



High-throughput Generation of Synthetic Antibodies from Highly Functional Minimalist Phage-displayed Libraries

Frederic A. Fellouse^{1†}, Kaori Esaki², Sara Birtalan¹, Demetrios Raptis²
Vincenzo J. Cancasci², Akiko Koide², Parkash Jhurani¹, Mark Vasser¹
Christian Wiesmann¹, Anthony A. Kossiakoff²
Shohei Koide^{2*} and Sachdev S. Sidhu^{1*}

¹Department of Protein Engineering, Genentech Inc
1 DNA Way, South San Francisco, CA 94080, USA

²Department of Biochemistry and Molecular Biology
The University of Chicago
929 East 57th Street, Chicago IL 60637, USA

Received 22 May 2007;
received in revised form 1 August 2007;
accepted 1 August 2007
Available online 19 August 2007

We have previously established a minimalist approach to antibody engineering by using a phage-displayed framework to support complementarity determining region (CDR) diversity restricted to a binary code of tyrosine and serine. Here, we systematically augmented the original binary library with additional levels of diversity and examined the effects. The diversity of the simplest library, in which only heavy chain CDR positions were randomized by the binary code, was expanded in a stepwise manner by adding diversity to the light chain, by diversifying non-paratope residues that may influence CDR conformations, and by adding additional chemical diversity to CDR-H3. The additional diversity incrementally improved the affinities of antibodies raised against human vascular endothelial growth factor and the structure of an antibody–antigen complex showed that tyrosine side-chains are sufficient to mediate most of the interactions with antigen, but a glycine residue in CDR-H3 was critical for providing a conformation suitable for high-affinity binding. Using new high-throughput procedures and the most complex library, we produced multiple high-affinity antibodies with dissociation constants in the single-digit nanomolar range against a wide variety of protein antigens. Thus, this fully synthetic, minimalist library has essentially recapitulated the capacity of the natural immune system to generate high-affinity antibodies. Libraries of this type should be highly useful for proteomic applications, as they minimize inherent complexities of natural antibodies that have hindered the establishment of high-throughput procedures. Furthermore, analysis of a large number of antibodies derived from these well-defined and simplistic libraries allowed us to uncover statistically significant trends in CDR sequences, which provide valuable insights into antibody library design and into factors governing protein–protein interactions.

© 2007 Published by Elsevier Ltd.

Edited by I. Wilson

Keywords: phage display; protein engineering; combinatorial mutagenesis; antibody library; vascular endothelial growth factor

*Corresponding authors. E-mail addresses: skoide@uchicago.edu; sidhu@gene.com.

Present address: F. A. Fellouse, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada M5G 1X5.

Abbreviations used: BSA, bovine serum albumin; CDR, complementarity determining region; CDR-H_n, (where *n*=1, 2, or 3), heavy chain CDR 1, 2, or 3; CDR-L3, light chain CDR3; ELISA, enzyme-linked immunosorbent assay; Fab, antigen-binding fragment; HTP, high-throughput pipeline; hVEGF, human vascular endothelial growth factor; mAb, monoclonal antibody; scFv, single-chain variable fragment.

Introduction

The natural immune system can produce antibodies that recognize essentially any antigen with high affinity and specificity, and thus, antibodies produced by animal immunization and hybridoma methods¹ are indispensable tools for biological research. As a powerful alternative to hybridoma technology, phage display allows for the display of libraries of antigen-binding fragments (Fabs) or single-chain variable fragments (scFvs) on the surfaces of bacteriophage particles that encapsulate the encoding DNA.^{2–6} These libraries can be cycled through binding selections *in vitro* to select antigen-specific antibodies. Selected antibodies are inherently monoclonal, and their amino acid sequences can be readily decoded from the DNA. Furthermore, the use of *in vitro* methods allows for exquisite control over selection conditions, thus allowing for tailored selection strategies for obtaining antibodies with extremely precise specificities, and also, for dealing with difficult antigens, such as integral membrane proteins. Importantly, phage display selection and analysis protocols are well suited for adaptation to a high-throughput pipeline (HTP), and consequently, the technology is likely to play an important role in proteomics research.^{7–9}

A particularly promising branch of phage-displayed library technology is represented by the so-called “synthetic” antibody libraries, which contain antigen-binding sites constructed entirely from man-made diversity.^{10,11} Libraries of this type are ideal for HTP applications, since frameworks can be chosen for optimal stability and performance, and the defined nature of the repertoires allows for rapid downstream sequence analysis and protein purification. In addition, the use of well-defined frameworks makes it possible to incorporate modular design features that enable facile affinity maturation or reformatting between different vector systems.

Over the last few years, significant advances in synthetic antibody engineering have been enabled by an enhanced understanding of the relationships between antibody structure and function. Analysis of antibody sequences and structures has revealed that antigen recognition is primarily mediated by the six complementarity determining regions (CDRs), or hypervariable loops, which form a functional antigen-binding site that is supported by more conserved framework regions.^{12,13} Even within the CDRs, there are considerable biases in spatial and chemical diversity that is compatible with antigen recognition; antibody–antigen complex structures show that residues at certain CDR positions most often make contact with antigen and certain types of amino acids are most often involved in productive binding contacts.^{14–17} Furthermore, while natural repertoires use combinations of dozens of different heavy and light chain frameworks to support CDR diversity, certain frameworks are much more stable than others, and consequently appear to be used more often in functional antibodies.¹⁸ Highly functional synthetic antibody libraries have been

constructed by using these natural clues to guide the choice of frameworks that are the best for supporting naïve diversity and to determine which types of chemical diversity should be introduced at which positions within the antigen-binding site.^{10,11}

