doi:10.1016/j.jmb.2007.02.029 J. Mol. Biol. (2007) 368, 1024–1041

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Evolution of an Interloop Disulfide Bond in High-Affinity Antibody Mimics Based on Fibronectin Type III Domain and Selected by Yeast Surface Display: Molecular Convergence with Single-Domain Camelid and Shark Antibodies

Daša Lipovšek[1]*, Shaun M. Lippow[2], Benjamin J. Hackel[2] Melissa W. Gregson[3], Paul Cheng[2], Atul Kapila[3] and K. Dane Wittrup[1,2]*

[1]*Biological Engineering Division, Massachusetts Institute of Technology Cambridge, MA 02139, USA*

[2]*Department of Chemical Engineering, Massachusetts Institute of Technology Cambridge, MA 02139, USA*

[3]*Department of Biology Massachusetts Institute of Technology, Cambridge MA 02139, USA*

**Corresponding authors*

The 10th human fibronectin type III domain (^{10}Fn3) is one of several protein scaffolds used to design and select families of proteins that bind with high affinity and specificity to macromolecular targets. To date, the highest affinity ^{10}Fn3 variants have been selected by mRNA display of libraries generated by randomizing all three complementarity-determining region -like loops of the ^{10}Fn3 scaffold. The sub-nanomolar affinities of such antibody mimics have been attributed to the extremely large size of the library accessible by mRNA display (10^{12} unique sequences).

Here we describe the selection and affinity maturation of ^{10}Fn3-based antibody mimics with dissociation constants as low as 350 pM selected from significantly smaller libraries (10^7–10^9 different sequences), which were constructed by randomizing only 14 ^{10}Fn3 residues. The finding that two adjacent loops in human ^{10}Fn3 provide a large enough variable surface area to select high-affinity antibody mimics is significant because a smaller deviation from wild-type ^{10}Fn3 sequence is associated with a higher stability of selected antibody mimics. Our results also demonstrate the utility of an affinity-maturation strategy that led to a 340-fold improvement in affinity by maximizing sampling of sequence space close to the original selected antibody mimic.

A striking feature of the highest affinity antibody mimics selected against lysozyme is a pair of cysteines on adjacent loops, in positions 28 and 77, which are critical for the affinity of the ^{10}Fn3 variant for its target and are close enough to form a disulfide bond. The selection of this cysteine pair is structurally analogous to the natural evolution of disulfide bonds found in new antigen receptors of cartilaginous fish and in camelid heavy-chain variable domains. We propose that future library designs incorporating such an interloop disulfide will further facilitate the selection of high-affinity, highly stable antibody mimics from libraries accessible to phage and yeast surface display methods.

Keywords: fibronectin; ^{10}Fn3; antibody mimic; yeast surface display; interloop disulfide

Abbreviations used: ^{10}Fn3, 10th human fibronectin type III domain; BSA, bovine serum albumin; CDR, complementarity-determining regions; EDTA, ethylenediaminetetraacetic acid; FACS, fluorescence-activated cell sorts; MACS, magnetic-activated cell sorts; PBS, phosphate-buffered saline.

E-mail addresses of the corresponding authors: dlipovsek@ML1.net; wittrup@MIT.edu

Introduction

The 10th human fibronectin type III domain (^{10}Fn3) is one of the protein scaffolds used in recent years to design and select, *in vitro*, families of proteins that bind with high affinity and specificity to a variety of macromolecular targets.[1–3] Such protein scaffolds, also known as antibody mimics, are designed to play many of the roles of antibodies and their fragments in research, diagnostics, and therapy; in addition, they often have the advantages of small size, inexpensive microbial production, and extreme stability. Wild-type ^{10}Fn3, for example, is only 10 kDa in size, is expressed at a high level in *Escherichia coli*, and has a melting temperature of 82 °C.[4] Despite its lack of sequence homology with antibody domains, ^{10}Fn3 has an immunoglobulin-like fold;[5,6] its solvent-accessible loops BC, DE, and FG are reminiscent of complementarity-determining regions (CDR) 1, 2, and 3, respectively. Wild-type ^{10}Fn3 contains no disulfide bonds, which allows it to preserve its stability under reducing conditions.

Variants of ^{10}Fn3 that bind protein targets with high affinity and specificity have been selected using phage display,[7–9] yeast two-hybrid,[10] and mRNA display.[4,11] In addition to binding their targets in solution and when immobilized on a solid surface,[11] ^{10}Fn3-based antibody mimics have been shown to work as Western-blot detection reagents[9] and affinity-purification reagents,[12] to block the interaction between a ligand and its natural receptor,[4] and to recognize specific conformations of their target.[10]

Dissociation constants of target-binding ^{10}Fn3 variants selected to date have varied from picomolar to micromolar. For example, ^{10}Fn3 variants with 20 pM affinity for TNF-α[11] and with 340 pM affinity for vascular endothelial growth factor receptor 2[4] were selected by mRNA display from libraries of 10^{12} unique sequences, which had been constructed by randomizing 21 residues in three loops (7 randomized residues in loop BC, 4 randomized residues in loop DE, and 10 randomized residues in loop FG). In contrast, ^{10}Fn3 variants with 1.3 μM and 250 nM affinity for the Src SH3 domain were selected by phage display from a library of 2×10^9 unique sequences, which had been constructed by randomizing 10 residues in two loops (5 randomized residues in loop BC and 5 randomized residues in loop FG, which also contained a 3-residue deletion). Whereas it appears likely that the 500-fold lower complexity of the phage-display library precluded the sampling of highest affinity ^{10}Fn3 variants,[9] an alternative explanation is that the randomization of two loops did not create a large enough target-binding surface for high-affinity binding. In this study, we set out to determine whether ^{10}Fn3 variants with sub-nanomolar affinity could be selected from libraries with only one or two randomized loops. This question is salient because a larger number of mutations from wild-type ^{10}Fn3 is associated with lowered stability and solubility,[4] as well as, presumably, a higher risk of immunogenicity. Stability, solubility, and immunogenicity of ^{10}Fn3-based antibody mimics will greatly affect their suitability for therapeutic applications.

We constructed ^{10}Fn3-based libraries of 10^7–10^8 unique, full-length sequences by randomizing 7 residues in loop BC (library BC7), 7 residues in loop FG (library FG7), or the same 14 residues in both loops BC and FG (libraries 2L14 and BF14, which differ only in the complexity of the library). We then used yeast surface display[13] to select ^{10}Fn3 variants that bind hen egg white lysozyme, an antigen used extensively in structural studies of antibody–antigen binding.[14–17] As expected, we found that antibody mimics selected from libraries with two randomized loops (libraries BF14 and 2L14) had higher affinity for lysozyme than those selected from libraries with a single randomized loop (libraries BC7 and FG7) and that antibody mimics selected from a library with a higher complexity (library BF14) had higher affinity for lysozyme than those selected from a library of the same design with a lower complexity (library 2L14). A striking feature of the highest affinity antibody mimics selected from a naive library is a pair of cysteines in structurally adjacent positions on the two loops, which appear to form a disulfide bond reminiscent of those found in shark[17,18] and camel[16,19] antibody variable domains. The antibody mimic with the lowest K_d, 350 pM, was obtained by an affinity maturation method similar to CDR walking originally reported for antibodies.[20]

Results

Library construction

We constructed four different libraries based on the wild-type human ^{10}Fn3 sequence by replacing codons in DNA stretches that encode 7 residues in each of two CDR-like loops, BC and FG, with mixtures of codons that encode all 20 amino acids. Library BC7 was constructed by randomizing the 7 residues (23–29) in loop BC; library FG7 was constructed by randomizing 7 residues (77–83) in loop FG; and libraries 2L14 and BF14 were constructed by simultaneously randomizing the 14 residues in loops BC and FG (23–29 and 77–83). The only difference between libraries 2L14 and BF14 is that the complexity of library BF14 is 10 times higher than the complexity of library 2L14.

The libraries were constructed by incorporation of triplet codons into oligonucleotides for polymerase-directed gene assembly, followed by transformation of yeast by homologous recombination-driven plasmid gap repair. The number of unique clones in each library, as estimated from the number of *Saccharomyces cerevisieae* transformants obtained,[21,22] varied from 3×10^7 (library FG7, 7 randomized residues in the FG loop) to 6×10^8 (library BF14, 14 randomized residues, 7 in the BC loop and 7 in the FG loop). Approximately 40% of sequenced clones from each library conformed to library design, whereas approximately 50% of the clones contained frameshift mutations that resulted in gene truncation, and

approximately 10% of the clones contained mutations in the ^{10}Fn3 sequence outside randomized loops. Deletions and insertions in randomized loops that were not associated with frameshift mutation were seen in only approximately 1% of sequenced clones. Functional library complexity, or the estimated number of unique, full-length clones in the seven libraries constructed, is listed in Table 1. This number is half the number of transformants obtained during library construction. Also listed in Table 1 is the proportion of all possible permutation of the positions randomized that can be sampled by each library. For the five libraries with 7 to 8 randomized residues, BC7, FG7, BFs1, BFs2, and BFs3 (the last three libraries are anti-lysozyme affinity maturation intermediates), this number ("sampling") varied from 1 to 4%; for the two libraries with 14 randomized residues, 2L14 and BF14, it was between 10^{-11} and 10^{-10}.

The distribution of amino acid residues in the randomized positions in BC and FG loops of naive libraries (Figure 1(a), white bars) varied from the expected frequency of 5% for each amino acid residue. The three residues overrepresented by more than 2% were Trp, Phe, and Tyr, with frequencies of 11, 10, and 8.2%, respectively. The three residues underrepresented by more than 2% were Leu, Arg, and Asp, with frequencies of 2.1, 2.3, and 2.5%, respectively.

Selection of antibody mimics that bind hen egg white lysozyme

The first round of selection, which used magnetic beads conjugated to anti-biotin antibody (magnetic-activated cell sorts (MACS)), captured 0.02–0.06% of yeast populations displaying libraries BC7, FG7, 2L14, and BF14 and 0.2–0.3% of yeast populations displaying affinity-maturation libraries BFs1, BFs2, and BFs3.

Three consecutive fluorescence-activated cell sorts (FACS) of MACS-sorted populations from the BC7 library, where yeast was labeled with 1 μM biotinylated lysozyme (HEL-b), showed no HEL-b–specific binding and no enriched sequences (data not shown); consequently, selection of lysozyme-binding antibody mimics from library BC7 was aborted.

FACS sorting of the other three first-generation libraries, FG7, 2L14, and BF14, led to an increase of HEL-b–specific signal and to enrichment of HEL-b–binding sequences as selection progressed (Table 2, Figure 2). For all three libraries, three FACS sorts with labeling at 1 μM HEL-b were sufficient to identify antibody mimics that had the highest affinity for HEL-b. From library FG7, which had been constructed by randomizing seven residues in loop FG of human ^{10}Fn3, four closely related clones were selected. The two related sequences present in more than 10% of all sequenced clones, FG4.1 and FG4.2, had dissociation constants of approximately 610 and 750 nM, respectively (Figure 2). From libraries 2L14 and BF14, which had been constructed by simultaneously randomizing seven residues in loop BC and seven residues in loop FG of human ^{10}Fn3, clones 2L4.1, BF4.4, BF4.9, and BF4.1 were selected as 10% or more of the sequenced population; their dissociation constants were 170, 120, 140, and 180 nM, respectively. Four additional rounds of FACS sorting with increasing stringency after labeling yeast at 100 and 10 nM HEL-b did not identify any higher-affinity lysozyme-binding clones, but did reduce the fraction of the lowest-affinity clone selected from library BF14, BF4.3, to a level where it could not be detected by sequencing 27 clones randomly picked from the pool selected after eight rounds of sorting.

