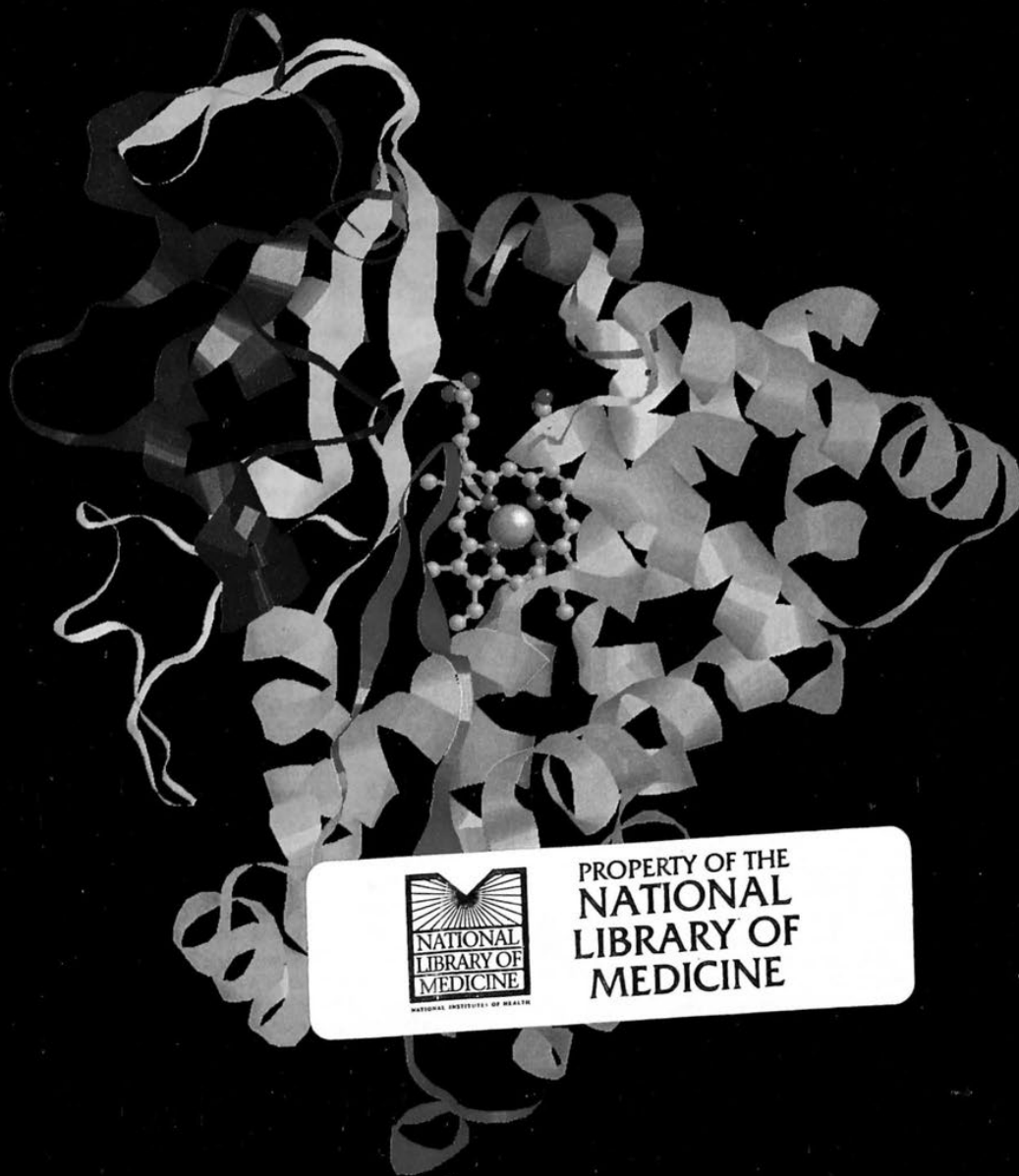


Protein engineering.
JUP - General Collection
N1 PR787P
v. 14, no. 10
Oct. 2001

PROTEIN ENGINEERING

Volume 14 number 10 October 2001



PROPERTY OF THE
NATIONAL
LIBRARY OF
MEDICINE

engineering p450
for biotransformation

OXFORD
UNIVERSITY
PRESS

ISSN 0269-2139 Coden PRENE9

PROTEIN ENGINEERING

Executive editors

A.R.Rees
School of Biology and Biochemistry, University of Bath, Bath, UK
G.A.Petsko
Rosenstiel Center, Brandeis University, Waltham, MA, USA

Associate editors

C.Craik
*Department of Pharmaceutical Chemistry and Biochemistry/
Biophysics, University of California, San Francisco, CA, USA*
T.Imoto
Faculty of Pharmaceutical Sciences, Kyushu University, Fukuoka, Japan
H.Nakamura
Institute for Protein Research, Osaka University, Osaka, Japan

Communications editor

M.Sternberg
ICRF, London UK

Editorial board

T.Alber, *University of California, Berkeley, CA, USA*
F.Arnold, *California Institute of Technology, Pasadena, CA, USA*
T.Beppu, *University of Tokyo, Tokyo, Japan*
T.L.Blundell, *Department of Biochemistry, Cambridge, UK*
M.Caruthers, *University of Colorado, Boulder, CO, USA*
C.Chothia, *MRC, Cambridge, UK*
G.M.Clore, *NIH, Bethesda, MD, USA*
F.E.Cohen, *University of California, San Francisco, CA, USA*
W.F.DeGrado, *University of Pennsylvania School of Medicine,
Philadelphia, PA, USA*
K.Dill, *UCSF, San Francisco, CA, USA*
A.R.Fersht, *University of Cambridge, Cambridge, UK*
A.Fontana, *CRIBI, Biotechnology Centre, Padua, Italy*
J.Gerlt, *University of Maryland, College Park, MD, USA*
B.Gutte, *Biochemisches Institut der Universität Zürich, Zürich,
Switzerland*
M.Ikehara, *Protein Engineering Research Institute, Osaka, Japan*
M.Karplus, *Harvard University, Cambridge, MA, USA*
S.B.H.Kent, *The Scripps Research Institute, La Jolla, CA, USA*
Y.Kyogoku, *Osaka University, Osaka, Japan*
B.Matthews, *University of Oregon, Eugene, OR, USA*
C.R.Matthews, *Pennsylvania State University, Pennsylvania, PA, USA*
R.E.Offord, *Centre Médicale Universitaire, Geneva, Switzerland*
S.B.Petersen, *University of Aalborg, Aalborg, Denmark*
R.T.Sauer, *MIT, Cambridge, MA, USA*
J.Schlessinger, *NYU, New York, USA*
J.Shi, *Chinese Biochemical Society, Shanghai, China*
G.P.Smith, *College of Arts and Science, Columbia, MI, USA*
D.Söll, *Yale University, New Haven, CT, USA*
M.J.E.Sternberg, *Imperial Cancer Research Fund, London, UK*
R.Stroud, *University of California, San Francisco, CA, USA*
J.Taylor, *Rutgers University, Piscataway, NJ, USA*
J.M.Thornton, *University College, London, UK*
J.E.Villafranca, *Agouron Pharmaceuticals, San Diego, CA, USA*
J.Wells, *Genentech Inc., San Francisco, CA, USA*
R.Wetzel, *University of Tennessee Medical Centre, Knoxville, TN, USA*
G.Winter, *MRC, Cambridge, UK*
M.Zoller, *Ariad Pharmaceuticals, Cambridge, MA, USA*

Cover design

The structure of *P. putida* cytochrome P450_{cam} (CYP101). CYP enzymes catalyse the oxidation of endogenous and xenobiotic compounds in biosynthetic and metabolic pathways in almost all living organisms. The C-H bond oxidation activity of CYP enzymes has no analogue in classical synthetic methodologies. Engineered CYP enzymes with broadened substrate specificity have enormous potential in biotechnological applications such as biotransformation and bioremediation. Understanding the role of active site residues in substrate recognition and the construction of *in vivo* substrate oxidation systems are important steps to these goals. For further details see Bell *et al.*, pages 797-802.