We have pursued a particularly simple approach to antibody engineering by using a single, highly stable Fab framework to support CDR diversity restricted to a binary code of only Tyr and Ser, two amino acids that are highly abundant in natural antigen-binding sites.^{19–21} Surprisingly, these chemically minimalist libraries have proven to be highly effective in generating specific antibodies against a wide array of antigens, thus demonstrating a dominant role for Tyr in antigen recognition and providing a rationale for the high abundance of Tyr in the CDRs of the natural immune repertoire. Recently, we have shown that the effectiveness of the Tyr/Ser binary code in mediating molecular recognition is likely an intrinsic property not restricted to antibody scaffolds, as libraries of this type were also effective in providing specific molecular recognition in the context of a much smaller, single-domain scaffold of the fibronectin type III domain.²²

These studies have established a benchmark for the minimal requirements for generating molecular recognition from naïve repertoires. However, while the Tyr/Ser binary libraries provided high affinity recognition in the nanomolar range for some antigens, in other cases affinities were in the micromolar range, suggesting that additional diversity is required for high affinity binding to a greater proportion of antigens. In the original Tyr/Ser binary Fab libraries, only positions that are expected to make a direct contact with the antigen (“paratope” residues) were diversified.²⁰ The natural antibody repertoire contains two additional types of CDR diversity.^{12,13} First, amino acid diversity is present at CDR positions that do not make direct contact with the antigen. Such diversity at “non-paratope” residues contributes to generating conformational diversity of the paratope residues. Second, additional chemical diversity beyond the Tyr/Ser binary code is likely to augment affinity and fine-tune specificity.

Here, we have systematically augmented the diversity in the minimalist Tyr/Ser binary code to produce simple, yet highly functional synthetic antibodies. We introduced additional CDR diversity in a stepwise manner, first at non-paratope positions and then in all CDR-H3 positions, and evaluated the effects systematically using human vascular endothelial growth factor (hVEGF) as the test antigen. The introduction of additional diversity incrementally improved the effectiveness of the Fab libraries. Our most complex Fab library was further evaluated with a diverse panel of antigens using newly established, HTP procedures for Fab generation. The library produced antibodies with tight affinities in the single-digit nanomolar range against all of the antigens tested, indicating that this fully

synthetic library, with highly biased amino acid codes built on a single Fab scaffold, has essentially recapitulated the capacity of the natural immune system to generate high-performance antibodies. Libraries of this type should be highly useful for HTP applications, as they use simple synthetic antibody design but incorporate enough additional conformational and chemical diversity to mediate high-affinity recognition of diverse antigens. Furthermore, the defined and simplistic nature of the library designs allowed us to detect statistically significant trends in sequence diversity amongst antigen-specific antibodies relative to the naïve repertoire, and these trends provide valuable insights into successful aspects of design that can be reinforced in future designs to further improve library performance.

Results

Library designs

We designed four synthetic antibody libraries to systematically study the functional consequences of adding complexity into a minimalist diversity design. All of the libraries used a similar humanized Fab as the scaffold (Figure 1(a)) and diversity was restricted to defined positions in the CDRs (Figure 1(b)). In the simplest library (library A), which has been described,²⁰ only the three heavy chain CDRs were diversified using a binary code restricted to Tyr/Ser. Solvent accessible paratope positions in CDR-H1 and -H2 were randomized, and seven positions in CDR-H3 were replaced by random-sequence loops of variable lengths. In the second library (library B), also described,²⁰ the complexity relative to library A was expanded by introducing diversity into the light chain at five paratope positions in CDR-L3.

For the next two libraries, the scaffold was modified slightly based on structural analysis of Fab-YSd1, an antibody derived from library B, which revealed that certain side-chains in CDR-L1 and -L2 were close to or in contact with bound antigen.²⁰ Although CDR-L1 and -L2 were not diversified in the libraries, we replaced four light chain residues (Asp28, Asn30, Thr31 and Phe53) with Ser residues, reasoning that the small, neutral Ser side-chains would be less likely to make detrimental contacts with antigens. In addition, we also replaced Ser93 in the heavy chain with Ala, because Ala is the most common residue at this position in natural antibodies¹² and we observed that this substitution improved the yield of Fab protein (data not shown). Using this new scaffold, we constructed a third library (library C) that built upon the complexity of library B by allowing for three different lengths in CDR-L3 and by randomizing eight non-paratope positions that likely influence the conformations of the CDRs (two each in CDR-H1, -H2, -H3 and -L3). Each of these positions was subjected to limited randomization

using degenerate codons that encode two or four amino acids that are commonly found at each position amongst natural antibody sequences.¹² Library D, the fourth and most complex library, was identical to library C, except that the chemical diversity of CDR-H3 was greatly increased. In contrast with the other libraries, which allowed for only binary Tyr/Ser diversity, the CDR-H3 design of library D was chemically complex; we used a tailored oligonucleotide synthesis strategy that biased the sequences in favor of Tyr, Ser and Gly but allowed for 19 of the 20 genetically encoded amino acids (only Cys was excluded).

Each of the libraries was constructed with optimized mutagenesis techniques^{23,24} that yielded highly diverse repertoires containing greater than 10¹⁰ unique members each. The four libraries differ greatly in terms of the amount of coverage provided by the actual diversities relative to the theoretical diversities, that is, the maximum number of unique combinatorial possibilities for the CDR diversity designs (Figure 1(c)). For the simplest libraries A and B, the theoretical diversities are lower than the actual diversities, and thus, these diversity designs are well sampled by the constructed repertoires. In library C, the addition of length diversity in CDR-L3 and chemical diversity at non-paratope residues produces a case where the theoretical diversity exceeds the actual diversity by approximately 1000-fold. In the case of the most complex library D, the use of highly diverse chemical diversity in CDR-H3 produces an extremely under-sampled situation in which the theoretical diversity exceeds the actual diversity by 20 orders of magnitude.