Libraries 2L14 and BF14 differed only in their complexity, 5×10^7 and 6×10^8, respectively; as expected, the best clones selected from the larger library had the higher affinity for hen egg white

Table 1. Design and complexity of ^{10}Fn3-based antibody-mimic libraries used in the selection for lysozyme binding

Name	BC[1]	FG[2]	Library complexity[3]	Theoretical complexity[4]	Sampling[5]
^{10}Fn3[a]	DAPAVTV	GRGDSPA	1	1	1
BC7	XXXXXXX[d]	GRGDSPA	2×10^7	1×10^9	2×10^{-2}
FG7	DAPAVTV	XXXXXXX[d]	2×10^7	1×10^9	2×10^{-2}
2L14	XXXXXXX[d]	XXXXXXX[d]	3×10^7	2×10^{18}	2×10^{-11}
BF14	XXXXXXX[d]	XXXXXXX[d]	3×10^8	2×10^{18}	2×10^{-10}
FG4.1[b]	DAPAVTV	WWDWEQA	1	1	1
BFs1	XXXXXXX[d]	WWDWEQA	5×10^7	1×10^9	4×10^{-2}
BF4.4[c]	WTAHYCD	CYSVFNY	1	1	1
BFs2	XXXXXXX[d]	OYSVFNY[e]	8×10^7	3×10^9	3×10^{-2}
BFs3	WTAHYOD[e]	XXXXXXX[d]	3×10^7	3×10^9	1×10^{-2}

[1]Sequence in loop BC, amino acid residues 23–29.
[2]Sequence in loop FG, amino acid residues 77–83.
[3]Estimated number of unique, full-length clones in a library.
[4]Number of theoretically possible permutation of sequences resulting from the randomization scheme, also known as theoretical sequence space.
[5]Library complexity divided by theoretical complexity.
[a] Wild-type ^{10}Fn3.
[b] Clone FG4.1 selected from library FG7.
[c] Clone BF4.4 selected from library BF14.
[d] X, equimolar mixture of all amino acids.
[e] O, equimolar mixture of serine and cysteine.

lysozyme, but the difference was slight (120 nM for BF4.4 *versus* 170 nM for 2L4.1). Clone 2L4.1 contained a rare deletion in loop BC.

A striking feature of four of the six sequenced antibody mimics selected from libraries 2L14 and BF14 is a pair of cysteines in adjacent loops BC and FG. Variant 2L4.1 contains cysteines in positions equivalent to 25/26 and 83 in full-length ^{10}Fn3; variant BF4.4 in positions 28 and 77; variant BF4.9 in positions 26, 82, and 83; and variant BF4.3 in positions 25 and 82.

Competition for binding to selected antibody mimics FG4.1 and BF4.4 among biotinylated hen egg white lysozyme, unbiotinylated hen egg white lysozyme, and biotin

A 1000-fold excess of biotin had no effect on binding of biotinylated lysozyme (HEL-b) to either FG4.1 or BF4.4 displayed on the yeast surface, and a 1000-fold excess of unbiotinylated lysozyme (HEL) had no effect on binding HEL-b to FG4.1, indicating the participation of biotinylation modification in the recognized epitope. In contrast, unbiotinylated HEL inhibited the binding of HEL-b to BF4.4, with a 10-fold excess of HEL preventing 80% of HEL-b from binding to yeast-displayed BF4.4 and a 100-fold excess of HEL preventing more than 95% of HEL-b from binding to yeast-displayed BF4.4. In addition, BF4.4 was shown to bind unbiotinylated HEL (as detected with polyclonal anti-HEL antibodies) with a dissociation constant indistinguishable from that for its binding to biotinylated HEL (data not shown), confirming that the epitope bound by BF4.4 is present on unbiotinylated lysozyme. In contrast, the lack of effect of either biotin or unbiotinylated lysozyme on the binding of FG4.1 to HEL-b suggests that the epitope bound by FG4.1 requires either both components of lysozyme and biotin or the caproyl linker between the two.

Affinity maturation of antibody mimics that bind hen egg white lysozyme

The design of second-generation antibody-mimic libraries was based on the selected sequences in the highest affinity variants from the first-generation library FG7 (highest affinity clone: FG4.1) and the first-generation library BF14 (highest affinity clone: BF4.4).

Library BFs1 contained the selected FG-loop sequence of variant FG4.1, WWDWEQA, and seven randomized residues in BC loop (Table 1). A round of MACS sorting followed by three rounds of FACS sorting at 1 μM HEL-b identified two variants with an improved affinity for lysozyme, BFs1c4.07 and BFs1c4.01, which have dissociation constants of 10 and 440 nM, respectively (Table 2). Four further FACS sorts under the more stringent conditions of 100 and 10 nM HEL-b reduced the fraction of the lower affinity variant, BFs1c4.01, to a level where it could not be detected by sequencing 24 clones randomly picked from the selected pool. Both variants BFs1c4.07 and BFs1c4.01 contain an insertion of a single amino acid in loop BC.

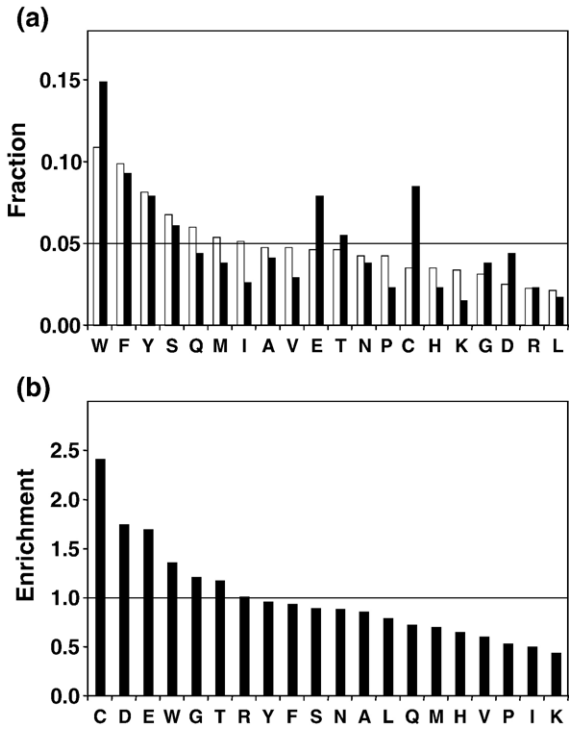

Figure 1. Distribution of amino-acid residues in ^{10}Fn3-based libraries and in selected and affinity-matured, lysozyme-binding antibody mimics. (a) Distribution of amino acid residues in randomized loop positions (residues 23–29 and 77–83) in 798 positions in libraries BC7, FG7, 2L14, and BF14 (white bars) and in 343 positions in selected loops listed in Table 1 (black bars). The residues in each selected loop were counted once regardless of the representation of the clone in a selected population. Residues in loops fixed during affinity maturation (i.e., BC loop in library BFs3 and FG loop in libraries BFs1 and BFs2) were not counted in the resulting affinity-matured clones. The horizontal line marks the 5% representation expected for all residues in a library made from a perfectly unbiased mixture of codons. (b) Ratio between the frequency of each amino acid residue in selected loops and in naive library loops ("Enrichment"). The horizontal line marks the ratio of 1 for residues with equal representation in library and selected loops.

The two parallel approaches to affinity-maturing variant BF4.4 involved fixing six of the seven selected residues in one of the two loops and re-randomizing all seven residues of the other loop. To test whether Cys28 and Cys77 of BF4.4 interacted (most likely by forming a disulfide bond), the codon for cysteine in the region encoding the "fixed" loop was replaced by a mixture of codons for Cys and Ser (Table 1). Thus, library BFs2 was constructed by setting position 77 to a mixture of Cys and Ser, by fixing residues 78–83 of the FG loop to the selected BF4.4 sequence of YSVFNY, and by re-randomizing loop BC (Table 1). Conversely, library BFs3 was constructed by setting position 28 to a mixture of Cys and Ser, by fixing residues 23–27 and 29 to

Table 2. Antibody-mimic clones selected for binding to hen egg white lysozyme

Clone[1]	Library[2]	BC[3]	FG[4]	Frame[5]	f4[6] (%)	f8[7] (%)	K_d[8] (nM)
[10]Fn3	–	DAPAVTV[a]	GRGDSPA	–	–	–	>10,000
FG4.1	FG7	DAPAVTV[a]	WWDWEQA	–	43	63	610±80
FG4.2	FG7	DAPAVTV[a]	WWDFEQE	–	50	38	750±90
FG4E	FG7	DAPAVTV[a]	WWDCEQE	–	3	0	ND[f]
FG4J	FG7	DAPAVTV[a]	WWDMDQW	G41V	3	0	ND[f]
2L4.1	2L14	–WPCQLA[b]	RMGMAEC	–	95	100	170±5
2L4.19	2L14	RWYEITV	FWDFMQA	–	5	0	ND[f]
BF4.4	BF14	WTAHYCD	CYSVFNY	–	50	85	120±10
BF4.9	BF14	QDSCTWV	RVYQKCC	–	18	15	140±20
BF4.1	BF14	HLDEWIA	RLTWVRK	–	27	0	180±20
BF4.3	BF14	EICQQFA	RVTERCW	–	5	0	ND[f]
BFs1c4.07	BFs1	WCCEYWYP[c]	WWDWEQA[a]	–	17	100	10±3
BFs1c4.01	BFs1	ARIWFTTN[c]	WWDWEQA[a]	–	50	0	440±30
BFs1c4.04	BFs1	YMGETSY	WWDWEQA[a]	–	4	0	ND[f]
BFs1c4.21	BFs1	YFGEFSM	WWDWEQA[a]	–	4	0	ND[f]
BFs1c4.02	BFs1	SFWPISM	WWDWEQA[a]	–	4	0	ND[f]
BFs1c4.10	BFs1	YKSPIIV	WWDWEQA[a]	S85R	4	0	ND[f]
BFs1c4.13	BFs1	YWAYTTS	WWDWEQA[a]	T94I	4	0	ND[f]
BFs1c4.16	BFs1	YWKYGPH	WWDWEQA[a]	–	4	0	ND[f]
BFs1c4.22	BFs1	EAWYTTQ	WWDWEQA[a]	–	4	0	ND[f]
BFs1c4.23	BFs1	MSWVATM	WWDWEQA[a]	D67E	4	0	ND[f]
BFs2_4.05	BFs2	VLDWFNA	RYSVFNY[e]	–	35	84	>1000
BFs2_4.06	BFs2	FLWPWWS	SYSVFNY[e]	–	22	16	>1000
BFs2_4.01	BFs2	CFYWNWI	SYSVFNY[e]	–	22	0	>1000
BFs2_4.02	BFs2	WTIFNYW	CYSVFNY[e]	–	4	0	ND[f]
BFs2_4.04	BFs2	FNFSWFS	CYSVFNY[e]	–	4	0	ND[f]
BFs2_4.12	BFs2	YHSECPN	CYSVFNY[e]	–	4	0	ND[f]
BFs2_4.19	BFs2	HGQWPWF	CYSVFNY[e]	–	4	0	ND[f]
BFs2_4.07	BFs2	KYSFEFE	CYSVFNY[e]	G41V	4	0	ND[f]
BFs3_8.01	BFs3	WTAHYCD[d]	CYDSEWW	–	0	92	0.35±0.03 0.49±0.05[g]
BFs3_4.06	BFs3	WTAHYCD[d]	CFNVNGW	–	30	8	1.6±0.1 4.8±1.0[g]
BFs3_4.02	BFs3	WTAHYCD[d]	CFTSQGW	–	17	0	2.5±0.1
BFs3_4.01	BFs3	WTAHYCD[d]	CFTSMGW	–	9	0	ND[f]
BFs3_4.08	BFs3	WTAHYCD[d]	CFSSMEW	–	9	0	ND[f]
BFs3_4.04	BFs3	WTAHYCD[d]	CFNTHGW	–	5	0	ND[f]
BFs3_4.26	BFs3	WTAHYCD[d]	CFSNEHW	–	5	0	ND[f]
BFs3_4.03	BFs3	WTAHYCD[d]	CFDTEYW	–	3	0	ND[f]
BFs3_4.09	BFs3	WTAHYCD[d]	CYNLGGW	–	3	0	ND[f]
BFs3_4.17	BFs3	WTAHYCD[d]	CFTQEEW	–	3	0	ND[f]
BFs3_4.19	BFs3	WTAHYCD[d]	CYSQYAW	–	3	0	ND[f]
BFs3_4.25	BFs3	WTAHYCD[d]	CFEMAGW	–	3	0	ND[f]
BFs3_4.28	BFs3	WTAHYCD[d]	CFHEDNW	–	3	0	ND[f]
BFs3_4.34	BFs3	WTAHYCD[d]	CFEDFWE	–	3	0	ND[f]
BFs3_4.14	BFs3	WTAHYCD[d]	CMDYFGY	–	3	0	ND[f]

[1]Name of selected clone.
[2]Name of library from which the clone was selected.
[3]Sequence in loop BC, amino acid residues 23–29. Residues found in 50% or more of the selected sequences in a particular selection are underlined.
[4]Sequence in loop FG, amino acid residues 77–83. Residues found in 50% or more of the selected sequences in a particular selection are underlined.
[5]Mutations from wild-type [10]Fn3 sequence outside randomized loops.
[6]Percentage of clones with this sequence in the population after four rounds of selection.
[7]Percentage of clones with this sequence in the population after eight rounds of selection.
[8]Dissociation constant. Errors in K_d measurements were estimated from results of two independent experiments, but may be higher for clones FG4.1 and FG4.2 due to their low affinity (Figure 2). Except where noted with [g], K_d was determined using yeast-surface–displayed antibody mimic.
[a] Sequences fixed during library construction.
[b] –, deletion of an amino acid residue.
[c] Loop with an insertion.
[d] Sequence fixed during selection, except for position 28, which was randomized between S and C.
[e] Sequence fixed during selection, except for position 77, which was randomized between S and C.
[f] Not determined.
[g] K_d determined by equilibrium competition titration using purified antibody mimic in solution (Figure 4).

those of the selected BC loop in BF4.4 sequence, WTAHY and D, respectively, and by re-randomizing loop FG.