Published by Oxford University Press
Oxford, UK

OXFORD
UNIVERSITY
PRESS

Production editor

Wendy Neale, *Oxford University Press*

Contents

Protein Engineering

Volume 14 number 10 October 2001

COMMUNICATION

Three-dimensional modeling of the I-TevI homing endonuclease catalytic domain, a GIY-YIG superfamily member, using NMR restraints and Monte Carlo dynamics

J.M.Bujnicki, P.Rotkiewicz, A.Kolinski and L.Rychlewski **717**

ORIGINAL ARTICLES

The building block folding model and the kinetics of protein folding

C.-J.Tsai and R.Nussinov **723**

Signature of quaternary structure in the sequences of legume lectins

N.Manoj and K.Suguna **735**

The investigation of the effects of counterions in protein dynamics simulations

P.Drabik, A.Liwo, C.Czaplewski and J.Ciarkowski **747**

Energetic analysis of binding of progesterone and 5 β -androstane-3,17-dione to anti-progesterone antibody DB3 using molecular dynamics and free energy calculations

M.Peräkylä and N.Nordman **753**

Lipid-facing correlated mutations and dimerization in G-protein coupled receptors

P.R.Gouldson, M.K.Dean, C.R.Snell, R.P.Bywater, G.Gkoutos and C.A.Reynolds **759**

Binding of external ligands onto an engineered virus capsid

U.Schmidt, R.Rudolph and G.Böhm **769**

Polyionic fusion peptides function as specific dimerization motifs

S.A.Richter, K.Stubenrauch, H.Lilie and R.Rudolph **775**

Mutating aspartate in the calcium-binding site of α -lactalbumin: effects on the protein stability and cation binding

S.E.Permyakov, V.N.Uversky, D.B.Veprintsev, A.M.Cherskaya, C.L.Brooks, E.A.Permyakov and L.J.Berliner **785**

Increased proteolytic resistance of ribonuclease A by protein engineering

Y.Markert, J.Köditz, J.Mansfeld, U.Arnold and R.Ulbrich-Hofmann **791**

Engineering the CYP101 system for *in vivo* oxidation of unnatural substrates

S.G.Bell, C.F.Harford-Cross and L.-L.Wong **797**

Anti-c-myc 9E10: epitope key positions and variability characterized using peptide spot synthesis on cellulose

K.Hilpert, G.Hansen, H.Wessner, G.Küttner, K.Welfle, M.Seifert and W.Höhne **803**

Cytidine deaminase from two extremophilic bacteria: cloning, expression and comparison of their structural stability

A.Cambi, S.Vincenzetti, G.De Sanctis, J.Neuhard, P.Natalini and A.Vita **807**

Continued overleaf

Optimized linker sequences for the expression of monomeric and dimeric bispecific single-chain diabodies

T.Völkel, T.Korn, M.Bach, R.Müller and R.E.Kontermann

815

Please visit the journal's World Wide Web site at www.protein.oupjournals.org

SUBSCRIPTIONS

A subscription to *Protein Engineering* comprises 12 issues, with an Annual Author and Subject Index. Subscriptions are entered on a calendar year basis only. Prices include postage by surface mail or, for subscribers in the USA and Canada by air freight or in Japan, Australia, New Zealand and India by air speeded post. Claims for non-receipt must be made within four months of dispatch/order (whichever is later).

Subscription rates for Volume 14

Institutional: UK and Europe £475.00; Rest of World US\$775.00. Single issues: UK and Europe £49; Rest of World US\$81. *Personal**: UK and Europe £142.00; Rest of World US\$233.00. Single issues £15 in the UK and Europe, US\$24 elsewhere. Online only: UK and Europe £427; Rest of World US\$697.00.

**Personal rates apply only when copies are sent to a private address and payment is by personal cheque or credit card.*

Orders. Orders and payments from, or on behalf of, subscribers in the various geographical areas shown below should be sent to the Oxford University Press office indicated.

The Americas: Oxford University Press, 2001 Evans Road, Cary, NC 27513, USA.

Rest of the World: Journals Subscriptions Department, Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, UK. Tel: (01865) 267907; fax: (01865) 267485; jnl.orders@oup.co.uk

The current volume plus 2 years' backstock are available from the publisher. Previous volumes are available from Swets and Zeitlinger, Backsets Department, PO Box 830, 2160 SZ Lisse, The Netherlands (tel: 2521 35111). Available in 16 mm microfilm, 35 mm microfilm and 105 microfiche from University Microfilms International, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, USA. Copies of articles published are also available from UMI.

Advertising. To advertise in *Protein Engineering* contact PRC Associates, The Annexe, Fitznells Manor, Chessington Road, Ewell Village, Surrey KT17 1TF, UK.

© Oxford University Press 2001

All rights reserved; no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without either the prior written permission of the Publishers, or a licence permitting restricted copying issued in the UK by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1P 9HE, or in the USA by the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. Special requests, such as copying for general distribution or for advertising or promotional purposes, should be addressed to the Production Editor, *Protein Engineering*, Oxford University Press, Walton Street, Oxford OX2 6DP, UK.

Protein Engineering (ISSN 0269-2139) is published monthly by Oxford University Press, Oxford, UK. Annual subscription price is US\$775.00. *Protein Engineering* is distributed by Mercury International, 365 Blair Road, Avenel, New Jersey, NJ 07001, USA. Periodical postage paid at Rahway, New Jersey, USA and additional entry points.

US POSTMASTER: send address changes to *Protein Engineering*, c/o Mercury International, 365 Blair Road, Avenel, New Jersey, NJ 07001, USA.

Typeset and printed by Information Press Ltd, Oxford, UK.

ANNOUNCEMENT

Communication, Comments and Protocols

The Editors would like to encourage submissions to this section, which it is felt is needed to publish material that is of wide interest to readers but does not, for a variety of reasons, fit within the more formal structure of a full article.

The types of papers should include, but not be restricted to:

- Descriptions of novel protocols
- Notes on evaluation techniques
- Commentaries on workshops or conferences
- Short articles reviewing a hot topic
- Software reviews

Please submit such papers to:

Dr Mike Sternberg
Professor of Structural Bioinformatics and
Director Imperial College Centre for Bioinformatics
Biochemistry Building
Department of Biological Sciences
Imperial College
London SW7 2AZ, UK
e-mail: m.sternberg@ic.ac.uk

Optimized linker sequences for the expression of monomeric and dimeric bispecific single-chain diabodies

Tina Völkel, Tina Korn, Miriam Bach, Rolf Müller and Roland E. Kontermann¹

Institut für Molekularbiologie und Tumorforschung, Philipps-Universität Marburg, Emil-Mannkopff-Strasse 2, 35033 Marburg, Germany

¹To whom correspondence should be addressed.
E-mail: rek@imt.uni-marburg.de

Bispecific single-chain diabodies (scDb) consist of the variable heavy and light chain domains of two antibodies connected by three linkers. The structure of an scDb in the V_H–V_L orientation is V_HA–linkerA–V_LB–linkerM–V_HB–linkerB–V_LA, with linkers A and B routinely chosen to be 5–6 residues and linker M 15–20 residues. Here, we applied display of scDb on filamentous phage to analyse the composition of optimal linker sequences. The three linkers were randomized in length and sequence using degenerated triplets coding for only six hydrophilic or aliphatic amino acids (Thr, Ser, Asp, Asn, Gly, Ala). Antigen-binding clones were then isolated by one to two rounds of selection on the two different antigens recognized by the bispecific scDb. Using an scDb directed against carcinoembryonic antigen (CEA) and β -galactosidase (Gal), we found that monomeric scDb had a preferred length of 15 or more amino acid residues for the middle linker M and of 3–6 residues for the linkers A and B. No obvious bias towards a preferred linker sequence was observed. Reduction of the middle linker below 13 residues led to the formation of dimeric scDb, which most likely results from interchain pairing between all the V_H and V_L domains. Dimeric scDb were also formed by fragments possessing a long linker M and linkers A and B of 0 or 1 residue. We assume that these dimeric scDb are formed by intrachain pairing of the central variable domains and interchain pairing of the flanking variable domains. Thus, the latter molecules represent a novel format of bispecific and tetravalent molecules. The described strategy allows for the isolation of both optimized and minimal linker sequences for the assembly of monomeric or dimeric single-chain diabodies.

Keywords: antigen-binding site/bispecific/homodimer/linker/multivalent/phage display/single-chain diabodies

Introduction

Diabodies (Db) are bivalent or bispecific antibody fragments generated by the dimerization of two V_H–V_L or V_L–V_H fragments (Holliger *et al.*, 1993). Bispecific diabodies are produced by heterodimerization of two fragments of the structure V_HA–V_LB and V_HB–V_LA expressed in the same cell. Applications of bispecific diabodies include the recruitment of enzymes and effector molecules or cells of the immune system (Holliger *et al.*, 1996, 1997; Kontermann *et al.*, 1997a,b; Krebs *et al.*, 1998); reviews are available (Holliger and Winter, 1997; Hudson and Kortt, 1999).