Selection of antibodies against hVEGF

To compare the relative effectiveness of the different libraries in generating specific antibodies, we chose hVEGF as a model antigen, because this angiogenic hormone is of great biological importance and has proven to be highly amenable to structural analysis in complex with natural and synthetic Fabs.^{21,25,26} All four libraries generated anti-hVEGF Fabs, and sequencing of approximately 100 clones from each selection revealed a total of 90 unique sequences representing 11, 52, 14 or 13 clones from libraries A, B, C or D, respectively (Figure 2).

For each unique clone, we used a single-point competitive phage enzyme-linked immunosorbent assay (ELISA) to obtain estimates of affinity.²¹ Phage were incubated with 100 nM hVEGF to establish an equilibrium between the phage-displayed Fabs and hVEGF. Uncomplexed Fab-phage were then detected by capture on plates coated with hVEGF. By comparing the amount of Fab-phage captured in the presence of solution-phage hVEGF to the amount captured in the absence of hVEGF, we could rank the clones on the basis of relative affinities (data not shown). Of the 90 unique clones, only 11 showed a greater than 50% reduction in binding in the presence of 100 nM hVEGF, and these included nine clones from library D and one clone

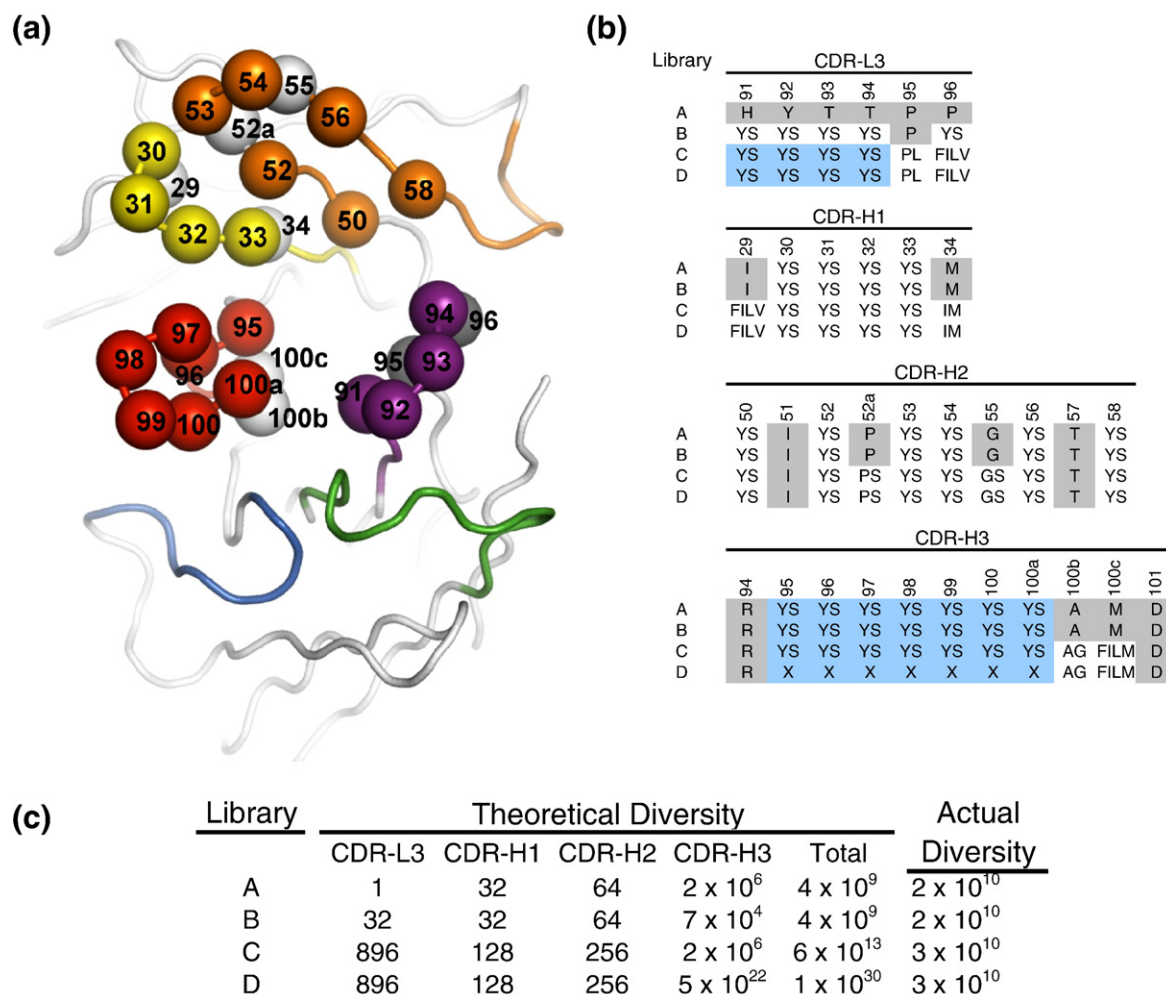


Figure 1. Library designs. (a) The backbones of the humanized 4D5 heavy and light chain variable domains are shown as tubes. The frameworks are colored grey, and the CDR loops defined by Kabat *et al.*¹² are colored as follows: CDR-H1, yellow; CDR-H2, orange; CDR-H3, red; CDR-L1, green; CDR-L2, blue; CDR-L3, purple. Positions that were diversified in the libraries are shown as spheres. Spheres colored according to the CDR coloring scheme represent paratope positions that were randomized in all four libraries (except CDR-L3, which was not randomized in library A). Grey spheres represent non-paratope positions that were only randomized in libraries C and D. Positions that were fixed as the parental 4D5 sequence are shaded grey. At each diversified position, the allowed amino acids are denoted by the single-letter code. X indicates positions at which 19 of the 20 genetically encoded amino acids (Cys was excluded) were allowed as described in Materials and Methods. Positions shaded blue indicate regions that were replaced by random loops of varying length. In the case of CDR-L3, the four positions were replaced by loops containing four, five or six residues. In the case of CDR-H3, the seven positions were replaced by all loop lengths between seven and 20 residues (library A), seven and 15 residues (library B) or six and 17 residues (libraries C and D). (b) CDR diversity designs. Positions that were fixed as the parental 4D5 sequence are shaded grey. At each diversified position, the allowed amino acids are denoted by the single-letter code. X indicates positions at which 19 of the 20 genetically encoded amino acids (Cys was excluded) were allowed as described in Materials and Methods. Positions shaded blue indicate regions that were replaced by random loops of varying length. In the case of CDR-L3, the four positions were replaced by loops containing four, five or six residues. In the case of CDR-H3, the seven positions were replaced by all loop lengths between seven and 20 residues (library A), seven and 15 residues (library B) or six and 17 residues (libraries C and D). (c) Theoretical and actual library diversities. The theoretical diversities are the number of combinatorial possibilities allowed in each library design, and the actual diversities are the number of unique clones obtained from the library construction process.