After the first four rounds of selection from library BFs3, which comprised a MACS sort followed by three FACS sorts at 1 μM and 100 nM HEL-b

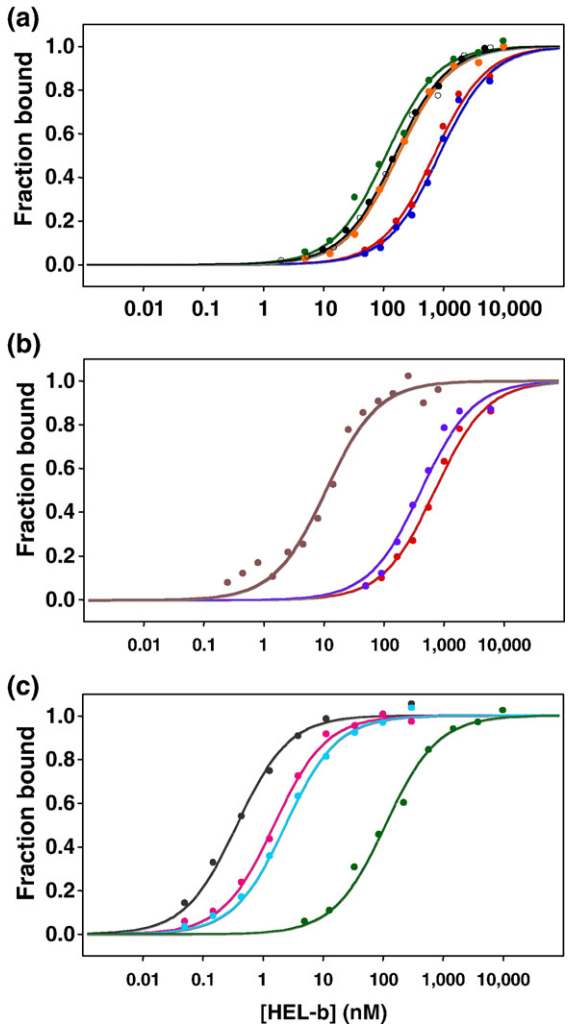

Figure 2. Binding of selected antibody mimics to biotinylated hen egg white lysozyme. The antibody mimics were displayed on yeast surface, and biotinylated lysozyme was supplied free in solution; binding was quantified using flow cytometry. Only one set of data per variant, which is representative of two or more sets of data collected for each clone, is shown. The dissociation constants listed in Table 2 are averages of values obtained in at least two independent experiments. (a) Antibody mimics selected from first-generation libraries. Blue, FG4.2; red, FG4.1; orange, 2L4.1; black filled circles, BF4.1; black open circles, BF4.9; green, BF4.4. (b) Antibody mimics selected by affinity maturation of clone FG4.1. Red, FG4.1; purple, BFs1c4.01; brown, BFs1c4.07. (c) Antibody mimics selected by affinity maturation of clone BF4.4. Green, BF4.4; light blue, BFs3_4.02; pink, BFs3_4.06; gray, BFs3_8.01.

(Figure 3(a), with the negative control in Figure 3(b)), sequencing of 34 clones yielded 14 unique sequences. All 14 sequences contained a cysteine in position 28 (which had been randomized between cysteine and serine in library BFs3) and in position 77 (which had been randomized among the 20 amino acid residues), demonstrating a strong bias for the two cysteines in these two positions. The two sequences found in at least 10% of the clones sequenced after four rounds of selection, BFs3_4.06 and BFs3_4.02, had dissociation constants of 1.6 and 2.5 nM, respectively. Four additional high-stringency FACS sorts, where the concentration of HEL-b was reduced to 10 and 1 nM (Figure 3(c), with the negative control in Figure 3(d)), yielded a new variant, BFs3_8.01, with a dissociation constant of 350 pM, and reduced the fraction of variant BFs3_4.02 to a level where it could not be detected by sequencing 24 clones randomly picked from the selected pool (Table 2). To demonstrate that the titration on the yeast surface is representative of the interaction with soluble protein, the dissociation constants of the two highest affinity clones were also determined by equilibrium competition titration using purified antibody mimic. The affinities measured by equilibrium competition were 490 pM and 4.8 nM for BFs3_8.01 and BFs3_4.06, respectively (Figure 4).

In contrast to library BFs3, the first four rounds of selection from library BFs2, which comprised a MACS sort followed by three FACS sorts at 1 μM, yielded no clones with cysteines in positions 28 and 77. None of the three sequences found in at least 10% of the clones sequenced, BFs2_4.05, BFs2_4.01, and BFs2_4.06, contained cysteines in either position 28 or 77, and the most frequently selected clone, BFs2_4.05, contained Arg77, a rare mutation from the original library design that was not observed after sequencing 22 clones from unselected library BFs2 (data not shown). Whereas BFs2_4.05, BFs2_4.01, and BFs2_4.06 could be observed to bind HEL-b at high nanomolar and micromolar concentrations, their dissociation constants were at least an order of magnitude higher than that of their parent clone, BF4.4, and were too high to be measured (data not shown). Four additional FACS sorts at 1 μM and 100 nM reduced the population to BFs2_4.05 and BFs2_4.06, but did not select any new sequences with tighter binding to lysozyme.

Sequences of selected and affinity-matured antibody mimics

Sequences of BC and FG loops in selected and affinity-matured, lysozyme-binding antibody mimics are listed in Table 2, and the representation of different amino acid side-chains in the selected loops is shown in Figure 1. Twelve of the twenty kinds of amino acid residues are found in selected loops at a frequency between 60 and 140% of the frequency of the same residue in the naive library. The selected clones show a distinct bias for cysteine, aspartic acid, and glutamic acid, with enrichment factors of 242, 175, and 170%, respectively, as well as a bias against Lys, Ile, and Pro, with enrichment factors of 44, 51, and 54%, respectively (Figure 1(b)).

Whereas most selected antibody-mimic sequences conformed to library design, several selected clones contained mutations outside the intentionally randomized loops or insertions or deletions in the randomized loops. Mutations in regions outside

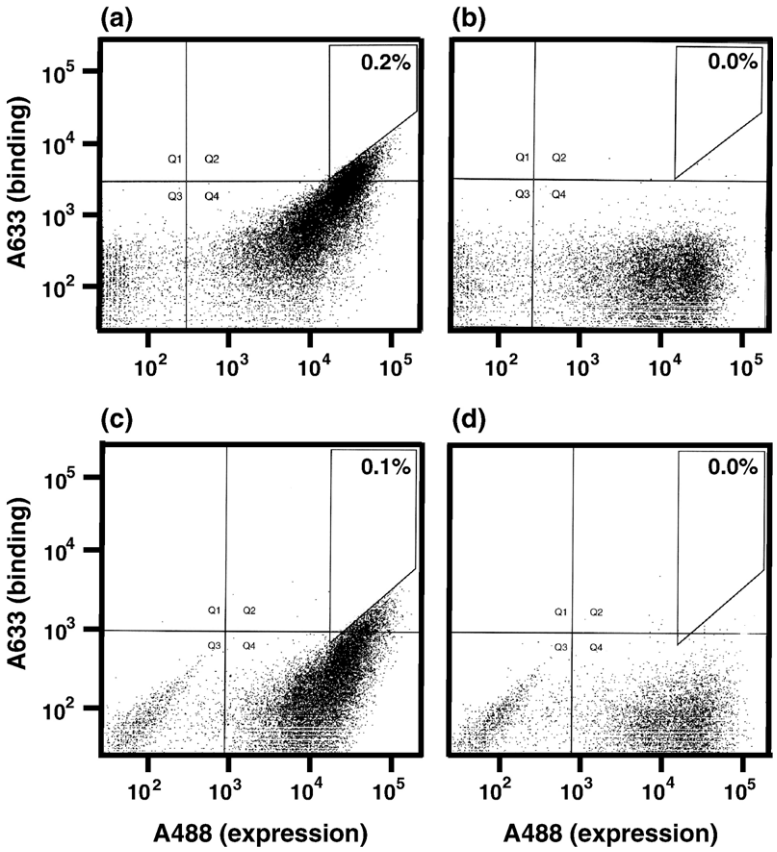

Figure 3. FACS sorting of lysozyme-binding yeast during affinity maturation BFs3. (a) Sort number 4; yeast labeled at 100 nM biotinylated lysozyme. (b) Negative control for sort number 4; same yeast labeled with detection reagents only, no lysozyme. (c) Final sort (number 8) of yeast labeled at 1 nM biotinylated lysozyme. (d) Negative control for sort number 8; same yeast labeled with detection reagents only, no lysozyme. A488, fluorescence due to Alexa 488, which detects the presence of C-terminal c-*myc* epitope on the yeast surface; A633, fluorescence due to Alexa 633, which detects the presence of biotin on biotinylated lysozyme.

randomized BC and FG loops were found in 1 of the 4 selected lysozyme-binding clones from library FG7, in 3 of the 10 clones selected from library BFs1, and in 1 of the 8 clones selected from library BFs2.

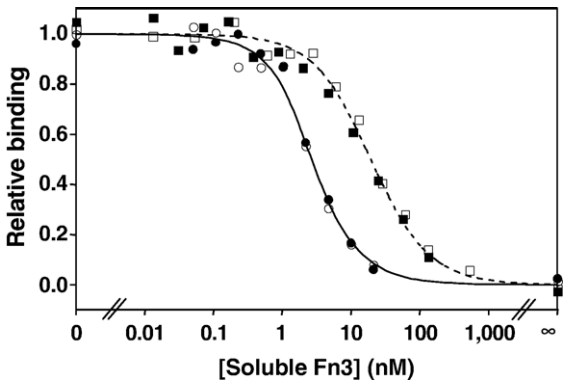

Figure 4. Equilibrium competition titration of antibody mimics. The antibody mimic BFs3_8.01 was displayed on the yeast surface; purified antibody mimic (circles with solid line, BFs3_8.01; squares with dashed line, BFs3_4.06) and 1 nM biotinylated lysozyme were supplied free in solution. Binding of biotinylated lysozyme to the yeast-surface–displayed antibody mimic BFs3_8.01 was quantified using flow cytometry and normalized between minimum and maximum fluorescence as determined by least-squares analysis. Data sets represented with filled and empty symbols were collected in two separate experiments. ∞, samples with no biotinylated lysozyme.

None of these sequences was found in 10% or more of the clones found in a selected population, and presumably these clones had relatively low affinity for HEL-b. The overall frequency of mutations outside randomized loops in selected antibody mimics, 5/43 = 12%, is similar to the 10% frequency seen in unselected antibody-mimic libraries.

A deletion was found in 1 of the 2 selected lysozyme-binding clones from library 2L14, and insertions were found in 2 of the 10 clones selected from library BFs1. All clones with deletions and insertions were associated with sequences found in 10% or more of the clones found in a selected population. The overall frequency of deletions and insertions in selected antibody mimics, 3/43 = 7%, is significantly higher than the approximately 1% frequency found in unselected antibody-mimic libraries.