We and others have recently modified the original dimeric diabody format by joining the two fragments by an additional middle linker (Brüsselbach *et al.*, 1999; Kipriyanov *et al.*, 1999). The overall structure of these single-chain diabody (scDb) molecules is V_HA–linkerA–V_LB–linkerM–V_HB–linkerB–V_LA, with linkers A and B consisting routinely of 5–6 residues and linker M of 15–20 residues. The middle linker (linker M) resembles those connecting the two variable domains in scFv molecules. Monomeric single-chain diabodies (scDb) are readily assembled in bacterial and mammalian cells and show improved stability under physiological conditions (Brüsselbach *et al.*, 1999; Kipriyanov *et al.*, 1999). We could demonstrate that scDb molecules can be applied to recruit prodrug-converting enzymes and cytotoxic T lymphocytes to tumour cells (Brüsselbach *et al.*, 1999; T.Korn and R.E.Kontermann unpublished work). Furthermore, scDb can be expressed intracellularly in active form in the secretory pathway or displayed on the surface of mammalian cells opening additional therapeutic applications (Kontermann and Müller, 1999). The functional affinity of scDb could be improved by fusion to the Fc or C_H3 region of the immunoglobulin γ 1 chain generating IgG-like tetravalent and bispecific molecules (Alt *et al.*, 1999).

Recent findings indicate that the length and composition of the linkers used to connect the variable domains strongly influence the formation and stability of single-chain Fv fragments (scFv). Most scFv fragments are generated using a 15 amino acid residue linker of composition (Gly₄Ser)₃ (Huston *et al.*, 1988). Several approaches have been described to obtain optimized linkers by phage display technology. These resulted in the selection of various new linkers, which improved secretion and antigen binding activity of the scFv fragments (Tang *et al.*, 1996; Turner *et al.*, 1997).

The length of the linker joining the two variable domains also dictates the formation of multimeric scFv molecules. Reduction of the length of the linker joining the two variable domains from 15 amino acid residues, generally used to produce single-chain Fv fragments (scFv), below 8–12 residues favours dimeric assembly of the V_H–V_L fragments generating diabodies (Holliger *et al.*, 1993; Kortt *et al.*, 1994; Afthan *et al.*, 1995). Further reduction of the linker sequence to less than five amino acids has been shown to result in the generation of tri- or tetrameric molecules (triabodies, tetrabodies) (Iliades *et al.*, 1997; Kortt *et al.*, 1997; Pei *et al.*, 1997; Le Gall *et al.*, 1999; Dolezal *et al.*, 2000); a review is also available (Hudson and Kortt, 1999).

Recently, Kipriyanov *et al.* showed that dimeric single-chain diabody molecules with four antigen-binding sites are formed by expressing polypeptides with a middle linker of 12 residues (Kipriyanov *et al.*, 1999). It has been proposed that these so-called tandem diabodies (Tandabs) form by interchain pairing of the four variable domains of each chain. These molecules possess an increased functional affinity and improved pharmacokinetics compared with monomeric scDb (Cochlovius *et al.*, 2000).

Table I. Oligonucleotides for construction of scDb linker M library

ML1For:	5'-CGG TAA <u>GGC GCG CCC</u> ABY ACC TAG GAC GGT CAG CTT GGT CCC TCC-3'
ML3For:	5'-CGG TAA <u>GGC GCG CCC</u> ABY ABY ABY ACC TAG GAC GGT CAG CTT GGT CCC TCC-3'
ML5For:	5'-CGG TAA <u>GGC GCG CCC</u> ABY ABY ABY ABY ABY ACC TAG GAC GGT CAG CTT GGT CCC TCC-3'
ML7For:	5'-CGG TAA <u>GGC GCG CCC</u> ABY ABY ABY ABY ABY ABY ABY ACC TAG GAC GGT CAG CTT GGT CCC TCC-3'
ML1Back:	5'-CTA <u>GGG CGC GCC</u> TCG RVT CAG GTG CAG CTG GTG CAA TCT GGG GGA-3'
ML3Back:	5'-CTA <u>GGG CGC GCC</u> TCG RVT RVT RVT CAG GTG CAG CTG GTG CAA TCT GGG GGA-3'
ML5Back:	5'-CTA <u>GGG CGC GCC</u> TCG RVT RVT RVT RVT RVT CAG GTG CAG CTG GTG CAA TCT GGG GGA-3'
ML7Back:	5'-CTA <u>GGG CGC GCC</u> TCG RVT RVT RVT RVT RVT RVT RVT CAG GTG CAG CTG GTG CAA TCT GGG GGA-3'

The *AscI* site is underlined. R = A, G; Y = C, T; V = A, C, G; B = T, G, C.

In the present study, we applied phage display to isolate single-chain diabodies with optimized linker sequences. We first demonstrated that single-chain diabody molecules can be displayed on filamentous phage by fusion to gene 3, generating a single-gene encoded scDb-fusion protein. Libraries of scDb molecules containing randomized linker regions were then applied to isolate molecules with linkers optimized in length and sequence. Furthermore, we could show that functional dimeric scDb molecules are formed by reducing the lengths of either the middle linker M or the linkers A and B below a certain threshold.

Materials and methods

Materials

β -Galactosidase was purified as described (Kontermann *et al.*, 1997a). CEA was kindly provided by Professor Sedlacek (HMR, Marburg, Germany) and anti-Myc-tag antibody 9E10 by Professor Eilers (IMT, Marburg, Germany). HRP-conjugated anti-M13 antibody was purchased from Amersham-Pharmacia (Freiburg, Germany) and HRP-conjugated anti-mouse IgG antibodies from Sigma (Deisenhofen, Germany).

Oligonucleotides

For LMB2, LMB3, fdSeq1 see Kontermann *et al.* (Kontermann *et al.*, 1997a). CEAVHSfiBack, 5'-CTC GCG GCC CAG CCG GCC ATG GCC CAG GTG AAA-3'; CEAVLNotFor, 5'-TTC TGC GGC CGC CCG TTT CAG CTC CAG CTT GGT-3'; GalVLSeqBack, 5'-TCC TGG TAC CAA CAG CAC CCA GGC-3'; GalVHSeqBack, 5'-GCG AAA GAT CGC ATA GCA GGG-3'; CEAVHSeqBack, 5'-TGG ATT GAT CCT GAG AAT GGT-3'. Oligonucleotides for constructing the scDb linker M and scDb linker AB library are summarized in Tables I and II.

Phage display of single-chain diabodies

For phage display of single-chain diabody scDb CEAGal, the coding region was isolated from plasmid pAB1 scDb CEAGal (Brüsselbach *et al.*, 1999) digested with *SfiI* and *NotI* and cloned into phagemid vector pHEN2 (Krebs *et al.*, 1998). For display of diabody Db CEAGal, the coding region was isolated from plasmid pAB1 Db CEAGal (Kontermann *et al.*, 1997a) by digestion with *NotI* and partial digestion with *SfiI* and cloning into pHEN2. Positive clones were identified by polymerase chain reaction (PCR) screening with primers LMB2 and LMB3.

Construction of a bispecific scDb linker M phage display library

Plasmid pAB1 scDb CEAGal was digested with *SfiI* and *AscI* (fragment 1) or with *AscI* and *NotI* (fragment 2). Fragment 1 was PCR amplified with primers CEAVHSfiBack and ML1For, ML3For, ML5For or ML7For. The products were digested

with *SfiI* and *AscI*. Fragment 2 was PCR amplified with primers CEAVLNotFor and ML1Back, ML3Back, ML5Back or ML7Back and digested with *AscI* and *NotI*. PCR products either were ligated by combining all eight fragments in one ligation reaction or the fragments were generated by combining ML1For and ML1Back, ML3For and ML3Back, ML5For and ML5Back or ML7For and ML7Back in four separate ligation reactions. The fragments were cloned into phagemid vector pHEN2 and six transformation reactions were performed using electrocompetent TG1 (Stratagene). Transformants were plated on two large (24×24 cm) LB plates containing 100 μ g/ml ampicillin and 1% glucose. Phage were rescued as described (Marks *et al.*, 1991). Diversity of linkers was confirmed by sequence analysis with primer GalVHSeqBack.

Construction of a bispecific scDb linker AB phage display library

Various positive clones isolated from the linker M library were digested with *BstEII* and *SacI* and the resulting GalVH-linkerM-GalVL fragment was amplified with primers A0Back, A1Back, A2Back, A3Back, A4Back, A5Back or A6Back and B0For, B1For, B2For, B3For, B4For, B5For or B6For in all 49 possible combinations. PCR fragments were pooled and digested with *BstEII* and *SacI* and cloned into plasmid pHEN2 scDb CEAGal digested with *BstEII* and *SacI*. Diversity of linkers was confirmed by sequence analysis with primers GalVLSeqBack, GalVHSeqBack and CEAVHSeqBack.