each from libraries B and C (Figure 2). To identify the highest affinity clones from libraries A and B, we repeated the ELISA in the presence of 1 μ M hVEGF and found two additional clones from library B and one clone from library A that exhibited greater than 50% reduction in binding. Thus, we identified 14 clones that included the highest affinity anti-hVEGF Fabs from each library, and we next analyzed these Fabs in detail as purified proteins.

The binding kinetics of the purified Fabs were studied by surface plasmon resonance (Figure 2). Four of the nine Fabs from library D exhibited extremely tight affinities ($K_d < 10$ nM), and while

library C yielded only one high-affinity clone, the affinity of this Fab (Fab-C1, $K_D = 9.7$ nM) was comparable to that of the best Fab from library D (Fab-D2, $K_D = 4.4$ nM). The affinities of the best Fabs from libraries B (Fab-B1, $K_d = 30$ nM) or A (Fab-A1, $K_d = 300$ nM) were approximately one or two orders of magnitude weaker, respectively, than that of Fab-D2.

The CDR-H3 sequences obtained from the different libraries differed greatly in terms of length and sequence. All of the clones from library A had CDR-H3 loops with identical lengths and homologous sequences. In contrast, the clones from library B were of differing lengths and exhibited homology

Figure 2. Sequences and binding parameters of anti-hVEGF Fabs. CDR sequences are shown for Fabs isolated from (a) library A, (b) library B, (c) library C, and (d) library D. The numbering is according to the nomenclature of Kabat *et al.*¹² Tyr, Ser and Gly residues are shown in yellow, red or green, respectively. Residues in grey were not randomized. Dashes indicate gaps in the alignment and X denotes ambiguous sequence. Asterisks (*) or double asterisks (**) indicate that Fab-phage binding was inhibited at least 50% by solution-phase hVEGF at a concentration of 100 nM or 1 μ M, respectively. Binding parameters (k_{on} , k_{off} , K_d) were determined from kinetic analysis of Fabs binding to immobilized hVEGF measured by surface plasmon resonance.

amongst each other at the C-terminal end, but were not homologous to the clones from library A. The clones from library C appeared to be a mixture of sequences with some homology to clones from either library A or B, but differed significantly from both groups. Notably, the highest affinity clone from library C (Fab-C1) appeared to be homologous to the clones from library A at the N-terminal end of CDR-H3 but differed at the C-terminal end. The clones from library D were the most diverse in terms of length and sequence.

To characterize different CDR-H3 loops in detail, we subjected one clone each from libraries B, C and D to shotgun scanning mutagenesis, a method which assesses the functional importance of individual side-chains^{27,28} (Figure 3). Shotgun scanning revealed that the antigen binding function of Fab-B1 depends mainly on the C-terminal end of CDR-H3, consistent with the homology observed amongst clones from library B. In contrast, Fab-C1 depends on the N-terminal end of CDR-H3, again consistent with the homology of this clone with clones from library A. Three residues in the short CDR-H3 loop of Fab-D1 are important for antigen binding, and these include two Tyr and a Gly residue. Comparison of the results from libraries A, B and C demonstrates that, even when only binary diversity is allowed, very different CDR-H3 loops are used for the recognition of hVEGF, depending on the nature of the diversity allowed in the other CDRs. Furthermore, comparison of the results from libraries C and D shows that, even though increased chemical diversity in CDR-H3 does not significantly enhance the maximum affinity, it does significantly expand the range of CDR-H3 loop lengths and sequences that are used for binding, suggesting that the CDR-H3 loops contained in library D may offer a wider range of binding solutions.

Structural characterization of anti-hVEGF Fab-D1

To gain insights into the structural basis for antigen recognition, the crystal structure of Fab-D1 in complex with hVEGF was solved and refined to 3.5 Å resolution with R_{factor} and R_{free} of 29.1 and 31.2%, respectively (Figure 4 and Table 1). Despite

the modest resolution, the quality of the structure is quite good with 98.9% of the non-proline and non-glycine residues in the most favored or additionally allowed parts of the Ramachandran plot.²⁹ Importantly, the maps were of sufficient quality to unambiguously trace all backbone regions, including the CDR loops, and also, to visualize the intermolecular interactions at the interface between the antibody and the antigen.