Contribution of Cys28–Cys77 pair to lysozyme binding

In the first experimental test of the 28–77 disulfide, an affinity-maturation library was designed that preserved all selected residues in the BF4.4 BC loop except for Cys28, which was replaced by a mixture of codons that encode either serine or cysteine (library BFs3, Table 1). The seven residues in the FG loop in library BFs3 were randomized among all 20 amino acid residues. After being sorted under increasingly stringent conditions, all selected clones contained cysteines in both positions 28 and 77,

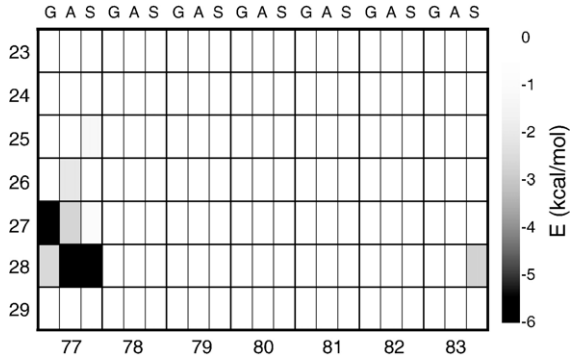

Figure 5. Modeling of disulfide bonds between cysteines in BC and FG loops. The numbers on the vertical axis refer to the position of the cysteine in the BC loop, and the numbers on the horizontal axis refer to the position of the cysteine in the FG loop. Energies of disulfide formation for each pair of BC and FG cysteines were calculated following restrained minimization using CHARMM,[39] with all the remaining residues in the BC and FG loops set to glycine (first column for a given pair of residues), alanine (second column), or serine (third column). Calculated energies are color coded, with a darker shade of gray corresponding to a lower (more negative) energy and thus to a higher likelihood of formation of the disulfide bond. White, 0 kcal/mol or higher; black, −6 kcal/mol or lower.

confirming a strong selective advantage of this cysteine pair.

In the second experimental test of the 28–77 disulfide, Cys28 and Cys77 of BFs3_8.01, the ^{10}Fn3 variant with the highest affinity for lysozyme, were mutated to serine, both individually and simultaneously, resulting in clones BFs3_8.01 (C28S), BFs3_8.01(C77S), and BFs3_8.01(C28S, C77S). The K_d for mutant BFs3_8.01(C28S) was 330 (±60) nM, approximately 1000-fold higher than that of BFs3_8.01. Yeast cells displaying mutants BFs3_8.01(C77S) and BFs3_8.01(C28S,C77S) showed negligible binding to lysozyme below 1 μM and only a slight increase in fluorescence between 1 μM and 10 μM HEL-b (data not shown), leading to a rough estimate of $K_d > 10$ μM. This dramatic effect of relatively conservative Cys to Ser mutations is consistent with the existence of a 28–77 disulfide bridge, which is critical for binding to hen egg white lysozyme.

Modeling of interloop disulfide bonds

Disulfide bonds were modeled among all 49 pairs of BC loop and FG loop positions using an NMR average structure of ^{10}Fn3 as a template.[5] For each pair of positions, two free cysteines were modeled for reference in addition to the disulfide-bonded form. Due to the lack of experimental data on side-chain conformation in the selected loops, the remaining 12 BC-loop and FG-loop positions were modeled as all glycine, all alanine, or all serine. Each structure was subjected to a restrained minimization protocol. The final total energy of the disulfide bond relative to the free cysteines is shown in Figure 5, and structural statistics for models with favorable disulfides are presented in Table 3.

A disulfide bond between the BC loop and the FG loop was calculated to be favorable for only five pairs of positions. Only for the pair 28–77 was a disulfide bond favorable across all three side-chain backgrounds, and this disulfide was top ranked for both the alanine and the serine backgrounds. The glycine background most favored a disulfide between positions 27 and 77.

The model of the 28–77 disulfide bond with alanine at the other 12 BC-loop and FG-loop positions is shown in Figure 6. Before minimization, this modeled structure was deformed due to a long sulfur–sulfur bond length, but otherwise it was free of steric clashes. A standard disulfide conformation was achieved through the restrained minimization, primarily by rotation of the protein backbone at position 77.

Table 3. Disulfide bonds between BC and FG loops of ^{10}Fn3 that are predicted to be energetically favorable

		C_α - C_α distance[3] (Å)			RMSD[4] (Å)			
Cys[1]	Back-ground[2]	^{10}Fn3[6]	DiS[7]	Ref[8]	^{10}Fn3/diS	^{10}Fn3/Ref	DiS/Ref	Energy[5] (kcal/mol)
27–77	Gly	9.9	5.8	9.1	1.18	1.10	0.62	−9.4
28–77	Ser	8.1	5.6	7.8	1.18	1.11	0.52	−6.9
28–77	Ala	8.1	5.6	7.8	1.20	1.12	0.53	−5.6
28–83	Ser	12.1	6.4	11.0	1.26	1.11	0.73	−1.4
27–77	Ala	9.9	6.0	9.6	1.20	1.13	0.67	−1.3
28–77	Gly	8.1	5.7	6.8	1.17	1.12	0.54	−1.2
26–77	Ala	13.2	6.1	13.3	1.27	1.12	0.95	−0.8
25–77	Ser	12.6	6.3	12.4	1.29	1.12	0.98	−0.3
27–77	Ser	9.9	6.2	10.3	1.19	1.11	0.72	−0.2

[1]Positions of the two modeled cysteines in loops BC and FG.
[2]Amino acid residues assigned to the remaining 12 positions in the BC and FG loops.
[3]Distance between the C_α atoms at the positions identified in the column labeled Cys.
[4]Root mean square distance between two structures, including all equivalent atoms.
[5]Calculated energy of a disulfide bond relative to the free cysteines.
[6]NMR solution structure of wild-type human ^{10}Fn3.
[7]Modeled structure with a disulfide between the cysteines in the two positions listed under Cys.
[8]Modeled structure with two free cysteines in the two positions listed under Cys.

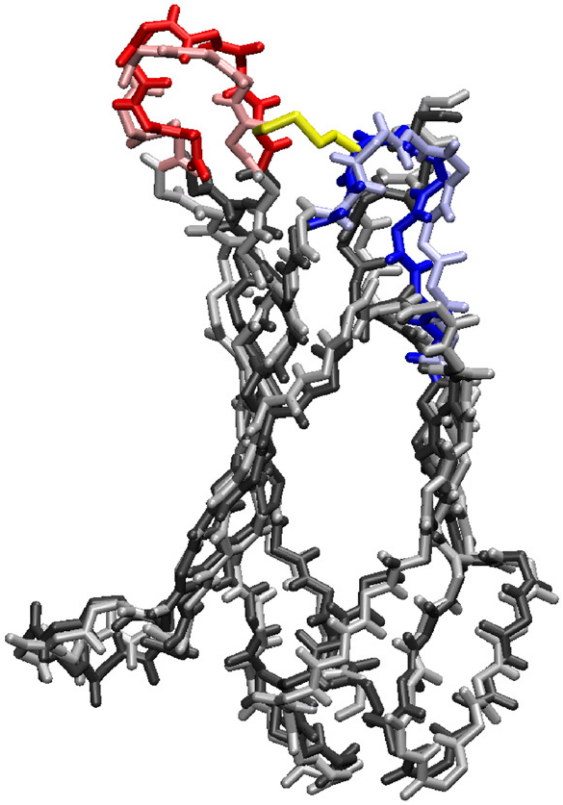

Figure 6. Model of a ^{10}Fn3 variant with a disulfide bond between Cys28 and Cys77, superimposed on the NMR structure of wild-type ^{10}Fn3. The disulfide model was generated using harmonically restrained minimization of the NMR structure with a disulfide bond built between cysteines at positions 28 and 77 and with alanine at the other 12 positions in BC and FG loops. Dark red, FG loop in the C28–C77 model; dark blue, BC loop in the C28–C77 model; dark gray, the rest of the main chain of the C28–C77 model; yellow, disulfide bond between Cys28 and Cys77 in the C28–C77 model, excluding hydrogen atoms; light red, FG loop in wild-type ^{10}Fn3; light blue, BC loop in wild-type ^{10}Fn3; light gray, the rest of main chain of wild-type ^{10}Fn3. Side-chains are not shown for clarity.

The goal of the disulfide-bond modeling was to suggest pairs of positions that are capable of forming a disulfide bond, not to rule out pairs of positions that cannot form a disulfide bond. Several simplifications used in the calculations may result in missing a low-energy conformation that would favor a particular disulfide bond. The glycine, alanine, and serine backgrounds used in modeling are only three of 4×10^{15} possible sequences for each disulfide, and the minimization protocol undersamples side-chain and backbone conformational space. Nevertheless, the modeling demonstrates that disulfide bonds can be introduced between several positions on the BC and FG loops without disrupting the ^{10}Fn3 fold. The cysteine pair found in clone BF4.4, Cys28 and Cys77, corresponds to the lowest disulfide-bond energy in the alanine and serine backgrounds and to one of the two favorable disulfide-bond energies in the glycine background. Interestingly, Cys27 and Cys77, which correspond to the lowest calculated disulfide-bond energy in the glycine background, are found in another selected clone, BFs2_4.12.

Effect of loop length on affinity for hen egg white lysozyme of selected antibody mimics

Two lysozyme-binding antibody mimics with BC loops different in length from wild-type human ^{10}Fn3, 2L4.1 and BFs1c4.07, were mutated to recreate wild-type–length, seven-residue BC loops. In the case of 2L4.1, which has a deletion in the BC loop, a glycine was inserted either in position 23 (2L4.1_nG) or in position 29 (2L4.1_cG). In the case of BFs1c4.07, which has an insertion in the BC loop, Trp23 (BFs1c4.07_nX) or Pro3 (BFs1c4.07_cX) was deleted. All four mutants bound hen egg white lysozyme too weakly for their binding to be quantified, but it is clear that their dissociation constants are at least 100-fold higher than those of the parent mutants (Figure 7).

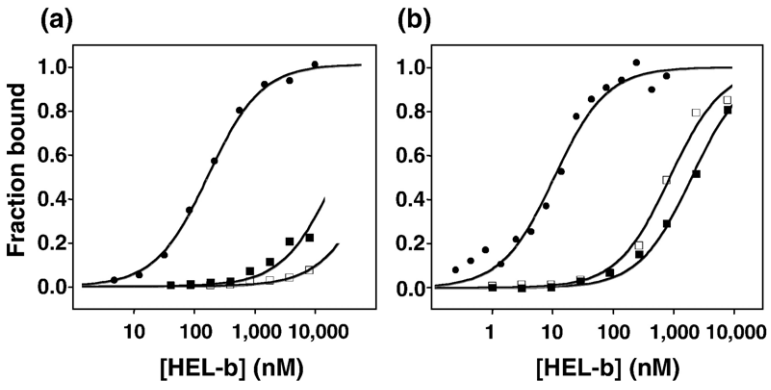

Figure 7. Binding to biotinylated hen egg white lysozyme of selected antibody mimics with variant lengths of BC loop and of their derivatives with wild-type–length BC loop. Binding was quantified using flow cytometry. The BC-loop sequence of each variant is shown in brackets following its name. (a) Filled circles, 2L4.1 (WPCQLA). Filled squares, 2L4.1_nG (GWPCQLA). Open squares, 2L4.1_cG (WPCQLAG). (b) Filled circles, BFs1c4.07 (WCCEYWYP). Filled squares, BFs1c4.07_nX (CCEYWYP). Open squares, BFs1c4.07_cX (WCCEYWY).

Yeast surface display as a method for determination of dissociation constants

The display of target-binding proteins on the yeast surface makes possible rapid and accurate characterization of their binding properties without the need for separate production and purification of individual clones.[23] Previous work that has established the agreement between K_d determination by yeast surface display and other methods is summarized in Figure 8.[23–31] In this study we confirmed this agreement by determining dissociation constants of the two highest affinity antibody mimics, BFs3_8.01 and BFs3_4.06, both using the standard yeast surface display protocol and competition with the purified, soluble antibody mimics added in solution (Figures 2 and 4). The K_d values determined for BFs3_8.01 were 350(±30) pM by yeast surface display and 490(±50) pM by soluble competition; for BFs3_4.06 the K_d values were 1.6(±0.1) nM by yeast surface display and 4.8(±1.0) nM by soluble competition.

Discussion

The comparative success of the four naive, ^{10}Fn3-based libraries in the selection of lysozyme-binding antibody mimics provides circumstantial evidence of the relative importance of thorough sequence-space sampling *versus* paratope size. Compared to the library made by randomizing 7 FG residues, the libraries made by randomizing 14 residues sample a much lower fraction of their theoretical sequence space: 2×10^{-11} and 2×10^{-10} for libraries 2L14 and BF14, respectively, in contrast with 2×10^{-2} for library FG7 (Table 1); nevertheless, antibody mimics with approximately 5-fold higher affinity were selected from libraries 2L14 and BF14 than from library FG7. The doubling of the diversified area in the two-loop libraries relative to the one-loop library more than offsets the disadvantages of a 10^9-fold scarcer sampling of the sequence space.