Selections

Antigens were coated in 1 ml of PBS on Nunc Maxisorb immunotubes at a concentration of either 10 μ g/ml for CEA or 25 μ g/ml β -galactosidase and selections were performed as described (Kontermann, 2001). One selection was performed using CEA in round 1 and β -galactosidase in round 2 (selections 1a and 2a) and a second selection was performed using β -galactosidase in round 1 and CEA in round 2 (selections 1b and 2b). Eluted phage were infected into suppressor strain TG1 for production of phage.

Screening and expression of soluble single-chain diabodies

Clones from round 1 and 2 were screened by ELISA for antigen binding of soluble scDb as described (Kontermann, 2001) using CEA and β -galactosidase as antigens. Alternatively, phage-displayed scDb were produced by rescue with helper phage of single TG1 colonies grown on microtitre plates (Marks *et al.*, 1991). Clones isolated from the scDb linker M library were sequenced with primer GalVHSeqBack and clones isolated from the scDb linker AB library with primers CEAVHSeqBack and GalVHSeqBack. For soluble expression, scDb fragments were subcloned into pAB1 as *SfiI/NotI* fragment. Purification of scDb fragments was performed from 3 l of induced TG1 cultures by immobilized metal affinity

Table II. Oligonucleotides for construction of scDb linker AB library

B0For:	5'-GTG GGT <u>GAG CTC</u> GAT GTC CGA GAC GGT TAC TAG GGT ACC-3'
B1For:	5'-GTG GGT <u>GAG CTC</u> GAT GTC ABY CGA GAC GGT TAC TAG GGT ACC-3'
B2For:	5'-GTG GGT <u>GAG CTC</u> GAT GTC ABY ABY CGA GAC GGT TAC TAG GGT ACC-3'
B3For:	5'-GTG GGT <u>GAG CTC</u> GAT GTC ABY ABY ABY CGA GAC GGT TAC TAG GGT ACC-3'
B4For:	5'-GTG GGT <u>GAG CTC</u> GAT GTC ABY ABY ABY ABY CGA GAC GGT TAC TAG GGT ACC-3'
B5For:	5'-GTG GGT <u>GAG CTC</u> GAT GTC ABY ABY ABY ABY ABY CGA GAC GGT TAC TAG GGT ACC-3'
B6For:	5'-GTG GGT <u>GAG CTC</u> GAT GTC ABY ABY ABY ABY ABY ABY CGA GAC GGT TAC TAG GGT ACC-3'
A0Back:	5' ACC ACG <u>GTC ACC</u> GTC TCC CAG TCT GTT CTG ACT CAG CTT CCC-3'
A1Back:	5' ACC ACG <u>GTC ACC</u> GTC TCC RVT CAG TCT GTT CTG ACT CAG CTT CCC-3'
A2Back:	5' ACC ACG <u>GTC ACC</u> GTC TCC RVT RVT CAG TCT GTT CTG ACT CAG CTT CCC-3'
A3Back:	5' ACC ACG <u>GTC ACC</u> GTC TCC RVT RVT RVT CAG TCT GTT CTG ACT CAG CTT CCC-3'
A4Back:	5' ACC ACG <u>GTC ACC</u> GTC TCC RVT RVT RVT RVT CAG TCT GTT CTG ACT CAG CTT CCC-3'
A5Back:	5' ACC ACG <u>GTC ACC</u> GTC TCC RVT RVT RVT RVT RVT CAG TCT GTT CTG ACT CAG CTT CCC-3'
A6Back:	5' ACC ACG <u>GTC ACC</u> GTC TCC RVT RVT RVT RVT RVT RVT CAG TCT GTT CTG ACT CAG CTT CCC-3'

The *SacI* site (For primers) and the *BstEII* site (Back primer) are underlined. R = A, G; Y = C, T; V = A, C, G; B = T, G, C.

chromatography (IMAC) as described (Kontermann *et al.*, 1997a). Purified fragments were dialysed against PBS and the concentrations were determined by spectrophotometry. Purity of fragments was analysed by 10% SDS-PAGE under non-reducing conditions and by immunoblotting experiments using an anti-His-tag antibody. Gel filtration of scDb molecules was performed by FPLC on a Superose 12 column (Pharmacia) with a flow-rate of 0.33 ml/min in PBS and loading 200 µl of the purified proteins (corresponding to ~30 µg of protein).

ELISA and recruitment experiments

ELISA and recruitment experiments were performed as described (Kontermann *et al.*, 1997a) using CEA coated in PBS on to Maxisorp microtitre plates (Nunc) at 10 µg/ml. For ELISA, bound proteins were detected with an HRP-conjugated anti-His-tag antibody (Santa Cruz). Recruitment experiments were performed with 50 µg/ml β-galactosidase and bound enzyme was detected with *o*-nitrophenyl-β-D-galactopyranoside (ONPG). The simultaneous binding of two molecules of the same antigen by dimeric single-chain diabodies was analysed in recruitment experiments using horseradish peroxidase-conjugated antigen (CEA, β-galactosidase) labelled with EZ-Link Plus-activated HRP according to the manufacturer's protocol (Pierce). Antibody molecules at a concentration of 20 µg/ml were preincubated for 1 h with HRP-conjugated antigen at a molar ratio of antibody to antigen of 1:1. The solution was then added to immobilized CEA or β-galactosidase coated at a concentration of 20 or 50 µg/ml, respectively. Bound antigen was directly detected with 3,3',5,5'-tetramethylbenzidine-H₂O₂ as used for ELISA.

Results

Phage display of bispecific single-chain diabodies

Both phage-displayed single-chain diabody scDb CEAGal and the dimeric diabody Db CEAGal bound with identical efficacy to the two antigens applying equal amounts of phage (Figure 1). Thus, single-chain diabody can be displayed on phage without interfering with antigen binding. However, in recruitment experiments no binding was observed with phage-displayed scDb CEAGal or Db CEAGal using 10¹² t.u./ml, in contrast to soluble antibody fragments which efficiently recruited β-galactosidase to immobilized CEA (not shown) (see also Brüsselbach *et al.*, 1999). This finding indicates that simultaneous binding of both antigens under the applied experimental conditions is abolished by phage display.

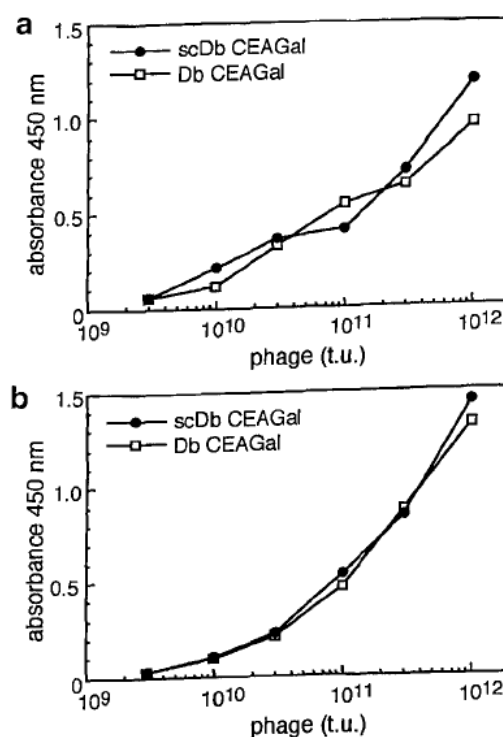


Fig. 1. Phage ELISA of Db CEAGal and scDb CEAGal with immobilized CEA (a) or β-galactosidase (b). Binding was compared using identical phage titres. Bound phage were detected with an HRP-conjugated anti-M13 antibody.

Selection of sequences from a single-chain diabody library with a randomized middle linker (linker M)

A library of scDb fragments containing a randomized middle linker (linker M) of varying length was generated by a two fragment ligation procedure (see Figure 2a for details). For randomization we used triplets with the sequence -RVT- (R = A or G, V = A, C or G), with the six possible triplets coding for only six amino acids (T, N, S, D, G, A), thus avoiding all hydrophobic and the positively charged amino acids and reducing the total number of possible permutations. The randomized linker M was inserted between glycine 108 of the anti-β-galactosidase V_Lλ domain and glutamine 1 of the anti-β-galactosidase V_H domain. Based on homology to V_Lκ domains, where arginine or lysine 107 was defined as a C-

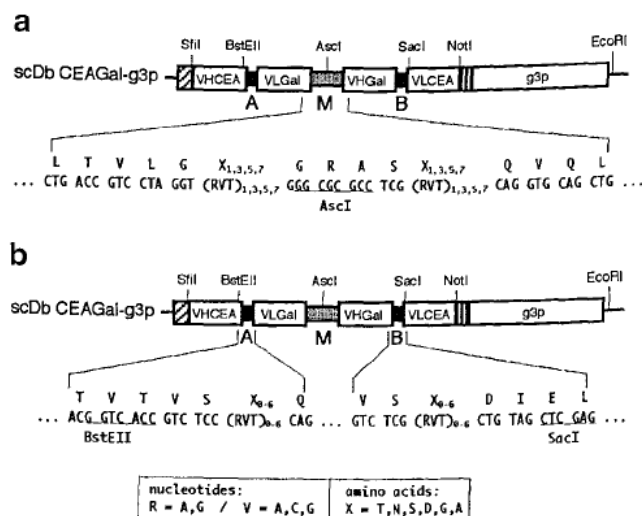


Fig. 2. Construction of scDb linker M (a) and linker AB (b) libraries. The linker regions and the positions randomized in length and sequence are indicated. The *AscI* site used for construction of scDb molecules is located in the middle of linker M and encodes the sequence –GRAS–. This sequence is conserved in all clones.