The structure of Fab-D1 is similar to that of the parental Fab-4D5. Superposition of the variable domains reveals significant differences only in the main-chain conformations of CDR-H3 and -L3 (Figure 4(a)), and these differences are consistent with the fact that these loops differ drastically between the two antibodies, both in terms of length and sequence. Furthermore, these conformational differences are relevant to antigen recognition, since both loops make substantial contacts with hVEGF (see below). The structure of the hVEGF homodimer is also very similar to previous structures in complex with other Fabs or a natural receptor domain (Flt-1_{D2}).¹¹ Each hVEGF homodimer binds to two Fab-D1 molecules (Figure 4(b)), which recognize a structural epitope that overlaps with the structural epitope for Flt-1_{D2} (Figure 4(c)). Calculations of solvent-accessible surface areas³⁰ reveal that complexation results in the burial of approximately 850 Å² on both hVEGF and on Fab-D1, which uses both the heavy (485 Å²) and light (358 Å²) chains (Figure 4(d)). All three heavy chain CDR loops make contact with the antigen, exclusively through residues located at positions that were randomized in the library. In the light chain, contacts involve residues at randomized positions in CDR-L3, but also, three Ser residues in the non-randomized CDR-L1. It is striking that nine of 11 Tyr residues selected at randomized positions form contacts with the antigen, and consequently, 69% of the buried surface on the paratope side is contributed by Tyr residues (Figure 4(e)). An additional 18% of the structural paratope is contributed by Ser residues and the remaining 13% is from contacts involving other residue types in CDR-H3 (Gly97 and Thr98), CDR-L3 (Pro95) and the light chain framework (Arg66). Gly99 in CDR-H3 is not in direct contact with the antigen, but the main-chain adopts an unusual

Fab	CDR-H3													
	95	96	97	98	99	100	100a	100b	100c	100d	100e	100f	100g	100h
B1	Y >100	Y 3	S 2	S 3	S 0.7	Y 3	Y 3	Y 1	S >100	Y 58	Y 80	Y >100	A 0.8	M 1
C1	S 2	Y 7	Y >35	S >35	Y >35	S 3	S 3	S 2	S 2	S 2	S 0.9	S 0.7	G 2	L 0.6
D1	Y 12	Y 28	G 1	T 2	G 25	A 4	M 0.7	-	-	-	-	-	-	-

(indicated by a greater than sign). The effects of the following mutations were tested: Tyr → Ser, Ser/Gly → Ala, Ala → Gly, Met → Leu, Leu → Met. See Materials and Methods for details.

Figure 3. Shotgun scanning analysis of anti-hVEGF CDR-H3 loops. The F value for each side-chain is shown below each sequence, and positions at which mutations are predicted to be deleterious to antigen-binding ($F > 5$) are shown in bold text. In cases where no mutations were observed amongst the antigen-binding clones, only a lower limit could be defined for the F value

