPE: DNA
<213> ORGANISM: Bovine adeno-associated virus

<400> SEQUENCE: 259

gtggcactcc cccccctgtc gcgttcgtc gtctcgtggc tcgattgggg ggggtggcagc      60
tcaaagagct gccagacgac ggccctcttg gccgtcgccc cccaatcga gccagcgaac      120
gagcgaacgc gacagggggg ggagtggcac actctctagc aaggggggtt tgt          173

<210> SEQ ID NO 260
<211> LENGTH: 173
<212> TYPE: DNA
<213> ORGANISM: Bovine adeno-associated virus

<400> SEQUENCE: 260

acaaaacccc cttgctagag agtgtggcac tccccccct gtgcgttcg ctcgttcgct      60
ggctcgattg gggggggcgac ggcccagagg gccgtcgtct ggcagctctt tgagctgcca      120
ccccccaat cgagccagcg aacgagcgaa cgcgacaggg gggggagtgc cac          173

<210> SEQ ID NO 261
<211> LENGTH: 150
<212> TYPE: DNA
<213> ORGANISM: Bat adeno-associated virus

<400> SEQUENCE: 261

aagaaacctc taatcatcat ctgtactttg tgaggggtgtg gtcgtgagaa cgggcatgcc      60
cgggcatgtc ctgttaatca ttaaccagaa tggacttttg acttaataag ttacgtatta      120
atgggtgtgt atataagtaa gtgtatttcc                                150

<210> SEQ ID NO 262
<211> LENGTH: 150
<212> TYPE: DNA
<213> ORGANISM: Bat adeno-associated virus

<400> SEQUENCE: 262

ggaaatacac ttacttatat acacacccat taatacgtaa cttattaagt ccaaagtcca      60

```

-continued

ttctgggttaa tgattaacag gacatgcccg ggcgatgcccg ttctcagcag cacaccotca 120

caaagtacag atgatgatta gaggtttctt 150

<210> SEQ ID NO 263

<211> LENGTH: 154

<212> TYPE: DNA

<213> ORGANISM: Snake parvovirus

<400> SEQUENCE: 263

cgccccaccc ctagtgatcg cgcgcgctct ctcttggggc ctgacggccg aaggcgtca 60

gctgccgagc ttcgctcggc aggccccaaag agagagcgcg cgcgatcact aggggtgggg 120

cgagtgcctt gctcaacggg ttttttgggtg ggcg 154

<210> SEQ ID NO 264

<211> LENGTH: 154

<212> TYPE: DNA

<213> ORGANISM: Snake parvovirus

<400> SEQUENCE: 264

cgccccacaa aaaaccggtt gacaggggca ctcgccccac ccctagtgat cgcgcgcgct 60

ctctcttggg gcctgccgag cgaagctcgg cagctgacgg ccttcggccg tcaggcccca 120

agagagagcg cgcgcgatca ctagggtgtg ggcg 154

<210> SEQ ID NO 265

<211> LENGTH: 536

<212> TYPE: PRT

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 265

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
1 5 10 15Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
20 25 30Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
35 40 45Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
50 55 60Thr Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
65 70 75 80Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Met His Val Leu Val Glu
85 90 95Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
100 105 110Arg Glu Lys Leu Ile Gln Arg Ile Tyr Arg Gly Ile Glu Pro Thr Leu
115 120 125Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
130 135 140Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
145 150 155 160Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Gln Tyr Leu
165 170 175Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val Ala Gln His
180 185 190

-continued

```

Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
    195                                200                                205

Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
    210                                215                                220

Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys
    225                                230                                235                                240

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
    245                                250                                255

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
    260                                265                                270

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln
    275                                280                                285

Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu
    290                                295                                300

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
    305                                310                                315                                320

Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
    325                                330                                335

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
    340                                345                                350

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
    355                                360                                365

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
    370                                375                                380

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
    385                                390                                395                                400

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
    405                                410                                415

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
    420                                425                                430

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
    435                                440                                445

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
    450                                455                                460

Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val
    465                                470                                475                                480

Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala
    485                                490                                495

Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val
    500                                505                                510

Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp
    515                                520                                525

Arg Leu Ala Arg Gly His Ser Leu
    530                                535

```

<210> SEQ ID NO 266

<211> LENGTH: 621

<212> TYPE: PRT

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 266

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp

-continued

1	5	10	15
Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu	20	25	30
Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile	35	40	45
Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu	50	55	60
Thr Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val	65	70	75
Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Met His Val Leu Val Glu	85	90	95
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile	100	105	110
Arg Glu Lys Leu Ile Gln Arg Ile Tyr Arg Gly Ile Glu Pro Thr Leu	115	120	125
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly	130	135	140
Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys	145	150	155
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Gln Tyr Leu	165	170	175
Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val Ala Gln His	180	185	190
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn	195	200	205
Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr	210	215	220
Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys	225	230	235
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala	245	250	255
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys	260	265	270
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln	275	280	285
Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu	290	295	300
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala	305	310	315
Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala	325	330	335
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro	340	345	350
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp	355	360	365
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala	370	375	380
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg	385	390	395
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val	405	410	415

-continued

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
 420 425 430
 Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
 435 440 445
 Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
 450 455 460
 Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val
 465 470 475 480
 Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala
 485 490 495
 Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val
 500 505 510
 Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp
 515 520 525
 Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
 530 535 540
 Phe Pro Cys Arg Gln Cys Glu Arg Met Asn Gln Asn Ser Asn Ile Cys
 545 550 555 560
 Phe Thr His Gly Gln Lys Asp Cys Leu Glu Cys Phe Pro Val Ser Glu
 565 570 575
 Ser Gln Pro Val Ser Val Val Lys Lys Ala Tyr Gln Lys Leu Cys Tyr
 580 585 590
 Ile His His Ile Met Gly Lys Val Pro Asp Ala Cys Thr Ala Cys Asp
 595 600 605
 Leu Val Asn Val Asp Leu Asp Asp Cys Ile Phe Glu Gln
 610 615 620

<210> SEQ ID NO 267

<211> LENGTH: 312

<212> TYPE: PRT

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 267

Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys
 1 5 10 15
 Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
 20 25 30
 Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
 35 40 45
 Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln
 50 55 60
 Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu
 65 70 75 80
 Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
 85 90 95
 Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
 100 105 110
 Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
 115 120 125
 Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
 130 135 140
 Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala

-continued

145	150	155	160
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg	165	170	175
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val	180	185	190
Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser	195	200	205
Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe	210	215	220
Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln	225	230	235
Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val	245	250	255
Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala	260	265	270
Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val	275	280	285
Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp	290	295	300
Arg Leu Ala Arg Gly His Ser Leu	305	310	

<210> SEQ ID NO 268
 <211> LENGTH: 397
 <212> TYPE: PRT
 <213> ORGANISM: Adeno-associated virus
 <400> SEQUENCE: 268

Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys	1	5	10	15
Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala	20	25	30	
Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys	35	40	45	
Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln	50	55	60	
Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu	65	70	75	80
Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala	85	90	95	
Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala	100	105	110	
Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro	115	120	125	
Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp	130	135	140	
Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala	145	150	155	160
Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg	165	170	175	
Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val	180	185	190	

-continued

Ile	Val	Thr	Ser	Asn	Thr	Asn	Met	Cys	Ala	Val	Ile	Asp	Gly	Asn	Ser
	195						200					205			
Thr	Thr	Phe	Glu	His	Gln	Gln	Pro	Leu	Gln	Asp	Arg	Met	Phe	Lys	Phe
	210					215					220				
Glu	Leu	Thr	Arg	Arg	Leu	Asp	His	Asp	Phe	Gly	Lys	Val	Thr	Lys	Gln
225					230					235					240
Glu	Val	Lys	Asp	Phe	Phe	Arg	Trp	Ala	Lys	Asp	His	Val	Val	Glu	Val
				245					250					255	
Glu	His	Glu	Phe	Tyr	Val	Lys	Lys	Gly	Gly	Ala	Lys	Lys	Arg	Pro	Ala
			260					265					270		
Pro	Ser	Asp	Ala	Asp	Ile	Ser	Glu	Pro	Lys	Arg	Val	Arg	Glu	Ser	Val
		275					280					285			
Ala	Gln	Pro	Ser	Thr	Ser	Asp	Ala	Glu	Ala	Ser	Ile	Asn	Tyr	Ala	Asp
	290					295					300				
Arg	Tyr	Gln	Asn	Lys	Cys	Ser	Arg	His	Val	Gly	Met	Asn	Leu	Met	Leu
305					310					315					320
Phe	Pro	Cys	Arg	Gln	Cys	Glu	Arg	Met	Asn	Gln	Asn	Ser	Asn	Ile	Cys
				325					330					335	
Phe	Thr	His	Gly	Gln	Lys	Asp	Cys	Leu	Glu	Cys	Phe	Pro	Val	Ser	Glu
			340					345					350		
Ser	Gln	Pro	Val	Ser	Val	Val	Lys	Lys	Ala	Tyr	Gln	Lys	Leu	Cys	Tyr
		355					360					365			
Ile	His	His	Ile	Met	Gly	Lys	Val	Pro	Asp	Ala	Cys	Thr	Ala	Cys	Asp
	370					375					380				
Leu	Val	Asn	Val	Asp	Leu	Asp	Asp	Cys	Ile	Phe	Glu	Gln			
385					390					395					

<210> SEQ ID NO 269

<211> LENGTH: 610

<212> TYPE: PRT

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 269

Met	Ala	Thr	Phe	Tyr	Glu	Val	Ile	Val	Arg	Val	Pro	Phe	Asp	Val	Glu
1				5					10					15	
Glu	His	Leu	Pro	Gly	Ile	Ser	Asp	Ser	Phe	Val	Asp	Trp	Val	Thr	Gly
		20					25						30		
Gln	Ile	Trp	Glu	Leu	Pro	Pro	Glu	Ser	Asp	Leu	Asn	Leu	Thr	Leu	Val
	35					40					45				
Glu	Gln	Pro	Gln	Leu	Thr	Val	Ala	Asp	Arg	Ile	Arg	Arg	Val	Phe	Leu
	50				55					60					
Tyr	Glu	Trp	Asn	Lys	Phe	Ser	Lys	Gln	Glu	Ser	Lys	Phe	Phe	Val	Gln
65				70					75					80	
Phe	Glu	Lys	Gly	Ser	Glu	Tyr	Phe	His	Leu	His	Thr	Leu	Val	Glu	Thr
				85				90						95	
Ser	Gly	Ile	Ser	Ser	Met	Val	Leu	Gly	Arg	Tyr	Val	Ser	Gln	Ile	Arg
			100					105					110		
Ala	Gln	Leu	Val	Lys	Val	Val	Phe	Gln	Gly	Ile	Glu	Pro	Gln	Ile	Asn
		115					120					125			
Asp	Trp	Val	Ala	Ile	Thr	Lys	Val	Lys	Lys	Gly	Gly	Ala	Asn	Lys	Val
	130					135					140				
Val	Asp	Ser	Gly	Tyr	Ile	Pro	Ala	Tyr	Leu	Leu	Pro	Lys	Val	Gln	Pro
145					150					155					160

-continued

Glu	Leu	Gln	Trp	Ala	Trp	Thr	Asn	Leu	Asp	Glu	Tyr	Lys	Leu	Ala	Ala	
				165					170					175		
Leu	Asn	Leu	Glu	Glu	Arg	Lys	Arg	Leu	Val	Ala	Gln	Phe	Leu	Ala	Glu	
			180					185					190			
Ser	Ser	Gln	Arg	Ser	Gln	Glu	Ala	Ala	Ser	Gln	Arg	Glu	Phe	Ser	Ala	
		195					200					205				
Asp	Pro	Val	Ile	Lys	Ser	Lys	Thr	Ser	Gln	Lys	Tyr	Met	Ala	Leu	Val	
	210					215					220					
Asn	Trp	Leu	Val	Glu	His	Gly	Ile	Thr	Ser	Glu	Lys	Gln	Trp	Ile	Gln	
225					230					235				240		
Glu	Asn	Gln	Glu	Ser	Tyr	Leu	Ser	Phe	Asn	Ser	Thr	Gly	Asn	Ser	Arg	
			245						250				255			
Ser	Gln	Ile	Lys	Ala	Ala	Leu	Asp	Asn	Ala	Thr	Lys	Ile	Met	Ser	Leu	
		260						265					270			
Thr	Lys	Ser	Ala	Val	Asp	Tyr	Leu	Val	Gly	Ser	Ser	Val	Pro	Glu	Asp	
		275					280					285				
Ile	Ser	Lys	Asn	Arg	Ile	Trp	Gln	Ile	Phe	Glu	Met	Asn	Gly	Tyr	Asp	
	290					295					300					
Pro	Ala	Tyr	Ala	Gly	Ser	Ile	Leu	Tyr	Gly	Trp	Cys	Gln	Arg	Ser	Phe	
305					310					315				320		
Asn	Lys	Arg	Asn	Thr	Val	Trp	Leu	Tyr	Gly	Pro	Ala	Thr	Thr	Gly	Lys	
			325						330					335		
Thr	Asn	Ile	Ala	Glu	Ala	Ile	Ala	His	Thr	Val	Pro	Phe	Tyr	Gly	Cys	
		340						345					350			
Val	Asn	Trp	Thr	Asn	Glu	Asn	Phe	Pro	Phe	Asn	Asp	Cys	Val	Asp	Lys	
		355					360					365				
Met	Leu	Ile	Trp	Trp	Glu	Glu	Gly	Lys	Met	Thr	Asn	Lys	Val	Val	Glu	
	370					375					380					
Ser	Ala	Lys	Ala	Ile	Leu	Gly	Gly	Ser	Lys	Val	Arg	Val	Asp	Gln	Lys	
385					390					395				400		
Cys	Lys	Ser	Ser	Val	Gln	Ile	Asp	Ser	Thr	Pro	Val	Ile	Val	Thr	Ser	
			405						410					415		
Asn	Thr	Asn	Met	Cys	Val	Val	Val	Asp	Gly	Asn	Ser	Thr	Thr	Phe	Glu	
			420					425						430		
His	Gln	Gln	Pro	Leu	Glu	Asp	Arg	Met	Phe	Lys	Phe	Glu	Leu	Thr	Lys	
		435					440					445				
Arg	Leu	Pro	Pro	Asp	Phe	Gly	Lys	Ile	Thr	Lys	Gln	Glu	Val	Lys	Asp	
	450					455					460					
Phe	Phe	Ala	Trp	Ala	Lys	Val	Asn	Gln	Val	Pro	Val	Thr	His	Glu	Phe	
465					470					475				480		
Lys	Val	Pro	Arg	Glu	Leu	Ala	Gly	Thr	Lys	Gly	Ala	Glu	Lys	Ser	Leu	
				485					490					495		
Lys	Arg	Pro	Leu	Gly	Asp	Val	Thr	Asn	Thr	Ser	Tyr	Lys	Ser	Leu	Glu	
			500					505					510			
Lys	Arg	Ala	Arg	Leu	Ser	Phe	Val	Pro	Glu	Thr	Pro	Arg	Ser	Ser	Asp	
		515					520					525				
Val	Thr	Val	Asp	Pro	Ala	Pro	Leu	Arg	Pro	Leu	Asn	Trp	Asn	Ser	Arg	
	530					535					540					
Tyr	Asp	Cys	Lys	Cys	Asp	Tyr	His	Ala	Gln	Phe	Asp	Asn	Ile	Ser	Asn	
545					550					555					560	

-continued

Lys Cys Asp Glu Cys Glu Tyr Leu Asn Arg Gly Lys Asn Gly Cys Ile
565 570 575

Cys His Asn Val Thr His Cys Gln Ile Cys His Gly Ile Pro Pro Trp
580 585 590

Glu Lys Glu Asn Leu Ser Asp Phe Gly Asp Phe Asp Asp Ala Asn Lys
595 600 605

Glu Gln
610

<210> SEQ ID NO 270

<211> LENGTH: 623

<212> TYPE: PRT

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 270

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
1 5 10 15

Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Ser Trp Val Ala Glu
20 25 30

Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
35 40 45

Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
50 55 60

Val Gln Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
65 70 75 80

Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Leu His Ile Leu Val Glu
85 90 95

Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
100 105 110

Arg Asp Lys Leu Val Gln Thr Ile Tyr Arg Gly Ile Glu Pro Thr Leu
115 120 125

Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
130 135 140

Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
145 150 155 160

Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Glu Tyr Ile
165 170 175

Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
180 185 190

Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Leu Asn
195 200 205

Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
210 215 220

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
225 230 235 240

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
245 250 255

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
260 265 270

Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ala
275 280 285

Pro Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
290 295 300

-continued

```

Asn Gly Tyr Glu Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
305                      310                      315                      320

Gln Lys Arg Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
                      325                      330                      335

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Ala Val Pro
                      340                      345                      350

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
                      355                      360                      365

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
370                      375                      380

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
385                      390                      395                      400

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
                      405                      410                      415

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
                      420                      425                      430

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe
                      435                      440                      445

Glu Leu Thr Arg Arg Leu Glu His Asp Phe Gly Lys Val Thr Lys Gln
450                      455                      460

Glu Val Lys Glu Phe Phe Arg Trp Ala Gln Asp His Val Thr Glu Val
465                      470                      475                      480

Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Asn Lys Arg Pro Ala
                      485                      490                      495

Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro Ser Val
500                      505                      510

Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp Phe Ala
515                      520                      525

Asp Arg Tyr Gln Asn Lys Cys Ser Arg His Ala Gly Met Leu Gln Met
530                      535                      540

Leu Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Asn Phe Asn Ile
545                      550                      555                      560

Cys Phe Thr His Gly Thr Arg Asp Cys Ser Glu Cys Phe Pro Gly Val
565                      570                      575

Ser Glu Ser Gln Pro Val Val Arg Lys Arg Thr Tyr Arg Lys Leu Cys
580                      585                      590

Ala Ile His His Leu Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser Ala
595                      600                      605

Cys Asp Leu Val Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln
610                      615                      620

```

<210> SEQ ID NO 271

<211> LENGTH: 624

<212> TYPE: PRT

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 271

```

Met Pro Gly Phe Tyr Glu Ile Val Leu Lys Val Pro Ser Asp Leu Asp
1                      5                      10                      15

Glu Arg Leu Pro Gly Ile Ser Asn Ser Phe Val Asn Trp Val Ala Glu
20                      25                      30