Table III. Linker M sequences of clones isolated from the linker M library

Clone	Linker M sequence	Length
wt	GGGGGSGGGRASGGGGGS	16
M13	GAGRASADN	9
M2	GTGSNTTDDGRAST	13
M4	GATTAADGRAAGAD	15
M9, M20	GNDATGNNGRASASD	15
M19	GDTGSDNNGRASDGG	15
M1	GSDSNAGRASAGNTS	15
M11	GGTNTDGRASAGNAA	15
M17	GNDNTAGRASDDTDN	15
M3	GTDTTADTGRASGDNTT	17
M14	GGDGGDGGRASDDDDNNSN	18

Linker M sequences were aligned with the conserved central –GRAS– motif.

terminal residue (Malby *et al.*, 1998; Dolezal *et al.*, 2000), leucine 107 can be defined as the C-terminus of the V_L domain (P.Hudson, personal communication). We therefore considered glycine 108 as part of the linker sequence. Thus, the combination of the possible linker fragments resulted in middle linkers of 7, 9, 11, 13, 15, 17 and 19 amino acid residues in length. The scDb linker M library with a total diversity of 1.7×10^6 clones was subjected to two rounds of selections on CEA and β -galactosidase (see Table III). Between 95 and 100% of the clones of round 2 recognized both antigens in ELISA. Over 80% of the sequenced clones contained a linker M with a total length of 15 or more residues. The majority (63%) had a linker M of 15 residues (Table III). Of 11 clones analysed only two had linkers with a length <15 residues (13 and 9 residues, respectively). In contrast, sequencing of randomly picked clones of the unselected library showed that the majority of clones (6/9) had a linker M with a length <15 residues. One clone (M14) possessed an unexpected linker length of 18 residues, which was most likely generated by mispriming at the 5' end of the linker-encoded sequence during PCR. Except for a slight under-representation

Table IV. Linker A, B and M sequences of clones isolated from the AB library

Clone	Linker A	Linker B	Linker M
wt	SGGGGS	SGGGGS	wt
AB16	N	DSSSTT	M3
AB8	DNA	DNGSTD	M1
AB6	GGD	TATGAD	M3
AB22	DSSD	NADST	M3
AB7	DDAA	NDSN	M3
AB3	GADSG	NGTN	M3
AB24	DSNTG	GTTNN	M1
AB12	GGDGA	NDASTN	M9
AB20	AGNSA	GSNNNSN	M3
AB13	GDGADAS	NTGT	M3

of serine residues, there was no significant deviation from the statistical usage of the six amino acids (16.7%) in the selected clones.

Selection of sequences from a single-chain diabody library with randomized linkers A and B

The linkers A and B of several scDb molecules (clones M1, M3, M4, M9, M11 and M19; see Table III) isolated from the linker M library were randomized in length and sequence introducing between 0 and 6 amino acids into the two linker regions in all possible combinations and the same degenerated triplets as for the linker M library (Figure 2b). Serine 112 was defined as the C-terminus of the V_H domains and residue 1 as the N-terminus of the V_L domains, as suggested earlier (Hudson and Kortt, 1999) (see Figure 2b). This linker AB library with a diversity of 2.7×10^7 was subjected to two rounds of alternating selection on CEA and β -galactosidase. All clones analysed from round 1 and 2 recognized both antigens in ELISA and 100% of the clones of round 2 were positive. Some differences in length were observed for linkers A and B (Table IV). The length of linker A was diverse, varying from one to seven amino acid residues. However, except for one clone, all linkers had a length of three or more amino acids. The seven amino acid residue linker of clone AB13 was, as for clone M14, most likely generated by mispriming during PCR, since this length was not originally present in the primers used for cloning. For linker B, all clones had a length of four, five or six amino acid residues. Differences were also observed for the amino acid usage. In linker A, only 3% threonine residues were found, whereas all other amino acids were within the expected range. The three clones with a linker A of one or three amino acids (clones AB16, AB6, AB8) all possessed a linker B of the maximum possible length of six residues. These data led us to suggest that a short linker at one site forces a longer linker at the other site in order to retain correct folding and full antigen binding activity. In order to investigate the assumption further, we generated a library of clones having linker M3 as middle linker and either a linker A of one residue and a linker B of 0–6 residues (linker A_1B_x library) or a linker B of one residue and a linker A of 0–6 residues (linker A_xB_1 library). These libraries with diversities of 6×10^4 and 4×10^4 , respectively, were subjected to one round of selection on CEA or β -galactosidase. Interestingly, 68% of the clones selected from the A_1B_x library with CEA and 13% of the clones selected on β -galactosidase reacted with both antigens, whereas fewer positive clones were isolated

Table V. Linker A and B sequences of clones isolated from the A₁B_x and A_xB₁ library

Library	Clone	Linker A	Linker B
A ₁ B _x	A ₁ 3	D	N D G S D S
	A ₁ 10	S	N T D S N T
	A ₁ 11	G	S N A A D S
	A ₁ 4	D	A A N A N
	A ₁ 5	D	T A T G N
A _x B ₁	A ₁ 12	D	G S A D T
	B ₁ 29	G	N
	B ₁ 17	G G	G
	B ₁ 26	T G G	N
	B ₁ 19	D T A T	D
	B ₁ 21	A D D A	D
	B ₁ 30	G S G A D	N

In library A₁B_x the length of linker A was fixed to one residue and linker B was 0–6 residues. In library A_xB₁ the length of linker B was fixed to one residue and linker A was 0–6 residues.

Table VI. Expression of monomeric and dimeric scDb fragments

Clone	Linker M (aa)	Linker A (aa)	Linker B (aa)	Yield (μg/l culture)	Format
wt	16	6	6	42	M
M3	17	6	6	35	M
M9	15	6	6	45	M
M2	13	6	6	50	M
M13	9	6	6	40	DM
M132	7	6	6	35	D
AB3	17	5	4	43	M
AB16	17	1	6	48	M
B ₁ 30	17	5	1	37	M
B ₁ 19	17	4	1	40	M
B ₁ 26	17	3	1	33	M
B ₁ 17	17	2	1	33	M
B ₁ 29	17	1	1	50	D/M
A ₀ B _{1D}	17	0	1	40	D/(M)
A ₀ B ₀	17	0	0	Insoluble	–

Soluble scDb molecules were purified by IMAC from periplasmic extracts of 3 l of bacterial culture. For linker sequences of the individual clones see Tables III, IV and V. M132 was constructed by combining the N-terminal part of linker M of clone M13 with the C-terminal part of linker M of clone M2, thus M132 contained a seven-residue linker M (GAGRAST). Clone A₀B_{1D} possessed an asparagine residue as linker B. M = monomer; D = dimer.

from the A_xB₁ library (26% from the CEA selection and 0% from the β-galactosidase selection). All clones isolated from the A₁B_x library had a linker B of five or six residues (Table V). These clones had preferentially (four out of six) an aspartic acid residue in the linker A₁. Different results were obtained from the selection of the A_xB₁ library. Here, the selected linkers A had a length between one and five residues with no observed preference for a certain length. A preference for glycine residues was found for linker A of 1–3 amino acid residues. Asparagine or aspartic acid residues were mainly found at the B₁ linker.