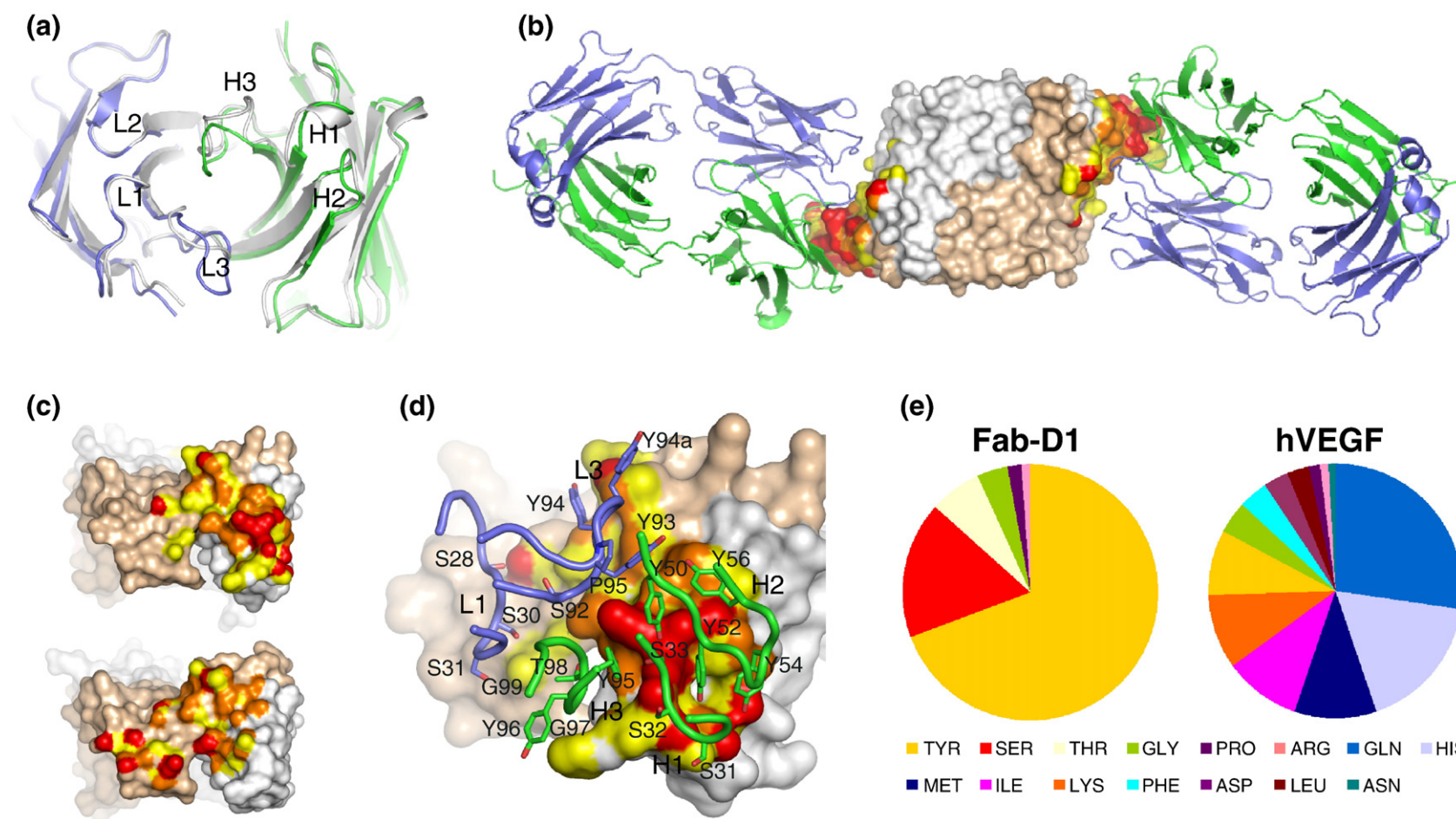


Figure 4. Structure of Fab-D1 in complex with hVEGF. (a) Superposition of the variable domains of Fab-D1 (light chain blue, heavy chain green) and humanized Fab-4D5 (1N8Z, white). Note the large differences in CDR-H3 and -L3. (b) Overall structure of the hVEGF:(Fab-D1)₂ complex. hVEGF is depicted in a surface representation with one molecule in white and the other in beige, the Fabs are shown as ribbon diagrams and colored as in (a). All atoms of hVEGF in contact with the Fab are colored yellow, orange or red if closer than 4.5 Å, 4.0 Å, or 3.5 Å to the Fab, respectively. (c) The structural epitopes for hVEGF binding to Fab-D1 (top) or Flt-1D₂ (bottom). hVEGF is depicted in a surface representation colored as in (b). (d) Detailed picture of the Fab-D1 residues in contact with hVEGF. All residues that contact hVEGF are depicted as sticks and labeled. The molecules are colored as for (a) and (b). (e) The amino acid composition of the interface between Fab-D1 and hVEGF. Each circle represents the surface area on Fab-D1 or hVEGF that becomes buried upon complex formation. The colors indicate the proportion of the buried surface area contributed by each amino acid type.

Table 1. Data collection and refinement statistics for the Fab-D1:hVEGF complex

A. Diffraction data	
Space group	$P2_1$
Unit cell constants (Å)	$a=154.4, b=90.2, c=205.6, \beta=90.85^\circ$
Resolution (Å)	50–3.5 (3.63–3.5) ^a
R_{sym}^b	0.97 (0.44) ^a
$\langle I/\sigma I \rangle$	10.7 (2.3) ^a
No. of reflections	233,837
No. of unique reflections	71,940
Completeness (%)	96.1 (93.4) ^a
B. Refinement	
Resolution (Å)	30–3.5
No. of reflections	65,259
$R_{\text{work}}, R_{\text{free}}^c$	0.291, 0.312
No. of residues	4200
No. of non-H atoms	21,356
rmsd bond length (Å)	0.007
rmsd angles (deg.)	1.0
Average isotropic B (Å ²)	65.6
rmsd B (bonded atoms) (Å ²)	2.9
Ramachandran plot (%) ^d	85.0, 13.9, 0.6, 0.5

^a Numbers in parentheses refer to the highest resolution shell.

^b $R_{\text{sym}} = \sum |I - \langle I \rangle| / \sum I$, where $\langle I \rangle$ is the average intensity of symmetry-related observations of a unique reflection.

^c $R_{\text{work}} = \sum |F_o - F_c| / \sum F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. R_{free} is the R -factor for a randomly selected 10% of the reflections excluded from all refinement.

^d Percentage of residues in the most favored, additionally allowed, generously allowed, and disallowed regions of a Ramachandran plot.²⁹

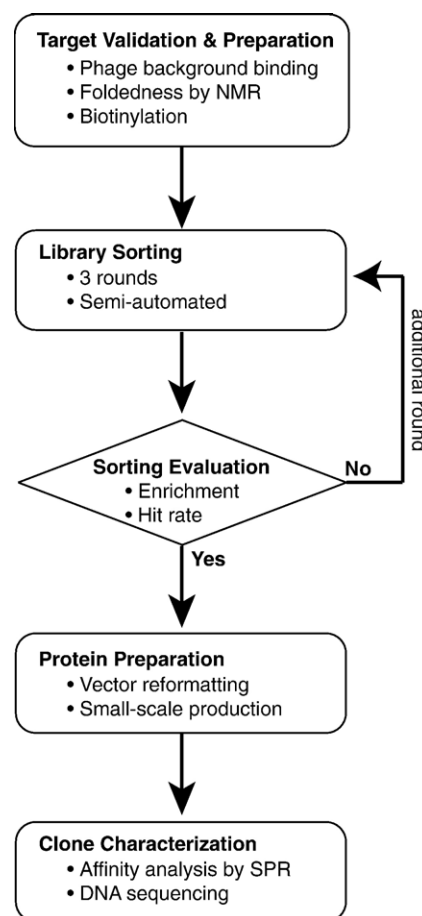
conformation that is only allowed for Gly in the Ramachandran plot,²⁹ explaining the intolerance of this residue to Ala substitution (Figure 3). In contrast with the chemically homogenous paratope, which is dominated by Tyr residues, the antigen side of the interface is constituted by buried surface from a wide diversity of amino acid types. Thus, as was observed previously for two anti-hVEGF Fabs¹⁹ and an anti-hDR5 Fab,²⁰ the complex of Fab-D1 with hVEGF reveals that Tyr residues, when presented in a suitable conformational context, are sufficient to mediate most of the interactions required for the recognition of antigen with high affinity.

Evaluation of library D by HTP Fab generation

As library D was effective in generating a number of high-affinity antibodies against hVEGF, we next explored the general effectiveness of this library using a wide variety of protein antigens. To facilitate this evaluation, we established HTP methods for library selection, protein purification and affinity analysis (Figure 5). The goal of the procedure design was to obtain diverse populations of high-affinity Fabs with minimal rounds of library sorting and with a minimal level of characterization of individual antibodies, particularly during intermediate steps. As described in detail in Materials and Methods, we implemented a magnetic bead-based selection strategy in which antigen concentration was gradually and systematically reduced in later selection rounds. We attached biotin moieties to each

antigen using chemical modification, and thus, the procedures do not require a genetically encoded affinity tag and are generally applicable to most proteins including those that cannot be produced in a recombinant form. Furthermore, the biotinylation reagent contains a cleavable bond, which allows for detachment of an immobilized target by an addition of a reducing reagent.³¹ This selective elution method maximizes the enrichment of Fab-phages that bind to the antigen, rather than to other materials present in the sorting solution. An automated magnetic bead handler was employed to increase throughput and reproducibility.