The theoretical sequence space associated with the two-loop libraries is likely to contain many more lysozyme-binding clones than the relatively small physical representation of the two-loop (BF14) library. The two libraries constructed by randomizing 14 residues simultaneously, 7 in the BC loop and 7 in the FG loop, differed only in their complexity, which was 10-fold higher for library BF14 than for library 2L14. As expected from the 10-fold larger number of sequences tested in the selection from library BF14, the antibody mimic from this library that bound lysozyme with the highest affinity (clone BF4.4; K_d = 120 nM) had a somewhat higher affinity

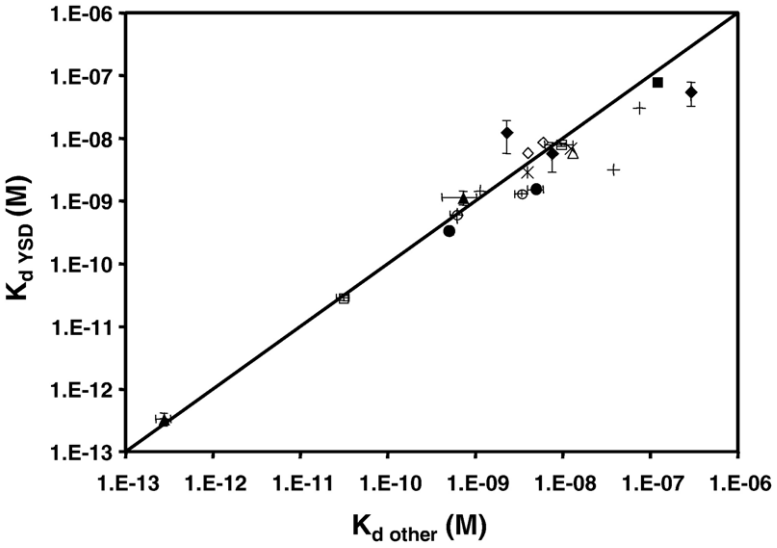

Figure 8. Comparison of dissociation constants determined by yeast surface display and other methods. $K_{d\ YSD}$, dissociation constant determined by yeast surface display; $K_{d\ other}$, dissociation constant determined by a different method. Open circles, single-chain antibodies binding to hen egg white lysozyme; K_d determined by yeast surface display[24] and fluorescence quench titration.[25] Filled circles, ^{10}Fn3-based antibody mimics binding to hen egg white lysozyme described in this report; K_d determined by yeast surface display (Figure 2, Table 2) and equilibrium competition titration with purified antibody mimics (Figure 4, Table 2). Open squares, single-chain antibody binding to carcinoembryonic antigen; K_d determined by yeast surface display[26] and titration of mammalian-cell–displayed carcinoembryonic antigen with soluble single-chain antibody (Mike Schmidt, unpublished results). +, neutralizing single-chain antibodies binding to botulinum neurotoxin[27]; K_d determined by yeast surface display and surface plasmon resonance. Filled triangles, single-chain antibodies binding to fluorescein; K_d determined by yeast surface display[28] and fluorescence quenching.[29] Open diamonds, single-chain antibodies binding to p53 peptides[23]; K_d determined by yeast surface display and surface plasmon resonance. *, single-chain antibody binding to epidermal growth factor (EGF)[23]; K_d determined by yeast surface display and surface plasmon resonance. Filled square, single-chain antibody binding to xeroderma pigmentosum-complementing protein group A[23]; K_d determined by yeast surface display and surface plasmon resonance. ×, single-chain antibody binding to heparin-binding EGF[23]; K_d determined by yeast surface display and surface plasmon resonance. Filled diamonds, soluble T-cell receptors binding to staphylococcal enterotoxin C3[30]; K_d determined by yeast surface display and surface plasmon resonance. Open triangle, soluble T-cell receptor binding to toxic shock syndrome toxin-1[31]; K_d determined by yeast surface display and surface plasmon resonance.

than the antibody mimic selected from the smaller library (clone 2L4.1; K_d = 170 nM). However, both the BF14 and the 2L14 libraries undersample theoretical diversity by over 10 orders of magnitude.

Three lysozyme-binding variants with K_d for lysozyme lower than 200 nM were found among the 3×10^8 clones in the BF14 library, suggesting that three in 3×10^8, or 10^{-8}, is the approximate fraction of improved clones throughout the theoretical sequence space. The fact that three separate high-affinity clones were selected makes it unlikely that their appearance was a lucky event not representative of the fraction of clones in the theorectical sequence space. Consequently, we expect that there were in the order of 10^{10} such improved clones among the 2×10^{18} sequences theoretically possible in this randomization scheme. It would be reasonable to expect that, of these 10^{10} variants with K_d lower than 200 nM, some would in fact have substantially higher affinities.

We tested this hypothesis by exploring the sequence space close to that of the BF4.4 clone, using an affinity maturation approach analogous to CDR walking in antibodies.[20] When the 7 BC residues in clone BF4.4 were essentially fixed and 7 residues in loop FG were randomized again, the resulting affinity-maturation library yielded ^{10}Fn3 variants with substantially higher affinity for lysozyme, including clone BFs3_8.01 with a K_d of 350 pM, a 340-fold improvement. Clearly, at least in the case of hen egg white lysozyme, extremely high-affinity antibody mimics can be obtained by randomizing no more than 14 residues in loops BC and FG and using libraries of modest size, as long as the sampling problem is circumvented by affinity maturation. A similar affinity-maturation strategy applied to clone FG4.1, in which 7 selected residues on the FG loop were fixed and 7 residues on the BC loop were randomized, led to a smaller but significant improvement in affinity of up to 60-fold, to the lowest K_d of 10 nM.

Since libraries BC7 and FG7 were constructed by randomizing seven residues in loops BC and FG, respectively, and since they each sampled about 2% of all possible permutations of 20 amino acids in seven positions ("theoretical sequence space"), the difference in outcome of selections from these two libraries can be ascribed to the difference in structural context between the two sets of residues randomized. The fact that no lysozyme-binding ^{10}Fn3 variants were selected from library BC7, but that several related clones with dissociation constants in the order of 700 nM were selected from library FG7, suggests that the greater flexibility and accessibility of the FG loop may be an advantage in binding to lysozyme. Nevertheless, the possibility that library FG7 had a better luck of the draw cannot be eliminated due to the low sampling of their theoretical sequence space by both libraries.

The most striking feature in the sequences in lysozyme-binding antibody mimics is the prevalence of cysteine, which is present in selected clones at more than twice its frequency in naive libraries (Figure 1). In particular, of the six antibody mimics selected from libraries with two randomized loops, 2L14 and BF14, four had at least one cysteine in both loop BC and loop FG (67% frequency). Given the 3.5% cysteine frequency in naive libraries, the probability of at least one cysteine being found, by chance, in a single loop is $(1-(1-0.035)^7) = 22\%$, and the probability of at least one cysteine being found, by chance, in both loops is $0.22^2 = 4.9\%$. The fact that the observed frequency of cysteines in both loops is more than 10-fold higher than that expected to occur by chance is consistent with a strong selective advantage of such cysteine pairs. We hypothesized that two cysteines on adjacent CDR-like loops of ^{10}Fn3 form a disulfide bond, which contributes to both thermodynamic stability of the ^{10}Fn3 fold and its binding to hen egg white lysozyme. We tested the hypothesis that two cysteines on adjacent CDR-like loops of ^{10}Fn3 form a disulfide bond using structural modeling, affinity-maturation experiments, and site-directed mutagenesis; all three approaches yielded results consistent with such a disulfide bond. We expect a final confirmation of the existence and role of the 28–77 disulfide to come from structural studies of variants with this cysteine pair in complex with lysozyme.

The most likely mechanism in which the 28–77 disulfide contributes to lysozyme binding is entropic, by reducing the flexibility of loops BC and FG in unbound BFs3_8.01. Nevertheless, it is also possible that the disulfide itself is a part of the lysozyme-binding surface on the antibody-mimic molecule. Indirect support for this possibility comes from the presence of two adjacent cysteines in the BC loop of variant BFs1c4.07 (Table 1, K_d = 10 nM), which may form a vicinal disulfide bond.[32] Because the affinity maturation strategy that fixed the BF4.4 FG loop (except for position 77, which was randomized between serine and cysteine) and re-randomized the BC loop (in library BFs2) did not yield any mutants with a Cys28–Cys77 pair, this suggests that most of the binding energy of variant BF4.4 to lysozyme was provided by its selected BC loop or by the combination of its BC loop and the 28–77 disulfide.

Our results have several implications for future design of ^{10}Fn3-based antibody libraries. First, they demonstrate that the randomization of two CDR-like loops of ^{10}Fn3, followed by affinity maturation, is sufficient to generate variants with sub-nanomolar binding to a target protein. Second, they opens the possibility of a modified, constrained ^{10}Fn3-based scaffold, which would include cysteines in positions 28 and 77. Only residues 23–27, 29, and 78–83 would be randomized to generate such constrained libraries. Whereas antibody mimics selected from these libraries would require oxidizing conditions for optimal binding, the presence of the interloop disulfide in clones BFs3_8.01 and BFs3_4.06 still allowed their expression in *E. coli* at 0.1–0.3 mg/l. Both a reduction in the number of residues randomized and the presence of the 28–77 interloop disulfide in a constrained-loop library would be expected to increase the stability and

solubility of selected antibody mimics. In addition, because such a library would require randomization of two fewer residues than the BF14 library, a physical library of the same size would achieve 400-fold more thorough sampling of theoretical sequence space. We expect that a ^{10}Fn3-based library with a fixed disulfide, as previously selected in the same scaffold, has a higher chance of success than did a library of human variable domains with a disulfide designed based on camel VHH structures,[33] although the compatibility of the 28–77 disulfide with binding of targets other than lysozyme remains to be demonstrated. Third, our results suggest that a seven-residue BC loop may not be optimal for lysozyme binding. Lysozyme-binding clones with several single amino acid insertions and deletions in the BC loop were selected, despite the fact that clones with non–wild-type loop length were rare in the corresponding naive libraries. The affinity for lysozyme of these clones was shown to depend on the non–wild-type loop length. Future selections against this target should vary the length of the BC loop as well as its sequence. More generally, it would be useful to vary loop length as well as loop sequence in ^{10}Fn3-based libraries.

Whereas this is the first report of an interloop disulfide bond selected in ^{10}Fn3-based antibody mimics, analogous disulfides have been discovered in naturally evolved variable antibody domains from camelids (llamas and camels)[16,19] and in new antigen receptors from cartilaginous fish (sharks, skates, rays, and chimaeras).[17,18] In contrast to variable domains produced by the immune systems of most other animals, which bind antigens in V_H/V_L pairs, camelid and new-antigen-receptor variable domains bind antigens singly. A possible explanation for the prevalence of interloop disulfides in these domains is that the added structural constraint of an interloop disulfide contributes to domain stability,[18,34–36] rigidity of the antigen-binding site,[16,34,36] or structural diversity,[17,34,35] all of which may contribute to antigen binding. The evolutionary distance between camelids and cartilaginous fish and the rarity of interloop disulfides in non-camelid mammals suggest that the interloop disulfides present in both types of single-variable-domain antibodies arose by convergent evolution.[18,36]

Our *in vitro* selection of ^{10}Fn3-based antibody mimics is the third example of convergent evolution to an interloop disulfide under pressure of antigen binding by a single-domain, immunoglobulin-like structure. Furthermore, this is the first system in which the role of such a disulfide bond has been tested experimentally by demonstrating the loss of antigen binding when the disulfide is disrupted by site-directed mutagenesis and the tendency of the system to re-evolve the disulfide after the cysteine positions are randomized during affinity maturation. We expect that the analogous interloop disulfides in natural camelid and shark antibodies will prove similarly critical for their antigen binding.

Materials and Methods

Library and vector construction

Oligonucleotides used in the construction of ^{10}Fn3-based libraries are listed in Table 4. Those oligos that contained randomized regions were synthesized by TriLink (San Diego, CA) from mixtures of triphosphoramidites designed to equalize the probability of incorporating a codon for each of the 20 amino acid residues (Glen Research, Sterling, VA). The remaining oligonucleotides were purchased from ThermoElectron (Ulm, Germany) or MWG Biotech (High Point, NC). The strategy used to construct the library is outlined in Figure 9(d). Pairs of oligonucleotides were annealed and extended using Hot Start KOD polymerase (Novagen, San Diego, CA; 1 U/50 μL) in 1× reaction buffer, 0.2 mM dNTPs, 1 mM $MgS0_4$, 1 M betaine, 3% dimethylsulfoxide. The thermocycling program included 2 min at 95 °C; 10 cycles of 30 s at 94 °C, 30 s at 58 °C, and 1 min at 68 °C; and 10 min at 68 °C.