Expression and antigen binding of scDb molecules with new linker sequences

Similar yields of purified protein were obtained for various clones (wt, M3, M9, M2, M13, AB16, AB3, B₁29 and B₁30) containing linkers of different length and composition (Table VI). Thus, the selected linker sequences did not significantly improve the yields of soluble protein. SDS-PAGE and immuno-

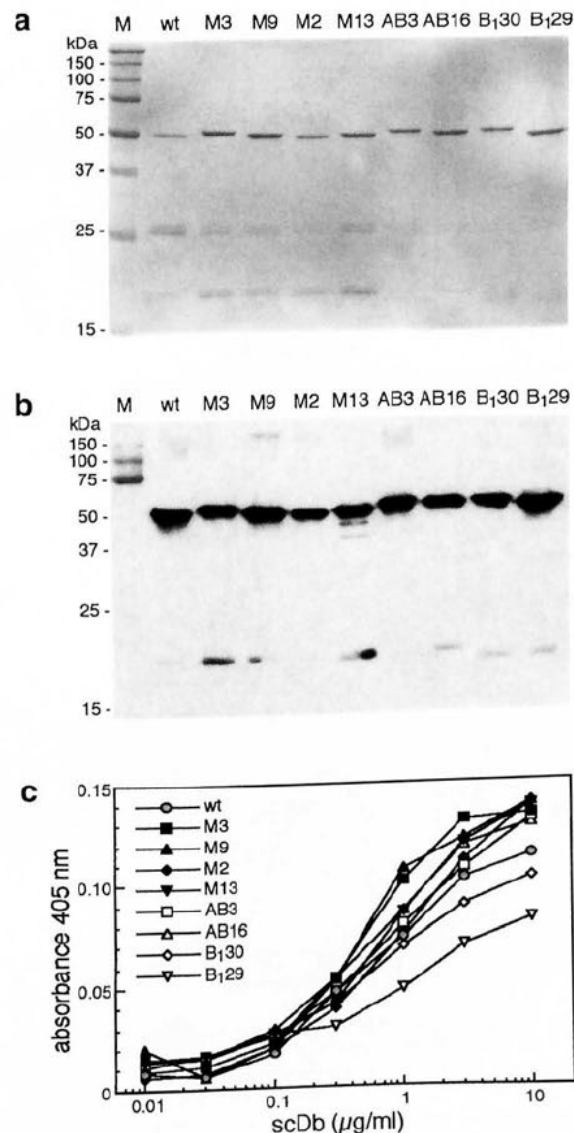


Fig. 3. SDS-PAGE and functional analysis of scDb. Purified proteins were analysed by 10% SDS-PAGE and stained with Coomassie Brilliant Blue (a) or immunoblotted with an anti-His-tag antibody (b). The dual specificity of the purified molecules was analysed by scDb-mediated recruitment of β-galactosidase (50 μg/ml) to plastic-bound CEA and detection of β-galactosidase with ONPG (c).

blotting experiments confirmed the correct size of the purified scDb polypeptides (Figure 3a and b). All the purified scDb molecules were active in ELISA with plastic-bound CEA (not shown). Simultaneous binding of both antigens was demonstrated in experiments recruiting β-galactosidase to CEA by the scDbs and the detection of bound enzyme by substrate conversion (Figure 3c).

Monomeric and dimeric assembly of scDb

Wild-type scDb and also clones M3, M9, M2, AB3, AB16 and B₁30 showed a predominant peak after separation by size exclusion chromatography on a Superose 12 column corresponding in size to monomeric molecules (~55 kDa) (Table VI). In contrast, scDb M13, possessing a nine-residue middle linker and scDb B₁29, with one-residue linkers A and B, showed two peaks corresponding in size to monomeric

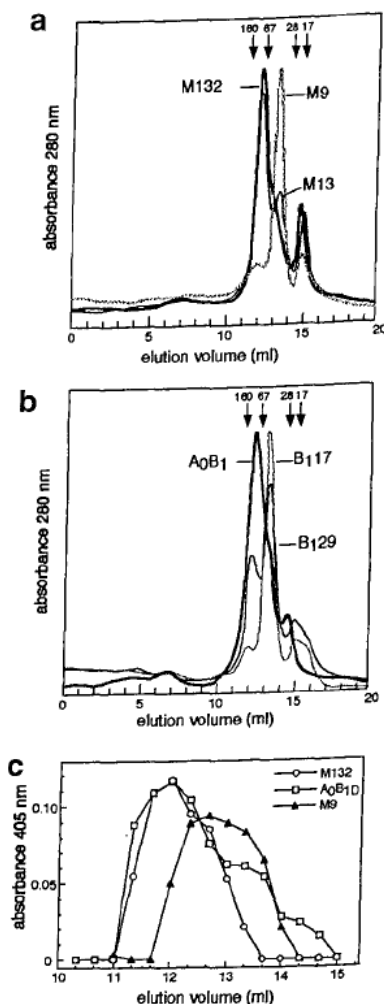


Fig. 4. (a, b) Gel filtration of scDb possessing different linker sequences. 200 μ l of purified scDb (see Table VI), corresponding to ~30–40 μ g protein, were separated by FPLC size exclusion chromatography on a Superose 12 column and 333 μ l fractions were collected. IgG, BSA, an scFv fragment and myoglobin were used as molecular weight standards. An overlay of the elution profiles of clones M9, M13, M132 is shown in (a) and of clones B117, B129, A0B1D in (b). The purified scDb contained to varying extents non-functional proteins with molecular weights <28 kDa. (c) Recruitment experiment with fractionated proteins of clones M132, A0B1D and M9 (see Table VI). Experiments were performed as described (Figure 3) using 50 μ l of each fraction and β -galactosidase at a concentration of 50 μ g/ml.

and dimeric molecules (Figure 4). Fractions of monomeric and dimeric scDb contained functional molecules as demonstrated by ELISA and recruitment experiments (not shown). The purified scDb also contained to various extents some low molecular weight contaminants (>20 kDa) with no antigen-binding activity.

In order to analyse whether further reduction of linker M below a length of nine residues has an influence on the production of dimeric molecules, we generated an additional construct (M132) combining the N-terminal part of linker M of clone M13 with the C-terminal part of linker M of clone 2. Thus, M132 contains a seven-residue linker M of the sequence GAGRAST. Size exclusion chromatography showed that M132 contained almost exclusively dimeric molecules and there was a clear shift from monomeric to dimeric molecules compared with M13 (Figure 4a; Table VI).

In order to analyse further the influence of the length of linkers A and B on the transition from monomeric to dimeric molecules, we analysed clones B117, B126 and B119 containing a linker A of two, three or four residues, respectively (see Table V). In addition, we generated two constructs (A0B1D, A0B0) containing a zero-linker A and either a one-residue (Asp) or no-residue linker B. Thus, in construct A0B0 the V_H domains are directly linked to the V_L domains without the presence of any additional residues. Size exclusion chromatography of purified proteins showed monomeric molecules for B117, B126 and B119 (Figure 4b; Table VI). In contrast, A0B1D contained mainly dimeric molecules. Only a small shoulder at the right side of the main peak indicated the presence of some monomeric molecules. The A0B0 preparation did not lead to significant amounts of soluble protein and was therefore not further analysed by size exclusion chromatography.

Recruitment experiments of the fractionated samples of M132 and A0B1D confirmed the presence of active bispecific molecules in the fractions containing dimeric molecules (Figure 4c). These could be clearly separated from the active fractions of M9, which were included as a representative sample of monomeric scDb.

In order to analyse further whether the dimeric scDb molecules exhibit an increased binding to their antigens compared with monomeric scDb, we performed additional recruitment experiments. A titration of antibody molecules showed an identical recruitment efficiency for M9 and M132, whereas that of A0B1D was slightly reduced compared with M9 (Figure 5a). This finding indicates that simultaneous binding of the two identical antigen-binding sites of dimeric scDbs is sterically hindered under the applied assay conditions. We therefore analysed the possibility of the antibody molecules recruiting soluble horseradish peroxidase (HRP)-conjugated antigen to immobilized antigen. This experiment showed that the two dimeric scDbs M132 and A0B1D were able to recruit HRP-conjugated β -galactosidase to plastic-bound β -galactosidase, whereas the monomeric scDb M9 did not mediate any recruitment (Figure 5b). Hence this experiment confirmed the presence of two active β -galactosidase binding sites in scDb M132 and A0B1D. As expected, recruitment of β -galactosidase to plastic-bound CEA, i.e. bispecific interaction, was observed for all three constructs. Similar results were found for HRP-conjugated CEA, although some weak recruitment of CEA-HRP to immobilized CEA was also seen with M9 (Figure 5c).