Lys Glu Trp Asp Val Pro Pro Asp Ser Asp Met Asp Pro Asn Leu Ile

```

-continued

35					40					45					
Glu 50	Gln	Ala	Pro	Leu	Thr	Val 55	Ala	Glu	Lys	Leu	Gln 60	Arg	Glu	Phe	Leu
Val 65	Glu	Trp	Arg	Arg	Val 70	Ser	Lys	Ala	Pro	Glu 75	Ala	Leu	Phe	Phe	Val 80
Gln	Phe	Glu	Lys	Gly 85	Glu	Thr	Tyr	Phe	His 90	Leu	His	Val	Leu	Ile 95	Glu
Thr	Ile	Gly	Val 100	Lys	Ser	Met	Val	Val 105	Gly	Arg	Tyr	Val 110	Ser	Gln	Ile
Lys	Glu	Lys 115	Leu	Val	Thr	Arg	Ile 120	Tyr	Arg	Gly	Val 125	Glu	Pro	Gln	Leu
Pro 130	Asn	Trp	Phe	Ala	Val	Thr 135	Lys	Thr	Arg	Asn	Gly 140	Ala	Gly	Gly	Gly
Asn 145	Lys	Val	Val	Asp	Asp 150	Cys	Tyr	Ile	Pro	Asn 155	Tyr	Leu	Leu	Pro	Lys 160
Thr	Gln	Pro	Glu	Leu 165	Gln	Trp	Ala	Trp	Thr 170	Asn	Met	Asp	Gln	Tyr 175	Leu
Ser	Ala	Cys	Leu 180	Asn	Leu	Ala	Glu	Arg 185	Lys	Arg	Leu	Val 190	Ala	Gln	His
Leu	Thr	His 195	Val	Ser	Gln	Thr	Gln 200	Glu	Gln	Asn	Lys 205	Glu	Asn	Gln	Asn
Pro 210	Asn	Ser	Asp	Ala	Pro	Val 215	Ile	Arg	Ser	Lys 220	Thr	Ser	Ala	Arg	Tyr
Met 225	Glu	Leu	Val	Gly 230	Trp	Leu	Val	Asp	Arg	Gly 235	Ile	Thr	Ser	Glu	Lys 240
Gln	Trp	Ile	Gln 245	Glu	Asp	Gln	Ala	Ser	Tyr 250	Ile	Ser	Phe	Asn	Ala 255	Ala
Ser	Asn	Ser	Arg 260	Ser	Gln	Ile	Lys	Ala 265	Ala	Leu	Asp	Asn 270	Ala	Ser	Lys
Ile	Met	Ser 275	Leu	Thr	Lys	Thr	Ala 280	Pro	Asp	Tyr	Leu 285	Val	Gly	Ser	Asn
Pro 290	Pro	Glu	Asp	Ile	Thr	Lys 295	Asn	Arg	Ile	Tyr 300	Gln	Ile	Leu	Glu	Leu
Asn 305	Gly	Tyr	Asp	Pro	Gln 310	Tyr	Ala	Ala	Ser	Val 315	Phe	Leu	Gly	Trp	Ala 320
Gln	Lys	Lys	Phe 325	Gly	Lys	Arg	Asn	Thr	Ile 330	Trp	Leu	Phe	Gly	Pro 335	Ala
Thr	Thr	Gly	Lys 340	Thr	Asn	Ile	Ala	Glu 345	Ala	Ile	Ala	His 350	Ala	Val	Pro
Phe	Tyr	Gly 355	Cys	Val	Asn	Trp	Thr 360	Asn	Glu	Asn	Phe 365	Pro	Phe	Asn	Asp
Cys 370	Val	Asp	Lys	Met	Val	Ile 375	Trp	Trp	Glu	Glu 380	Gly	Lys	Met	Thr	Ala
Lys 385	Val	Val	Glu	Ser	Ala 390	Lys	Ala	Ile	Leu	Gly 395	Gly	Ser	Lys	Val	Arg 400
Val	Asp	Gln	Lys	Cys 405	Lys	Ser	Ser	Ala	Gln 410	Ile	Glu	Pro	Thr	Pro 415	Val
Ile	Val	Thr	Ser 420	Asn	Thr	Asn	Met	Cys 425	Ala	Val	Ile	Asp 430	Gly	Asn	Ser
Thr	Thr	Phe 435	Glu	His	Gln	Gln	Pro 440	Leu	Gln	Asp	Arg 445	Met	Phe	Glu	Phe

-continued

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln
 450 455 460
 Glu Val Lys Asp Phe Phe Arg Trp Ala Ser Asp His Val Thr Asp Val
 465 470 475 480
 Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Lys Lys Arg Pro Ala
 485 490 495
 Ser Asn Asp Ala Asp Val Ser Glu Pro Lys Arg Glu Cys Thr Ser Leu
 500 505 510
 Ala Gln Pro Thr Thr Ser Asp Ala Glu Ala Pro Ala Asp Tyr Ala Asp
 515 520 525
 Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu
 530 535 540
 Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Ile Ser Asn Val Cys
 545 550 555 560
 Phe Thr His Gly Gln Arg Asp Cys Gly Glu Cys Phe Pro Gly Met Ser
 565 570 575
 Glu Ser Gln Pro Val Ser Val Val Lys Lys Lys Thr Tyr Gln Lys Leu
 580 585 590
 Cys Pro Ile His His Ile Leu Gly Arg Ala Pro Glu Ile Ala Cys Ser
 595 600 605
 Ala Cys Asp Leu Ala Asn Val Asp Leu Asp Asp Cys Val Ser Glu Gln
 610 615 620

<210> SEQ ID NO 272

<211> LENGTH: 623

<212> TYPE: PRT

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 272

Met Pro Gly Phe Tyr Glu Ile Val Leu Lys Val Pro Ser Asp Leu Asp
 1 5 10 15
 Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Ser Trp Val Ala Glu
 20 25 30
 Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
 35 40 45
 Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Glu Phe Leu
 50 55 60
 Val Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
 65 70 75 80
 Gln Phe Glu Lys Gly Asp Ser Tyr Phe His Leu His Ile Leu Val Glu
 85 90 95
 Thr Val Gly Val Lys Ser Met Val Val Gly Arg Tyr Val Ser Gln Ile
 100 105 110
 Lys Glu Lys Leu Val Thr Arg Ile Tyr Arg Gly Val Glu Pro Gln Leu
 115 120 125
 Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
 130 135 140
 Asn Lys Val Val Asp Asp Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
 145 150 155 160
 Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Asp Gln Tyr Ile
 165 170 175
 Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His

-continued

180						185						190					
Leu	Thr	His	Val	Ser	Gln	Thr	Gln	Glu	Gln	Asn	Lys	Glu	Asn	Gln	Asn		
195						200						205					
Pro	Asn	Ser	Asp	Ala	Pro	Val	Ile	Arg	Ser	Lys	Thr	Ser	Ala	Arg	Tyr		
210						215						220					
Met	Glu	Leu	Val	Gly	Trp	Leu	Val	Asp	Arg	Gly	Ile	Thr	Ser	Glu	Lys		
225						230						235					
Gln	Trp	Ile	Gln	Glu	Asp	Gln	Ala	Ser	Tyr	Ile	Ser	Phe	Asn	Ala	Ala		
245						250						255					
Ser	Asn	Ser	Arg	Ser	Gln	Ile	Lys	Ala	Ala	Leu	Asp	Asn	Ala	Ser	Lys		
260						265						270					
Ile	Met	Ser	Leu	Thr	Lys	Thr	Ala	Pro	Asp	Tyr	Leu	Val	Gly	Gln	Asn		
275						280						285					
Pro	Pro	Glu	Asp	Ile	Ser	Ser	Asn	Arg	Ile	Tyr	Arg	Ile	Leu	Glu	Met		
290						295						300					
Asn	Gly	Tyr	Asp	Pro	Gln	Tyr	Ala	Ala	Ser	Val	Phe	Leu	Gly	Trp	Ala		
305						310						315					
Gln	Lys	Lys	Phe	Gly	Lys	Arg	Asn	Thr	Ile	Trp	Leu	Phe	Gly	Pro	Ala		
325						330						335					
Thr	Thr	Gly	Lys	Thr	Asn	Ile	Ala	Glu	Ala	Ile	Ala	His	Ala	Val	Pro		
340						345						350					
Phe	Tyr	Gly	Cys	Val	Asn	Trp	Thr	Asn	Glu	Asn	Phe	Pro	Phe	Asn	Asp		
355						360						365					
Cys	Val	Asp	Lys	Met	Val	Ile	Trp	Trp	Glu	Glu	Gly	Lys	Met	Thr	Ala		
370						375						380					
Lys	Val	Val	Glu	Ser	Ala	Lys	Ala	Ile	Leu	Gly	Gly	Ser	Lys	Val	Arg		
385						390						395					
Val	Asp	Gln	Lys	Cys	Lys	Ser	Ser	Ala	Gln	Ile	Asp	Pro	Thr	Pro	Val		
405						410						415					
Ile	Val	Thr	Ser	Asn	Thr	Asn	Met	Cys	Ala	Val	Ile	Asp	Gly	Asn	Ser		
420						425						430					
Thr	Thr	Phe	Glu	His	Gln	Gln	Pro	Leu	Gln	Asp	Arg	Met	Phe	Lys	Phe		
435						440						445					
Glu	Leu	Thr	Lys	Arg	Leu	Glu	His	Asp	Phe	Gly	Lys	Val	Thr	Lys	Gln		
450						455						460					
Glu	Val	Lys	Asp	Phe	Phe	Arg	Trp	Ala	Ser	Asp	His	Val	Thr	Glu	Val		
465						470						475					
Thr	His	Glu	Phe	Tyr	Val	Arg	Lys	Gly	Gly	Ala	Arg	Lys	Arg	Pro	Ala		
485						490						495					
Pro	Asn	Asp	Ala	Asp	Ile	Ser	Glu	Pro	Lys	Arg	Ala	Cys	Pro	Ser	Val		
500						505						510					
Ala	Gln	Pro	Ser	Thr	Ser	Asp	Ala	Glu	Ala	Pro	Val	Asp	Tyr	Ala	Asp		
515						520						525					
Arg	Tyr	Gln	Asn	Lys	Cys	Ser	Arg	His	Val	Gly	Met	Asn	Leu	Met	Leu		
530						535						540					
Phe	Pro	Cys	Arg	Gln	Cys	Glu	Arg	Met	Asn	Gln	Asn	Val	Asp	Ile	Cys		
545						550						555					
Phe	Thr	His	Gly	Val	Met	Asp	Cys	Ala	Glu	Cys	Phe	Pro	Val	Ser	Glu		
565						570						575					
Ser	Gln	Pro	Val	Ser	Val	Val	Arg	Lys	Arg	Thr	Tyr	Gln	Lys	Leu	Cys		
580						585						590					

-continued

Pro Ile His His Ile Met Gly Arg Ala Pro Glu Val Ala Cys Ser Ala
 595 600 605

Cys Glu Leu Ala Asn Val Asp Leu Asp Asp Cys Asp Met Glu Gln
 610 615 620

<210> SEQ ID NO 273
 <211> LENGTH: 623
 <212> TYPE: PRT
 <213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 273

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
 1 5 10 15

Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
 20 25 30

Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
 35 40 45

Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
 50 55 60

Val Gln Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
 65 70 75 80

Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Leu His Val Leu Val Glu
 85 90 95

Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
 100 105 110

Arg Glu Lys Leu Val Gln Thr Ile Tyr Arg Gly Val Glu Pro Thr Leu
 115 120 125

Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
 130 135 140

Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
 145 150 155 160

Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Glu Tyr Ile
 165 170 175

Ser Ala Cys Leu Asn Leu Ala Glu Arg Lys Arg Leu Val Ala Gln His
 180 185 190

Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Leu Asn
 195 200 205

Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
 210 215 220

Met Glu Leu Val Gly Trp Leu Val Asp Arg Gly Ile Thr Ser Glu Lys
 225 230 235 240

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
 245 250 255

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
 260 265 270

Ile Met Ala Leu Thr Lys Ser Ala Pro Asp Tyr Leu Val Gly Pro Ser
 275 280 285

Leu Pro Ala Asp Ile Lys Thr Asn Arg Ile Tyr Arg Ile Leu Glu Leu
 290 295 300

Asn Gly Tyr Asp Pro Ala Tyr Ala Gly Ser Val Phe Leu Gly Trp Ala
 305 310 315 320

Gln Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala

-continued

325					330					335					
Thr	Thr	Gly	Lys	Thr	Asn	Ile	Ala	Glu	Ala	Ile	Ala	His	Ala	Val	Pro
			340						345				350		
Phe	Tyr	Gly	Cys	Val	Asn	Trp	Thr	Asn	Glu	Asn	Phe	Pro	Phe	Asn	Asp
		355					360					365			
Cys	Val	Asp	Lys	Met	Val	Ile	Trp	Trp	Glu	Glu	Gly	Lys	Met	Thr	Ala
	370					375					380				
Lys	Val	Val	Glu	Ser	Ala	Lys	Ala	Ile	Leu	Gly	Gly	Ser	Lys	Val	Arg
385						390					395				400
Val	Asp	Gln	Lys	Cys	Lys	Ser	Ser	Ala	Gln	Ile	Asp	Pro	Thr	Pro	Val
			405						410					415	
Ile	Val	Thr	Ser	Asn	Thr	Asn	Met	Cys	Ala	Val	Ile	Asp	Gly	Asn	Ser
		420					425						430		
Thr	Thr	Phe	Glu	His	Gln	Gln	Pro	Leu	Gln	Asp	Arg	Met	Phe	Lys	Phe
		435					440					445			
Glu	Leu	Thr	Arg	Arg	Leu	Glu	His	Asp	Phe	Gly	Lys	Val	Thr	Lys	Gln
450						455					460				
Glu	Val	Lys	Glu	Phe	Phe	Arg	Trp	Ala	Ser	Asp	His	Val	Thr	Glu	Val
465						470					475				480
Ala	His	Glu	Phe	Tyr	Val	Arg	Lys	Gly	Gly	Ala	Ser	Lys	Arg	Pro	Ala
			485					490						495	
Pro	Asp	Asp	Ala	Asp	Ile	Ser	Glu	Pro	Lys	Arg	Ala	Cys	Pro	Ser	Val
			500					505					510		
Ala	Asp	Pro	Ser	Thr	Ser	Asp	Ala	Glu	Gly	Ala	Pro	Val	Asp	Phe	Ala
		515					520					525			
Asp	Arg	Tyr	Gln	Asn	Lys	Cys	Ser	Arg	His	Ala	Gly	Met	Ile	Gln	Met
530						535					540				
Leu	Phe	Pro	Cys	Lys	Thr	Cys	Glu	Arg	Met	Asn	Gln	Asn	Phe	Asn	Ile
545						550					555				560
Cys	Phe	Thr	His	Gly	Val	Arg	Asp	Cys	Leu	Glu	Cys	Phe	Pro	Gly	Val
			565					570						575	
Ser	Glu	Ser	Gln	Pro	Val	Val	Arg	Lys	Lys	Thr	Tyr	Arg	Lys	Leu	Cys
			580					585					590		
Ala	Ile	His	His	Leu	Leu	Gly	Arg	Ala	Pro	Glu	Ile	Ala	Cys	Ser	Ala
		595				600						605			
Cys	Asp	Leu	Val	Asn	Val	Asp	Leu	Asp	Asp	Cys	Val	Ser	Glu	Gln	
	610					615					620				

<210> SEQ ID NO 274

<211> LENGTH: 625

<212> TYPE: PRT

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 274

Met	Pro	Gly	Phe	Tyr	Glu	Ile	Val	Ile	Lys	Val	Pro	Ser	Asp	Leu	Asp
1				5					10					15	
Glu	His	Leu	Pro	Gly	Ile	Ser	Asp	Ser	Phe	Val	Asn	Trp	Val	Ala	Glu
		20						25					30		
Lys	Glu	Trp	Glu	Leu	Pro	Pro	Asp	Ser	Asp	Met	Asp	Arg	Asn	Leu	Ile
		35					40					45			
Glu	Gln	Ala	Pro	Leu	Thr	Val	Ala	Glu	Lys	Leu	Gln	Arg	Asp	Phe	Leu
	50					55					60				

-continued

Val 65	Gln	Trp	Arg	Arg	Val 70	Ser	Lys	Ala	Pro	Glu 75	Ala	Leu	Phe	Phe	Val 80
Gln	Phe	Glu	Lys	Gly	Glu	Ser	Tyr	Phe	His	Leu	His	Val	Leu	Val	Glu
			85					90					95		
Thr	Thr	Gly	Val	Lys	Ser	Met	Val	Leu	Gly	Arg	Phe	Leu	Ser	Gln	Ile
		100						105					110		
Arg	Glu	Lys	Leu	Gly	Pro	Asp	His	Leu	Pro	Ala	Gly	Ser	Ser	Pro	Thr
	115					120						125			
Leu	Pro	Asn	Trp	Phe	Ala	Val	Thr	Lys	Asp	Ala	Val	Met	Ala	Pro	Ala
	130					135					140				
Gly	Gly	Asn	Lys	Val	Val	Asp	Glu	Cys	Tyr	Ile	Pro	Asn	Tyr	Leu	Leu
145				150						155					160
Pro	Lys	Thr	Gln	Pro	Glu	Leu	Gln	Trp	Ala	Trp	Thr	Asn	Met	Glu	Glu
			165						170					175	
Tyr	Ile	Ser	Ala	Cys	Leu	Asn	Leu	Ala	Glu	Arg	Lys	Arg	Leu	Val	Ala
		180						185					190		
Gln	His	Leu	Thr	His	Val	Ser	Gln	Thr	Gln	Glu	Gln	Asn	Lys	Glu	Asn
		195					200					205			
Leu	Asn	Pro	Asn	Ser	Asp	Ala	Pro	Val	Ile	Arg	Ser	Lys	Thr	Ser	Ala
	210					215					220				
Arg	Tyr	Met	Glu	Leu	Val	Gly	Trp	Leu	Val	Asp	Arg	Gly	Ile	Thr	Ser
225				230						235					240
Glu	Lys	Gln	Trp	Ile	Gln	Glu	Asp	Gln	Ala	Ser	Tyr	Ile	Ser	Phe	Asn
			245						250					255	
Ala	Ala	Ser	Asn	Ser	Arg	Ser	Gln	Ile	Lys	Ala	Ala	Leu	Asp	Asn	Ala
			260					265					270		
Gly	Lys	Ile	Met	Ala	Leu	Thr	Lys	Ser	Ala	Pro	Asp	Tyr	Leu	Val	Gly
		275					280					285			
Pro	Ser	Leu	Pro	Ala	Asp	Ile	Thr	Gln	Asn	Arg	Ile	Tyr	Arg	Ile	Leu
	290					295					300				
Ala	Leu	Asn	Gly	Tyr	Asp	Pro	Ala	Tyr	Ala	Gly	Ser	Val	Phe	Leu	Gly
305				310						315					320
Trp	Ala	Gln	Lys	Lys	Phe	Gly	Lys	Arg	Asn	Thr	Ile	Trp	Leu	Phe	Gly
			325						330					335	
Pro	Ala	Thr	Thr	Gly	Lys	Thr	Asn	Ile	Ala	Glu	Ala	Ile	Ala	His	Ala
		340						345					350		
Val	Pro	Phe	Tyr	Gly	Cys	Val	Asn	Trp	Thr	Asn	Glu	Asn	Phe	Pro	Phe
		355					360					365			
Asn	Asp	Cys	Val	Asp	Lys	Met	Val	Ile	Trp	Trp	Glu	Glu	Gly	Lys	Met
	370					375					380				
Thr	Ala	Lys	Val	Val	Glu	Ser	Ala	Lys	Ala	Ile	Leu	Gly	Gly	Ser	Lys
385				390						395					400
Val	Arg	Val	Asp	Gln	Lys	Cys	Lys	Ser	Ser	Ala	Gln	Ile	Asp	Pro	Thr
			405						410					415	
Pro	Val	Ile	Val	Thr	Ser	Asn	Thr	Asn	Met	Cys	Ala	Val	Ile	Asp	Gly
		420						425					430		
Asn	Ser	Thr	Thr	Phe	Glu	His	Gln	Gln	Pro	Leu	Gln	Asp	Arg	Met	Phe
		435					440					445			
Lys	Phe	Glu	Leu	Thr	Arg	Arg	Leu	Glu	His	Asp	Phe	Gly	Lys	Val	Thr
	450					455					460				
Lys	Gln	Glu	Val	Lys	Glu	Phe	Phe	Arg	Trp	Ala	Ser	Asp	His	Val	Thr

-continued

465	470	475	480
Glu Val Ala His Glu Phe Tyr Val Arg Lys Gly Gly Ala Ser Lys Arg	485	490	495
Pro Ala Pro Asp Asp Ala Asp Lys Ser Glu Pro Lys Arg Ala Cys Pro	500	505	510
Ser Val Ala Asp Pro Ser Thr Ser Asp Ala Glu Gly Ala Pro Val Asp	515	520	525
Phe Ala Asp Arg Tyr Gln Asn Lys Cys Ser Arg His Ala Gly Met Leu	530	535	540
Gln Met Leu Phe Pro Cys Lys Thr Cys Glu Arg Met Asn Gln Asn Phe	545	550	555
Asn Ile Cys Phe Thr His Gly Val Arg Asp Cys Ser Glu Cys Phe Pro	565	570	575
Gly Val Ser Glu Ser Gln Pro Val Val Arg Lys Arg Thr Tyr Arg Lys	580	585	590
Leu Cys Ala Ile His His Leu Leu Gly Arg Ala Pro Glu Ile Ala Cys	595	600	605
Ser Ala Cys Asp Leu Val Asn Val Asp Leu Asp Asp Cys Val Ser Glu	610	615	620
Gln			
625			

<210> SEQ ID NO 275

<211> LENGTH: 663

<212> TYPE: PRT

<213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 275

Met Arg Ser Tyr Tyr Glu Val Ile Val Gln Leu Pro Asn Asp Val Glu	1	5	10	15
Ser Gln Val Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Ile Thr Ser	20	25	30	
Arg Glu Trp Thr Leu Pro Glu Asp Ala Asp Trp Asp Leu Asp Gln Val	35	40	45	
Asp Gln Val Gln Leu Thr Leu Gly Asp Lys Ile Gln Arg Glu Ile Arg	50	55	60	
Asn His Trp Gly Thr Met Ala Lys Glu Pro Asp Phe His Tyr Phe Ile	65	70	75	80
Gln Leu Glu Gln Gly Glu Val Phe Phe His Leu His Val Leu Leu Glu	85	90	95	
Thr Cys Ser Val Lys Pro Met Val Leu Gly Arg Tyr Ile Arg His Ile	100	105	110	
Gln Gln Lys Ile Val Ser Lys Val Tyr Cys Gly His Glu Pro Ala Met	115	120	125	
Glu Gly Trp Met Arg Val Thr Lys Thr Lys Asn Phe Gly Gly Ala Asn	130	135	140	
Lys Val Arg Ala Glu Ser Tyr Ile Pro Ala Tyr Leu Ile Pro Lys Gln	145	150	155	160
Gln Pro Glu Val Gln Trp Ala Trp Thr Asn Val Pro Glu Tyr Ile Lys	165	170	175	
Ala Cys Leu His Arg Glu Leu Arg Ala Ser Leu Ala Arg Leu His Phe	180	185	190	

-continued

Glu	Glu	Ala	Gly	Leu	Ser	Gln	Ser	Lys	Glu	Asn	Leu	Ala	Arg	Thr	Ala
		195					200					205			
Asp	Gly	Ala	Pro	Val	Ile	Ala	Thr	Arg	Val	Ser	Lys	Arg	Tyr	Met	Glu
	210					215					220				
Leu	Val	Asp	Trp	Leu	Val	Glu	Lys	Gly	Ile	Thr	Thr	Glu	Lys	Glu	Trp
	225				230					235					240
Leu	Leu	Glu	Asn	Arg	Glu	Ser	Phe	Arg	Ser	Phe	Gln	Ala	Ser	Ser	Asn
			245						250					255	
Ser	Ala	Arg	Gln	Ile	Lys	Thr	Ala	Leu	Gln	Gly	Ala	Ile	Gln	Glu	Met
			260					265					270		
Leu	Leu	Thr	Lys	Thr	Ala	Glu	Asp	Tyr	Leu	Val	Gly	Lys	Glu	Pro	Val
		275					280					285			
Ser	Asp	Asp	Glu	Ile	Arg	Gln	Asn	Arg	Ile	Tyr	Lys	Ile	Leu	Glu	Leu
	290					295					300				
Asn	His	Tyr	Asp	Pro	Ala	Tyr	Val	Gly	Ser	Ile	Leu	Val	Gly	Trp	Cys
	305				310					315					320
Gln	Lys	Lys	Trp	Gly	Lys	Arg	Asn	Thr	Leu	Trp	Leu	Phe	Gly	Pro	Ala
			325						330					335	
Thr	Thr	Gly	Lys	Thr	Asn	Ile	Ala	Glu	Ala	Ile	Ala	His	Ala	Val	Pro
			340					345					350		
Phe	Tyr	Gly	Cys	Val	Asn	Trp	Thr	Asn	Glu	Asn	Phe	Pro	Phe	Asn	Asp
		355					360					365			
Cys	Val	Glu	Lys	Met	Ile	Ile	Trp	Trp	Glu	Glu	Gly	Lys	Met	Thr	Ala
	370					375					380				
Lys	Val	Val	Glu	Thr	Ala	Lys	Ala	Ile	Leu	Gly	Gly	Ser	Arg	Val	Arg
	385				390					395					400
Val	Asp	Gln	Lys	Cys	Lys	Ala	Ser	Val	Pro	Ile	Glu	Pro	Thr	Pro	Val
			405						410					415	
Ile	Ile	Thr	Ser	Asn	Thr	Asn	Met	Cys	Tyr	Val	Ile	Asp	Gly	Asn	Thr
		420						425					430		
Thr	Thr	Phe	Glu	His	Lys	Gln	Pro	Leu	Glu	Asp	Arg	Met	Phe	Lys	Leu
		435					440					445			
Glu	Leu	Leu	Thr	Arg	Leu	Pro	Asp	Asp	Phe	Gly	Lys	Val	Thr	Lys	Gln
	450					455					460				
Glu	Val	Arg	Gln	Phe	Phe	Arg	Trp	Ser	Gln	Asp	His	Leu	Thr	Pro	Val
	465				470					475					480
Ile	Pro	Glu	Phe	Leu	Val	Arg	Lys	Ala	Glu	Ser	Arg	Lys	Arg	Pro	Ala
			485						490					495	
Pro	Ser	Gly	Glu	Gly	Tyr	Ile	Ser	Pro	Thr	Lys	Arg	Pro	Ala	Leu	Ala
		500						505					510		
Glu	Gln	Gln	Gln	Ala	Ser	Glu	Ser	Ala	Glu	Pro	Val	Pro	Thr	Arg	Tyr
		515						520				525			
Arg	Ile	Lys	Cys	Ser	Lys	His	Cys	Gly	Met	Asp	Lys	Met	Leu	Phe	Pro
	530					535					540				
Cys	Gln	Ile	Cys	Glu	Ser	Met	Asn	Arg	Asn	Ile	Asn	Ile	Cys	Ala	Ile
	545				550					555					560
His	Lys	Thr	Thr	Glu	Cys	Lys	Glu	Cys	Phe	Pro	Glu	Tyr	Gly	Asp	Lys
			565					570					575		
Asp	Thr	Val	Pro	Glu	Leu	Pro	Pro	Cys	Thr	Glu	His	Asn	Val	Ser	Arg
		580						585					590		
Cys	Tyr	Gln	Cys	His	Ser	Gly	Glu	Leu	Tyr	Arg	Val	Thr	Ser	Asp	Ser

-continued

595	600	605
Asp Glu Lys Pro Ala Pro	Glu Ser Asp Glu Gly Thr	Glu Pro Ser Tyr
610	615	620
Ala Pro Cys Thr Ile His	His Leu Met Gly Lys Ser	Arg Gly Leu Val
625	630	635
Ser Cys Ala Ala Cys Arg	Leu Lys Asn Ser Thr	Leu His Asp Asp Leu
645	650	655
Asp Asp Gly Asp Leu Glu Gln		
660		

<210> SEQ ID NO 276
 <211> LENGTH: 662
 <212> TYPE: PRT
 <213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 276

Met Arg Ser Tyr Tyr Glu Val Ile Val Gln Leu Pro Asn Asp Val Glu
1 5 10 15
Ser Gln Val Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Ile Thr Ser
20 25 30
Arg Glu Trp Thr Leu Pro Glu Asp Ala Asp Trp Asp Leu Asp Gln Val
35 40 45
Asp Gln Val Gln Leu Thr Leu Gly Asp Lys Ile Gln Arg Glu Ile Arg
50 55 60
Thr His Trp Gly Thr Met Ala Lys Glu Pro Asp Phe His Tyr Phe Ile
65 70 75 80
Gln Leu Glu Gln Gly Glu Val Phe Phe His Leu His Val Leu Leu Glu
85 90 95
Thr Cys Ser Val Lys Pro Met Val Leu Gly Arg Tyr Ile Arg His Ile
100 105 110
Gln Gln Lys Ile Val Ser Lys Val Tyr Cys Ala Thr Ser Leu Arg Trp
115 120 125
Lys Asp Gly Cys Val Val Thr Lys Thr Lys Asn Phe Gly Gly Ala Asn
130 135 140
Lys Val Arg Ala Glu Ser Tyr Ile Pro Ala Tyr Leu Ile Pro Lys Gln
145 150 155 160
Gln Pro Glu Val Gln Trp Ala Trp Thr Asn Val Pro Glu Tyr Ile Lys
165 170 175
Ala Cys Leu His Arg Glu Leu Arg Ala Ser Leu Ala Arg Leu His Phe
180 185 190
Glu Glu Ala Gly Val Ser Gln Ser Lys Glu Asn Leu Ala Arg Thr Ala
195 200 205
Asp Gly Ala Pro Val Met Pro Thr Arg Val Ser Lys Arg Tyr Met Glu
210 215 220
Leu Val Asp Trp Leu Val Glu Lys Gly Ile Thr Thr Glu Lys Glu Trp
225 230 235 240
Leu Leu Glu Asn Arg Glu Ser Phe Arg Ser Phe Gln Ala Ser Ser Asn
245 250 255
Ser Ala Arg Gln Ile Lys Thr Ala Leu Gln Gly Ala Ile Gln Glu Met
260 265 270
Leu Leu Thr Lys Thr Ala Glu Asp Tyr Leu Val Gly Lys Asp Pro Val
275 280 285