We applied these procedures to a panel of soluble proteins generously provided by the Midwest Center for Structural Genomics (MCSG). These proteins, ranging in length from 98 to 320 residues, were originally selected for structure determination trials based on low sequence identity (<30%) to proteins with known three-dimensional structures. Therefore, the panel represents a diverse array of proteins that are particularly suitable for evaluating the performance of the antibody library. We tested these proteins for the absence of non-specific binding to phage particles and then confirmed their structural integrity using one-dimensional NMR spectroscopy. Ultimately we chose a set of 13 targets

**Figure 5.** Flow chart summary of the high-throughput Fab generation process. See the text for details.

		K_{on} $\times 10^4 \text{ s}^{-1} \text{ M}^{-1}$	K_{off} $(\times 10^{-3} \text{ s}^{-1})$	K_D (nM)	CDR-L3	CDR-H1	CDR-H2	CDR-H3
					61 92 93 94 94b 95 96	29 30 31 32 33 34	50 51 52 53 54 55 56 57 58	94 95 96 97 98 99 100 100a 100b 100c 100d 100e 100f 100g 100h 100i 100m 101
(a)	Fab	18	<0.1	<1	Y Y Y Y Y Y Y Y	Y Y Y Y Y Y Y	Y Y Y Y Y Y Y Y	Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D10	9.0	<0.1	<1	Y Y Y Y Y Y Y Y	Y Y Y Y Y Y Y	Y Y Y Y Y Y Y Y	Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D11	20	0.47	2.3	S S S Y Y Y S S	L F V Y Y S S S M	S S Y Y S S S S S Y	R E Y T T Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D12	21	0.70	6.6	S S S Y Y Y S S	L F V Y Y S S S M	S S Y Y S S S S S Y	R E Y T T Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(b)	D13	11	7.7	71	S S S Y Y Y S S	R I L Y Y Y Y Y I	S S Y Y S S S S S Y	R L Y I E W N Y S Y V D Y A M Y Y Y Y Y Y
	D14	29	<0.1	<1	X X X X X X X X	V Y Y S S S I	S I Y Y Y Y S S S T S	R S S W Y Y S Y S Y R G W Y G L Y Y Y Y Y Y
	D15	20	<0.1	<1	Y S S Y Y Y Y Y	I S S S S S I	S I Y Y Y Y S S S T S	R Y G L Y I Y S Y Q F Y F Y M Y G L Y Y Y Y Y Y
	D16	20	0.85	4.2	S S S Y Y Y Y Y	P F F Y Y S S S I	S I Y Y Y Y S S S T S	R X X X X X X X X X X X X X X X X X X X X X X
(c)	D17	17	1.5	8.8	Y Y Y Y S S Y Y	L L I Y Y S S S I	S I Y Y Y Y S S S T S	R Y G Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D18	49	0.56	1.1	Y Y S S Y Y Y L	V F Y Y S S I	S I Y Y S S S S S T S	R O W W Y Y H Y I Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D19	28	1.4	4.9	S Y S Y Y Y Y Y	F Y Y S S S I	S I Y Y S S S S S T S	R P R Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D20	9.3	0.49	5.3	S S S Y Y Y Y Y	L F L Y Y S S S I	S I Y Y S S S S S T S	R Y H Q Y V R Y Y F W Y Y Y Y Y Y Y Y Y Y Y
(d)	D21	17	1.1	6.3	S Y Y Y S S Y Y	P I L Y Y S S S I	S I Y Y S S S S S T S	R W W L Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D22	6.1	0.43	7.0	Y Y Y Y S S Y Y	L L L Y Y S S S I	S I Y Y S S S S S T S	R F Y R H P Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D23	13	1.2	8.7	Y Y Y Y S S Y Y	P V I Y Y S S S I	S I Y Y S S S S S T S	R Y G G Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D24	26	2.7	10	S Y S Y Y Y Y Y	P F I Y Y S S S I	S I Y Y S S S S S T S	R Y Y E W N N Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(e)	D25	34	5.0	15	S Y S Y Y Y Y Y	P F V Y Y S S S I	S I Y Y S S S S S T S	R Y Y T W P Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D26	35	0.67	1.9	S Y Y Y Y Y Y Y	P I V Y Y S S S I	S I Y Y S S S S S T S	R E F Y Y L W Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D27	43	0.84	1.9	Y Y S S Y Y Y Y	L F F Y Y S S S I	S I Y Y S S S S S T S	R E Y Y W Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D28	39	1.9	4.8	S Y Y Y S S Y Y	L L F Y Y S S S I	S I Y Y S S S S S T S	R E R Y D W Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(f)	D29	37	2.2	5.9	Y Y S S Y Y Y Y	L F F Y Y S S S I	S I Y Y S S S S S T S	R E N H W Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D30	25	1.5	6.0	Y Y S S Y Y Y Y	L L F Y Y S S S I	S I Y Y S S S S S T S	R E R F I Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D31	61	4.8	7.9	Y Y S S Y Y Y Y	L F F Y Y S S S I	S I Y Y S S S S S T S	R E Y M N Y M L Q E A F Y Y Y Y Y Y Y Y Y Y
	D32	20	2.2	11	Y S S Y S S Y Y	L I V Y Y S S S I	S I Y Y S S S S S T S	R E H I Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(g)	D33	13	5.5	44	Y S S Y S S Y Y	L I V Y Y S S S I	S I Y Y S S S S S T S	R K R Y K M M Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D34	86	2.1	2.5	S S Y Y Y S S L	F Y Y S S S I	S I Y Y S S S S S T S	R S S S Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D35	61	1.8	3.0	S Y Y Y S S Y Y	P I L Y Y S S S I	S I Y Y S S S S S T S	R I O S W Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D36	103	3.3	3.2	S S S Y S S Y Y	L V Y Y Y S S I	S I Y Y S S S S S T S	R F F L W K I D Y W Y Y Y Y Y Y Y Y Y Y Y Y