The first set of annealing/extension reactions was performed in four tubes of 50 μL each, with two oligonucleotides in each tube: Tube a contained 20 pmol a1 and 10 pmol a2; tube b contained 10 pmol b3 (or b3Rt) and 20 pmol b4; tube c contained 20 pmol c5 and 10 pmol c6; and tube d contained 10 pmol d7 (or d7Rt) and 20 pmol d8. Randomized oligo b3Rt was used in the construction of libraries BC7, 2L14, BF14, and BFs2 and randomized oligo d7Rt was used in the construction of libraries FG7, 2L14, BF14, BFs1, and BFs3. In the second set of annealing/extension reactions, the resulting contents of tube a were combined with those of tube b and the contents of tube c were combined with those of tube d, and the annealing/extension thermocycling protocol was repeated. The products from tubes ab and cd were diluted 1/10 with dH_2O. In the third set of annealing/extension reactions, 40 μL each of the two diluted products was combined with fresh PCR reagents in a total volume of 200 μL and then annealed and extended as above, yielding 2–20 μg of full-length, double-stranded DNA encoding a ^{10}Fn3-derived library. Finally, 2 μg of the product was amplified 10 to 100-fold using a 100× molar excess of PCR primers p1 and p8, in a total volume of 1.5–2 mL, using the thermocycling program described above with 20 annealing/extension cycles.

Plasmid pCTf1f4 includes the constant regions of ^{10}Fn3 upstream of the BC loop and downstream of the FG loop, separated by recognition sites of restriction enzymes NcoI, SmaI, and NdeI. It was constructed by annealing and extending oligonucleotides v1 and v2 (Table 4) as described above, digesting the product with NheI and BamHI, and ligating it between the NheI and BamHI sites of vector pCTcon.[21,22]

To transform each DNA-fragment library into *S. cerevisiae*, 20 μg of pCTf1f4 digested by NcoI, SmaI, and NdeI was combined with 60 μg of the DNA fragment encoding a full-length library and 1 mL electrocompetent EBY100 yeast. After electroporation in 200-μL aliquots, the yeast culture was grown in 500 mL of selective medium SD-CAA, pH 4.5, 50 μg/mL kanamycin, 100 U/mL penicillin G, and 200 U/mL streptomycin and characterized as described previously.[21,22] The number of transformants obtained for each library was 3.8×10^7 for library BC7; 4.0×10^7 for library FG7; 5.3×10^7 for library 2L14; 5.2×10^8 for library BF14; 9.3×10^7 for library BFs1; 1.6×10^8 for library BFs2; and 5.2×10^7 for library BFs3.

Between 20 and 60 independent clones from each library were sequenced to confirm library design and to characterize the distribution of amino acid residues in the

Table 4. Oligonucleotides used in construction of ^{10}Fn3-based antibody-mimic libraries and mutants

Name	Sequence	Use
v1	GACTGAGCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAG TTGTTGCTGCGACCCCCACCAGCCTACTGATCAGCTGGCC ATGGATCTGCCCCGGG	pCTf1f4
v2	CGTCATGGATCCCTGGGATGGTTTGTCAATTTCTGTTCGGT AATTAATGGAAATTGGCTTGCTGCTCATATGCGTCGCCCC GGGGCAGATCCATGG	pCTf1f4
a1	GACTGAGCTAGCGTTTCTGATGTTCCGAGGGACCTGGAAG TTGTTGCTGC	All libraries
a2	CCAGCTGATCAGTAGGCTGGTGGGGGTCGCAGCAACAAC TTCCAGGTC	All libraries
b3	CCAGCCTACTGATCAGCTGGGATGCTCCTGCTGTCACAGTG AGATATTACAGGATCACTTACGG	FG7
b4	GGCACAGTGAACTCCTGGACAGGGCTATTTCCTCCTGTTTC TCCGTAAGTGATCCTGTAATATC	All libraries
c5	GTCCAGGAGTTCACTGTGCCTGGGAGCAAGTCTACAGCTA CCATCAGCGGCCTTAAACCTG	All libraries
c6	GTGACAGCATACACAGTGATGGTATAATCAACTCCAGGTT TAAGGCCGCTGATG	All libraries
d7	CCATCACTGTGTATGCTGTCACTGGCCGTGGAGACAGCCCCG CAAGCAGCAAGCCAATTTCCATTAATTAC	BC7
d8	CGTCATGGATCCCTGGGATGGTTTGTCAATTTCTGTTCGGT AATTAATGGAAATTGGCTTGC	All libraries
p1	GACTGAGCTAGCGTTTCTGATG	All libraries
p8	CGTCATGGATCCCTGGGATG	All libraries
b3Rt	CCAGCCTACTGATCAGCTGGXXXXXXXAGATATTACAGGAT CACTTACGG[1]	BC7, BF14, 2L14, BFs1, BFs2
d7Rt	CCATCACTGTGTATGCTGTCACTXXXXXXXAGCAGCAAGCC AATTTCCATTAATTAC[1]	FG7, BF14, 2L14, BFs3
d7s1	CCATCACTGTGTATGCTGTCACTTGGTGGGACTGGGAACA GGCTAGCAGCAAGCCAATTTCCATTAATTAC	BFs1
d7s2	CCATCACTGTGTATGCTGTCACTTSCTACTCGGTCTTCAACT ATAGCAGCAAGCCAATTTCCATTAATTAC[2]	BFs2
b3s3	CCAGCCTACTGATCAGCTGGTGGACTGCTCATTACTSCGAC AGATATTACAGGATCACTTACGG[2]	BFs3
b3_gwpcqla	CCAGCCTACTGATCAGCTGGGGATGGCCTTGTCAACTAGC TAGATATTACAGGATCACTTACGG	2L4.1_nG
b3_wpcqlag	CCAGCCTACTGATCAGCTGGTGGCCTTGTCAACTAGCTGG AAGATATTACAGGATCACTTACGG	2L4.1_cG
d7_rmgmaec	CCATCACTGTGTATGCTGTCACTCGTATGGGTATGGCTGA ATGTAGCAGCAAGCCAATTTCCATTAATTAC	2L4.1_nG, 2L4.1_cG
b3_cceywyp	CCAGCCTACTGATCAGCTGGTGCTGTGAATACTGGTACCC TAGATATTACAGGATCACTTACGG	BFs1c4.07_nX
b3_wcceywy	CCAGCCTACTGATCAGCTGGTGGTGCTGTGAATACTGGTA CAGATATTACAGGATCACTTACGG	BFs1c4.07_cX
d7_wwdweqa	CCATCACTGTGTATGCTGTCACTTGGTGGGACTGGGAACA AGCAAGCAGCAAGCCAATTTCCATTAATTAC	BFs1c4.07_nX, BFs1c4.07_cX
b3_wtahycd	CCAGCCTACTGATCAGCTGGTGGACTGCACACTACTGTGA TAGATATTACAGGATCACTTACGG	BFs3_8.01(C77S)
b3_wtahysd	CCAGCCTACTGATCAGCTGGTGGACTGCACACTACTCAG ATAGATATTACAGGATCACTTACGG	BFs3_8.01 (C28S), BFs3_8.01 (C28S,C77S)
d7_cydseww	CCATCACTGTGTATGCTGTCACTTGTTACGATTCGGAATG GTGGAGCAGCAAGCCAATTTCCATTAATTAC	BFs3_8.01 (C28S),
d7_sydseww	CCATCACTGTGTATGCTGTCACTTCATACGATTCGGAATG GTGGAGCAGCAAGCCAATTTCCATTAATTAC	BFs3_8.01 (C77S), BFs3_8.01 (C28S,C77S)

[1]X, equimolar mixture of triphosphoramidites encoding all 20 amino acids.
[2]S, equimolar mixture of C and G.

randomized positions of naive libraries. This distribution of amino acid residues is determined by the distribution of triphosphoramidites in the randomized positions in oligonucleotides b3Rt, which encodes the randomized loop BC, and d7Rt, which encodes the randomized loop FG (Table 4). Consistent with the use of the same triphosphoramidite mixture to synthesize the two randomized oligos, the codon distribution was the same for both randomized loops. The distribution shown in the white bars of Figure 1(a) was derived by combining the information from all randomized clones sequenced, including 798 positions evenly distributed between the two randomized oligonucleotides.

The complexity of each library was estimated based on the number of transformants obtained and on the proportion of full-length, ^{10}Fn3-like clones identified by sequencing (Results, Table 1).

Selection for binding to hen egg white lysozyme using magnetic bead sorting

For all selections, yeast cultures harboring libraries of ^{10}Fn3-based antibody mimics were induced for 18 h at 30 °C in galactose-containing medium (90% SG-CAA/10% SD-CAA, 50 μg/mL kanamycin, 100 U/mL penicillin G, 200 U/mL streptomycin).[21,22]

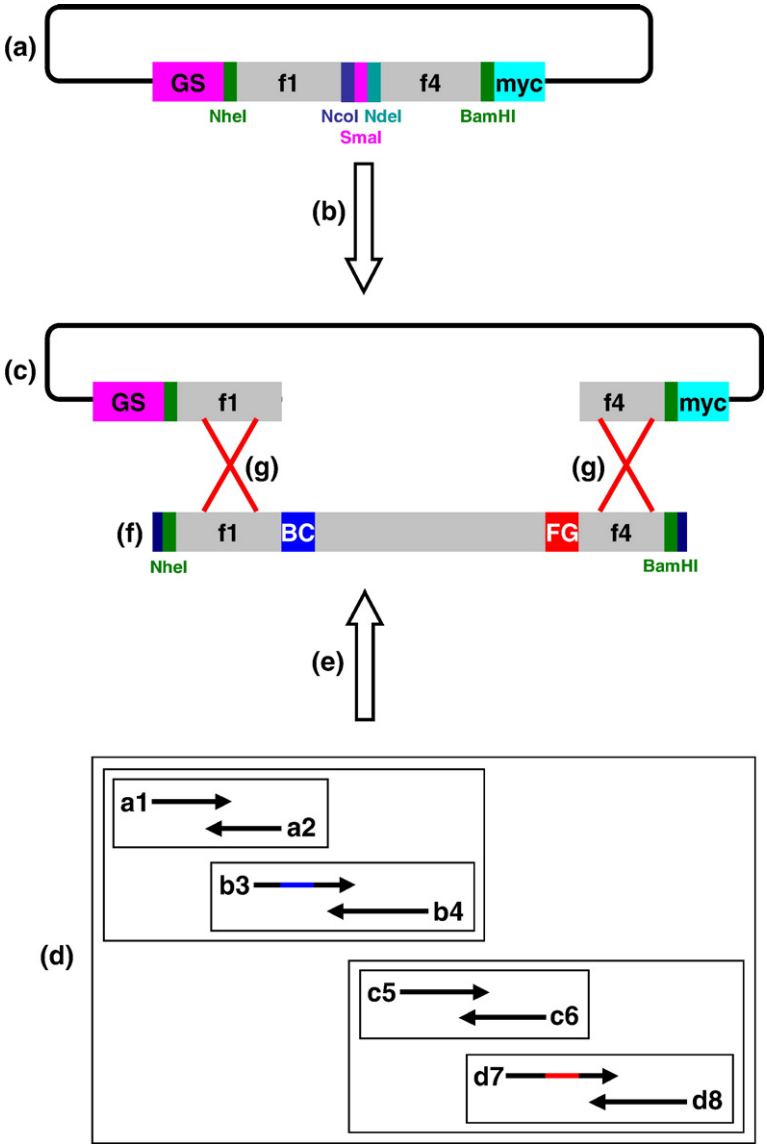

Figure 9. Construction of antibody-mimic libraries based on human ^{10}Fn3 for yeast surface display. (a) Yeast surface display vector pCTf1f4 contains the invariant human ^{10}Fn3 sequences N-terminal of the BC loop (f1) and C-terminal of the FG loop (f4), separated by NcoI, NdeI, and SmaI restriction sites. Digestion of pCTf1f4 with NcoI, NdeI, and SmaI (b) generates a linearized, gapped version of the plasmid (c). (d) Overlapping construction oligonucleotides a1–d8 were used in a series of PCR-like annealing/extension reactions (e) to yield a synthetic piece of DNA that encodes a full-length, ^{10}Fn3-like construct (f). The blue section of primer b3 and the red section of primer d7 indicate the sequence encoding the stretches of seven amino acid residues that were randomized in the BC and FG loops, respectively, to generate libraries of variant sequences. The nested boxes in (d) show the order in which construction oligos and DNA fragments resulting from their extension were combined. Linearized plasmid pCTf1f4 (c) and synthetic, ^{10}Fn3-derived insert (f) were co-transformed into electrocompetent EBY100 *S. cerevisiae*, resulting in *in vivo* homologous recombination (g) driven by the 40-nucleotide overlaps between the plasmid and the insert in f1 and f4 regions and thus in the incorporation of the insert into the plasmid.