```
<210> SEQ ID NO 277
<211> LENGTH: 610
<212> TYPE: PRT
```

-continued

<213> ORGANISM: Bovine adeno-associated virus

<400> SEQUENCE: 277

```

Met Ala Thr Phe Tyr Glu Val Ile Val Arg Val Pro Phe Asp Val Glu
 1           5           10           15
Glu His Leu Pro Gly Ile Ser Asp Asn Phe Val Asp Trp Val Thr Gly
          20           25           30
Gln Ile Trp Glu Leu Pro Pro Glu Ser Asp Leu Asn Leu Thr Leu Ile
          35           40           45
Glu Gln Pro Gln Leu Thr Val Ala Asp Arg Ile Arg Arg Val Phe Leu
          50           55           60
Tyr Glu Trp Asn Lys Phe Ser Lys Gln Glu Ser Lys Phe Phe Val Gln
 65           70           75           80
Phe Glu Lys Gly Ser Glu Tyr Phe His Leu His Thr Leu Val Glu Thr
          85           90           95
Ser Gly Ile Ser Ser Met Val Leu Gly Arg Tyr Val Ser Gln Ile Arg
          100          105          110
Ala Gln Leu Val Lys Val Val Phe Gln Asn Ile Glu Pro Arg Ile Asn
          115          120          125
Asp Trp Val Ala Ile Thr Lys Val Lys Lys Gly Gly Ala Asn Lys Val
          130          135          140
Val Asp Ser Gly Tyr Ile Pro Ala Tyr Leu Leu Pro Lys Val Gln Pro
145          150          155          160
Glu Leu Gln Trp Ala Trp Thr Asn Leu Glu Glu Tyr Lys Leu Ala Ala
          165          170          175
Leu Asn Leu Glu Glu Arg Lys Arg Leu Val Ala Gln Phe Gln Leu Glu
          180          185          190
Ser Ser Gln Arg Ser Gln Glu Ala Ser Ser Gln Arg Asp Val Ser Ala
          195          200          205
Asp Pro Val Ile Lys Ser Lys Thr Ser Gln Lys Tyr Met Ala Leu Val
          210          215          220
Ser Trp Leu Val Glu His Gly Ile Thr Ser Glu Lys Gln Trp Ile Gln
225          230          235          240
Glu Asn Gln Glu Ser Tyr Leu Ser Phe Asn Ser Thr Gly Asn Ser Arg
          245          250          255
Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Ser Lys Ile Met Ser Leu
          260          265          270
Thr Lys Ser Ala Ser Asp Tyr Leu Val Gly Gln Thr Val Pro Glu Asp
          275          280          285
Ile Ser Glu Asn Arg Ile Trp Gln Ile Phe Asp Leu Asn Gly Tyr Asp
          290          295          300
Pro Ala Tyr Ala Gly Ser Val Leu Tyr Gly Trp Cys Thr Arg Ala Phe
305          310          315          320
Gly Lys Arg Asn Thr Val Trp Leu Tyr Gly Pro Ala Thr Thr Gly Lys
          325          330          335
Thr Asn Ile Ala Glu Ala Ile Ser His Thr Val Pro Phe Tyr Gly Cys
          340          345          350
Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp Cys Val Glu Lys
          355          360          365
Met Leu Ile Trp Trp Glu Glu Gly Lys Met Thr Ser Lys Val Val Glu
          370          375          380

```

-continued

Pro	Ala	Lys	Ala	Ile	Leu	Gly	Gly	Ser	Arg	Val	Arg	Val	Asp	Gln	Lys
385					390					395					400
Cys	Lys	Ser	Ser	Val	Gln	Val	Asp	Ser	Thr	Pro	Val	Ile	Ile	Thr	Ser
				405					410					415	
Asn	Thr	Asn	Met	Cys	Val	Val	Val	Asp	Gly	Asn	Ser	Thr	Thr	Phe	Glu
			420					425					430		
His	Gln	Gln	Pro	Leu	Glu	Asp	Arg	Met	Phe	Arg	Phe	Glu	Leu	Met	Arg
			435				440					445			
Arg	Leu	Pro	Pro	Asp	Phe	Gly	Lys	Ile	Thr	Lys	Gln	Glu	Val	Lys	Asp
	450					455					460				
Phe	Phe	Ala	Trp	Ala	Lys	Val	Asn	Gln	Val	Pro	Val	Thr	His	Glu	Phe
465					470					475					480
Met	Val	Pro	Lys	Lys	Val	Ala	Gly	Thr	Glu	Arg	Ala	Glu	Thr	Ser	Arg
				485					490					495	
Lys	Arg	Pro	Leu	Asp	Asp	Val	Thr	Asn	Thr	Asn	Tyr	Lys	Ser	Pro	Glu
			500					505					510		
Lys	Arg	Ala	Arg	Leu	Ser	Val	Val	Pro	Glu	Thr	Pro	Arg	Ser	Ser	Asp
		515						520				525			
Val	Pro	Val	Glu	Pro	Ala	Pro	Leu	Arg	Pro	Leu	Asn	Trp	Ser	Ser	Arg
	530					535					540				
Tyr	Glu	Cys	Arg	Cys	Asp	Tyr	His	Ala	Lys	Phe	Asp	Ser	Val	Thr	Gly
545					550					555					560
Glu	Cys	Asp	Glu	Cys	Glu	Tyr	Leu	Asn	Arg	Gly	Lys	Asn	Gly	Cys	Ile
				565					570					575	
Phe	His	Asn	Ala	Thr	His	Cys	Gln	Ile	Cys	His	Ala	Val	Pro	Pro	Trp
			580					585					590		
Glu	Lys	Glu	Asn	Val	Ser	Asp	Phe	Asn	Asp	Phe	Asp	Asp	Cys	Asn	Lys
			595				600					605			
Glu	Gln														
	610														

<210> SEQ ID NO 278

<211> LENGTH: 613

<212> TYPE: PRT

<213> ORGANISM: Bat adeno-associated virus

<400> SEQUENCE: 278

Met	Glu	Phe	Tyr	Ser	Ile	Val	Leu	Arg	Leu	Pro	Gly	Asp	Phe	Asp	Ser
1				5					10					15	
Glu	Val	Pro	Gly	Leu	Gln	Asp	Ser	Phe	Tyr	Lys	Trp	Ile	Ser	Gly	Pro
		20						25					30		
Arg	Arg	Glu	Leu	Pro	Glu	Trp	Ser	Asp	Leu	Asp	Pro	Gly	Gln	Ile	Glu
		35				40						45			
Ser	Ala	Tyr	Gln	Ile	Leu	Ala	Asp	Lys	Leu	Val	Arg	Glu	Phe	Ala	Gln
	50				55					60					
Lys	Trp	Ala	Ala	Phe	Ser	Glu	Asp	Pro	Arg	Ala	Pro	Tyr	Phe	Ala	Gln
65					70				75					80	
Leu	Glu	Lys	Gly	Arg	Glu	Asn	Phe	His	Val	His	Val	Leu	Ala	Ser	Ser
			85					90					95		
Lys	Lys	Val	Gly	Ser	Phe	Val	Val	Gly	Arg	Tyr	Val	Arg	Lys	Met	Arg
		100						105					110		
Gln	His	Leu	Val	Asp	Val	Val	Phe	Arg	Lys	Cys	Glu	Pro	Val	Asp	Ala
		115					120					125			

-continued

Asp	Trp	Leu	Gln	Val	Gln	Lys	Ser	Gly	Asn	His	Lys	Ser	Asn	Glu	Ile
130						135					140				
Lys	Asp	Glu	Gly	Phe	Ile	Pro	Ala	Tyr	Leu	Leu	Pro	Lys	Arg	Gln	Ser
145					150					155					160
Glu	Leu	Gln	Trp	Ala	Trp	Thr	Asn	Ile	Glu	Lys	Tyr	Glu	Arg	Ala	Thr
				165					170					175	
Leu	Ser	Val	Ala	Glu	Arg	Ala	Arg	Leu	Val	Glu	Glu	Trp	Lys	Arg	Ser
			180					185					190		
Leu	Ala	Ala	Glu	Glu	Ser	Asp	Pro	Ala	Glu	Pro	Glu	Arg	Arg	Pro	Arg
			195				200					205			
Lys	Ser	Thr	Lys	Ser	Ala	Ser	Glu	Tyr	Met	Ala	Leu	Val	Arg	Trp	Leu
	210					215					220				
Val	Asp	Asn	Gly	Ile	Ala	Thr	Glu	Arg	Glu	Trp	Met	Arg	Glu	Asp	Ser
225					230					235					240
Asp	Gly	Tyr	Leu	Ser	Tyr	Asn	Ala	Thr	Gly	Ala	Thr	Arg	Ala	Gln	Ile
				245					250					255	
Lys	Ala	Ala	Leu	Asp	Asn	Ala	Ala	Arg	Ile	Met	Val	Asn	Thr	Lys	Thr
			260					265					270		
Ala	Ala	Asp	Tyr	Leu	Val	Gly	Arg	Asn	Pro	Pro	Leu	Asp	Val	Glu	Asp
			275				280					285			
Asn	Arg	Ile	Tyr	Arg	Leu	Phe	Arg	Met	Asn	Gly	Tyr	Asp	Pro	Ala	Tyr
	290					295					300				
Ala	Gly	Ser	Val	Leu	Leu	Gly	Trp	Cys	Arg	Thr	Gly	Phe	Gly	Lys	Arg
305					310					315					320
Asn	Thr	Val	Trp	Leu	Phe	Gly	Pro	Ala	Thr	Thr	Gly	Lys	Thr	Asn	Leu
				325					330					335	
Ala	Glu	Ala	Ile	Ser	His	Ser	Val	Pro	Phe	Tyr	Gly	Cys	Val	Asn	Trp
			340					345					350		
Thr	Asn	Glu	Asn	Phe	Pro	Phe	Asn	Asp	Cys	Val	Asp	Lys	Met	Ile	Ile
		355					360					365			
Trp	Trp	Glu	Glu	Gly	Lys	Met	Thr	Ser	Lys	Val	Val	Glu	Ser	Ala	Lys
	370					375					380				
Ala	Ile	Leu	Gly	Gly	Ser	Lys	Val	Arg	Val	Asp	Gln	Lys	Cys	Lys	Asn
385					390					395					400
Ser	Gln	Gln	Ile	Glu	Pro	Thr	Pro	Val	Ile	Ile	Thr	Ser	Asn	Thr	Asn
				405					410					415	
Met	Cys	Glu	Val	Val	Asp	Gly	Asn	Ser	Thr	Thr	Phe	Glu	His	Arg	Gln
			420					425					430		
Pro	Leu	Glu	Asp	Arg	Met	Phe	Lys	Phe	Glu	Leu	Thr	Val	Arg	Leu	Gln
			435				440					445			
Pro	Thr	Phe	Gly	Lys	Ile	Thr	Lys	Gln	Glu	Val	Arg	Glu	Phe	Phe	Lys
					450		455				460				
Trp	Ala	Glu	Leu	Asn	Ala	Val	Asp	Val	Glu	Tyr	Asp	Phe	Leu	Val	Arg
465					470					475					480
Lys	Ile	Asn	Gln	Ser	Asp	Thr	Gly	Gly	Gly	Val	Lys	Arg	Gly	Ala	Glu
				485					490					495	
Pro	Thr	Lys	Asp	Glu	Pro	Pro	Ala	Lys	Arg	Val	Phe	Phe	Tyr	Gly	Ala
			500					505					510		
Thr	Ser	Glu	Gly	Glu	Asp	Val	Arg	Glu	Gly	Ala	Pro	Gly	Glu	Ser	Asp
			515				520					525			

-continued

Ser Val Asn Phe Ala Glu Arg Tyr Val Ser Lys Cys Ser Lys His Leu
530 535 540

Ser Trp Ser Asn Met Arg Tyr Pro Cys Arg Ala Cys Glu Arg Met Asn
545 550 555 560

Ala Asp Val Asn Val Cys Thr Pro His Gly Cys Arg Asp Cys Pro Glu
565 570 575

Cys Phe Pro Arg Pro Ala Pro Val Pro Ile Ala Glu His Asp Leu Cys
580 585 590

Leu Ala Pro Ile Glu Asp Ser Asp Phe Tyr Val Gly Cys Ile Asp Asp
595 600 605

Val Asn Lys Glu Gln
610

<210> SEQ ID NO 279
 <211> LENGTH: 562
 <212> TYPE: PRT
 <213> ORGANISM: Snake parvovirus

<400> SEQUENCE: 279

Met Ala Phe Tyr Glu Val Val Phe Arg Leu Pro Arg Asp Asn Asn Asn
1 5 10 15

Leu Leu Asp Glu Asp Arg Tyr Gln Pro Glu Leu Lys Glu Glu Asp Asp
20 25 30

Trp Pro Glu Glu Tyr Leu Thr Ser Glu Asp Ala Ser Phe Ile Gly Leu
35 40 45

Ala Tyr Ala Val Leu Ser Glu Ile Arg Arg Phe Phe Gly Lys Glu Leu
50 55 60

Gln Trp Phe Ala Gln Val Glu Trp Cys Pro Thr Ala Gly Tyr His Met
65 70 75 80

His Val Leu Leu Asn His Pro Lys Leu Ser Asn Gln Thr Tyr Gly Arg
85 90 95

Lys Val Asn Glu Leu Ala Cys Arg Ile Val Asp Thr Phe Gly Leu Ile
100 105 110

Asn Pro Glu Glu Val Ile Ser Thr His Tyr Val Lys Ser Asn Tyr Gly
115 120 125

His Lys Lys Val Arg Val Ile His Leu Glu Ser Tyr Leu Lys Asn Tyr
130 135 140

Phe Phe Arg Lys Thr Leu Ala Pro Pro Asn Tyr Thr Glu Glu Gly Asp
145 150 155 160

Tyr Lys Arg Glu Glu Glu Val Val Leu Trp Ala Phe Thr Asn Ile Val
165 170 175

Ala Trp Lys Pro Phe Val Arg Asn Leu Ile Lys Arg Ser Glu Leu Ala
180 185 190

Thr Val Pro Lys Gln Pro Glu Asn Pro Ala Gly Asp Gly Pro Ala Pro
195 200 205

Arg Val Thr Ala Gly Thr Arg His Phe Met Glu Thr Ile Asp Trp Leu
210 215 220

Val Lys His Gly Ile Thr Thr Glu Arg Glu Phe Cys His Ala Asn Arg
225 230 235 240

Pro Leu Tyr Leu Ser Met Leu Ala Ser Thr Ser Gly Ala Gly Gln Ile
245 250 255

Lys Arg Ala Leu Asp Gln Ala Lys His Met Met Thr Ser Thr Met Ser
260 265 270

-continued

Ala Glu Asp Tyr Leu Thr Thr Glu Glu Asp Val Ile Glu Pro Pro Thr
 275 280 285

Glu Asn Arg Ile Tyr Lys Ile Met Lys Leu Asn Arg Tyr Asp Pro Glu
 290 295 300

Leu Ala Ala Ala Leu Phe Tyr Gly Trp Thr Cys Lys Asn Phe Gly Lys
 305 310 315 320

Arg Asn Thr Ile Trp Leu Tyr Gly Pro Ala Thr Thr Gly Lys Thr Ile
 325 330 335

Ile Ala Gln Ala Ile Ala His Ala Val Lys Leu Phe Ala Gly Val Asn
 340 345 350

Trp Thr Asn Glu Asn Phe Pro Phe Cys Asn Cys Pro Gly Lys Leu Leu
 355 360 365

Ile Trp Trp Glu Glu Gly Lys Met Thr Asn Lys Met Val Glu Thr Ala
 370 375 380

Lys Cys Ile Leu Gly Gly Ser Ala Val Pro Val Asp Ile Lys Gly Lys
 385 390 395 400

Pro Ala Glu Met Cys Pro Gln Thr Pro Cys Ile Ile Thr Ser Asn Thr
 405 410 415

Asn Met Cys Gln Val Tyr Asp Gly Asn Ser Ser Ser Phe Glu His Gln
 420 425 430

Glu Pro Leu Glu Glu Arg Met Phe Met Phe Arg Leu Asn Thr Lys Leu
 435 440 445

Pro Ser Thr Phe Gly Lys Ile Thr Glu Glu Glu Val Lys Gln Phe Ile
 450 455 460

Thr Trp Gly Arg Ser Leu Lys Val Gln Val Pro His Gln Phe Arg Val
 465 470 475 480

Pro Thr Thr Gly Glu Tyr Lys Arg Pro Ala Pro Glu Ala Lys Ala His
 485 490 495

Ser Ser Asp Glu Pro Pro Lys Glu Lys Val Ala Arg Ile Asp Asp Ser
 500 505 510

Leu Thr Arg Tyr Val Asn Asn Ile Asp Glu Ser Ala Thr Ser Arg Glu
 515 520 525

Met Phe Leu Glu Ile Ala Asn Thr Asn Gln Cys Met Leu His His Cys
 530 535 540

Phe Ser Cys Thr Glu Cys Tyr Pro Glu Leu Leu Asp Asp Met Asp Lys
 545 550 555 560

Glu Gln

<210> SEQ ID NO 280
 <211> LENGTH: 4718
 <212> TYPE: DNA
 <213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 280

```

ttgcccactc cctctctgcg cgctcgctcg ctcggtgggg cctgcggacc aaaggtccgc    60
agacggcaga gctctgctct gccggcccca ccgagcgagc gacgcgcag agagggagtg    120
ggcaactcca tctactagggg taatcgcgaa gcgcctccca cgctgccgcg tcagcgctga    180
cgtaaattac gtcatagggg agtggctctg tattagctgt cacgtgagtg cttttgcgac    240
atthttgcgac accacgtggc catttagggt atatatggcc gagtgcgca gcaggatctc    300
cattttgacc gcgaatttg aacgagcagc agccatgccg ggcttctacg agatcgtgat    360

```

-continued

caaggtgccc agcgacctgg acgagcacct gccgggcatt tctgactcgt ttgtgagctg	420
ggtggccgag aaggaatggg agctgcccc ggattctgac atggatctga atctgattga	480
gcaggcacccc ctgacctggtg ccgagaagct gcagcgcgac ttcttggtcc aatggcgccc	540
cgtgagtaag gccccggagg cctctcttct tgttcagttc gagaagggcg agtcctactt	600
ccacctccat attctggtgg agaccacggg ggtcaaatcc atggtgctgg gccgcttctt	660
gagtcagatt agggacaagc tgggtgcagac catctaccgc gggatcgagc cgacctgcc	720
caactggttc gcggtgacca agacgcgtaa tggcgccgga ggggggaaca agtggtgga	780
cgagtgtac atcccaact acctctgcc caagactcag cccgagctgc agtgggcgtg	840
gactaacatg gaggagtata taagcgctg tttgaacctg gccgagcgca aacggctcgt	900
ggcgagcac ctgacctacg tcagccagac ccaggagcag aacaaggaga atctgaacct	960
caattctgac gcgcctgtca tccggtcaaa aacctccgcg cgctacatgg agctggtcgg	1020
gtggctggtg gaccggggca tcacctccga gaagcagtg atccaggagg accaggcctc	1080
gtacatctcc ttcaacgcg cttccaactc gcggtcccag atcaaggccg ctctggacaa	1140
tgcgggcaag atcatggcgc tgaccaaact gcgccccgac tacctggtag gccccgctcc	1200
gccccgggac attaaaacca accgcactca ccgcactcct gagctgaacg gctacgaacc	1260
tgcctacgcc ggctccgtct ttctcggtg ggcacagaaa aggttcggga agcgcaacac	1320
catctggctg tttgggcccg ccaccacggg caagaccaac atcgcggaag ccatcgccca	1380
cgccgtgccc ttctacggct gcgtcaactg gaccaatgag aactttccct tcaatgattg	1440
cgtcgacaag atggtgatct ggtgggagga gggcaagatg acggccaagg tcgtggagtc	1500
cggcaaggcc attctcggcg gcagcaaggt gcgctgggac caaaagtga agtcgtccgc	1560
ccagatcgac cccacccccg tgatcgtcac ctccaacacc aacatgtgcg ccgtgattga	1620
cgggaacagc accaccttgc agcaccagca gccgttgag gaccggatgt tcaaatttga	1680
actcaccgc cgtctggagc atgactttgg caaggtgaca aagcaggaag tcaaagagtt	1740
cttcgctgg gcgcaggatc acgtgaccga ggtggcgcat gagttctacg tcagaaaggg	1800
tggagccaac aaaagaccg ccccgatga cgcggataaa agcgagccca agcgggctg	1860
cccctcagtc gcggtccat cgacgtcaga cgcggaagga gctccggtgg actttgccga	1920
caggtagcaa aacaaatgtt ctctgcacgc gggcatgctt cagatgctgt ttccctgcaa	1980
gacatgcgag agaataaatc agaatttcaa catttgcttc acgcacggga cgagagactg	2040
ttcagagtgc tttcccgccg tgcagaatc tcaaccggtc gtcagaaaga ggacgtatcg	2100
gaaactctgt gccattcatc atctgctggg gcgggctccc gagattgctt gctcggcctg	2160
cgatctggtc aacgtggacc tggatgactg tgtttctgag caataaatga cttaaacag	2220
gtatggctgc cgatggttat cttccagatt ggctcgagga caacctctct gagggcattc	2280
gcgagtggg ggaactgaaa cctggagccc cgaagcccaa agccaaccag caaaagcagg	2340
acgacggccg gggctctggtg cttcctggct acaagtacct cggacccttc aacggactcg	2400
acaaggggga gcccgtaac gcggcgagc cagcgccctc cgagcacgac aaggcctacg	2460
accagcagct caaagcgggt gacaatccgt acctgcggta taaccacgcc gacgcgagt	2520
ttcaggagcg tctgcaagaa gatacgtctt ttgggggcaa cctcgggcga gcagtcttcc	2580
aggccaagaa gcgggtcttc gaacctctcg gtctggttga ggaaggcgct aagacggctc	2640

-continued