Discussion

The formation of monomeric or dimeric single-chain diabodies critically depends on the length of the three linkers joining the four variable domains of one chain. Selections of the middle linker, which resembles the linker used in scFv fragments, confirmed previous findings that linkers with a length of 13 or more amino acid residues are necessary for correct pairing of adjacent variable domains to form an antigen-binding site (Kortt *et al.*, 1994; Aflthan *et al.*, 1995). Interestingly, although our library contained linkers with a length of up to 19 amino acid residues, the preferred linker length was 15 residues. The optimal linker length may, however, vary between different V_H and V_L domains. In addition, the orientation of the two domains (V_H - V_L or V_L - V_H) can influence the optimal linker length. The distance between the C-terminus of V_L and the N-terminus of V_H is ~39–43 Å, whereas that between the C-terminus of V_H and the N-terminus of V_L is only 32–34 Å

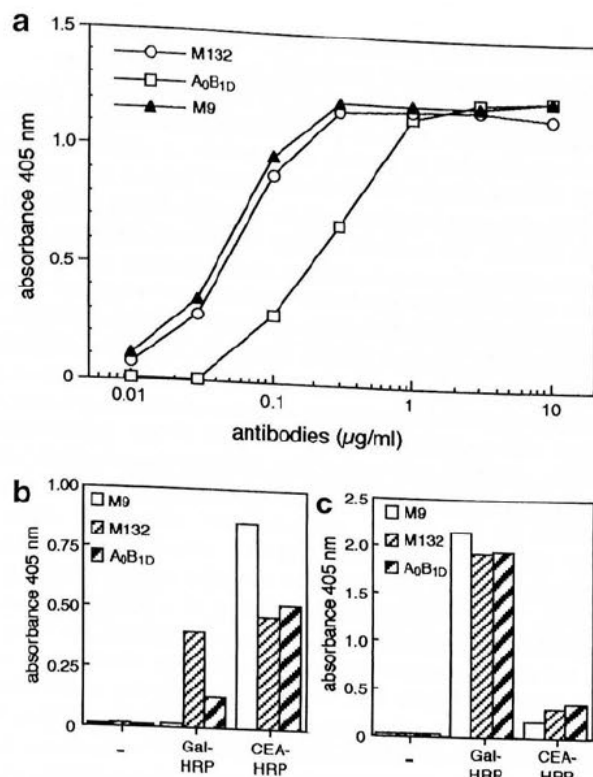


Fig. 5. (a) Recruitment of β -galactosidase to plastic-bound CEA mediated by monomeric scDb M9 and dimeric scDbs M132 and A0B1D as described in Figure 4c. (b, c) ScDb-mediated recruitment of HRP-conjugated β -galactosidase (Gal-HRP) or CEA (CEA-HRP) to plastic-bound β -galactosidase (b) or CEA (c). HRP-conjugated antigens were preincubated with scDb M9, M132 or A0B1D and were then added to immobilized antigens. Bound antigens were detected by addition of HRP substrate. This experiment confirmed bispecific binding of all three constructs and demonstrated bivalent interaction with the of β -galactosidase or CEA for M132 and A0B1D.

(Huston *et al.*, 1988; Plückthun and Pack, 1997). For this reason, other groups used 20 amino acid residues, instead of 15 residues, to link the V_L domain to a V_H domain in order to avoid formation of diabodies (Krebbel *et al.*, 1997). In the scDb format used in our study the middle linker connects a V_L domain with a V_H domain. Our findings indicate that 13–15 residues are sufficient to span this distance.

One of the selected clones (M13) possessed a middle linker of nine residues. These nine residues can span a maximum distance of ~ 32 Å, which is probably too short to form an Fv domain from the adjacent variable domains. In accordance with this finding, M13 contained a significant amount of dimeric molecules. Further reduction of linker M to seven residues increased the amount of dimeric molecules. The observed transition from monomeric to dimeric scDb by reduction of the length of linker M from 13 to seven residues resembles the transition from monomeric scFv to bivalent diabodies. This indicates that the core region of dimeric scDb, with a short linker M connecting a V_L and a V_H domain, may resemble the structure of a bivalent diabody (see Figure 6). Hence the overall structure of these dimeric molecules would be different from that postulated by Kipriyanov *et al.* (Kipriyanov *et al.*, 1999). They generated dimeric bispecific single-chain diabodies possessing a middle linker of 12 residues and linkers A and B of nine residues. They presented a model of homodimers (assigned tandem diabodies or Tandabs) with all

Monomeric and dimeric scDb

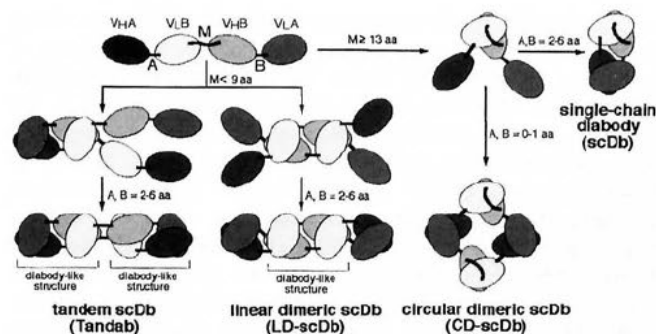


Fig. 6. Overview of the putative multi-step assembly processes leading to bispecific monomeric or dimeric scDb molecules. The assembly of the four variable domains into a monomeric scDb requires a middle linker M of 13 or more residues and linkers A and B of 2–6 residues. If linkers A and B are too short (i.e. of 0–1 residue in length) to allow intrachain pairing of the V_{H1} and V_{L1} domains, dimeric scDb are formed by interchain pairing of these domains. These circular dimeric scDb (CD-scDb) have most likely a similar overall structure as the previously described tetrameric tetraabodies (Hudson and Kortt, 1999). Intrachain pairing of the V_{L1} and V_{H1} domains is disfavoured by a linker M of ≤ 9 amino acids. This results in the formation of dimeric scDb by interchain pairing of all variable domains. We propose two putative pathways for assembly of these dimeric scDb, which requires linkers A and B of ≥ 2 residues. In the first pathway, assembly takes place from the end of the scDb polypeptides. This stepwise assembly leads to the formation of two diabody-like structures connected by two linkers M. Such a structure has been proposed previously for a single-chain diabody with a 12-residue middle linker (Kipriyanov *et al.*, 1999) and was assigned as tandem diabody (Tandab). In the second pathway, the central domains assemble first into a diabody-like structure with two identical binding sites. Subsequently, the flanking domains assemble into antigen-binding sites of the second specificity. We assigned these molecules as linear dimeric scDb (LD-scDb).

four variable domains of one chain interacting with the variable domains of the second chain. In this model, two diabody-like structures are connected in a flexible manner by the middle linkers (see Figure 6). Assuming a diabody-like structure for the core region, it is not possible, however, for the adjacent variable domains to fold with the core domains into a diabody-like structure owing to the occupation of the V_H – V_H interface by this core structure (Peresic *et al.*, 1994). Thus, in a model with a central diabody-like core structure, the flanking variable domains must fold into a molecule resulting in a yet unknown overall structure. We assigned these molecules linear dimeric single-chain diabodies (LD-scDb).

The formation of tandem diabodies or LD-scDb may depend on the order of assembly (Figure 6). While assembly of the variable domains of two chains from the ends may result in tandem diabodies, the assembly of the central variable domains first followed by assembly of the flanking domains into antigen-binding sites may lead to LD-scDb. However, this assumption is speculative and further studies are needed to show the correct structure of these molecules.

For linkers A and B, which are usually chosen to be 5–6 amino acid residues for expression of bivalent or bispecific diabodies or single-chain diabodies, we found a preferred length of 3–6 residues. This is in agreement with various findings obtained with bivalent diabodies which demonstrated that a reduction of these linkers below 2–3 residues results in the formation of tri- or tetrameric molecules (Atwell *et al.*, 1999; Hudson and Kortt, 1999; Le Gall *et al.*, 1999). Again, this was dependent on the antibody fragment used and the orientation of the V_H and V_L domains (Dolezal *et al.*, 2000).

Reduction of the length of linkers A and B below two

residues results in the formation of dimeric molecules. Since linker M is of sufficient length, the two variable domains of one chain connected by linker M can fold into an antigen-binding site. However, owing to short linkers A and B, the flanking variable domains within one chain cannot assemble into antigen-binding sites leading to interchain pairing of these domains (Figure 6). We assume that the structure of these dimeric single-chain diabodies closely resembles that of tetrameric scFv molecules (tetrabodies) (Hudson and Kortt, 1999). Hence these molecules are structurally different from the above-described linear dimeric single-chain diabody molecules. We assigned these molecules as circular dimeric single-chain diabodies (CD-scDb). Compared with tetrabodies, which are monospecific homotetramers, CD-scDb are bispecific homodimers. Owing to random association of the four polypeptide chains forming a tetra-body, the generation of bispecific tetrabodies is limited and has not yet been described. Hence the CD-scDb format allows the generation of small bispecific and tetravalent molecules with well-defined antigen-binding sites. It will, of course, also be possible to use this format to generate monospecific tetravalent molecules. As has been shown for various multimeric scFv molecules and also tandem diabodies, such molecules possess an increased functional affinity due to di- or multivalent antigen binding and improved pharmacokinetic properties (Hudson and Kortt, 1999; Kipriyanov *et al.*, 1999). No improved functional affinity was observed in our experiments analysing the binding of the scDb molecules to plastic-bound antigen. This indicates that under these conditions only one binding site of the molecules can interact at the same time. However, our experiments analysing the recruitment of soluble HRP-conjugated antigen to immobilized antigen demonstrated that the dimeric scDb molecules possessed two functional and accessible binding sites for each antigen, thus confirming the tetravalent and bispecific nature of these molecules.