The first round of selection was performed using magnetic beads (MACS), which allowed the sorting of cultures of >10^9 yeast cells. The starting cultures contained the larger of 10^9 or 10× representation of the library of induced yeast cells. The yeast was washed with 25 mL of ice-cold phosphate-buffered saline (PBS), pH 7.4, 2 mM ethylenediaminetetraacetic acid (EDTA), 0.5% bovine serum albumin (BSA) and then incubated in 5 mL of the same buffer containing 1 μM biotinylated hen egg white lysozyme (Sigma, St. Louis, MO) for 1 h at room temperature with gentle rotation. Following the incubation, the sample was chilled on ice, washed with 25 mL of ice-cold PBS, pH 7.4, 2 mM EDTA, 0.5% BSA and resuspended in 2.5 mL of the same buffer. A 100-μL aliquot of Anti-Biotin MicroBeads (Miltenyi Biotec, Auburn, CA) was added to the yeast and incubated on ice for 10 min. Ice-cold PBS, pH 7.4, 2 mM EDTA, 0.5% BSA was added to the sample to a total volume of 25 mL immediately before it was subjected to separation on an AutoMACS Cell Separator (Miltenyi Biotec), using the preset program for positive selection of rare cells (possel_s). Selected cells were collected in 6 mL SD-CAA, pH 4.5, 50 μg/mL kanamycin, 100 U/mL penicillin G, 200 U/mL streptomycin; quantified by serial dilution followed by plating on SD-CAA agar plates; and grown in 50 mL of the same medium for 2 days at 30 °C.

Selection for binding to hen egg white lysozyme using fluorescence-activated cell sorting

The remaining seven rounds of selection were performed by FACS, starting with 2×10^6 to 3×10^6 induced yeast cells. The yeast was washed with 1 mL PBS, pH 7.4, 0.1% BSA, resuspended in 100 μL of the same buffer containing biotinylated hen egg white lysozyme (Lysozyme-Biotin Caproyl; Sigma), and incubated at room temperature with gentle rotation for 3 h (for 1 nM HEL-b) or for 1 h (for all other concentrations of HEL-b).

After being washed with 1 mL of ice-cold PBS, pH 7.4, 0.1% BSA, the cells were labeled with primary and secondary antibodies as described previously.[21,22] All antibody and streptavidin-based reagents were obtained from Molecular Probes (Eugene, OR). Chicken anti-c-*myc* primary antibody and fluorescentlabeled goat anti-chicken secondary antibody were used to label the yeast for surface display of c-*myc*–tagged antibody mimics, and fluorescent- labeled streptavidin or anti-biotin antibody

was used to label HEL-b associated with lysozyme-binding antibody mimics. The first FACS sort was performed on yeast cells labeled with Alexa-633–conjugated goat anti-chicken antibody and Alexa-488–conjugated murine anti-biotin antibody, and the remaining FACS sorts were performed on yeast cells labeled with Alexa-488–conjugated goat anti-chicken antibody and Alexa-633 conjugated streptavidin (Molecular Probes).

Double-labeled yeast cells were sorted on a Becton–Dickinson Aria high-speed cell sorter with 488 and 635 nM lasers, at 6000–10,000 cells/s. Gates were adjusted to collect the yeast cells with the highest 0.1–1% of HEL-b–associated signal (Alexa 488 in the first FACS sort and Alexa 633 in the remaining six FACS sorts) and in the top half of expression-associated signal (Alexa 633 in the first FACS sort and Alexa 488 in the remaining six FACS sorts), as shown in Figure 3. Duplicate samples labeled with the same antibody and streptavidin reagents, but in the absence of HEL-b (e.g., Figure 3(b) and (d)), were used to avoid selecting the cells that bound detection reagents instead of lysozyme.

For all libraries, the first two FACS sorts were performed on yeast labeled with 1 μM HEL-b. Once a population of cells was observed that was labeled with Alexa 633 in the presence (e.g., Figure 3(a) and (c)) but not in the absence (e.g., Figure 3(b) and (d)) of HEL-b, the concentration of HEL-b in the subsequent round was decreased by an order of magnitude. For library BC7, only three FACS sorts were performed, with all induced yeast samples labeled with 1 μM HEL-b. For libraries FG7, 2L14, BF14, and BFs1, the seven FACS sorts were performed on yeast labeled at the following concentrations of HEL-b: 1 μM, 1 μM, 1 μM, 100 nM, 100 nM, 10 nM, 10 nM. For library BFs2, the first five FACS sorts were performed on yeast labeled at 1 μM HEL-b, followed by two FACS sorts on yeast labeled with 100 nM HEL-b. For library BFs3, the seven FACS sorts were performed on yeast labeled at the following concentrations of HEL-b: 1 μM, 1 μM, 100 nM, 10 nM, 1 nM, 1 nM, 1 nM.

Selected cells were collected in 0.5 mL of SD-CAA, pH 4.5, 50 μg/mL kanamycin, 100 U/mL penicillin G, and 200 U/mL streptomycin. The collected cells were grown to saturation in 5 mL of the same medium, with shaking, for 2 days at 30 °C, before being induced and labeled for the next round of sorting.

Sequencing of selected antibody mimics

Sequences of selected populations were determined both after four rounds of selection (one MACS sort followed by three FACS sorts) and after eight rounds of selection (one MACS sort followed by seven FACS sorts). Plasmid DNA was extracted from 1 mL of saturated culture of each selected yeast population using a Zymoprep kit (Zymo Research, Orange, CA), and 1 μL of the plasmid was transformed into competent XL1-Blue *E. coli* (Stratagene, La Jolla, CA). Plasmid minipreps were prepared and sequenced for between 20 and 34 colonies for each selected population (Table 2). Plasmids with sequence that was found in more than 10% of all sequenced clones in a given selected population were transformed back into EBY100 *S. cerevisiae* for further characterization of single clones.

Expression and purification of selected antibody mimics

The genes encoding residues 1–101 of clones BFs3_8.01 and BFs3_4.06 followed by a C-terminal hexahistidine purification tag were cloned into the pET-24b vector (Novagen, San Diego, CA) and the construct was transformed into BL21(DE3) pLysS *E. coli* (Invitrogen, Carlsbad, CA). Five hundred milliliters of Luria–Bertani medium with 50 μg/mL kanamycin and 34 μg/mL chloramphenicol was inoculated with an overnight culture and grown at 37 °C with rotation at 225 rpm to an A_{600} of 0.1–0.2. Protein expression was induced with 0.5 mM IPTG for 18–24 h at 30 °C at 225 rpm. Cells were collected by centrifugation at 3000*g* for 20 min, resuspended in 50 mL of lysis buffer (50 mM sodium phosphate, pH 8.0, 0.5 M NaCl, 5% glycerol, 5 mM CHAPS, 25 mM imidazole, and 1× Complete EDTA-free protease inhibitor cocktail (Roche, Indianapolis, IN)), and sonicated on ice for 90 s. Cell lysate was centrifuged at 19,000*g* for 40 min to isolate the soluble fraction. Five milliliters of TALON Superflow Metal Affinity Resin (Clontech, Mountain View, CA) was pre-equilibrated with wash buffer No. 1 (50 mM sodium phosphate, pH 8.0, 0.5 M NaCl, 5% glycerol, and 25 mM imidazole), added to the soluble fraction, and mixed for 1 h at 4 °C. The loaded resin was washed with 50 mL each of wash buffer No. 1 and wash buffer No. 2 (PBS, pH 7.4, and 25 mM imidazole). Histidine-tagged ^{10}Fn3-based antibody mimic was eluted with 10-mL fractions of elution buffer (PBS, pH 7.4, and 500 mM imidazole), dialyzed extensively against PBS, pH 7.4, and concentrated to 0.5 mL with an Amicon Ultra centrifugal filter with a 5-kDa membrane (Millipore, Billerica, MA). Monomeric protein was isolated via gel filtration on a Superdex 75 HR10/300 column (Amersham Pharmacia Biotech, Piscataway, NJ). The yield of BFs3_8.01 was 0.3 mg/L of bacterial culture, and the yield of BFs3_4.06 was 0.1 mg/L.

Measurement of affinity for hen egg white lysozyme for selected antibody mimics

Affinity of selected antibody mimics was determined using yeast surface display according to established protocols.[21,22] Briefly, yeast cultures harboring pCT-derived plasmids encoding single selected clones of ^{10}Fn3-based antibody mimics were induced and washed as described above. Aliquots of induced yeast cultures were incubated for 14–18 h at 25 °C, with shaking, with 0.05–10,000 nM HEL-b. After incubation, the samples were chilled on ice, washed with 500 μL of ice-cold PBS, pH 7.4, 0.1% BSA, and incubated for 15 min on ice, shielded from light, in 50 μL of the same buffer containing 10 μg/mL R-phycoerythrin–conjugated streptavidin (Molecular Probes). The samples were then washed with 500 μL of ice-cold PBS, pH 7.4, 0.1% BSA, and analyzed on a Coulter Epics XL flow cytometer. *F*_dis, the mean fluorescence of the yeast population that displays the ^{10}Fn3 variant at a particular concentration of HEL-b, was used as a measure of [Fn3:HEL-b], the concentration of the complex between the yeast-displayed Fn3 and biotinylated lysozyme:

$$\begin{aligned}\text{Fractional occupancy} &= [\text{Fn3:HEL-b}]/[\text{Fn3}]_0 \\ &= (F\_\text{dis} - F\_\text{dis\_min})/F\_\text{dis\_max}\end{aligned}$$

$$[\text{Fn3:HEL-b}]/[\text{Fn3}]_0 = [\text{HEL-b}]/(K_\text{d} + [\text{HEL-b}])$$

$$\begin{aligned}&\Rightarrow (F\_\text{dis} - F\_\text{dis\_min})/F\_\text{dis\_max} \\ &= [\text{HEL-b}]/(K_\text{d} + [\text{HEL-b}])\end{aligned}$$

$$\begin{aligned}&\Rightarrow F\_\text{dis} = F\_\text{dis\_min} + F\_\text{dis\_max}^*[\text{HEL-b}] \\ &/(K_\text{d} + [\text{HEL-b}])\end{aligned}$$

Non-specific binding of lysozyme at concentrations above 10 μM to yeast cells (data not shown) and limited

availability of biotinylated lysozyme precluded accurate determination of dissociation constants higher than 500 nM.

The affinity of selected antibody mimics was also determined by an equilibrium competition titration, in which purified antibody mimic was titrated for its ability to compete for binding to biotinylated lysozyme with a yeast-displayed antibody mimic. Yeast harboring a pCT-derived plasmid encoding for BFs3_8.01 was induced and washed as described above. Cells were resuspended in 100 μL of PBS, pH 7.4, 0.1% BSA, 1.0 nM HEL-b, and 0.01–500 nM purified ^{10}Fn3-based antibody mimic. Samples were incubated at 25 °C for 24 h, labeled with R-phycoerythrin–conjugated streptavidin, as described above for yeast surface display titrations, and analyzed by flow cytometry.

The equilibrium competition experiment consists of two equilibria, the equilibrium between the surface-displayed ^{10}Fn3 variant BFs3_8.01 and HEL-b, and the equilibrium between purified, soluble ^{10}Fn3 variant (BFs3_8.01 or BFs3_4.06) and HEL-b. Assuming a two-state binding model for both interactions, mean fluorescence data were fit with four parameters: K_d of surface-displayed Fn3, K_d of soluble Fn3, minimum fluorescence, and maximum fluorescence. The K_d of surface-displayed Fn3 variant BFs3_8.01 was previously measured as described above, and the remaining three parameters were fit using a least-squares method.

Competition for binding to selected antibody mimics FG4.1 and BF4.4 among biotinylated hen egg white lysozyme, unbiotinylated hen egg white lysozyme, and biotin

To compare the binding epitopes of lysozyme-binding variants FG4.1 and BF4.4, yeast displaying each variant was incubated with HEL-b at the concentration equal to its K_d, as well as with an equimolar amount, 10×, 100×, or 1000× excess of either unbiotinylated HEL or biotin. After a 2-h incubation at 25 °C, the samples were washed and analyzed on a Coulter Epics XL flow cytometer as described above.