```

ctggaaagaa acgtccggta gagcagtcgc cacaagagcc agactcctcc tcgggcatcg 2700
gcaagacagg ccagcagccc gctaaaaaga gactcaattt tggtcagact ggcgactcag 2760
agtcagtcct cgatccacaa cctctcggag aacctccagc aacccccgct gctgtgggac 2820
ctactacaat ggcttcaggc ggtggcgcac caatggcaga caataacgaa ggcgcgcagc 2880
gagtgggtaa tgctcagga aattggcatt gcgattccac atggctgggc gacagagtca 2940
tcaccaccag caccgcgacc tgggccttgc ccacctacaa taaccacctc tacaagcaaa 3000
tctccagtgc ttcaacgggg gccagcaacg acaaccacta ctteggctac agcaccacct 3060
gggggtatct tgatttcaac agattccact gccacttttc accacgtgac tggcagcgac 3120
tcatcaacaa caattgggga ttccggccca agagactcaa cttcaaacctc ttcaacatcc 3180
aagtcaagga ggtcacgacg aatgatggcg tcacaacatc cgctaataac cttaccagca 3240
cgggtcaagt cttctcggac tcggagtacc agcttccgta cgtcctcggc tctgcgcacc 3300
agggctgcct ccctccgttc ccggcggacg tgttcatgat tccgcaatac ggctacctga 3360
cgctcaacaa tggcagccaa gccgtgggac gttcactctt ttactgcctg gaatatttcc 3420
cttctcagat gctgagaacg ggcaacaact ttaccttcag ctacaccttt gaggaagtgc 3480
ctttccacag cagctacgcg cacagccaga gcctggaccg gctgatgaat cctctcatcg 3540
accaatacct gtattacctg aacagaactc aaaatcagtc cggaagtgcc caaaacaagg 3600
acttgctgtt tagccgtggg tctccagctg gcatgtctgt tcagccaaa aactggctac 3660
ctggaccctg ttatcggcag cagcgcgttt ctaaaacaaa aacagacaac aacaacagca 3720
attttacctg gactggtgct tcaaaatata acctcaatgg gcgtgaatcc atcatcaacc 3780
ctggcactgc tatggcctca cacaaagacg acgaagacaa gttctttccc atgagcggtg 3840
tcatgatttt tggaaaagag agcgccggag cttcaaacac tgcattggac aatgtcatga 3900
ttacagacga agaggaaatt aaagccacta acctgtgggc caccgaaaga tttgggaccg 3960
tggcagtcac tttccagagc agcagcacag acctgcgac cggagatgtg catgctatgg 4020
gagcattacc tggcatggtg tggcaagata gagacgtgta cctgcagggt cccatttggg 4080
ccaaaattcc tcacacagat ggacactttc acccgtctcc tcttatgggc ggetttggac 4140
tcaagaacct gcctcctcag atcctcatca aaaacacgcc tgttctcgcg aatcctccgg 4200
cggagttttc agctacaaag tttgcttcat tcatcaccca atactccaca ggacaagtga 4260
gtgtggaaat tgaatgggag ctgcagaaag aaaacagcaa gcgtggaat cccgaagtgc 4320
agtacacatc caattatgca aaatctgcc aacgttgattt tactgtggac aacaatggac 4380
tttatactga gcctcgcccc attggcaccg gttaccttac ccgtcccctg taattacgtg 4440
ttaatcaata aaccgggtga ttcggttcag ttgaactttg gtctcctgta cttcttatct 4500
tatcggttac catggttata gcttacacat taactgcttg gttgcgcttc gcgataaaag 4560
acttacgtca tcgggttacc cctagtgatg gagttgcccc ctccctctct gcgcgctcgc 4620
tcgctcggtg gggcctcgcg accaaaggtc cgcagacggc agagctctgc tctgcgggcc 4680
ccaccgagcg agcgagcgcg cagagaggga gtgggcaa 4718

```

<210> SEQ ID NO 281

<211> LENGTH: 4679

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

-continued

<400> SEQUENCE: 281

ttggccactc cctctctgcg cgctcgcgcg ctcactgagg ccgggcgacc aaaggtegcc	60
cgacgccccg gctttgcccc ggccggcctca gtgagcgagc gagcgcgag agaggggagt	120
gccaaactcca tcaactagggg ttcttgaggg ggtggagtcg tgacgtgaat tacgtcatag	180
ggttaggagg gtctgttatt agaggtcacg tgagtgtttt gcgacatttt gcgacaccat	240
gtggtcacgc tgggtattta agcccagtg agcacgcagg gtctccattt tgaagcggga	300
ggtttgaaac gcgagccgcc atgccgggtg ttacgagat tgtgattaag gtccccagcg	360
accttgacga gcactcgccc ggcatctctg acagctttgt gaactgggtg gccgagaagg	420
aatgggagtt gccgccagat tctgacatgg atctgaatct gattgagcag gcacccctga	480
ccgtggccga gaagctgcag cgcgactttc tgacggaatg gcgccgtgtg agtaaggccc	540
cggaggccct tttctttgtg caatttgaga agggagagag ctacttccac atgcacgtgc	600
tcgtggaaac caccgggggtg aaatccatgg ttttgggacg tttctgagt cagattcgcg	660
aaaaactgat tcagagaatt taccgcgga tcgagccgac tttgcaaac tggttcgcg	720
tcacaaagac cagaaatggc gccggaggcg ggaacaagg ggtggatgag tgctacatcc	780
ccaattactt gctcccaaaa acccagcctg agctccagtg ggctgggact aatatggaac	840
agtattttaag cgctgtttt aatctcacgg agcgtaaac gttgggtggc cagcatctga	900
cgcacgtgtc gcagacgcag gagcagaaca aagagaatca gaatcccaat tctgatgcgc	960
cggatgatcag atcaaaaact tcagccaggt acatggagct ggtcgggtgg ctcgtggaca	1020
aggggattac ctcggagaag cagtggatcc agggaggacca ggctcatac atctccttca	1080
atgcggcctc caactcgcg tcccaaatca aggtgcctt ggacaatgcg ggaaagatta	1140
tgagcctgac taaaaccgcc cccgactacc tgggtggcca gcagcccggtg gaggacattt	1200
ccagcaatcg gatattataa attttggaa taaacgggta cgatcccaaa tatgcggctt	1260
ccgtctttct gggatgggccc acgaaaaagt tcggcaagag gaacaccatc tggtgtttg	1320
ggcctgcaac tacccgggaag accaacatcg cggaggccat agcccacact gtgcccttct	1380
acgggtgcgt aaactggacc aatgagaact ttcccttcaa cgaactgtgc gacaagatgg	1440
tgatctggtg ggaggagggg aagatgaccg ccaaggctcg ggagtcggcc aaagccattc	1500
tcggaggaag caaggtgcgc gtggaccaga aatgcaagtc ctcggcccag atagaccga	1560
ctcccgatg cgtcacctcc aacaccaaca tgtgcgccgt gattgacggg aactcaacga	1620
ccttcgaaca ccagcagccg ttgcaagacc ggatgttcaa atttgaactc acccgccgtc	1680
tggatcatga ctttgggaag gtcaccaagc aggaagtcaa agacttttcc cgggtggcaa	1740
aggatcacgt ggttgagggt gagcatgaat tctacgtcaa aaagggtgga gccaaagaaa	1800
gaccgcgccc cagtgcgcga gatataagt agcccaaacg ggtgcgcgag tcagttgcgc	1860
agccatcgac gtcagacgcg gaagcttcga tcaactacgc agacaggtag caaaacaaat	1920
gttctcgta cgtgggcatg aatctgatgc tgtttccctg cagacaatgc gagagaatga	1980
atcagaattc aaatatctgc ttcactcacg gacagaaaga ctgttttagag tgctttcccg	2040
tgtcagaatc tcaaccggtt tctgtcgtca aaaaggcgta tcagaaactg tgetacattc	2100
atcatatcat gggaaaaggc ccagacgctt gcaactgcctg cgatctgggtc aatgtggatt	2160
tggatgactg catctttgaa caataaatga tttaaatcag gtatggctgc cgatggttat	2220

-continued

cttcagatt ggctcgagga cactctctct gaaggaataa gacagtgggtg gaagtc meta 2280
cctggccac caccac meta gccgcagag cggcataagg acgacagcag gggctctgtg 2340
cttctgggt acaagtacct cggaccttc aacggactcg acaagggaga gccgg meta 2400
gaggcagacg ccgcgccct cgagcacgac aaagcctacg accggcagct cgacagcga 2460
gacaacctcg acctcaagta caaccacgcc gacgcggagt ttcaggagcg ccttaaagaa 2520
gatacgtctt ttgggggcaa cctcggacga gcagtcttc aggc meta 2580
gaacctctgg gcctgggtga ggaacctgtt aagacggctc cgg meta 2640
gagcactctc ctgtggagcc agactcctcc tcgggaaccg gaaaggcggg ccagcagcct 2700
gcaag meta gattgaattt tggtcagact ggagacgcag actcagtaacc tgaccccgag 2760
cctctcggac agccaccagc agccccctct ggtctgggaa ctaatacgat ggctacaggc 2820
agtggcgcac caatggcaga caataacgag ggcgcgcagc gagtgggtaa ttcctcggga 2880
aattggcatt gcgattccac atggatgggc gacagagtca tcaccaccag caccgaacc 2940
tgggccccgc ccacct meta caaccacctc taca meta 3000
tcgaacgaca atcactactt tggtacagc accccttggg ggtat meta 3060
ttccactgcc acttttcacc acgtgactgg caaagactca tcaaca meta 3120
cgacccaaga gactcaactt caagctcttt aacattcaag tcaagagggt cagcagaaat 3180
gacggtacga cgacgattgc caataacctt accagcagcg ttcagggtgt tactgactcg 3240
gagtaccagc tcccgtaagt cctcggctcg gcgcatacag gatgcctccc gccgttccca 3300
gcagacgtct tcattgggtgc acagtatgga tacctcacc tcga meta 3360
gtaggacgct cttcatttta ctgcctggag tactttcctt ctcagatgct gcgtaccgga 3420
aacaacttta ccttcagcta cacttttgag gacgttcctt tccacagcag ctacgctcac 3480
agccagagtc tggaccgtct catgaatcct ctcatcgacc agtacctgta ttacttgagc 3540
aga meta ctccaagtgg aaccaccacg cagtcaaggc ttcagtttcc tcaggccgga 3600
gcgagt meta ttcgggacca gtctaggaac tggcttcctg gacctgtta ccgccagcag 3660
cgagtatcaa agacatctgc ggataacaac aacagtgaat actcgtggac tggagctacc 3720
aagtaccacc tcaatggcag agactctctg gtgaatccgg gcccgccat ggcaagccac 3780
aaggcagatg aagaaaagtt ttttctcag agcgggggtc tcactcttgga gaagcaaggc 3840
tcagag meta caaatgtgga catt meta 3900
acaaccaatc ccgtggctac ggagcagtat ggttctgtat ctaccaacct ccagagaggc 3960
aacagacaag cagctaccgc agatgtcaac acacaaggcg ttcttccagg catggctctgg 4020
caggacagag atgtgtacct tcagggggccc atctgggcaa agattccaca cacggacgga 4080
cattttcacc cctctccct catgggtgga ttcggactta aacacctcc tccacagatt 4140
ctcatcaaga acacccgggt acctgcgaat ccttcgacca ccttcagtgc ggcaagttt 4200
gcttccttca tcacacagta ctccacggga caggtcagcg tggagatcga gtgggagctg 4260
cagaagg meta acagcaaacg ctggaatccc gaaattcagt acacttccaa ctaca meta 4320
tctgttaatg tggactttac tgtggacact aatggcgtgt attcagagcc tcgccccatt 4380
ggcaccagat acctgactcg taatctgtaa ttgcttgta atcaataaac cgtttaattc 4440
gtttcagttg aactttggtc tctgcgtatt tctttcttat ctagtttcca tggtacgta 4500

-continued

gataagtagc atggcggggtt aatcattaac tacaaggaac ccctagtgat ggagttggcc	4560
actccctctc tgcgcgctcg ctcgctcact gaggcgggc gaccaaaggt cgcccgacgc	4620
ccgggctttg cccggggggc ctcagtgagc gagcgagcgc gcagagaggg agtggccaa	4679

<210> SEQ ID NO 282

<211> LENGTH: 4726

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 282

ttggccactc cctctatgcg cactcgtcgc ctcggtgggg cctggcgacc aaaggtcgcc	60
agacggacgt gctttgcacg tccggcccca ccgagcgagc gagtgcgcat agagggagtg	120
gccaactcca tctactagagg tatggcagtg acgtaacgcg aagcgcgaga agcgagacca	180
cgcctaccag ctgcgtcagc agtcaggtga cctttttgcg acagtttgcg acaccacgtg	240
gccgctgagg gtatatattc tgcagtgagc gaaccaggag ctccattttg accgcgaaat	300
ttgaacgagc agcagccatg ccgggggttct acgagattgt cctgaaggtc ccgagtgacc	360
tggacgagcg cctgccgggc atttctaact cgtttgttaa ctgggtggcc gagaaggaat	420
gggacgtgcc gccggattct gacatggatc cgaatctgat tgagcaggca cccctgaccg	480
tggccgaaaa gcttcagcgc gagtctctgg tggagtggcg ccgctgagtg aaggccccgg	540
aggccctctt ttttgtccag ttcgaaaagg gggagaccta cttccacctg cagtgctga	600
ttgagaccat cgggggtcaa tccatggtgg tcggccgcta cgtgagccag attaaagaga	660
agctgggtgac ccgcatctac cgcgggggtcg agccgcagct tccgaactgg ttcgcggtga	720
ccaaaacgcg aaatggcgcc gggggcgga acaagggtgt ggacgactgc tacatcccca	780
actacctgct ccccaagacc cagcccgagc tccagtgggc gtggactaac atggaccagt	840
atttaagcgc ctgtttgaat ctcgctggagc gtaaacggct ggtggcgag catctgacgc	900
acgtgtcgca gacgcaggag cagaacaaag agaatacaga cccaattct gacgcgccgg	960
tcatcaggtc aaaaacctca gccaggtaca tggagctggt cgggtggctg gtggaccgcg	1020
ggatcacgtc agaaaagcaa tggattcagg aggaccaggc ctctgacatc tcttcaacg	1080
ccgcctccaa ctcgctggtc cagatcaagg ccgctgtgga caatgcctcc aagatcatga	1140
gcctgacaaa gacgggtccg gactacctgg tgggcagcaa cccgcggag gacattacca	1200
aaaatcggat ctaccaaactc ctggagctga acgggtacga tccgcagtac gcggcctccg	1260
tcttctctgg ctgggcgcaa aagaagtctg ggaagaggaa caccatctgg ctctttgggc	1320
cggccacgac gggtaaaacc aacatcgcg aagccatcgc ccacgccgtg cccttctacg	1380
gctgcgtaaa ctggaccaat gagaactttc ctttcaacga ttgcgtcgac aagatggtga	1440
tctggtggga ggagggcaag atgacggcca aggtcgtgga gagcgccaag gccattctgg	1500
gcggaagcaa ggtgcgcgtg gacaaaaagt gcaagtcacg ggcccagatc gaacccactc	1560
ccgtgatcgt cacctccaac accaacaatgt gcgctgtgat tgacgggaac agcaccacct	1620
tcgagcatca gcagccgctg caggaccgga tgtttgaatt tgaacttacc cgccgtttgg	1680
accatgactt tgggaaggte accaaacagg aagtaaagga ctttttccgg tgggcttccg	1740
atcacgtgac tgacgtggct catgagttct acgtcagaaa ggggtggagct aagaaacgcc	1800
ccgcctccaa tgacgcggat gtaagcgagc caaaacggga gtgcacgtca cttgcgcagc	1860

-continued

cgacaacgtc agacgcggaa gcaccggcgg actacgcgga caggtaccaa aacaaatgtt	1920
ctcgtcacgt gggcatgaat ctgatgcttt ttccctgtaa aacatgcgag agaataatc	1980
aaatttccaa tgtctgtttt acgcatggtc aaagagactg tggggaatgc ttccctggaa	2040
tgtcagaatc tcaaccggtt tctgtcgtca aaaagaagac ttatcagaaa ctgtgtccaa	2100
ttcatcatat cctgggaagg gcacccgaga ttgcctgttc ggctgcgat ttggccaatg	2160
tggacttgga tgactgtgtt tctgagcaat aaatgactta aaccaggatg ggctgctgac	2220
ggttatcttc cagattggct cgaggacaac ctttctgaag gcattcgtga gtgggtgggt	2280
ctgaaacctg gagtccctca acccaaagcg aaccaacaac accaggacaa ccgtcggggg	2340
cttgtgcttc cgggttccaa atacctcgga cccggtaacg gactcgacaa aggagagccg	2400
gtcaacgagg cggacgcggc agccctcgaa cagacaaaag cttacgacca gcagctcaag	2460
gccggtgaca acccgtaact caagtacaac cagcgcgacg ccgagtttca ggagcgtctt	2520
caagaagata cgtcttttgg gggcaacctt ggcagagcag tcttcagggc caaaaagagg	2580
atccttgagc ctcttggtct ggttgaggaa gcagctaaaa cggctcctgg aaagaagggg	2640
gctgtagatc agtctctca ggaaccggac tcatcatctg gtgttgcaa atcgggcaaa	2700
cagctcgcca gaaaaagact aaatttcggt cagactggag actcagagtc agtcccagac	2760
cctcaacctc tcggagaacc accagcagcc cccacaagtt tgggatctaa tacaatggct	2820
tcaggcggtg gcgcaccaat ggcagacaat aacgagggtg ccgatggagt gggtaattcc	2880
tcaggaaatt ggcattgcga ttcccaatgg ctgggcgaca gactcatcac caccagcacc	2940
agaacctggg cctgcccac ttacaacaac catctctaca agcaaatctc cagccaatca	3000
ggagcttcaa acgacaacca ctactttggc tacagcacc cttgggggta tttgacttt	3060
aacagattcc actgccactt ctcaccacgt gactggcagc gactcattaa caacaactgg	3120
ggattccggc ccaagaaact cagcttcaag ctcttcaaca tccaagttag aggggtcacg	3180
cagaacgatg gcacgacgac tattgccaat aaccttacca gcacgggtca agtggttacg	3240
gactcggagt atcagctccc gtacgtgctc gggtcggcgc accaaggctg tctcccgcgc	3300
tttcacggc acgtcttcat ggtccctcag tatggatacc tcacctgaa caacggaagt	3360
caagcggtag gacgctcatc cttttactgc ctggagtact tcccttcgca gatgctaagg	3420
actggaaata acttccaatt cagctatacc ttcgaggatg taccttttca cagcagctac	3480
gctcacagcc agagtttggg tcgcttgatg aatcctctta ttgatcagta tctgtactac	3540
ctgaacagaa cgcaaggaa aacctctgga acaaccaacc aatcacggct gctttttagc	3600
caggctgggc ctcagtctat gtctttgcag gccagaaatt ggctacctgg gccctgctac	3660
cggcaacaga gactttcaaa gactgctaac gacaacaaca acagtaactt tccctggaca	3720
gcggccagca aatatcatct caatggccgc gactcgtgg tgaatccagg accagctatg	3780
gccagtcaca aggacgatga agaaaaattt ttccctatgc acggcaatct aatatttggc	3840
aaagaaggga caacggcaag taacgcagaa ttagataatg taatgattac ggatgaagaa	3900
gagattcgta ccaccaatcc tgtggcaaca gagcagtatg gaactgtggc aaataacttg	3960
cagagctcaa atacagctcc cagcactgga actgtcaatc atcagggggc cttacctggc	4020
atggtgtggc aagatcgtga cgtgtacctt caaggaccta tctgggcaaa gattcctcac	4080
acggatggac actttcatcc ttctcctctg atgggaggct ttggactgaa acatccgcct	4140

-continued

```

cctcaaatca tgatcaaaaa tactccggtta cgggcaaatc ctccgacgac ttccagcccg 4200
gccaaagtgtg cttcatttat cactcagtag tccactggac aggtcagcgt ggaaattgag 4260
tgaggagctac agaagaaaaa cagcaaacgt tggaaatccag agattcagta cacttccaac 4320
tacaacaagt ctgttaaatgt ggactttact gtagacacta atggtgttta tagtgaacct 4380
cgccctattg gaaccccggtta tctcacacga aacttggtgaa tcctgggttaa tcaataaacc 4440
gtttaattcg ttccagttga actttggctc ttgtgcactt ctttatcttt atcttggttc 4500
catggctact gcgtagataa gcagcggcct cgggcgcttg cgttcgcggg tttaacaactg 4560
ctgggttaata ttttaactctc gccataacctc tagtgatgga gttggccact ccctctatgc 4620
gcactcgctc gctcgggtggg gcctggcgac caaaggctgc cagacggacg tgctttgcac 4680
gtccggcccc accgagcgag cgagtgcgca tagagggagt ggccaa 4726

```

<210> SEQ ID NO 283

<211> LENGTH: 4767

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 283