Selection of the A_1B_x library with a one-residue linker A resulted in the isolation of clones possessing a long linker B. Thus, a short linker A in our scDb molecule promotes the selection of a long linker B. This indicates that short linkers A and B in a monomeric molecule are disfavoured, presumably for steric reasons. Although we could show that polypeptides with short linkers A and B assemble into functional dimeric molecules, scDb which form exclusively dimeric molecules (e.g. M132) were not selected from phage display libraries. This indicates that phage display forces the selection of scDb with linkers which allow monomeric assembly.

The selection of the A_xB_1 library, with linker B restricted to one residue, did result in the isolation of clones with linkers A having a length of 1–5 residues. Thus, in contrast to the observation with a one-residue linker A, one-residue linker B does not necessitate a long linker A. This might be explained by the introduction of some structural constraints by a short linker B. For example, it might be possible that a short linker B leads to a tight back-to-back packing of the V_H domains, which also brings the two domains connected by linker A into close proximity. Thus linker A can be very short without interfering with the assembly of functional molecules.

It can be envisaged that the use of linkers of different length, with one linker being very short, results in scDb with a bend structure. Indeed, electron microscopy studies by Hudson's group in Australia revealed boomerang-shaped projections of bivalent diabodies. The angle between the two arms ranged from 60 to 180°, with a mean angle of 122° (Lawrence

et al., 1998). This variation in the angle between the arms points to a considerable flexibility in the linker region.

Our approach can be applied to identify a set of minimal linker sequences. This is particularly interesting for therapeutic applications to reduce the number of additional amino acids, which might be immunogenic. However, immunogenicity may also be influenced by the amino acid composition of the linkers. We isolated a large variety of different linkers using a minimal set of codons encoding only six different amino acids. Since the linker regions are exposed to the solvent, we focused in this study on small aliphatic, hydrophilic and negatively charged amino acids. In a similar approach, Hennecke *et al.* generated a library of an anti-fluorescein scFv with sequences at the beginning and end of the linker encoding β -turns and a randomized middle part of the linker where glycines, prolines and polar and charged amino acids were allowed to occur at certain positions (Hennecke *et al.*, 1998). Selection of this library applying the selectively infective phage (SIP) technology resulted in various non-repetitive sequences with no obvious bias towards certain amino acid residues. These new linkers performed in a manner identical with that of the original $(Gly_4Ser)_3$ linker. Similarly, no obvious bias towards a specific linker sequence was observed in our study and the linkers that we selected did not significantly improve the yields of soluble protein. This indicates that the variable domains of scDb CEAGal are responsible for the low expression levels observed in our study. Other groups used triplets coding for all 20 amino acids and phage display to isolate new linkers for the generation of scFv fragments. Although one study described that proline and positively charged amino acids were more abundant and negatively charged amino acids disfavoured (Tang *et al.*, 1996), another study did not result in the isolation of linkers with any amino acid preference or a significant sequence similarity (Turner *et al.*, 1997). Starting from poorly expressed scFv, in these two studies an improved secretion and solubility of scFv fragments were described compared with the parental scFv fragments containing a $(Gly_4Ser)_3$ linker.

Further applications of phage displayed scDb include the generation of naive libraries for the direct selection of bispecific scDb. Such libraries can be either generated by random assembly of all four variable domains or by keeping one antigen-binding site fixed and varying only the variable domains of the second antigen-binding site. For example, bispecific scDb used for the retargeting of cytotoxic T lymphocytes (T.Korn and R.E.Kontermann, unpublished work) could be isolated from a library containing the variable domains directed against CD3. These libraries could be directly used to select correctly folded scDb molecules against a variety of different target cells or molecules.

Acknowledgement

We thank Margitta Alt for excellent technical assistance.

References

- Aflthan, K., Takkinen, K., Sizmann, D., Söderlund, H. and Teeri, T.T. (1995) *Protein Eng.*, **8**, 725–731.
- Alt, M., Müller, R. and Kontermann, R.E. (1999) *FEBS Lett.*, **454**, 90–94.
- Atwell, J.L., Breheny, K.A., Lawrence, L.J., McCoy, A.J., Kortt, A.A., Hudson, P.J. (1999) *Protein Eng.*, **12**, 597–604.
- Brüsselbach, S., Korn, T., Völkel, T., Müller, R. and Kontermann, R.E. (1999) *Tumor Targeting*, **4**, 115–123.
- Cochlovius, B., Kipriyanov, S.M., Stassar, M.J., Schuhmacher, J., Benner, A., Moldenhauer, G. and Little, M. (2000) *Cancer Res.*, **60**, 4336–4341.

- Dolezal, O., Pearce, L.A., Lawrence, L.J., McCoy, A.J., Hudson, P.J. and Kortt, A.A. (2000) *Protein Eng.*, **13**, 565–574.
- Hennecke, F., Krebber, C. and Plückthun, A. (1998) *Protein Eng.*, **11**, 405–410.
- Holliger, P. and Winter, G. (1997) *Cancer Immunol. Immunother.*, **45**, 128–130.
- Holliger, P., Prospero, T.D. and Winter, G. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 6444–6448.
- Holliger, P., Brissinck, J., Williams, R.L., Thielemans, K. and Winter, G. (1996) *Protein Eng.*, **9**, 299–305.
- Holliger, P., Wing, M., Pound, J.D., Bohlen, H. and Winter, G. (1997) *Nat. Biotechnol.*, **15**, 632–636.
- Hudson, P.J. and Kortt, A.A. (1999) *J. Immunol. Methods*, **231**, 177–189.
- Huston, J.S. et al. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 5879–5883.
- Iliades, P., Kortt, A.A. and Hudson, P.J. (1997) *FEBS Lett.*, **409**, 437–441.
- Kipriyanov, S., Moldenhauer, G., Schumacher, J., Cochlovius, B., von der Lieth, C.-W., Matys, E.R. and Little, M. (1999) *J. Mol. Biol.*, **293**, 41–56.
- Kontermann, R.E. (2001) In Kontermann, R.E. and Dübél, S. (eds), *Antibody Engineering, a Laboratory Manual*. Springer, Heidelberg, pp. 137–148.
- Kontermann, R.E. and Müller, R. (1999) *J. Immunol. Methods*, **226**, 179–188.
- Kontermann, R.E., Martineau, P., Cummings, C.E., Karpas, A., Allen, D., Derbyshire, E. and Winter, G. (1997a) *Immunotechnology*, **3**, 137–144.
- Kontermann, R.E., Wing, M.G. and Winter, G. (1997b) *Nature Biotechnol.*, **15**, 629–632.
- Kortt, A.A. et al. (1994) *Eur. J. Biochem.*, **221**, 151–157.
- Kortt, A.A. et al. (1997) *Protein Eng.*, **10**, 423–433.
- Krebber, A., Bornhauser, S., Burmester, J., Honegger, A., Willuda, J., Bosshard, H.R. and Plückthun, A. (1997) *J. Immunol. Methods*, **201**, 35–55.
- Krebs, B., Ackermann, B. and Rose-John, S. (1998) *J. Interferon Cytokine Res.*, **18**, 783–791.
- Lawrence, L.J., Kortt, A.A., Iliades, P., Tulloch, P.A. and Hudson, P.J. (1998) *FEBS Lett.*, **425**, 479–484.
- Le Gall, F., Kipriyanov, S.M., Moldenhauer, G. and Little, M. (1999) *FEBS Lett.*, **453**, 164–168.
- Malby, R.L., McCoy, A.J., Kortt, A.A., Hudson, P.J. and Colman, P.M. (1998) *J. Mol. Biol.*, **279**, 901–910.
- Marks, J.D., Hoogenboom, H.R., Bonnert, T.P., McCafferty, J., Griffiths, A.D. and Winter, G. (1991) *J. Mol. Biol.*, **222**, 581–597.
- Pei, X.Y., Holliger, P., Murzin, A.G. and Williams, R.L. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 9637–9642.
- Peresic, O., Webb, P.A., Holliger, P., Winter, G. and Williams, R.L. (1994) *Structure*, **2**, 1217–1226.
- Plückthun, A. and Pack, P. (1997) *Immunotechnology*, **3**, 83–105.
- Tang, Y., Jiang, N., Parakh, C. and Hilvert, D. (1996) *J. Biol. Chem.*, **271**, 15682–15686.
- Turner, D.J., Ritter, M.A. and George, A.J.T. (1997) *J. Immunol. Methods*, **205**, 43–54.

Received February 9, 2001; revised June 26, 2001; accepted July 10, 2001