Construction of wild-type ^{10}Fn3 and of mutants of selected antibody mimics

Wild-type ^{10}Fn3 and mutants of selected antibody mimics were constructed from synthetic oligonucleotides using a process analogous to library construction described in Figure 9, except that defined oligonucleotides were used instead of oligonucleotides with randomized regions (Table 4).

Lysozyme-binding variant BFs3_8.01, which contains cysteines in positions 28 and 77, was mutated to replace either or both of the cysteines with serine, generating mutants BFs3_8.01(C28S), BFs3_8.01(C77S), and BFs3_8.01 (C28S,C77S).

In addition, lysozyme-binding variants 2L4.1, which contains a deletion in the BC loop, and BFs1c4.07, which contains an insertion in the BC loop, were mutated to wild-type BC loop length with minimal perturbation in loop sequence. Mutants 2L4.1_nG and 2L4.1_cG contain a single-glycine insertion at the N-terminal and C-terminal end of the 2L4.1 BC loop, resulting in BC loops GWPCQLA and WPCQLAG, respectively. Mutants BFs1c4.07_nX and BFs1c4.07_cX contain single-residue deletions at the N-terminal and C-terminal ends of the BFs1c4.07 BC loop, resulting in BC loops CCEYWYP and WCCEYWY, respectively.

Oligonucleotides used in construction of the mutants are listed in Table 4. Plasmids encoding the mutants were transformed into EBY100 *S. cerevisiae,* and lysozyme-binding properties of the mutants were quantified as described above.

Modeling of disulfide bonds

The NMR restrained energy-minimized average structure of the 10th type III module of human fibronectin (PDB code 1TTG)[5] was prepared for computational modeling by removing all water molecules and building on hydrogen atoms using the HBUILD facility[37] and the param22 all-atom parameter set[38] in CHARMM.[39] Disulfide bonds were modeled among all 49 pairs of BC-loop and FG-loop positions. For each pair of positions, two models were created: one with a disulfide bond and one with two free cysteines. The remaining 12 positions in the BC and FG loops were modeled as all glycine, all alanine, or all serine. The cysteine and serine side-chains were built in their default conformation.

The structure with two free cysteines and the structure with a disulfide bond were each energy-minimized using no non-bonded cut-offs and a 4r distance-dependent dielectric constant. The minimization protocol included a harmonic restraint of 0.1 kcal/mol per Å^2 on each atom in the protein to preserve the native structure where possible and to reduce the effects of the imperfect energy function. Each structure was minimized to convergence, the harmonic restraints were then reset for the new atomic coordinates, and the process was repeated for a total of three complete minimizations. Similar results were obtained with two or four rounds of minimization.

The energy reported for a disulfide bond between a particular pair of positions is the difference between total energies, after minimization, of the disulfide-bonded structure and the corresponding structure with two free cysteines. Similar results were obtained when directly using the final energy after minimization for each disulfide. RMSD calculations were performed between all atoms common to both structures, using the McLachlan algorithm[40] as implemented in the program ProFit V2.2 (Martin, A.C.R.)†. Figure 6 was drawn using Visual Molecular Dynamics.[41]

Acknowledgements

This project was funded by the DuPont-MIT Alliance (DL), CA96504, CA101830, a National Science Foundation Graduate Fellowship (SML), and a National Defense Science and Engineering Graduate Fellowship (BJH). We thank the MIT Flow Cytometry Core Facility for technical support and Mike Schmidt for permission to include his unpublished results in Figure 8.

References

1. Nygren, P. A. & Skerra, A. (2004). Binding proteins from alternative scaffolds. *J. Immunol. Methods*, **290**, 3–28.

† http://www.bioinf.org.uk/software/profit/

2. Binz, H. K. & Pluckthun, A. (2005). Engineered proteins as specific binding reagents. *Curr. Opin. Biotechnol.* **16**, 459–469.
3. Silverman, J., Liu, Q., Bakker, A., To, W., Duguay, A., Alba, B. M. *et al.* (2005). Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains. *Nature Biotechnol.* **23**, 1556–1561.
4. Parker, M. H., Chen, Y., Danehy, F., Dufu, K., Ekstrom, J., Getmanova, E. *et al.* (2005). Antibody mimics based on human fibronectin type three domain engineered for thermostability and high-affinity binding to vascular endothelial growth factor receptor two. *Protein Eng. Des. Sel.* **18**, 435–444.
5. Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. (1992). The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell*, **71**, 671–678.
6. Dickinson, C. D., Veerapandian, B., Dai, X. P., Hamlin, R. C., Xuong, N. H., Ruoslahti, E. & Ely, K. R. (1994). Crystal structure of the tenth type III cell adhesion module of human fibronectin. *J. Mol. Biol.* **236**, 1079–1092.
7. Koide, A., Bailey, C. W., Huang, X. & Koide, S. (1998). The fibronectin type III domain as a scaffold for novel binding proteins. *J. Mol. Biol.* **284**, 1141–1151.
8. Richards, J., Miller, M., Abend, J., Koide, A., Koide, S. & Dewhurst, S. (2003). Engineered fibronectin type III domain with a RGDWXE sequence binds with enhanced affinity and specificity to human alphavbeta3 integrin. *J. Mol. Biol.* **326**, 1475–1488.
9. Karatan, E., Merguerian, M., Han, Z., Scholle, M. D., Koide, S. & Kay, B. K. (2004). Molecular recognition properties of FN3 monobodies that bind the Src SH3 domain. *Chem. Biol.* **11**, 835–844.
10. Koide, A., Abbatiello, S., Rothgery, L. & Koide, S. (2002). Probing protein conformational changes in living cells by using designer binding proteins: application to the estrogen receptor. *Proc. Natl Acad. Sci. USA*, **99**, 1253–1258.
11. Xu, L., Aha, P., Gu, K., Kuimelis, R. G., Kurz, M., Lam, T. *et al.* (2002). Directed evolution of high-affinity antibody mimics using mRNA display. *Chem. Biol.* **9**, 933–942.
12. Huang, J., Koide, A., Nettle, K. W., Greene, G. L. & Koide, S. (2005). Conformation-specific affinity purification of proteins using engineered binding proteins: application to the estrogen receptor. *Protein Expr. Purif*, **47**, 348–354.
13. Boder, E. T. & Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnol.* **15**, 553–557.
14. Amit, A. G., Mariuzza, R. A., Phillips, S. E. & Poljak, R. J. (1985). Three-dimensional structure of an antigen–antibody complex at 6 A resolution. *Nature*, **313**, 156–158.
15. Bentley, G. A. (1996). The crystal structures of complexes formed between lysozyme and antibody fragments. *EXS*, **75**, 301–319.
16. Desmyter, A., Transue, T. R., Ghahroudi, M. A., Thi, M. H., Poortmans, F., Hamers, R. *et al.* (1996). Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nature Struct. Biol.* **3**, 803–811.
17. Stanfield, R. L., Dooley, H., Flajnik, M. F. & Wilson, I. A. (2004). Crystal structure of a shark single-domain antibody V region in complex with lysozyme. *Science*, **305**, 1770–1773.
18. Roux, K. H., Greenberg, A. S., Greene, L., Strelets, L., Avila, D., McKinney, E. C. & Flajnik, M. F. (1998). Structural analysis of the nurse shark (new) antigen receptor (NAR): molecular convergence of NAR and unusual mammalian immunoglobulins. *Proc. Natl Acad. Sci. USA*, **95**, 11804–11809.
19. Muyldermans, S., Atarhouch, T., Saldanha, J., Barbosa, J. A. & Hamers, R. (1994). Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Eng.* **7**, 1129–1135.
20. Barbas, C. F., 3rd, Hu, D., Dunlop, N., Sawyer, L., Cababa, D., Hendry, R. M. *et al.* (1994). In vitro evolution of a neutralizing human antibody to human immunodeficiency virus type 1 to enhance affinity and broaden strain cross-reactivity. *Proc. Natl Acad. Sci. USA*, **91**, 3809–3813.
21. Colby, D. W., Kellogg, B. A., Graff, C. P., Yeung, Y. A., Swers, J. S. & Wittrup, K. D. (2004). Engineering antibody affinity by yeast surface display. *Methods Enzymol.* **388**, 348–358.
22. Chao, G., Lau, W. L., Hackel, B. J., Sazinsky, S. L., Lippow, S. M. & Wittrup, K. D. (2006). Isolating and engineering human antibodies using yeast surface display. *Nature Protocols*, **1**, 755–768.
23. Feldhaus, M. J., Siegel, R. W., Opresko, L. K., Coleman, J. R., Feldhaus, J. M., Yeung, Y. A. *et al.* (2003). Flow-cytometric isolation of human antibodies from a nonimmune Saccharomyces cerevisiae surface display library. *Nature Biotechnol.* **21**, 163–170.
24. VanAntwerp, J. J. & Wittrup, K. D. (2000). Fine affinity discrimination by yeast surface display and flow cytometry. *Biotechnol. Prog.* **16**, 31–37.
25. Hawkins, R. E., Russell, S. J., Baier, M. & Winter, G. (1993). The contribution of contact and non-contact residues of antibody in the affinity of binding to antigen. The interaction of mutant D1.3 antibodies with lysozyme. *J. Mol. Biol.* **234**, 958–964.
26. Graff, C. P., Chester, K., Begent, R. & Wittrup, K. D. (2004). Directed evolution of an anti-carcinoembryonic antigen scFv with a 4-day monovalent dissociation half-time at 37 degrees C. *Protein Eng. Des. Sel.* **17**, 293–304.
27. Razai, A., Garcia-Rodriguez, C., Lou, J., Geren, I. N., Forsyth, C. M., Robles, Y. *et al.* (2005). Molecular evolution of antibody affinity for sensitive detection of botulinum neurotoxin type A. *J. Mol. Biol.* **351**, 158–169.
28. Midelfort, K. S. & Wittrup, K. D. (2006). Context-dependent mutations predominate in an engineered high-affinity single chain antibody fragment. *Protein Sci.* **15**, 324–334.
29. Boder, E. T., Midelfort, K. S. & Wittrup, K. D. (2000). Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc. Natl Acad. Sci. USA*, **97**, 10701–10705.
30. Kieke, M. C., Sundberg, E., Shusta, E. V., Mariuzza, R. A., Wittrup, K. D. & Kranz, D. M. (2001). High affinity T cell receptors from yeast display libraries block T cell activation by superantigens. *J. Mol. Biol.* **307**, 1305–1315.
31. Buonpane, R. A., Moza, B., Sundberg, E. J. & Kranz, D. M. (2005). Characterization of T cell receptors engineered for high affinity against toxic shock syndrome toxin-1. *J. Mol. Biol.* **353**, 308–321.
32. Carugo, O., Cemazar, M., Zahariev, S., Hudaky, I., Gaspari, Z., Perczel, A. & Pongor, S. (2003). Vicinal disulfide turns. *Protein Eng.* **16**, 637–639.
33. Tanha, J., Dubuc, G., Hirama, T., Narang, S. A. & MacKenzie, C. R. (2002). Selection by phage display of llama conventional V(H) fragments with heavy chain

antibody V(H)H properties. *J. Immunol. Methods*, **263**, 97–109.
34. Vu, K. B., Ghahroudi, M. A., Wyns, L. & Muyldermans, S. (1997). Comparison of llama VH sequences from conventional and heavy chain antibodies. *Mol. Immunol.* **34**, 1121–1131.
35. Diaz, M., Stanfield, R. L., Greenberg, A. S. & Flajnik, M. F. (2002). Structural analysis, selection, and ontogeny of the shark new antigen receptor (IgNAR): identification of a new locus preferentially expressed in early development. *Immunogenetics*, **54**, 501–512.
36. Streltsov, V. A., Varghese, J. N., Carmichael, J. A., Irving, R. A., Hudson, P. J. & Nuttall, S. D. (2004). Structural evidence for evolution of shark Ig new antigen receptor variable domain antibodies from a cell-surface receptor. *Proc. Natl Acad. Sci. USA*, **101**, 12444–12449.
37. Brunger, A. T. & Karplus, M. (1988). Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. *Proteins*, **4**, 148–156.
38. MacKerell, J., A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Jr, Evanseck, J. D., Field, M. J. *et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
39. Brooks, B. R., Bruccolerire, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217.
40. McLachlan, A. D. (1982). Rapid comparison of protein structures. *Acta Crystallog. A*, **38**, 871–873.
41. Humphrey, W., Dalke, A., Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38, 27–28.

Edited by F. Schmid

(Received 30 May 2006; received in revised form 6 February 2007; accepted 9 February 2007)
Available online 22 February 2007