```

ttggccactc cctctatgcg cgctcgctca ctcaactcggc cctggagacc aaaggtctcc 60
agactgccgg cctctggcgg gcaggggccga gtgagtgcgc gagcgcgcac agaggagtg 120
gccaaactcca tcacttaggt ttgccactg acgtcaatgt gacgtcctag ggtagggag 180
gtccctgtat tagcagtcac gtgagtgtcg tatttcgagg agcgtagcgg agcgcatacc 240
aagctgccac gtcacagcca cgtgggtcgt ttgcgacagt ttgcgacacc atgtggtcag 300
gagggtatat aaccgcgagt gagccagcga ggagctccat ttgcccgcg aattttgaac 360
gagcagcagc catgccgggg ttctacgaga tcgtgctgaa ggtgcccagc gacctggacg 420
agcacctgcc cggcatttct gactcttttg tgagctgggt ggccgagaag gaatgggagc 480
tgccgccgga ttctgacatg gacttgaatc tgattgagca ggcacccctg accgtggccg 540
aaaagctgca acgcgagttc ctggctcaggt ggcgcgcgct gagtaaggcc ccggaggccc 600
tcttctttgt ccagttcgag aaggggggaca gctacttcca cctgcacatc ctggtggaga 660
ccgtgggctg caaatccatg gtggtgggccc gctacgtgag ccagattaaa gagaagctgg 720
tgaccgcgat ctaccgcggg gtcgagccgc agcttccgaa ctggttcgcg gtgaccaaga 780
cgcgtaatgg cgcgggaggc ggaacaaggc tgggtggacg ctgtacatc cccaactacc 840
tgctcccaa gaccagccc gagctccagt gggcgtggac taacatggac cagtatataa 900
gcgctgtttt gaatctcgcg gagcgtaaac ggctgggtgg gcagcatctg acgcacgtgt 960
cgagagcgca ggagcagaac aaggaaaacc agaaccctaa ttctgacgag ccggtcatca 1020
ggtcaaaaac ctccgccagg tacatggagc tggtcgggtg gctggtggac cgcgggatca 1080
cgtcagaaaa gcaatggatc caggaggacc aggcgtccta catctccttc aacgcgcct 1140
ccaactcgcg gtcacaaatc aaggccgcgc tggacaatgc ctccaaaatc atgagcctga 1200
caaagacggc tccggactac ctgggtgggccc agaaccgcc ggaggacatt tccagcaacc 1260
gcatctaccg aatctcgag atgaacgggt acgatccgca gtacgcggcc tccgtcttcc 1320
tggtctgggc gcaaaagaag ttccggaaga ggaacaccat ctggctcttt gggccggcca 1380
cgacgggtaa aaccaacatc gcggaagcca tcgcccacgc cgtgcccttc tacggtcgcg 1440

```

-continued

tgaactggac	caatgagaac	tttccgttca	acgattgcgt	cgacaagatg	gtgatctggt	1500
gggaggaggg	caagatgacg	gccaaaggtcg	tagagagcgc	caaggccatc	ctgggcggaa	1560
gcaaggtgcg	cgtggacca	aagtgcaagt	catcggccca	gatcgacca	actcccgta	1620
tcgtcacctc	caacaccaac	atgtgcgcgg	tcategacgg	aaactcgacc	acettcgagc	1680
accaacaacc	actccaggac	cggatgttca	agttcgagct	caccaagcgc	ctggagcacg	1740
actttggcaa	ggtcaccaag	caggaagtca	aagacttttt	ccggtgggcg	tcagatcacg	1800
tgaccgaggt	gactcacgag	ttttacgtca	gaaagggtgg	agctagaaa	aggcccgccc	1860
ccaatgacgc	agatataagt	gagcccaagc	gggcctgtcc	gtcagttgcg	cagccatcga	1920
cgtcacagcg	ggaagctccg	gtggactacg	cggacaggt	ccaaaacaaa	tgttctcgtc	1980
acgtgggtat	gaatctgatg	ctttttccct	gccggcaatg	cgagagaatg	aatcagaatg	2040
tggacatttg	cttcacgcac	ggggctcatg	actgtgccga	gtgcttcccc	gtgtcagaat	2100
ctcaaccctg	gtctgtcgtc	agaaagcgga	cgtatcagaa	actgtgtccg	attcatcaca	2160
tcattggggag	ggcgcccgag	gtggcctgct	cggcctgcga	actggccaat	gtggacttgg	2220
atgactgtga	catggaacaa	taaatgactc	aaaccagata	tgactgacgg	ttaccttcca	2280
gattggctag	aggacaacct	ctctgaaggc	gttcgagagt	gggtggcgct	gcaacctgga	2340
gccccctaac	ccaaggcaaa	tcaacaacat	caggacaacg	ctcgggggtct	tgtgcttccg	2400
ggttacaaat	acctcggacc	cggcaacgga	ctcgacaagg	gggaaccctg	caacgcagcg	2460
gacgcggcag	ccctcgagca	cgacaaggcc	tacgaccagc	agctcaaggc	cggtgacaac	2520
ccctacctca	agtacaacca	cgccgacgcg	gagttccagc	agcggttcca	gggcgcacaca	2580
tcgtttgggg	gcaacctcgg	cagagcagtc	ttccaggcca	aaaagagggt	tcttgaacct	2640
cttggtctgg	ttgagcaagc	gggtgagacg	gctcctggaa	agaagagacc	gttgattgaa	2700
tccccccagc	agccccgactc	ctccacgggt	atcggcaaaa	aaggcaagca	gccggctaaa	2760
aagaagctcg	ttttcgaaga	cgaaactgga	gcaggcgacg	gacccccctga	gggatcaact	2820
tccggagcca	tgtctgatga	cagtgaagtg	cgtgcagcag	ctggcggagc	tgcatcgag	2880
ggcggacaag	gtgccgatgg	agtgggtaat	gcctcgggtg	attggcattg	cgattccacc	2940
tggcttgagg	gccacgtcac	gaccaccagc	accagaacct	gggtcttgcc	cacctacaac	3000
aaccacctct	acaagcgact	cggagagagc	ctgcagtcca	acacctacaa	cggattctcc	3060
acccccctgg	gatactttga	cttcaaccgc	ttccactgcc	actttctacc	acgtgactgg	3120
cagcgactca	tcaacaacaa	ctggggcatg	cgacccaaag	ccatgcgggt	caaaatcttc	3180
aacatccagg	tcaaggaggt	cacgacgtcg	aacggcgaga	caacgggtggc	taataacctt	3240
accagcacgg	ttcagatctt	tgcggactcg	tcgtacgaac	tgccgtacgt	gatggatgcg	3300
ggtaagagg	gcagcctgcc	tccttttccc	aacgacgtct	ttatggtgcc	ccagtaaggc	3360
tactgtggac	tggtagccgg	caacacttcg	cagcaacaga	ctgacagaaa	tgctttctac	3420
tgccctggagt	actttctctc	gcagatgctg	cggactggca	acaactttga	aattacgtac	3480
agttttgaga	agggtgcctt	ccactcgatg	tacgcgcaca	gccagagcct	ggaccggctg	3540
atgaacctc	tcategacca	gtacctgtgg	ggactgcaat	cgaccaccac	cggaaaccac	3600
ctgaatgccg	ggactgccac	caccaacttt	accaagctgc	ggcctaccaa	cttttccaac	3660
tttaaaaaga	actggctgcc	cgggccttca	atcaagcagc	agggcttctc	aaagactgcc	3720

-continued

```

aatcaaaact acaagatccc tgccaccggg tcagacagtc tcataaata cgagacgcac 3780
agcactctgg acggaagatg gactgacctg acccccgac ctccaatggc cacggctgga 3840
cctgcggaaca gcaagttcag caacagccag ctcatctttg cggggcctaa acagaacggc 3900
aacacggcca cgtaccogg gactctgac ttcacctctg aggaggagct ggcagccacc 3960
aacgccaccg atacggacat gtggggcaac ctacctggcg gtgaccagag caacagcaac 4020
ctgccgaccg tggacagact gacagccttg ggagccgtgc ctggaatggt ctggcaaac 4080
agagacatct actaccaggg tcccatttgg gccaaagatt ctcataccga tggacacttt 4140
caccctcac cgctgatttg tgggtttggg ctgaaacacc cgctcctca aatttttatc 4200
aagaacaccc cggtagctgc gaattctgca acgacctca gctctactcc ggtaaactcc 4260
ttcattactc agtacagcac tggccagggt tcggtgcaga ttgactggga gatccagaag 4320
gagcgggtcca aacgctggaa ccccgagggt cagtttacct ccaactacgg acagcaaac 4380
tctctgttgt gggctccoga tgcggctggg aaatacactg agcctagggt tatcggtacc 4440
cgctacctca cccaccacct gtaataacct gttaataaat aaaccggttt attcgtttca 4500
gttgaaactt ggtctccgtg tcttcttat ctatctctg ttccatgggt actgcgtaca 4560
taagcagcgg cctgcggcgc ttgcgcttcg cgggtttaca ctgccgggta atcagtaact 4620
tctggcaaac cagatgatgg agttggccac attagctatg cgcgctcgct cactcactcg 4680
gccctggaga ccaaaggtct ccagactgcc gccctctggc cggcagggcc gactgagtga 4740
gcgagcgcgc atagaggagg tggccaa 4767

```

<210> SEQ ID NO 284

<211> LENGTH: 4642

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 284

```

ctctcccccc tgcgcgcttc gctcgcctgc tggctcggtt ggggggggtg cagctcaaag 60
agctgccaga cgacggccct ctggccgtcg ccccccaaa cgagccagcg agcgagcgaa 120
cgcgacaggg gggagagtgc cacactctca agcaaggggg ttttgtaagc agtgatgtca 180
taatgatgta atgcttattg tcacgcgata gttaatgatt aacagtcatg tgatgtgttt 240
tatccaatag gaagaaagcg cgcgtatgag ttctcgcgag acttcggggg tataaaagac 300
cgagtgaacg agcccgccgc cattctttgc tctggactgc tagaggacct tcgctgccat 360
ggctaccttc tatgaagtca ttgttcgctg cccatttgac gtggaggaa atctgcctgg 420
aatttctgac agcttttggt actgggtaac tgggtcaaat tgggagctgc ctccagagtc 480
agatttaaat ttgactctgg ttgaacagcc tcagttgacg gtggctgata gaattcgccg 540
cgtgttcttg tacgagtgga acaaatcttc caagcaggag tccaaattct ttgtgcagtt 600
tgaaaaggga tctgaatatt ttcatctgca cagccttggt gagacctcg gcatctcttc 660
catggtcttc ggccgctacg tgagtcagat tcgcgcccag ctggtgaaag tggctctcca 720
gggaattgaa cccagatca acgactgggt cgccatcacc aaggtaaaga agggcggagc 780
caataagggt gtggattctg ggtatattcc cgcctacctg ctgccgaagg tccaaccgga 840
gcttcagtg gctgggacaa acctggacga gtataaattg gccgccctga atctggagga 900
gcgcaaacgg ctgcgcgcgc agtttctggc agaactctcg cagcgcctcg aggaggcggc 960

```

-continued

ttcgacgctg	gagttctcgg	ctgacccggg	catcaaaagc	aagacttccc	agaaatacat	1020
ggcgctcgtc	aactggctcg	tggagcacgg	catcacttcc	gagaagcagt	ggatccagga	1080
aaatcaggag	agctacctct	ccttcaactc	caccggcaac	tctcggagcc	agatcaaggc	1140
cgcgctcgac	aacgcgacca	aaattatgag	tctgacaaaa	agcgcggtgg	actacctcgt	1200
ggggagctcc	gttcccagag	acatttcaaa	aaacagaatc	tggcaaattt	ttgagatgaa	1260
tgggtacgac	cggcctacg	cgggatccat	cctctacggc	tgggtgtcagc	gctccttcaa	1320
caagaggaac	accgtctggc	tctacggacc	cggcacgacc	ggcaagacca	acatcgcgga	1380
ggccatcgcc	cacactgtgc	ccttttacgg	ctgcgtgaac	tggaccaatg	aaaactttcc	1440
ctttaatgac	tgtgtggaca	aaatgctcat	ttggtgggag	gagggaaaga	tgaccaacaa	1500
ggtggttgaa	tccgccaagg	ccatcctggg	gggtcaaaag	gtgcgggtcg	atcagaaatg	1560
taaatcctct	gttcaaattg	attctacccc	tgctattgta	acttccaata	caaacatgtg	1620
tgtggtgggtg	gatgggaatt	ccacgacctt	tgaacaccag	cagccgctgg	aggaccgcat	1680
gttcaaattt	gaactgacta	agcggtctcc	gccagatttt	ggcaagatta	ctaagcagga	1740
agtcaaggac	ttttttgctt	gggcaaaggt	caatcaggtg	cgggtgactc	acgagtttaa	1800
agttcccagg	gaattggcgg	gaactaaagg	ggcggagaaa	tctctaaaac	gcccactggg	1860
tgacgtcacc	aatactagct	ataaaagtct	ggagaagcgg	gccaggtctc	catttgttcc	1920
cgagacgcct	cgcagttcag	acgtgactgt	tgatcccgtc	cctctgcgac	cgctcaattg	1980
gaattcaagg	tatgattgca	aatgtgacta	tcatgctcaa	tttgacaaca	tttctaacaa	2040
atgtgatgaa	tgtgaatatt	tgaatcgggg	caaaaatgga	tgtatctgtc	acaatgtaac	2100
tcactgtcaa	atttgtcatg	ggattccccc	ctgggaaaag	gaaaacttgt	cagattttgg	2160
ggattttgac	gatgccaata	aagaacagta	aataaagcga	gtagtcatgt	cttttgttga	2220
tcaccttcca	gattggttgg	aagaagttgg	tgaaggctct	cgcgagtttt	tgggccttga	2280
agcgggcccc	ccgaaaccaa	aacccaatca	gcagcatcaa	gatcaagccc	gtggtcttgt	2340
gctgcctggt	tataactatc	tccgacccgg	aaacgggtctc	gatcgaggag	agcctgtcaa	2400
cagggcagac	gaggtcgcgc	gagagcacga	catctcgtac	aacgagcagc	ttgaggcggg	2460
agacaacccc	tacctcaagt	acaaccacgc	ggacgccgag	tttcaggaga	agctcgccga	2520
cgacacatcc	ttcgggggaa	acctcggaaa	ggcagtcctt	caggccaaga	aaagggttct	2580
cgaacctttt	ggcctggttg	aagaggggtgc	taagacggcc	cctaccggaa	agcggataga	2640
cgaccacttt	ccaaaaagaa	agaaggtctg	gaccgaagag	gactccaagc	cttccacctc	2700
gtcagacgcc	gaagctggac	ccagcggatc	ccagcagctg	caaatcccag	cccaaccagc	2760
ctcaagtttg	ggagctgata	caatgtctgc	gggaggtggc	ggccccattg	gcgacaataa	2820
ccaaggtgcc	gatggagtgg	gcaatgcctc	gggagattgg	cattgcgatt	ccacgtggat	2880
gggggacaga	gtcgtcacca	agtccacccg	aacctgggtg	ctgcccagct	acaacaacca	2940
ccagtaccga	gagatcaaaa	gcggctccgt	cgacggaagc	aacgccaacg	cctactttgg	3000
atacagcacc	ccctgggggt	actttgactt	taaccgcttc	cacagccact	ggagcccccg	3060
agactggcaa	agactcatca	acaactactg	gggtctcaga	ccccggtccc	tcagagtcaa	3120
aatcttcaac	attcaagtca	aagaggtcac	ggtgcaggac	tccaccacca	ccatcgccaa	3180
caacctcacc	tccaccgtcc	aagtgtttac	ggacgacgac	taccagctgc	cctacgtcgt	3240

-continued

```

cggcaacggg accgagggat gctgcccgc ctteccctcg caggtcttta cgctgccga 3300
gtacgggttac gcgacgctga accgcgacaa cacagaaaat cccaccgaga ggagcagctt 3360
cttctgccta gactactttc ccagcaagat gctgagaacg ggcaacaact ttgagtttac 3420
ctacaacttt gaggaggtgc ccttcactc cagcttcgct ccagtcaga acctgttcaa 3480
gctggccaac ccgctggtgg accagtactt gtaccgcttc gtgagcacia ataactctgg 3540
cggagtccag ttcaacaaga acctggcccg gagatacgcc aacacctaca aaaactgggt 3600
cccgggggcc atggggcgaa cccagggtg gaacctgggc tccggggtca accgcgccag 3660
tgtcagcgcc ttcgccacga ccaataggat ggagctcgag ggcgcgagtt accaggtgcc 3720
cccgagcgcc aacggcatga ccaacaacct ccagggcagc aacacctatg cctggagaa 3780
cactatgac ttcaacagcc agccggcgaa cccgggcacc accgccacgt acctcgagg 3840
caacatgctc atcaccagcg agagcgagac gcagccggtg aaccgcgtgg cgtacaacgt 3900
cggcgggcag atggccacca acaaccagag ctccaccact gccccgcga ccggcacgta 3960
caacctccag gaaatcgctc ccggcagcgt gtggatggag agggacgtgt acctccaagg 4020
acctatctgg gccaagatcc cagagacggg ggcgcacttt caccctctc cgcccatggg 4080
cggattcgga ctcaaacacc caccgccc atgctctc aagaacacgc ctgtgcccg 4140
aaatatcacc agcttctcgg acgtgcccgt cagcagcttc atcaccagt acagaccgg 4200
gcaggtcacc gtggagatgg agtgggagct caagaaggaa aactccaaga ggtggaaccc 4260
agagatccag tacacaaaca actacaacga ccccgagttt gtggactttg ccccgacag 4320
caccggggaa tacagaacca ccagacctat cggaacccga taccttacc gaccttta 4380
acctattcat gtcgcatacc ctcaataaac cgtgtattcg tgtcagtaaa atactgctc 4440
ttgtggtcat tcaatgaata acagcttaca acatctacaa aacctcctg cttgagagt 4500
tggcactctc cccctgtcg cgttcgctc ctcgctggc cgtttgggg ggtggcagct 4560
caaagagctg ccagacgacg gccctctggc cgtcgcccc ccaacgagc cagcgagcga 4620
gcgaacgcga caggggggag ag 4642

```

<210> SEQ ID NO 285

<211> LENGTH: 4721

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 285

```

ttggccactc cctctatgcg cgctcgctcg ctcggtgggg cctgcggacc aaaggtccgc 60
agacggcaga gctctgctct gccggcccca ccgagcgagc gagcgcgcat agagggagtg 120
gccaactcca tcactagggg tacccggaag cgctccccc cgtgcgcgt cagcgctgac 180
gtaaatcacg tcatagggga gtggtcctgt attagctgtc acgtgagtg ttttgcgaca 240
ttttgcgaca ccacgtggcc atttgaggta tatatggcg agtgagcgag caggatctcc 300
atthtgaccg cgaaatttga acgagcagca gccatgccgg gtttctacga gatcgtgatc 360
aaggtgccga gcgacctgga cgagcacctg ccgggcattt ctgactcgtt tgtgaactgg 420
gtggccgaga aggaatggga gctgccccg gattctgaca tggatctgaa tctgatcgag 480
caggcacccc tgaccgtggc cgagaagctg cagcgcgact tcctggtcca atggcgccgc 540
gtgagtaagg ccccgaggc cctgttcttt gttcagttcg agaaggcgga gagctacttc 600

```

-continued

caccttcacg	ttctggtgga	gaccacgggg	gtcaagtcca	tgggtctagg	ccgcttctcg	660
agtcagattc	gggagaagct	ggtccagacc	atctaccgcg	gggtcgagcc	cacgctgccc	720
aactgggtcg	cggtgaccaa	gacgcgtaat	ggcgccggcg	gggggaacaa	ggtggtggac	780
gagtgcctaca	tccccaacta	cctcctgccc	aagaccacgc	ccgagctgca	gtgggcgtgg	840
actaacatgg	aggagtatat	aagcgctgt	ttgaacctgg	ccgaacgcaa	acggctcgtg	900
gcgcagcacc	tgaccacgt	cagccagacg	caggagcaga	acaaggagaa	tctgaacccc	960
aattctgacg	cgcccgatg	caggtcaaaa	acctccgcgc	gctacatgga	gctggtcggg	1020
tggctggtgg	accggggcat	cacctccgag	aagcagtggg	tccaggagga	ccaggcctcg	1080
tacatctcct	tcaacgcgcg	ctccaactcg	cggctccaga	tcaaggccgc	gctggacaat	1140
gccggcaaga	tcatggcgct	gaccaaattc	gcgcccgact	acctggtggg	gccctcgctg	1200
cccgcgga	gttaaaaccaa	ccgcattctac	cgcattcctgg	agctgaacgg	gtacgatcct	1260
gcctacgcgc	gctccgtctt	tctcggtctg	gccagaaaa	agttcgggaa	gcgcaacacc	1320
atctggtctg	ttggggccgc	caccaccggc	aagaccaaca	ttgcggaagc	catcgccccc	1380
gccgtgccct	tctacggtcg	cgtaactgg	accaatgaga	actttccctt	caacgattgc	1440
gtcgacaaga	tgggtgatctg	gtgggaggag	ggcaagatga	cggccaaggt	cgtggagtcc	1500
gccaaggcca	ttctcgggcg	cagcaaggtg	cgcgtggacc	aaaagtgcga	gtcgtccgcc	1560
cagatcgacc	ccacccccgt	gatcgtcacc	tccaacacca	acatgtgcgc	cgtgattgac	1620
gggaacagca	ccaccttcga	gcaccagcag	ccgttgccag	accggatggt	caaatgtgaa	1680
ctcaccgcc	gtctggagca	cgactttggc	aaggtgacga	agcaggaagt	caaagagttc	1740
ttccgctggg	ccagtgatca	cgtgaccgag	gtggcgcatg	agttctacgt	cagaaagggc	1800
ggagccagca	aaagaccgcg	ccccgatgac	gcggatataa	gcgagcccaa	gcgggcctgc	1860
ccctcagtcg	cggatccatc	gacgtcagac	gcggaaggag	ctccggtgga	ctttgccgac	1920
aggtaccaaa	acaaatgttc	tcgtcacgcg	ggcatgattc	agatgctgtt	tccctgcaaa	1980
acgtgcgaga	gaatgaatca	gaatttcaac	atttgcttca	cacacggggt	cagagactgt	2040
ttagagtgtt	tccccggcgt	gtcagaatct	caaccggctg	tcagaaaaaa	gacgtatcgg	2100
aaactctcg	cgattcatca	tctgctgggg	cgggcgccc	agattgcttg	ctcggcctgc	2160
gacctggtca	acgtggacct	ggacgactgc	gtttctgagc	aataaatgac	ttaaaccagg	2220
tatggctgcc	gatggttatc	ttccagattg	gctcgaggac	aacctctctg	agggcattcg	2280
cgagtgggtg	gacctgaaac	ctggagcccc	gaaacccaaa	gccaaccagc	aaaagcagga	2340
caacggcccg	ggtctggtgc	ttcctggcta	caagtacctc	ggacccttca	acggactcga	2400
caagggggag	cccgtcaacg	cggcggacgc	agcggccctc	gagcacgaca	aggcctacga	2460
ccagcagctc	aaagcgggtg	acaatccgta	cctgcggtat	aaccacgcgc	acgccgagtt	2520
tcaggagcgt	ctgcaagaag	atacgtcatt	tgggggcaac	ctcgggcgag	cagtcttcca	2580
ggccaagaag	cgggttctcg	aacctctcgg	tctggttgag	gaaggcgcta	agacggctcc	2640
tgcaagaag	agaccggtag	agccgtcacc	tcagcgttcc	cccgaactct	ccacgggcat	2700
cggcaagaaa	ggccagcagc	ccgccagaaa	gagactcaat	ttcggtcaga	ctggcgactc	2760
agagtcagtc	cccgaacctc	aacctctcgg	agaacctcca	gcagcgcctc	ctagtgtggg	2820
atctggtaca	gtggctgcag	gcggtggcgc	accaatggca	gacaataacg	aaggtgccga	2880

-continued

cggagtgggt aatgcctcag gaaattggca ttgcgattcc acatggctgg gcgacagagt	2940
cattaccacc agcacccgaa cctgggcctt gccacacct aacaaccacc tctacaagca	3000
aatctccagt gaaactgcag gtagtaccac cgacaacacc tacttcggct acagcaccac	3060
ctgggggtat ttgacttta acagattcca ctgccacttc taccacgtg actggcagcg	3120
actcatcaac aacaactggg gattccggcc caagaagctg cggttcaagc tcttcaacat	3180
ccaggtaag gaggtcagc cgaatgacgg cgttacgacc atcgctaata accttaccag	3240
cacgattcag gtattctcgg actcggaata ccagctgcgg tacgtcctcg gctctgcgca	3300
ccagggtcgc ctgcctcgt tcccgcgga cgtcttcatt attcctcagt acggctacct	3360
gactctcaac aatggcagtc agtctgtggg acgttcctcc ttctactgcc tggagtactt	3420
ccccctcag atgctgagaa cgggcaacaa ctttgagttc agctacagct tcgaggacgt	3480
gcctttccac agcagctacg cacacagcca gagcctggac cggctgatga atcccccat	3540
cgaccagtac ttgtactacc tggccagAAC acagagtaac ccaggaggca cagctggcaa	3600
tcgggaactg cagttttacc agggcgggccc ttcaactatg gccgaacaag ccaagaattg	3660
gttacctgga ccttgcttcc ggcaacaaag agtctccaaa acgctggatc aaaacaacaa	3720
cagcaacttt gcttgactg gtgccaccaa atatcacctg aacggcagaa actcgttggt	3780
taatcccggc gtcgccatgg caactcacia ggacgacgag gaccgctttt tcccaccag	3840
cggagtcctg atttttggaa aaactggagc aactaacaaa actacattgg aaatgtgtt	3900
aatgacaaat gaagaagaaa ttctgcctac taatcctgta gccacggaag aatacgggat	3960
agtcagcagc aacttacaag cggctaatac tgcagcccag acacaagttg tcaacaacca	4020
gggagcctta cctggcatgg tctggcagaa ccgggacgtg tacctgcagg gtcccatctg	4080
ggccaagatt cctcacacgg atggcaactt tcacccgtct cctttgatgg gcggcttttg	4140
acttaaacat ccgcctctc agatcctgat caagaacact cccgttccc ctaatcctcc	4200
ggaggtgttt actcctgcca agtttgcttc gttcatcaca cagtacagca ccggacaagt	4260
cagcgtggaa atcgagtggg agctgcagaa ggaaaacagc aagcgtgga acccgagat	4320
tcagtacacc tccaactttg aaaagcagac tgggtgggac ttgtccgttg acagccaggg	4380
tgtttactct gagcctcgcc ctattggcac tcgttacctc acccgtaatc tgtaattgca	4440
tgttaatcaa taaaccggtt gattcgttcc agttgaactt tggctctctg tgcttcttat	4500
cttatcggtt tccatagcaa ctgggttacac attaactgct tgggtgcgct tcacgataag	4560
aacactgacg tcaccgcggg acccctagt atggagttgg ccaactccctc tatgcgcgct	4620
cgctcgctcg gtggggcctg cggaccaaag gtccgcagac ggacagagctc tgctctgccg	4680
gccccaccga gcgagcgagc gcgcatagag ggagtggcca a	4721

<210> SEQ ID NO 286

<211> LENGTH: 4393

<212> TYPE: DNA

<213> ORGANISM: Adeno-associated virus

<400> SEQUENCE: 286

cagagagggg gtggccaact ccatcactag gggtagcgcg aagcgctcc cagctgccg	60
cgtcagcgct gacgtaaatt acgtcatagg ggagtggccc tgtattagct gtcacgtgag	120
tgcttttgcg gcattttgcy acaccacgtg gccatttgag gtatatatgg ccgagtgagc	180

-continued

gagcaggatc	tccattttga	ccgcgaaatt	tgaacgagca	gcagccatgc	cgggcttcta	240
cgagatcgtg	atcaaggtgc	cgagcgacct	ggacgagcac	ctgcgggca	tttctgactc	300
gtttgtgaac	tgggtggcgc	agaaggaatg	ggagctgccc	cgggattctg	acatggatcg	360
gaatctgac	gagcaggcac	cctgacctgt	ggccgagaag	ctgcagcgcg	acttcctggt	420
ccaatggcgc	cgcgtgagta	aggccccgga	ggccctcttc	tttgttcagt	tcgagaaggg	480
cgagagctac	tttcacctgc	acgttctggt	cgagaccacg	gggttcaagt	ccatggtgct	540
aggccgcttc	ctgagtcaga	tccgggaaaa	gcttggtcca	gaccatctac	ccgcggggtc	600
gagccccacc	ttgcccaact	ggttcgcggt	gaccaaagac	gcggtaatgg	cgccggcggg	660
ggggaacaag	gtggtggacg	agtgtctacat	ccccaaactac	ctcctgcccc	agactcagcc	720
cgagctgcag	tgggcgtgga	ctaactgga	ggagtatata	agcgcgtgct	tgaacctggc	780
cgagcgcaaa	cggctcgtgg	cgcagcacct	gaccacgctc	agccagacgc	aggagcagaa	840
caaggagaat	ctgaacccca	attctgacgc	gcccgatgac	aggtcaaaaa	cctccgcgcg	900
ctatatggag	ctggtcgggt	ggtggtgga	cgggggcac	acctccgaga	agcagtggtat	960
ccaggaggac	caggcctcgt	acatctcctt	caacgcgcgc	tccaaactcgc	ggtcccagat	1020
caaggccgcg	ctggacaatg	cgggcaagat	catggcgctg	accaaateccg	cgcccgacta	1080
cctggtgggg	ccctcgtcgc	cgcgggacat	taccagaaac	cgcactctacc	gcactcctcgc	1140
tctcaacggc	tacgacctg	cctacgcgcg	ctccgtcttt	ctcggctggg	ctcagaaaaa	1200
gttcgggaaa	cgcaaacacca	tctggctggt	tggaccgcgc	accaccggca	agaccaacat	1260
tgcggaagcc	atgcgccacg	cgtgcccctt	ctacggctgc	gtcaactgga	ccaatgagaa	1320
ctttcccttc	aatgattgcg	tcgacaagat	ggtgatctgg	tgggaggagg	gcaagatgac	1380
ggccaaggtc	gtggagtccg	ccaaggccat	tctcggcggc	agcaagggtgc	gcgtggacca	1440
aaagtgcgaag	tcgtccgccc	agatcgaccc	cacccccgtg	atcgtcacct	ccaacaccaa	1500
catgtgcgcc	gtgattgacg	ggaacagcac	caccttcgag	caccagcagc	ctctccagga	1560
cggatgtttt	aagttcgaa	tcacccgcgc	tctggagcac	gactttggca	aggtgacaaa	1620
gcaggaagtc	aaagagttct	tccgctgggc	cagtgatcac	gtgaccgagg	tggcgcatga	1680
gttttacgtc	agaaagggcg	gagccagcaa	aagacccgcc	cccgatgacg	cggataaaaag	1740
cgagcccaag	cgggcctgcc	cctcagtcgc	ggatccatcg	acgtcagacg	cgggaaggagc	1800
tccggtggac	tttgccgaca	ggtaccaaaa	caaagtgtct	cgtcacgcgcg	gcatgcttca	1860
gatgctgttt	ccctgcaaaa	cgtgcgagag	aatgaatcag	aatttcaaca	tttgcttcac	1920
acacggggtc	agagactgct	cagagtgttt	ccccggcgctg	tcagaatctc	aaccggctcgt	1980
cagaaagagg	acgtatcgga	aactctgtgc	gattcatcat	ctgctggggc	gggctcccga	2040
gattgcttgc	tcggcctgcg	atctggtcaa	cgtggacctg	gatgactgtg	tttctgagca	2100
ataaatgact	taaaccaggt	atggctgccc	atggttatct	tccagattgg	ctcagaggaca	2160
acctctctga	gggcattcgc	gagtgtgtgg	cgtgaaacc	tggagccccg	aagcccaaaag	2220
ccaaccagca	aaagcaggac	gacggccggg	gtctggtgct	tcctggctac	aagtacctcg	2280
gaccttcaaa	cggactcgac	aagggggagc	cgtcaacgc	ggcggacgca	gcggccctcg	2340
agcacgacaa	ggcctacgac	cagcagctgc	aggcgggtga	caatccgtac	ctgcggtata	2400
accacgccga	cgccgagttt	caggagcgtc	tgcaagaaga	tacgtctttt	gggggcaacc	2460

-continued

```

tcgggcgagc agtcttccag gccagaagc gggttctcga acctctcggt ctggttgagg 2520
aaggcgctaa gacggctcct ggaaagaaga gaccggtaga gccatcacc cagcgttctc 2580
cagactcttc tacgggcctc ggcaagaaag gccaacagcc cgccagaaaa agactcaatt 2640
ttggtcagac tggcgactca gagtcagttc cagacctca acctctcgga gaacctccag 2700
cagcgccctc tgggttgagg cctaatacaa tggctgcagg cgggtggcgca ccaatggcag 2760
acaataacga aggcgcgcac ggagtgggta gttcctcggg aaattggcat tgcgattcca 2820
catggctggg cgacagagtc atcaccacca gcaaccgaac ctgggcccctg cccacctaca 2880
acaaccacct ctacaagcaa atctccaacg ggacatcggg aggagccacc aacgacaaca 2940
cctacttcgg ctacagcacc ccttgggggt attttgactt taacagattc cactgccact 3000
tttcaccacg tgactggcag cgactcatca acaacaactg gggattccgg cccaagagac 3060
tcagcttcaa gctcttcaac atccaggta aggaggtcac gcagaatgaa ggcaccaaga 3120
ccatcgccaa taacctcacc agcaccatcc aggtgtttac ggactcggag taccagctgc 3180
cgtaegtctt cggtctgccc caccagggct gcttgcctcc gttcccgccg gacgtgttca 3240
tgattcccca gtacggctac ctaacactca acaacggtag tcaggccgtg ggacgtcctc 3300
ccttctactg cctggaatac tttccttcgc agatgctgag aaccggcaac aacttccagt 3360
ttacttacac ctctgaggac gtgcctttcc acagcageta cgcccacagc cagagcttgg 3420
accggctgat gaatcctctg attgaccagt acctgtacta cttgtctcgg actcaaaaca 3480
caggaggcac ggcaaatcag cagactctgg gcttcagcca aggtgggcct aatacaatgg 3540
ccaatcaggc aaagaactgg ctgccaggac cctgttacgg ccaacaacgc gtctcaacga 3600
caaccgggca aaacaacaat agcaactttg cctggactgc tgggaccaa taccatctga 3660
atggaagaaa ttcattggct aatcctggca tcgctatggc aacacacaaa gacgacgagg 3720
agcgtttttt tccagtaac gggatcctga tttttggcaa acaaaatgct gccagagaca 3780
atgctgatta cagcgatgtc atgctcacca gcgaggaaga aatcaaaacc actaacctg 3840
tggctacaga ggaatacggc atcgtggcag ataacttga gcagcaaaac acggctcctc 3900
aaattggaac tgtcaacagc cagggggcct taccgggtat ggtctggcag aaccgggacg 3960
tgtacctgca ggggtccatc tgggccaaga ttcctcacac ggacggcaac ttcacccgt 4020
ctccgtgat gggcggtttt ggctgaaac atcctccgcc tcagatcctg atcaagaaca 4080
cgctgtacc tgcggatcct ccgaccacct tcaaccagtc aaagctgaac tctttcatca 4140
cgcaatacag caccggacag gtcagcgtgg aaattgaatg ggagctgcag aaggaaaaca 4200
gcaagcgtcg gaaccccgag atccagtaca cctccaacta ctacaaatct acaagtgtgg 4260
actttgctgt taatacagaa ggcgtgtact ctgaaccccg cccattggc acccgttacc 4320
tcaccgta tctgtaattg cctgttaatc aataaacggg ttgattcgtt tcagttgaac 4380
tttgtctct gcg 4393

```

<210> SEQ ID NO 287

<211> LENGTH: 4286

<212> TYPE: DNA

<213> ORGANISM: Bat adeno-associated virus

<400> SEQUENCE: 287

```

aagaaacctc taatcatcat ctgtactttg tgagggtgtg gtcgtgagaa cgggcatgcc 60

```

-continued

cgggcatgtc	ctgttaatca	ttaaccagaa	tggactttgg	acttaataag	ttaagtatta	120
atgggtgtgt	atataagtaa	gtgtatttcc	ggactcactc	attcgacttc	cggcggtcga	180
ggagcgaaga	cgcgaagcca	tggagtctta	cagcatcggt	ctccgcctgc	cggcgacttt	240
tgactcggag	gtgccggggc	tgcaggacag	cttctacaag	tggatatccg	ggccccggcg	300
ggagctgcgc	gagtggtcgg	acctggaccc	gggtcagatc	gagtcggcgt	atcagattct	360
cgccgacaag	ctggtccgcg	agtttgcgca	gaagtgggcc	gccttcagcg	aggacccgag	420
ggcgcgttac	tttgcgcagc	tgcagaagg	cggggagaac	tccacgtgc	acgtgctcgc	480
ctcctccaag	aaggtcgggt	cgttcgttgt	cggccgctac	gtcaggaaga	tgcgccagca	540
cctcgtcgac	gtcgtcttcc	ggaagtgcga	gcccgtcgac	gccgattggc	tgcaggtgca	600
aaagagcggg	aaccacaagt	ccaacgagat	caaggacgag	ggctttatcc	cgcctacct	660
gctgccgaag	cgccaatcgg	agctgcagt	ggcgtggacg	aacatagaaa	agtacgagcg	720
cgccacgctg	agcgtcgcag	agcgcgccag	actcgtcgag	gagtggaaag	ggagcctcgc	780
cgccgaggaa	tctgaccccg	cggaaaccga	acgcgctcgc	cgaagtgcga	ccaaatccgc	840
gtcagaatac	atggcgctcg	tcaggtggct	cgtggacaac	ggcatcgcca	ccgagcgcca	900
gtggatgcgc	gaggactcgg	acgggtacct	gagctacaac	gccacggggc	ccacgcgcgc	960
ccagatcaag	gccgcgctcg	acaacgcggc	ccgcacatcg	gtcaacacga	agacggccgc	1020
cgactacctg	gtcggcagga	accgcgccct	cgcgtggag	gacaacagga	tctaccggct	1080
gtttcgcctg	aacggctacg	acccgcctta	cgcgggcagc	gtgctgctcg	gctggtgccg	1140
caccggcttc	ggcaagcgca	acaccgtctg	gctcttcggg	cgggccacga	ccggcaagac	1200
gaacctcgcc	gaggccatca	gccactcggg	gcccttctac	ggctgcgtca	actggaccaa	1260
cgagaacttc	cccttcaacg	actgcgtgga	caagatgac	atctggtggg	aggagggcaa	1320
gatgacgtcc	aaggtcgtcg	agagcgccaa	ggccatcctc	ggcggcagca	aggtgcgcgt	1380
cgaccagaag	tgcaagaact	cgcagcagat	cgcgcacacg	cccgtcatca	tcaccagcaa	1440
caccaacatg	tgcgaggtcg	tgcacgggaa	ctcgaccacg	ttcgagcacc	ggcagccgct	1500
cgaggacagg	atgttcaagt	tcgagctcac	cgtcaggctg	cagcccacct	tcggcaagat	1560
caccaagcag	gaggtgcgcg	agttctttaa	gtgggcggag	ctcaacgccg	tcgacgtgga	1620
gtacgacttt	ctcgtgcgca	agatcaacca	atcggacact	gggggcgggg	ttaagcgccg	1680
ggcagagcct	acaaaggacg	agcctccgcg	gaagcgcggt	ttcttctacg	gcgcgacttc	1740
ggagggagag	gacgtccgcg	aggagacccc	gggggaatct	gactccgtga	acttcgccga	1800
gaggtacgtg	tcaaatgct	cgaagcacct	gtcctgggtca	aatatgcgct	atccctgccg	1860
cgcctgcgag	cgcataaag	cggacgtgaa	cgtctgcacc	ccccacggat	gcagagactg	1920
tcccagtgct	tttccccgcc	cgcgcgcggt	cccgatagcc	gagcacgacc	tgtgcctggc	1980
cccgatcgag	gactctgact	tttacgtcgg	ctgtatcgac	gacgtcaata	aagagcaata	2040
aatgatttta	cagcagtcac	gtcgtttgtc	gatcaccccc	cagactggct	cagggagatc	2100
ggagaggggc	tctccgagtt	catcggacta	gaggcgggac	cgcgaagcc	taagccgggc	2160
taccaggacc	gcgcgcgcgg	cctcgtcgtc	cccggtacga	agtaacctcg	gcccttcaac	2220
ggactcgaca	ggggcgagcc	cgtcaacgcg	gccgacgcgg	cggccaagaa	gcacgacgag	2280
gagtacgacc	ggctcctcaa	ggcgggggac	aaccgtacc	tcgcctacaa	ccacgcggac	2340

-continued

```

gccgagttcc agaaggacct cagcggcgac tccagcctag ccggcaacgc cgcgaacgcg 2400
ctcttccagg ccaagaagag ggtcctcgag cccttcggac tggtcgaggg agagccggag 2460
cccaagaaga cgccttcctg caagaggccg cagcctcgc cggactcgtc gagcggcgctc 2520
gggaagaagg gcgaccagcc cgcgcggaag aggtcgcact ttgggacgga gccgcgggt 2580
caagacggag cgggacgagc ggcccaggca gccggagata tggcatctgc tgagggtggt 2640
gcggttggtg gcggaccagt gggcgacgat gcacaagggt ccgatggagt gggtaatgcc 2700
tcaggaattt ggcattacga ttccgtttgg atggacggcg ctgtcatcac caagtccacc 2760
cgaacctgga gcctgcccgc ctacaacaac cacctctacc gccagatcca gtccagcggc 2820
accggagacg gcacgtactt tggttacagc acgccttggg gatacttcga tttcaatcga 2880
ttccactgcc actttttctc cagggactgg cagcggctca tcaacaacca ctggggcgtc 2940
cgccccaagg ggctccactt taagctcttc aacatccagg tcaaggaggt cagcagcacc 3000
gacggcacca cgaccatcgc caacaacctc accagcacca tacagggtctt tgccgacacg 3060
gagtagcagc tcccgtagt gctcggcaac gccacgagg gctgcctgcc gccgttcccg 3120
gcgagcgtct tcatgtctgc gcagtagcgc tacctgacgc tcaacgcca cccacgaac 3180
cagggcgctg ccgcgtctag cctgcgcgag agcgccttct actgcctgga gtactttccg 3240
agccagatgc tgaggaccgg gaacaacttc tcgttcagct acgagttcga gaagtgccc 3300
ttccactcga tgttcatgca cagccagagc ctggaccggc tcatgaacct gctcatcgac 3360
cagtacctct ggtacctcaa cgcaccagcc ggcaacaacc tctccttcaa caaggccggc 3420
gccaagaact tccccgagta ctcccgcaac tggctgccc ggcccggcca gcgcgtgcag 3480
cagtggagca cgatcggaac ccagaacaac gcgcagaccg gcacgtgggc cagcgccaac 3540
aagtggatcc tcatgggcag gtcacgcaag atggcccccg ggctggccca gccggtcagg 3600
aacgcccaga ccgtcaccaa cggtctgcag ctcatcttca acaacgagac catcaagggc 3660
tccaccgcga cggcctccac cgtacactcg gggctgctcg tcaccaacga gtcggagacc 3720
gccccgacca accccaactc ggcgacaaag tggggcgctc tgacggacaa ccagcagacc 3780
acctccacta cgcgcagcgt cagcgcagac ctcgaggcgc acgtcttccc gggcatggtc 3840
tggcaggacc gcgacatcta cctgcagggc ccgatatggg ccaagatccc ggagacggac 3900
ggccacttcc acccgctccc gctcatgggc ggcttcgggc tcaagaacct gccgcgcgag 3960
atcctcgta agaacacgcc cgtgcccgcg gtgcgcacca cgaccttcac accgcagaag 4020
gtcaactcct tcatcacgca gtactccacc gcccaggtea ccgtcgagat cgagtgggag 4080
ctgcgcaagg agaagagcaa gcgctggaac cccgagatcc agtacacctc caacttcgag 4140
aactcgacca acgtacagtt ctccgtcaac ggcgacggcg cgtacatcga gccgcgaccg 4200
atcggaacct ggtacctcac ccacaacctg taaataaatc cccaataaaa ccgtgtattc 4260
gtttcagttg aactttggtc tctgcg 4286

```

<210> SEQ ID NO 288

<211> LENGTH: 4693

<212> TYPE: DNA

<213> ORGANISM: Bovine adeno-associated virus

<400> SEQUENCE: 288

```

gtggcactcc cccccctgtc gcgttcgctc gtccgctggc tcgattgggg ggggtggcagc 60

```

-continued

tcaaagagct gccagacgac ggccctcttg gccgtcgccc ccccaatcga gccagcgaac	120
gagcgaacgc gacagggggg ggagtgccac actctctagc aaggggggtt ttaggtggt	180
gatgtcattg ttgatgtcat tatagtgtc acgcgatagt taatgattaa cagtcattgt	240
atgtgtgtta tccaatagga tgaagcgcg cgaatgagat ctccgcgagac ttccggggt	300
taaaaggggt gagtgaacga gcccgccgcc attctctgct ctggactgct agaggaccct	360
cgctgccatg gctaccttct atgaagtcac tggtcgcgtt ccatttgatg tggaagagca	420
cctgcctgga atttctgaca actttgtaga ctgggtaact ggtcaaattt gggagctgcc	480
tcccagagtc gatttgaatt tgactctgat tgagcagcct cagctgacgg tggctgacag	540
aattcgccgc gtgttctctg acgagtggaa caaattttcc aagcaggaga gcaaattctt	600
tgtgcagttt gaaaagggat ctgaatattt tcatctgcac acgctcgtgg agacctccgg	660
catctcttct atggtccttg gccgctacgt gagtcagatt cgcgccagc tggatgaagg	720
ggtgttcacg aacattgagc cgcggattaa cgactgggtc gccatcacca aggtaaagaa	780
gggaggagcc aataagggtg tggattctgg gtatattccc gcctacctgc tggcgaagg	840
ccaaccagag cttcagtggt cgtggactaa cctcgaagag tataaattgg ccgccctcaa	900
tctggaggag cgcaaacggc tcgtcgctca gtttcagctt gagtccctgc agcgtctgca	960
agaggcatct tcccagaggg acgtttcggc tgaccgggtc atcaagagca agacttccca	1020
gaaatacatg gcgctggtaa gctggctggt ggaacatggc atcaattccg agaagcagtg	1080
gattcaggag aatcaggaga gctacctgc cttcaactcc acgggaaact ctcgagacca	1140
gattaaagcc gcgcttgaca acgcgtcaaa aattatgagt ctgaccaaatt ctgcctcaga	1200
ctatctctgt ggacagactg ttccagagga catttctgaa aacagaatct ggcagatttt	1260
tgatctcaac ggctacgacc cggcatacgc gggctctgtt ctctacggct ggtgcactcg	1320
cgccttttga aagaggaaca cgtctcggct gtatggaccc gcgaccaccg gaaagaccaa	1380
catcgcgga gccatctctc acaccgtgcc cttttatggc tgggtgaaact ggactaatga	1440
gaactttccc tttaatgact gtgtggaaaa aatgttgatc tgggtggagg agggaaagat	1500
gaccagcaag gtggtggaac ccgccaaggc catcttgggg gggctctagag tacgagtgga	1560
tcaaaaatgt aaatcctctg tacaagtaga ctctaccccg gtgattatca cctccaatac	1620
taacatgtgt gtggtggtgg atgggaactc cacgaccttt gaacaccagc agccgctgga	1680
agaccgcatg ttcagatttg aactcatgcg gcggctcccg ccagattttg gcaagattac	1740
caagcaggaa gtcaaagact tttttgcttg ggcaaaggc aaccaggtgc cggtgactca	1800
cgagtttatg gttcccaaga aagtggcggg aactgagagg gcggagactt ctagaaaacg	1860
cccactggat gacgtcacca ataccaacta taaaagtccg gagaagcggg cccggctctc	1920
agttgttcct gagacgcctc gcagttcaga cgtgcctgta gagcccgctc ctctgcgacc	1980
tctcaactgg tcttcagggt atgaatgcag atgtgactat catgctaaat ttgactctgt	2040
aacgggggaa tgtgacgagt gtgaatattt gaatcggggc aaaaatggct gtatctttca	2100
taatgctaca cattgtcaaa tttgtcacgc tggtcctcca tgggaaaagg aaaatgtgtc	2160
agattttaat gattttgatg actgtaataa agagcagtaa ataaagtgag tagtcatgtc	2220
ttttgttgac caccctccag attggttggg atcgatcggc gacggcttct gtgaatttct	2280
cggccttgag gcgggtcccc cgaaacccaa ggccaatcaa cagaagcaag ataacgctcg	2340

-continued

aggtcttctg	cttctctgggt	acaagtatct	tggtcctggg	aacggccttg	ataagggcga	2400
tcctgtcaat	tttgcagacg	aggttgcccg	agagcacgac	ctctctacc	agaacagct	2460
tgaggcgggc	gataaccctt	acctcaagta	caaccacgcg	gacgcagagt	ttcaggagaa	2520
actcgcttct	gacacttctt	ttgggggaaa	ccttggaag	gctgttttcc	aggctaaaa	2580
gaggattctc	gaacctcttg	gcctgggtga	gacgcgggat	aaaacggcgc	ctgcggcaaa	2640
aaagaggcct	ctagagcaga	gtcctcaaga	gccagactcc	tcgagcggag	ttggcaagaa	2700
aggcaaacag	cctgccagaa	agagactcaa	ctttgacgac	gaacctggag	ccggagacgg	2760
gcctccccc	gaaggacct	ctccggagc	tatgtctact	gagactgaaa	tgctgcagc	2820
agctggcgga	aatggtggcg	atcggggaca	aggtgccgag	ggagtgggta	atgcctccgg	2880
tgattggcat	tgcgattcca	cttggtcaga	gagccacgtc	accaccacct	caaccgcac	2940
ctgggtcctg	ccgacctaca	acaaccacct	gtacctgcgg	ctcggtcga	gcaacgccag	3000
cgacaccttc	aacggattct	ccccccttg	gggatacttt	gactttaacc	gcttccactg	3060
ccacttctcg	ccaagagact	ggcaaaggct	catcaacaac	cactggggag	tgcgcccaa	3120
aagcatgcaa	gtccgcatct	tcaacatcca	agttaaggag	gtcacgacgt	ctaacgggga	3180
gacgaccgta	tccaacaacc	tcaccagcac	ggtccagatc	tttgccgaca	gcacgtacga	3240
gctcccgta	gtgatggatg	caggtcagga	gggcagcttg	cctcctttcc	ccaacgacgt	3300
gttcattggtg	cctcagtacg	ggtactgcgg	actggttaacc	ggaggcagct	ctcaaaacca	3360
gacagacaga	aatgccttct	actgtctgga	gtactttccc	agccagatgc	tgagaaccgg	3420
aaacaacttt	gagatggtgt	acaagtttga	aaactgccc	ttccactcca	tgtacgtcca	3480
cagccagagc	ctggataggc	tgatgaacct	gctgctggac	cagtacctgt	gggagctcca	3540
gtctaccacc	tctggaggaa	ctctcaacca	gggcaattca	gccaccaact	ttgccaagct	3600
gacaaaaaca	aactttttctg	gctaccgcaa	aaactggctc	ccggggccca	tgatgaagca	3660
gcagagattc	tccaagactg	ccagtcaaaa	ctacaagatt	cccagggaa	gaaacaacag	3720
tctgtcccat	tatgagacca	gaactacct	cgacggaaga	tgaggaatt	ttgccccggg	3780
aacggccatg	gcaaccgcag	ccaacgacgc	caccgacttc	tctcaggccc	agctcatctt	3840
tgcgggggcc	aacatcacccg	gcaacaccac	cacagatgcc	aataacctga	tgttcacttc	3900
agaagatgaa	cttagggcca	ccaacccccg	ggacactgac	ctgtttggcc	acctggcaac	3960
caaccagcaa	aacgccacca	ccgttcctac	cgtagacgac	gtggacggag	tcggcggtga	4020
cccggaatg	gtgtggcagg	acagagacat	ttactacca	gggcccattt	gggcaaaat	4080
tccacacacg	gatggacact	ttcaccctgc	tcctctcatt	ggcggaattg	gactgaaaag	4140
cccgctcca	caaatattca	tcaaaaacac	tcctgtaccc	gccaatcccc	caacgacctt	4200
ctctccggcc	agaatcaaca	gcttcatcac	ccagtacagc	accggacagg	tggtgtcaa	4260
aatagaatgg	gaaatccaga	aggagcggtc	caagagatgg	aaccagagg	tccagttcac	4320
gtccaaactac	ggagcacagg	actcgcttct	ctgggctccc	gacaacgccg	gagcctacaa	4380
agagcccagg	gccattggat	cccgatacct	caccaaccac	ctctagccca	attctgttgc	4440
ataccctcaa	taaacgtgtg	atcgctttca	gtaaaatact	gcctcttggtg	gtcattcggc	4500
gtacaacagc	ttacaacaac	aacaaaacc	ccttgctaga	gagtgtggca	ctccccccc	4560
tgctcgcttc	gctcgcttcg	tggtcgatt	gggggggtgg	cagctcaaag	agctgccaga	4620

-continued

cgacggccct ctgggcccgc gccccccaa tcgagccagc gaacgagcga acgcgacagg 4680

ggggggagtg cca 4693

<210> SEQ ID NO 289

<211> LENGTH: 4432

<212> TYPE: DNA

<213> ORGANISM: Snake parvovirus

<400> SEQUENCE: 289

cgccccaccc ctagtgatcg cgcgcgctct ctcttggggc ctgacggccg aaggccgtca 60

gctgccgagc ttcgctcggc agggcccaag agagagcgcg cgcgactact aggggtgggg 120

cgagtgcctt gctcaacggg ttttttggtg ggcggagcaa tgacgtcagc ggacatgtct 180

ggacatgtct ttgagcaagt ccatataagg agttccgccc gatatgcaaa tgagcaatcg 240

cgcaaagcat tttgggtagt caccatgaat aaaaaggaca gcaagaaaga tgacgcccc 300

taattttaat aggaatttta accatggcgt ttacgaggt tgtgtttcgt ttgccaagag 360

acaataacaa cttgttggat gaagatagat atcagccaga gttgaaaga gaagatgact 420

ggcctgagga atatttaacc agtgaagatg ccagctttat cggactagcg tatgctgtgc 480

taagtgaat tcggagattc tttggaagg aactacaatg gtttggccag gttgaatggt 540

gtcctactgc tggttaccac atgcatgttt tgttgaacca tcctaagctg agtaaccaga 600

cttatggaag aaaggtcaat gaactggcct gccgtatagt cgataccttt ggctaatta 660

atccagaaga agtcatcagt acccattatg ttaaaagcaa ctatggacat aaaaaggatga 720

gagtcattca cctagagtct tatttgaaga actacttttt cagaaagact ttagctcttc 780

ccaattatac cgaggaagga gactataaaa gagaggaaga agtcgtgctg tgggcattta 840

cgaatatcgt cgcttggaag ccattcgtgc ggaatctcat caagagatcg gagctagcga 900

ctgttcctaa gcaaccagag aatccggcgg gagacggacc ggcacctcga gtgactgcag 960

gaaccgcca ttttatggaa accatcgact ggttggtgaa acatggaatt actacagaac 1020

gagaattctg ccacgccaac cgccctttgt acctgtctat gctggcttct acttcgggtg 1080

ctgggcagat taaaagagcg ctggaccagg cgaaacacat gatgaccagc accatgtcag 1140

cagaggatta cctgacaaca gaagaggatg tgatcgaacc acctactgaa aatagaatct 1200

acaagattat gaaactgaat cgctatgatc cagaactagc agctgctctc ttctacggct 1260

ggacctgcaa gaactttggc aagagaaaca ccatctggct gtatggtcca gctactaccg 1320

gcaaaacat catcgctcaa gctattgcac atgctgttaa actgtttgct ggtgttaatt 1380

ggactaatga aaactttccc tttctgaact gtccagggaa actgcttacc tgggtgggag 1440

agggcaagat gacaaacaaa atgggtggaga cggctaaatg tatactgggg ggatctgctg 1500

tacctgtaga catcaaaggc aaaccgctg aaatgtgtcc tcaaacaccc tgtattatta 1560

ctagcaatc taacatgtgt caagtatatg atggtaatag ttctagcttt gagcaccag 1620

aaccctaga ggaacgcag tttatgttca gacttaatac taaactgcca tcgacctttg 1680

gcaagatcac agaagaggaa gtcaaacagt ttattacctg ggggaggagc ttaaagggtc 1740

aagttccaca tcagttcaga gtgcctacca caggagagta taaaaggcca gccccgag 1800

cgaaagctca ttcttcggat gagccgcaa aagagaaggt cgcgcgtatt gatgactctc 1860

taaccaggtg tgtaacaat attgatgagt cagctaccag tagagaaatg tttctagaga 1920

-continued

ttgctaatac	taatcaatgt	atgttgcatc	attgcttttc	ttgtaccgaa	tgttatcctg	1980
aattgcttga	tgacatggac	aaggaacaat	aaacttactg	ataacagata	tggtattttct	2040
cgatgatttc	tttgagata	aatataaaga	gactgttaac	gaactcggta	aaccgggtcaa	2100
tcctaaacct	gtaaaacaca	ttagcgaagc	tcactcgcaa	cctggcagca	ggaggggctt	2160
tgtggtgcct	gggtatcggt	atcttgggcc	tggtaatagc	ttggaccgtg	gaaagcccgt	2220
taacaaagca	gacgaggctg	ctaaaaagca	cgatcaagaa	tacgatcaac	agcttaaagc	2280
gggagacaat	ccctacataa	aatataatca	cgcggacgaa	cagttccaga	aagacctaca	2340
aggtgatacc	agtctagccg	gcaacgcggc	taacgctcta	tttcaaggca	aaaagactct	2400
actagcgcgc	cttggcctag	tagagacccc	tgctcgcaaa	acgtctgaaa	agcacaaatt	2460
agacgaatac	tatcctaaag	ctaaaaaggc	caaacaaggc	ttgcagatac	cagctccacc	2520
taaaggcgga	gaagaagaag	ctacatcgtc	acaatctgga	gggagcccag	caggttccga	2580
tactagcggc	acatctgtca	tggtctacag	aggaggcggt	ccgatggcag	acgataacca	2640
gggcgcgcag	ggagtgggta	attcctcagg	tgattggcat	tgcgatacca	agtggatggg	2700
agaccacgtc	attacaaagt	caaccagaac	ttgggtgctc	cccacttacg	ggaatcatct	2760
ctacgggcct	atcaactttg	acggcaccac	aggttcgggt	gctaatgcag	cctatgcagg	2820
atacaagact	ccctgggggt	actttgactt	caatcgattc	cattgccact	tctccccccg	2880
agactggcaa	agactcatca	acaaccacac	aggcatcagg	ccgaaaggac	tcaaatcaa	2940
agtctttaac	gtccaagtca	aagaagttac	aacacaagat	tcaacgaaaa	caattgccaa	3000
caatctcacc	agcaccgtag	agatctttgc	ggacgagaac	tacgacttac	catatgtatt	3060
aggcagtgct	acacaaggca	catttcctcc	atttcccaat	gatgtattta	tgttaccaca	3120
atatgcttat	tgtacacttc	aaggaaatcc	ggggaaatcc	gtagatagaa	gtgcctttta	3180
ttgttttagaa	tattttcctt	cacaaatgct	gagaacagga	aacaattttg	agttccagtt	3240
taaatttgaa	gaagtccctt	ttcattctgg	atgggcacag	agtcaaagcc	tagacagatt	3300
gatgaatccg	ttgcttgatc	aatatctgat	aggagactat	ggaacagatg	catcaggaaa	3360
ccttatttat	cacagagctg	gtccaaatga	tttgaatgaa	ttctacaaga	attgggcacc	3420
tgcaccctat	gaatgtatcc	agaatattaa	cagcagtgat	aataccaaga	atgctaattc	3480
tataaatggg	tcaaatctca	ccaacaaatg	gggactacaa	ggaagacaag	catgggatgc	3540
tccaggattt	gttcaagcta	gtacctatga	agggtgcagc	gcaggacaat	ctcttcttaa	3600
tggcgtactt	actttcgata	aaagttcagc	tactacttca	tctccagctg	ctactgcagt	3660
aaacagaaca	attgaagacg	aaatacaggg	taccaataat	tttggtaatg	ctagaaataa	3720
cattgttgct	atcaatcaac	aaacgaaagg	aacaaatcca	acaacaggta	gtacatctca	3780
atttgagaca	atgccaggta	tggtgtggtc	taatagagac	atttacttac	aggggcctat	3840
ttgggctaaa	attccaaata	cagatggaca	ttttcatcct	tctcccagaa	tggttggttt	3900
tggattaaaa	catcctccgc	ctatgattct	gatcaaaaat	acaccagttc	ctgctgatcc	3960
tccaactacc	ttcaatccaa	tgccacagac	tagtttcatt	actgaataca	gtacaggaca	4020
agtaactggt	gaaatgttgt	gggaggtaca	gaaagaatcc	tccaaaagat	ggaatccaga	4080
agtacagttt	acttccaatt	ttggaacttc	agatccagct	gttgatggaa	taccgtttgg	4140
aattaataat	ttgggtactt	atgttgaatc	tagacctatt	ggaactcggt	atatttctaa	4200

-continued

```

acacttgtaa ataataaaaa ttgtcaaatt tgcactaaga attgttgtca cgtggttgtt 4260
tacatgcttg ctaaaacacg cccaccaaaa aaccggttga gcagggcact cgccccaccc 4320
ctagtgatcg cgcgcgtctc ctcttggggc ctgccgagcg aagctcggca gctgacggcc 4380
ttcggccgtc aggccccaag agagagcgcg cgcgatcact aggggtgggg cg 4432

```

<210> SEQ ID NO 290

<211> LENGTH: 4682

<212> TYPE: DNA

<213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 290

```

tggccagttt ccaaacacagg ctgcctcgtc cactcggggc ccggcccca gggggccggt 60
agcgacgcct ttggcgtcgc ggcccagtg agcgagcgag cctgttttgg aaactggcca 120
gcactccggt gaggtaatgc cgtcacgtgg tcgggaacgg gaacgggaaa tctcgcgaga 180
acgtaaacaa atataagacg acgccacacg gcgctgcgtc atacgcgcgc gcgcaccggc 240
gagatgaggt cgtactacga ggtcatcggt caactgccga acgacgtcga gagtcaggta 300
cctggaatct ccgattcatt cgtcaactgg attacgtcgc gagaatggac gttgccggaa 360
gacgccgatt gggatttggg ccaggctgat caagttcaac tgacgctcgg cgacaaaatc 420
caacgggaga ttcgaaatca ctggggggacg atggcgaaa g aaccggactt tcaactatctt 480
atccaactgg aacaaggatga ggtgttcttt catttgacg tcctgctgga aacgtgttcc 540
gtgaagccga tggctactcg aagatatatt cgacatatc acaaaaaat tgtcagtaaa 600
gtctactcgc gccacgagcc cgcgatggaa ggatggatgc gggtgaccaa gacccaaaat 660
ttcgggggcg cgaacaaggt ccgggcccag tcgtatatc ccgcctacct gattccgaaa 720
cagcaaccgg aagtgcagtg ggcgtggact aacgtgccc agtatataa agcgtgcctg 780
caccgagaac tgcgtgccag tctcgcgcga cttcacttcg aggaggcggg gctctcgcaa 840
tccaaggaaa atctcgcgag aactgcagac ggcgctccc tgatcgcgac ccgcgtcagc 900
aaacgctaca tggagctcgt ggattggctc gtggagaagg ggatcaccac cgagaaggaa 960
tggtctgctg aaaacagaga aagctttcgg agcttccagg cctcgagcaa ctcggcgcgt 1020
cagatcaaga cggccctgca agcgcccatc caggagatgc ttctgaccaa gacggcggag 1080
gactacctcg tcggaaaaga acccgctcgc gacgacgaga ttctgcagaa ccgcattctc 1140
aagattttgg aactgaacca ctacgacca gcgtacgtg ggagtatttt ggtcgggtgg 1200
tgccagaaga aatggggcaa gcgaaacacg ctgtggctgt tcggacctgc gaccaccggc 1260
aagaccaaca tcgcggaggc tattgcccat gctgtgcctg tctatggatg cgttaactgg 1320
accaacgaga actttccggt caacgactgc gtcgaaaaga tgattatctg gtgggaagag 1380
ggcaaatga ccgccaaggt ggtggaaacg gctaaagcga ttctgggagg atctcgggtg 1440
agagtggacc aaaaatgcaa agcttcggtt ccgatcgaac cgacgcgggt cattattacc 1500
agtaacacca acatgtgtta tgcctcgac gggaacacga ctacgttcga gcataagcag 1560
ccgttgagg acaggatgtt taagctcgag ctgttgactc gtttaccga tgactttggt 1620
aaggtgacca aacaggaggt gcgtcaatc ttcagggtgt ccaggatca cctgaccct 1680
gtgatcccg aattcctggt gcggaaggcg gagtctcgca aaagaccgc cccctccggg 1740
gaaggctata taagcccgac aaagcggccc gcgctcgag agcagcagca ggcgtcggag 1800

```

-continued

agcgcggagc	cggttccac	caggtatcgt	atcaaatgct	cgaacattg	cgggatggat	1860
aaaatgttgt	ttccttgcca	aatttgtgaa	tcgatgaaca	gaaatattaa	tatatgtgcg	1920
attcataaaa	caacggaatg	taaagaatgt	tttccggaat	acggggataa	ggacactgtg	1980
ccagaactac	ccccctgtac	agaacataac	gtgtctcgtt	gttatcaatg	tcattcgggc	2040
gaattgtatc	gcgtgacttc	ggactctgac	gagaaacccg	cccccgagag	tgatgaaggc	2100
actgagccat	cctatgtccc	ctgcacgatt	caccatttga	tgggcaagag	tcgcgggtta	2160
gtgtcgtgtg	ctgcatgtcg	tttgaaaaat	agtacgttgc	atgatgactt	ggatgacggt	2220
gatcttgaac	aataaatgat	tgaatatag	ccatgtctct	catttctgat	gcgattccag	2280
attggttga	gcggttggtc	aaaaaggag	tgaatgctgc	ggctgatttc	taccatttgg	2340
aaagcggtec	tcctcatcct	aaggcaaatc	agcaaaccca	agaatctcct	gaaaaggacg	2400
attcgagagg	tctcgtgttc	ccgggttaca	agtatctagg	ccctttcaac	ggtctagata	2460
aaggaaaacc	cgtcaacgag	gcagacgctg	ccgccttaga	gcacgacaag	gcttacgacc	2520
tcgaactcaa	ggacgggcac	aaccctgact	ttgagtacaa	cgaggccgat	agacgtttcc	2580
aggaacgtct	caaagacgat	acctctttcg	gagggaatct	cggtaaagcc	atctttcagg	2640
ccaaaaagag	ggctctcgag	ccctttggtc	tgattgaaca	acccgacaac	acggccggga	2700
ccggggagaa	gcgtcccgaa	cgcgtcgacg	actttttccc	gaaaaagaag	aaggccaaga	2760
ccgagcaagg	caaagccctc	gctcaaacgg	gcgaagaccc	cggagaagga	acctcttcca	2820
acgctggatc	aagcgcctcc	tctagtgtgg	gatcatctgt	catggctgaa	ggagggtggc	2880
gtccaatggg	cgatgcaggc	caagggtgcc	acggagtggg	caattcctcg	ggaaattggc	2940
attgcgattc	ccaatggctg	gacaacggag	tcgttacccg	aaccactcga	acctgggtcc	3000
tgcccagcta	caacaaccac	ttgtacaagc	ggatccaagg	accgggagga	accgacccca	3060
acaataaatt	ctttggattc	agcaccctcc	gggggtactt	tgactacaac	cgattccact	3120
gccactttct	cccccgagac	tggaacgac	tcatacaaca	caactggggc	atccgaacca	3180
aagcgatcgc	ctttgacttc	tttaacatcc	agggtaaaga	agtcactgtc	caagactcca	3240
acaccaccat	cgccaacaac	ctcaccagca	cggccaagt	ctttgcggac	aaggactacc	3300
agctgccgta	cgtcctcgga	tcggctacag	agggcacctt	cccgcgttc	ccagcggata	3360
tctacacgat	cccgcagtat	ggttactgca	cgttaaaacta	caacaacgag	gcgggtggatc	3420
gttcggcctt	ctactgtcta	gactatttcc	cctcagacat	gctgcggaca	ggaaataact	3480
ttgaattcac	ttacacgttc	gaggacgttc	ctttccatag	catgtttgct	cacaaccaga	3540
cgttagaccg	gctgatgaat	cctctcgtcg	atcagtaacct	gtgggctttc	agttccgtca	3600
gccaacacag	ctcgtctgga	cgggcactca	attattcacg	cgcgaccaa	accaatatgg	3660
caaccacgta	cagaaactgg	ttacctggac	ccttcgtccg	ggatcagcaa	atctttacgg	3720
gggctagcaa	catcacccaa	aacaacgtgt	tcaacgtttg	ggataaaggc	aagcagtggg	3780
tgatagacaa	tcggatcaat	atgatgcagc	ccggccctgc	agcagcgacc	accttttagcg	3840
gagaacccga	cgtcaagcc	atgcaaaaaca	cgtgggcctt	tagtcggacg	gtctacgacc	3900
agacaaccag	tacgaccgat	cgttaaccagt	tgctcattac	caacgaagat	gaaatcagac	3960
ccaccaactc	ggtcggcatc	gacacgtggg	gagtagttcc	caacaacaac	cagtccaagg	4020
tgaccgcggg	cactcgcgcg	gccatcaaca	accaaggggc	gcttcccggg	atgggtgtggc	4080

-continued

```

aaaacagaga catttacctc caaggacca ttggggcaa aatccccgac acagacaatc 4140
acttccatcc gtccccgett attggcggtt ttggctgcaa gcateccccct ccccagattt 4200
tcattaaaaa cacaccgctc ccggccaacc ctccggaaac gtccagacg gccaaagggtg 4260
cctccttcac caaccagtac tcgaccggac agtgcaccgt cgaaatcttt tgggaactca 4320
agaaggaaac ctccaagcgc tggaaacccg aaatccagtt cacctccaac ttgggcaacg 4380
cgcccgacat ccagtttctg gtctccgaca ccggatccta ttccgaacct cgtcccatcg 4440
gtacccggtg ccttaccaaa cctctgtaaa ttaaaccctt caataaacgg tttatgcgta 4500
actgtacttc cgtctctgtg cgttattcag tcacatgata cggcattacc tcaccggagt 4560
gtggccaggt ttccaaaaca ggctcgctcg ctactcggg ccgcgacgcc aaaggcgctcg 4620
ctaccggccc ccttgggggc ggggcccagag tgagcgagcg agcctgtttt ggaaactggc 4680
ca 4682

```

<210> SEQ ID NO 291

<211> LENGTH: 4694

<212> TYPE: DNA

<213> ORGANISM: Avian adeno-associated virus

<400> SEQUENCE: 291

```

tggccagttt ccaagacagg ctgcctcgt cactcgggcc gggggcccaa agggggccct 60
agcgaccgct tcgcggtcgc ggcccagtg agcgagcgag cctgtcttgg aaactggcca 120
gcactccggt gaggtaatgc cgtcacgtgg tcgggaatgg gaacgggaaa tctcgcgaga 180
acgtaaacaa atataagacg gcgccacacg gcgctgcgtc atacgcgcgc gcgcaccggc 240
gagatgaggt cgtactacga ggtcatcggt cagctgcccc acgacgtcga gagtcaaggta 300
cctggaatct ccgattcgtt cgtcaactgg attacgtcgc gagaatggac gttgcctgag 360
gacgcccatt gggatttggg ccaggctgat caagttcaac tgacgctcgg cgacaaaatc 420
caacgggaga ttcgaactca ttgggggacg atggccaaag aaccggactt tcactatctt 480
atccaaactg aacaaggatg ggtgttcttt catttacacg tcctgctgga aacgtgttcc 540
gtaaagccga tgggtactcg aagatatatc cgacatattc aacaaaaaat tgtgagtaaa 600
gtctactcgc ccacgagcct acgatggaag gatggatcgc tggtagccaa gaccaaaaat 660
ttcgggggcg cgaacaaggt ccgggcccag tcgtatatcc ccgcctacct gatcccga 720
cagcaaccgg aagtgcagtg ggcgtggact aacgtgcccc agtatataaa agcgtgcttg 780
caccgagaac tgcgtgccag tctcgcgcga cttcacttcg aggaggcggg cgtctcgcaa 840
tccaaggaaa atctcgcgag aactgcagac ggcgctcccc tgatgccgac ccgcgtcagc 900
aaacgctaca tggagctcgt ggattggctc gtggagaagg ggatcaccac cgagaaggaa 960
tggctgctgg aaaacagaga aagctttcgg agctttcagg cctcgagcaa ctcggcgcgt 1020
cagatcaaga cggccctgca agcgccatt caggagatgc ttctgaccaa gacggcggag 1080
gactacctcg tcggaaagga tcccgtctcg gacgacgaca tccgtcagaa ccgcattctac 1140
aagattcttg aactgaacca ctacgaccca gcgtacgtgg ggagtatttt ggtcgggtgg 1200
tgccagaaga aatggggcaa gcgaaacacg ctgtggctgt tcggacatgc gaccaccggc 1260
aagaccaaca tcgcggaggc tattgcccac gctgtgccgt tctatggatg cgtaactgg 1320
accaacgaga actttccggt caacgactgc gtcgaaaaaa tgattatctg gtgggaggag 1380

```

-continued

ggcaaaatga	ccgccaaagt	ggtggaaaca	gccaaaggcga	ttctgggagg	atctcgggtg	1440
agagtggacc	aaaaatgcaa	agcttcggtt	ccgatcgaac	cgacgccggt	cattattacc	1500
agtaacacca	acatgtgtta	tgtcatcgac	gggaacacga	ccacgttcga	gcataagcag	1560
ccgttgagg	acaggatgtt	taagctcgaa	ttgctgactc	ggttgcctga	tgactttggt	1620
aaggtagcca	aacaggaggt	gcgtcaattc	ttcagggtgt	ctcaggatca	cctgaccctt	1680
gtgatcccg	aattcctagt	gcggaaggcg	gagtcctcga	aaagaccgcg	cccttcggg	1740
gaaggctata	taagcccgac	aaagcggccc	gcgctcgag	agcagcagca	ggcgtcgag	1800
agcgcggacc	cggttccac	caggatcgt	atcaaatgct	cgaaacattg	cggtatggat	1860
aaaatgttgt	ttccttgcca	aatttgtgaa	tcgatgaaca	gagatattaa	tatttgtgct	1920
attcataaaa	cgaccgactg	taaagagtgt	ttccccgact	acggggataa	agatgatgta	1980
gaactacccc	cctgtacaga	acacaacgtg	tctcgttgtt	atcaatgtca	ttcgggcgaa	2040
ttgtatcgcg	tgacttcgga	ctctgacgag	aaacctgccc	ccgagagtga	tgaaggcacc	2100
gagccatcct	atgctccctg	cacgattcac	cacctgatgg	gcaagagtca	cgggttagtc	2160
acttgcgcg	cgtgtcggtt	gaaaaatagt	acgttgcatg	atgacttgga	tgacggtgat	2220
ctcgaacaat	aatgattga	aatgtagcca	tgtctctcat	ttctgatgcg	attccagatt	2280
ggttgaggcg	gttggtcaaa	aaggagtgga	atgctgcagc	tgattttctac	catttgga	2340
gcggtcctcc	tcgtcctaag	gcaaatcagc	aaactcaaga	atctcttgaa	aaggacgatt	2400
cgagaggctc	cgtgttccca	ggctacaatt	atctaggccc	tttcaacggt	ctagataaag	2460
gagaaccctg	caacgaggca	gacgctgccg	ccttagaaca	cgacaaggct	tacgacctcg	2520
aaatcaagga	cgggcacaa	ccgtactttg	agtacaacga	ggccgacaga	cgtttccagg	2580
aacgtctcaa	agacgatacc	tcctttggag	gcaatttagg	taaagccatc	ttccaggcca	2640
aaaagagggt	tctcgaaacc	tttggctcgg	tggaagactc	aaagacggct	ccgaccggag	2700
acaagcggaa	aggcgaagac	gaacctcgtt	tgcccgcac	ttcttcacag	actccaaga	2760
aaaacaagaa	gcctcgcaag	gaaagacctt	ccggcggggc	agaagatccg	ggcgaaggca	2820
cctcttccaa	cgctggagca	gcagcacccg	cctctagtgt	gggatcatct	atcatggctg	2880
aaggaggtgg	cggcccagtg	ggcgtgcag	gccagggtgc	cgatggagtg	ggcaattcct	2940
ccggaaattg	gcattgcgat	tcccaatggc	tggaaaacgg	agtcgtcact	cgaaccaccc	3000
gaacctgggt	cttgcccagc	tacaacaacc	acctgtacaa	acgaatccaa	ggaccagcg	3060
gaggcgacaa	caacaacaaa	ttctttggat	tcagcaccac	ctggggatac	tttgactaca	3120
atcgattcca	ctgccacttt	tccccgcgag	actggcaacg	actcatcaac	aacaactggg	3180
gcacccgtcc	caaagcgatg	cgttttagac	tctttaacat	ccagggttaa	gagggtcacgg	3240
tccaagactt	caacaccacc	atcggaaca	acctcaccag	tacggtccag	gtctttgcgg	3300
acaaggacta	ccaactgccg	tacgtcctcg	gatcggctac	cgaaggcacc	ttcccgcgt	3360
tcccagcgga	tatctacacg	atcccgagct	acgggtactg	cacgctaaac	tacaacaacg	3420
aggcgggtgga	tcgttcggcc	ttctactgtc	tggaactactt	tcctcagac	atgctgcgga	3480
caggaaataa	ctttgagttt	acttacacct	tcgaggagct	tcctttccat	agcatgtttg	3540
cccacaacca	gacgctagac	cggctgatga	atccccctcg	ggatcagtac	ctctgggctt	3600
tcagctccgt	cagccaagca	ggctcatctg	gacgagctct	tcattactcg	cgggcgacta	3660

-continued

```

aaaccaacat ggcggtcaaa tataggaact ggtaacctgg gcctttcttc cgtgatcagc 3720
aaatctttac ggcgctagc aacatcacta aaaataacgt ctttagcggt tgggaaaaag 3780
gcaagcaatg ggaactcgac aatcggaacca acctaatagca gcccggtcct gcggcagcga 3840
ccacctttag cggagaacct gaccgtcaag ccatgcaaaa cagctgggt ttttagcagga 3900
ccgtctacga tcaaacgacc gccacgaccg atcgtaacca gatactcadc accaacgaag 3960
acgaaatcag accccaacaac tcggctcggt tgcacgcgtg gggagcagtt cccaccaaca 4020
accagtcatg cgtgaccccc ggcaactcgc cgcccgctca caatcaaggg gcgcttcccg 4080
ggatggtgtg gcaaaacaga gacatttacc ctacagggac ccatttggcc aaaattcccg 4140
acactgacaa tcacttccat ccgtccccgc ttattgggag gtttggtgc aagcatcccc 4200
ctccccagat ttctattaaa aacacacccg tccctgccaa cccttcggaa acgttccaga 4260
cggccaaagt ggctctcttc atcaaccagt actcgaccgg acagtgcacc gtcgaaatct 4320
tttggaact caagaaggaa acctccaagc gctggaaccc cgaaatccag ttcacctcca 4380
actttggcaa cgccggcgac atccagttg ccgtctccga cagggatcc tattccgaac 4440
ctcgtcccat cggtagccgt taccttacca aacctctgta aattaaaccc ttcaataaac 4500
cgtttatgag taactgtatt tccgtctcct gtcgttatcc agtcacatga tgcggcatta 4560
cctcacccga gtgtgggcca gtttccaaga caggctcgtg cgtcactcg ggccggggcc 4620
ccaaaggggc ccctagcgac cgcttcgagg tcggggcccg agtgagcgag cgagcctgtc 4680
ttggaactg gccca 4694

```

```

<210> SEQ ID NO 292
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (5)..(20)
<223> OTHER INFORMATION: This region may encompass 0-16 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

```

```

<400> SEQUENCE: 292

```

```

gctcnnnnnn nnnnnnnnnn 20

```

```

<210> SEQ ID NO 293
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(16)

```

-continued

```

<223> OTHER INFORMATION: This region may encompass 0-16 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
        description of substitutions and preferred embodiments

<400> SEQUENCE: 293

nnnnnnnnnn nnnnnngctc                                     20

<210> SEQ ID NO 294
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (9)..(20)
<223> OTHER INFORMATION: This region may encompass 0-12 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
        description of substitutions and preferred embodiments

<400> SEQUENCE: 294

gctcgctcnn nnnnnnnnnn                                     20

<210> SEQ ID NO 295
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(12)
<223> OTHER INFORMATION: This region may encompass 0-12 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
        description of substitutions and preferred embodiments

<400> SEQUENCE: 295

nnnnnnnnnn nngctcgctc                                     20

<210> SEQ ID NO 296
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (5)..(16)
<223> OTHER INFORMATION: This region may encompass 0-12 nucleotides
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (21)..(32)

```

-continued

<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (21)..(32)
<223> OTHER INFORMATION: This region may encompass 0-12 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 296

gctcnnnnnn nnnnnngctc nnnnnnnnnn nn 32

<210> SEQ ID NO 297
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(12)
<223> OTHER INFORMATION: This region may encompass 0-12 nucleotides
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (17)..(28)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (17)..(28)
<223> OTHER INFORMATION: This region may encompass 0-12 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 297

nnnnnnnnnn nngctcnnnn nnnnnnnngc tc 32

<210> SEQ ID NO 298
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(12)
<223> OTHER INFORMATION: This region may encompass 0-12 nucleotides
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (21)..(32)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (21)..(32)
<223> OTHER INFORMATION: This region may encompass 0-12 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 298

nnnnnnnnnn nngctcgctc nnnnnnnnnn nn 32

-continued

<210> SEQ ID NO 299
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (5)..(16)
<223> OTHER INFORMATION: This region may encompass 0-12 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 299

gctcnnnnnn nnnnnngctc 20

<210> SEQ ID NO 300
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (13)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (13)..(20)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 300

gctcgctgc tcnnnnnnnn 20

<210> SEQ ID NO 301
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 301

nnnnnnnngc tcgctcgctc 20

<210> SEQ ID NO 302
<211> LENGTH: 28

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (5)..(12)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (5)..(12)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (21)..(28)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (21)..(28)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 302

gctcnnnnnn nngctcgtc nnnnnnnn 28

<210> SEQ ID NO 303
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (13)..(20)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (13)..(20)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 303

nnnnnnnngc tcnnnnnnnn gctcgtc 28

<210> SEQ ID NO 304
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (9)..(16)

-continued

```

<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (21)..(28)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (21)..(28)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 304

gctcgcctcnn nnnnnngetc nnnnnnnn                                28

<210> SEQ ID NO 305
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (17)..(24)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (17)..(24)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 305

nnnnnnnnngc tcgctcnnnn nnnngctc                                28

<210> SEQ ID NO 306
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (21)..(28)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (21)..(28)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

```

-continued

<400> SEQUENCE: 306

nnnnnnnnngc tcgctcgcgc nnnnnnnn

28

<210> SEQ ID NO 307

<211> LENGTH: 28

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<220> FEATURE:

<221> NAME/KEY: modified_base

<222> LOCATION: (5)..(12)

<223> OTHER INFORMATION: a, c, t, g, unknown or other

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (5)..(12)

<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides

<220> FEATURE:

<221> NAME/KEY: modified_base

<222> LOCATION: (17)..(24)

<223> OTHER INFORMATION: a, c, t, g, unknown or other

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (17)..(24)

<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides

<220> FEATURE:

<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 307

gctcnnnnnn nngctcnnnn nnnngctc

28

<210> SEQ ID NO 308

<211> LENGTH: 16

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (1)..(16)

<223> OTHER INFORMATION: This sequence may encompass 1-4 'gctc'
repeating units

<220> FEATURE:

<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 308

gctcgcgcgc tcgctc

16

<210> SEQ ID NO 309

<211> LENGTH: 24

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (1)..(8)

<223> OTHER INFORMATION: This region may encompass 0-2 'gctc'
repeating units

<220> FEATURE:

<221> NAME/KEY: modified_base

<222> LOCATION: (9)..(16)

<223> OTHER INFORMATION: a, c, t, g, unknown or other

<220> FEATURE:

-continued

```

<221> NAME/KEY: misc_feature
<222> LOCATION: (9)..(16)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (17)..(24)
<223> OTHER INFORMATION: This region may encompass 0-2 'gctc'
repeating units
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 309

gctcgctcnn nnnnnngctc gctc                                     24

<210> SEQ ID NO 310
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1)..(8)
<223> OTHER INFORMATION: This region may encompass 0-2 'gctc'
repeating units
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)..(16)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (9)..(16)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (17)..(24)
<223> OTHER INFORMATION: This region may encompass 0-2 'gctc'
repeating units
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (25)..(32)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (25)..(32)
<223> OTHER INFORMATION: This region may encompass 0-8 nucleotides
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (33)..(40)
<223> OTHER INFORMATION: This region may encompass 0-2 'gctc'
repeating units
<220> FEATURE:
<223> OTHER INFORMATION: see specification as filed for detailed
description of substitutions and preferred embodiments

<400> SEQUENCE: 310

gctcgctcnn nnnnnngctc gctcnnnnnn nngctcgctc                       40

<210> SEQ ID NO 311
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
oligonucleotide

<400> SEQUENCE: 311

gcgcgctcgc tcgctc                                               16

```

-continued

```

<210> SEQ ID NO 312
<211> LENGTH: 13
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        oligonucleotide

<400> SEQUENCE: 312
gagttggcca ctc

```

13

1. A parvoviral ITR nucleotide sequence comprising an engineered Rep binding sequence region (eRBSR) wherein the parvoviral ITR nucleotide sequence is capable of forming a double-stranded Rep binding sequence region.

2. The parvoviral ITR nucleotide sequence of claim 1, wherein the parvoviral ITR nucleotide sequence comprises a Rep nicking stem loop.

3. The parvoviral ITR nucleotide sequence of claim 1, wherein the parvoviral ITR nucleotide sequence is from about 125 to about 165 nucleotides in length.

4. The parvoviral ITR nucleotide sequence of claim 1, wherein the parvoviral ITR nucleotide sequence is derived from an adeno-associated virus (AAV) ITR nucleotide sequence.

5. (canceled)

6. The parvoviral ITR nucleotide sequence of claim 1 comprising an oligonucleotide comprising an engineered Rep binding sequence region (eRBSR), wherein the eRBSR is from about 4 nucleotides to about 64 nucleotides in length.

7. (canceled)

8. (canceled)

9. The parvoviral ITR of claim 6, wherein the eRBSR comprises one or more consensus motif sequences of GCTC.

10. (canceled)

11. The parvoviral ITR of claim 9 claim 10, wherein the eRBSR comprises 1-5 consensus motif sequences of GCTC, and wherein two or more of the consensus motif sequences are contiguous with one another.

12. The parvoviral ITR of claim 9 claim 10, wherein the eRBSR comprises 1-5 consensus motif sequences of GCTC, and wherein none of the consensus motif sequences are contiguous with one another.

13. The parvoviral ITR of claim 6 wherein the eRBSR comprises one or more consensus motif sequences selected from the group consisting of NCTC, GNTC, GCNC, and GCTN, wherein N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine.

14. The parvoviral ITR of claim 9, wherein at least two of the nucleotides in one or more of the GCTC consensus motif sequences are modified.

15. The parvoviral ITR of claim 9 having the formula (GCTC)_x-N_y-(GCTC)_z (SEQ ID NO: 309), wherein N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine; wherein

- a. x and z are independently 0-5, with the proviso that when x=0, z=1-5, and when z=0, x=1-5; and
- b. y=0-16.

16. The parvoviral ITR of claim 13, wherein the eRBSR sequence is selected from the group consisting of SEQ ID NO: 68 (GCGCcgcatgcgGCTC), SEQ ID NO: 71 (gagcgggggcgcGCTC) and SEQ ID NO: 74 (ccgaccacgacgacgg).

17. The parvoviral ITR of claim 6, wherein the eRBSR comprises at least one nucleotide which is chemically modified.

18. (canceled)

19. (canceled)

20. (canceled)

21. The parvoviral ITR nucleotide sequence of claim 6, wherein the eRBSR comprises a Rep binding sequence region derived from a donor AAV sequence selected from the group consisting of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAVrh8, AAVrh10, AAV-DJ, and AAV-DJ8 with the proviso that the donor AAV sequence of the eRBSR is derived from a different AAV serotype than that of the ITR.

22. The parvoviral ITR nucleotide sequence of claim 6, wherein the binding affinity between a Rep protein and the eRBSR is decreased relative to the binding affinity between a Rep protein and a wild-type Rep binding sequence region.

23. (canceled)

24. (canceled)

25. (canceled)

26. The parvoviral ITR nucleotide sequence of claim 22, wherein the Rep protein is selected from the group consisting of Rep68, Rep78, Rep40, Rep52, Rep DA-1, V8-865, AAV1 Rep, AAV3 Rep, AAV4 Rep, AAV7 Rep, and AAV8 Rep.

27. The parvoviral ITR nucleotide sequence of claim 22, wherein the Rep protein is selected from the group consisting of Rep68, Rep78, Rep40, Rep52, Rep DA-1, V8-865, AAV1 Rep, AAV3 Rep, AAV4 Rep, AAV7 Rep, and AAV8 Rep with the proviso that Rep protein is from a different AAV serotype than that of the eRBSR.

28. The parvoviral ITR nucleotide sequence of claim 6, wherein nicking by a Rep protein is decreased relative to the nicking of an ITR sequence comprising a wild-type Rep binding sequence region.

29. (canceled)

30. The parvoviral ITR nucleotide sequence of claim 28, wherein the Rep protein is selected from the group consisting of Rep68, Rep78, Rep40, Rep52, Rep DA-1, V8-865, AAV1 Rep, AAV3 Rep, AAV4 Rep, AAV7 Rep, and AAV8 Rep.

31. (canceled)

32. (canceled)

33. (canceled)

34. (canceled)
35. (canceled)
36. An oligonucleotide comprising an engineered Rep binding sequence region (eRBSR), wherein the eRBSR is from about 4 nucleotides to about 64 nucleotides in length.
37. (canceled)
38. (canceled)
39. The oligonucleotide of claim 36, wherein the eRBSR comprises one or more consensus motif sequences of GCTC.
40. (canceled)
41. The oligonucleotide of claim 39, wherein the eRBSR comprises 1-5 consensus motif sequences of GCTC, and wherein two or more of the consensus motif sequences are contiguous with one another.
42. The oligonucleotide of claim 39, wherein the eRBSR comprises 1-5 consensus motif sequences of GCTC, and wherein none of the consensus motif sequences are contiguous with one another.
43. The oligonucleotide of claim 36, wherein the eRBSR comprises one or more consensus motif sequences selected from the group consisting of NCTC, GNTC, GCNC, and GCTN, wherein N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine.
44. The oligonucleotide of claim 39, wherein at least two of the nucleotides in one or more of the GCTC consensus motif sequences are modified.
45. The oligonucleotide of claim 39 having the formula $(GCTC)_x-N_y-(GCTC)_z$ (SEQ ID NO: 309), wherein N is any naturally or non-naturally occurring nucleoside selected from adenosine, guanosine, uridine, cytidine, thymidine; wherein
- x and z are independently 0-5, with the proviso that when x=0, z=1-5, and when z=0, x=1-5; and
 - y=0-16.
46. The oligonucleotide of claim 44, wherein the eRBSR a sequence selected from the group consisting of SEQ ID NO: 68 (GCGCgcgcgcGCTC), SEQ ID NO: 71 (gagcgggggcgcGCTC) and SEQ ID NO: 74 (ccgaccacgacgacgg).
47. The oligonucleotide of claim 36, wherein the eRBSR comprises at least one nucleotide which is chemically modified.
48. (canceled)
49. (canceled)
50. (canceled)
51. A polynucleotide comprising, in the 5' to 3' direction, a first parvoviral inverted terminal repeat sequence, a first heterologous sequence, a parvoviral inverted terminal repeat (ITR) nucleotide sequence comprising an engineered Rep binding sequence region, a second heterologous sequence, wherein the second heterologous sequence is complementary to the first heterologous sequence, and a second parvoviral inverted terminal repeat sequence.
52. The polynucleotide of claim 51, wherein said parvoviral (ITR) nucleotide sequence is selected from the parvoviral ITR nucleotide sequence of claim 6.
53. The polynucleotide of claim 51, wherein the first and second heterologous sequences comprise a transgene operatively linked to regulatory sequence(s) that controls expression of the transgene.
54. The polynucleotide of claim 53, wherein the transgene encodes a therapeutic polypeptide.
55. (canceled)
56. An adeno-associated viral (AAV) particle comprising the polynucleotide of claim 51.
57. (canceled)
58. (canceled)
59. (canceled)
60. (canceled)
61. A pharmaceutical composition comprising the AAV particle of claim 56 and one or more pharmaceutically acceptable excipients.
62. A method of producing a self-complementary adeno-associated viral particle comprising
- introducing into competent bacterial cells
 - a payload construct vector comprising a heterologous sequence flanked on each side by a parvoviral ITR sequence, wherein either of the parvoviral ITR sequences comprises an eRBSR, to produce a payload construct expression vector; and
 - one or more viral construct vector(s) comprising parvoviral rep and/or cap gene sequences under the control of one or more regulatory elements to produce a viral construct expression vector;
 - introducing into viral replication cells
 - the payload construct expression vector produced in step (a.i) to produce a payload construct particle; and
 - the viral construct expression vector(s) produced in step (a.ii) to produce a viral construct particle; and
 - co-infecting a viral replication cell with the payload construct viral particle produced in step (b.i) and the one or more viral construct viral particle(s) of step (b.ii) to produce a self-complementary adeno-associated viral (scAAV) particle.
63. The method of claim 62 further comprising isolation of the scAAV particle.
64. The method of claim 63 further comprising purification of the scAAV particle.
65. The method of claim 62, wherein the heterologous sequence in the payload construct vector is operatively linked to one or more regulatory sequence(s) that control the expression of the heterologous sequence.
66. The method of claim 62, wherein the payload construct expression vector of step (a.i) and the one or more viral construct expression vector(s) of step (a.ii) are produced via homologous recombination with a linearized baculovirus DNA.
67. The method of claim 62, wherein the payload construct expression vector of step (a.i) and the one or more viral construct expression vector(s) of step (a.ii) are produced via transposase-mediated transfer into a bacmid.
68. (canceled)
69. (canceled)
70. (canceled)
71. An scAAV particle produced by the method of claim 62.
72. (canceled)
73. (canceled)
- * * * * *