(h)	D37	56	2.4	4.2	S S S Y S S Y Y	P F L Y Y S S S I	S I Y Y S S S S S T S	R R S S Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D38	63	3.0	4.8	S Y Y Y S S Y Y	P I L Y Y S S S I	S I Y Y S S S S S T S	R Y G Y S W Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D39	33	3.0	9.0	S S S Y Y Y S Y	P V X X X X X X	S I Y Y S S S S S T S	R S S F W Y W P H Y D Y D P A F Y Y Y Y Y Y
	D40	79	7.6	9.6	S S Y Y Y Y S Y	P V F Y Y S S S I	S I Y Y S S S S S T S	R Y G F G Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(i)	D41	17	7.0	41	Y S S Y S S Y Y	L F V Y Y S S S I	S I Y Y S S S S S T S	R G G S F Y A Y E Y W V D G I Y Y Y Y Y Y
	D42	60	<0.1	<1	Y Y Y Y S S Y Y	L F F S S S S I	S I Y Y S S S S S T S	R E D Y V E Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D43	99	0.49	0.5	H Y T T Y Y Y Y	P P F S S S S I	S I Y Y S S S S S T S	R D Y Y F Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D44	78	0.59	0.7	H Y T T Y Y Y Y	P P L Y Y S S I	S I Y Y S S S S S T S	R Y S S Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(j)	D45	23	0.18	0.8	X X X X X X X X	L Y Y S S S I	S I Y Y S S S S S T S	R G G R I V D W M L Y Y Y Y Y Y Y Y Y Y Y
	D46	14	4.0	28	H Y T T Y Y Y Y	P P F S S S S I	S I Y Y S S S S S T S	R S F Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D47	15	4.5	31	H Y T T Y Y Y Y	P P F S S S S I	S I Y Y S S S S S T S	R S F Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D48	20	0.61	3.0	S S Y Y Y S S L	L L L Y Y S S I	S I Y Y S S S S S T S	R Y T F O Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(k)	D49	19	0.79	4.1	H Y T T Y Y Y Y	P P L Y Y S S I	S I Y Y S S S S S T S	R M K Y Y M P V W Y Y Y Y Y Y Y Y Y Y Y
	D50	19	1.2	6.3	Y S Y Y Y Y Y Y	L F F G Y Y S I	S I Y Y S S S S S T S	R R Y A G Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D51	10	0.83	8.7	Y S S S Y Y Y Y	L F V Y Y S S S I	S I Y Y S S S S S T S	R Q G V S A Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D52	38	3.3	8.8	Y Y S S Y Y Y Y	P I V Y Y S S S I	S I Y Y S S S S S T S	R E Q R Y N Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(l)	D53	25	2.3	9.1	Y Y S S S S Y Y	L I V Y Y S S S I	S I Y Y S S S S S T S	R V R Y A Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D54	11	1.3	12	Y Y S S S S Y Y	L I V Y Y S S S I	S I Y Y S S S S S T S	R M V H G Y A H Q Y Y Y Y Y Y Y Y Y Y Y
	D55	14	<0.1	<1	S S S Y Y Y Y Y	P I F Y Y S S S I	S I Y Y S S S S S T S	R Y G Y S R D F Y D Y Y F G A F Y Y Y Y Y
	D56	6.6	<0.1	<1	Y Y Y Y S S Y Y	P F L S S S S I	S I Y Y S S S S S T S	R S M K Y W Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(m)	D57	9.9	<0.1	<1	S S S Y Y Y S Y	L V L S S S S I	S I Y Y S S S S S T S	R F Y Y F H G Y Y Y Y Y Y Y Y Y Y Y Y Y
	D58	12	0.13	1.1	S S S Y Y Y S Y	L V L S S S S I	S I Y Y S S S S S T S	R T F S P G F W W Y Y Y Y Y Y Y Y Y Y Y
	D59	11	0.90	8.4	S S S Y Y Y S Y	L F L S S S S I	S I Y Y S S S S S T S	R F Y Y F H G Y Y Y Y Y Y Y Y Y Y Y Y Y
	D60	5.8	0.81	14	S S S Y Y Y S Y	L F L S S S S I	S I Y Y S S S S S T S	R F A Y G G W S Y Y Y Y Y Y Y Y Y Y Y Y
(n)	D61	6.8	1.9	29	S S S Y Y Y S Y	L F L S S S S I	S I Y Y S S S S S T S	R T Q Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D62	8.6	<0.1	<1	S S S Y Y Y Y Y	P I F Y Y S S S I	S I Y Y S S S S S T S	R Y S N W Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D63	12	0.28	2.4	S Y S Y Y Y Y Y	L X L S S S S I	S I Y Y S S S S S T S	R Y S N W Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D64	17	0.52	3.0	S S Y S Y Y S Y	P F I Y Y S S S I	S I Y Y S S S S S T S	R T N H P Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(o)	D65	13	0.50	3.9	S Y Y Y S S Y Y	P V F S S S S I	S I Y Y S S S S S T S	R M S F W N Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D66	28	1.2	4.3	S S Y S Y Y S Y	P I F S S S S I	S I Y Y S S S S S T S	R Y G G Y A Y A F Y Y Y Y Y Y Y Y Y Y Y
	D67	10	0.60	5.8	S S S Y Y Y Y Y	P L I G Y S S S I	S I Y Y S S S S S T S	R T Y R Y W Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D68	22	1.8	8.0	S S S Y Y Y Y Y	P F I G Y S S S I	S I Y Y S S S S S T S	R I P P Y L Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(p)	D69	10	0.81	8.1	S S S Y Y Y Y Y	P L I G Y S S S I	S I Y Y S S S S S T S	R K K Y W Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D70	39	0.46	1.2	S Y Y Y Y Y Y Y	P I F S S S S I	S I Y Y S S S S S T S	R F V Y Y A Y A I Y Y Y Y Y Y Y Y Y Y Y
	D71	22	0.46	2.0	S S Y Y Y Y Y Y	L V F S S S S I	S I Y Y S S S S S T S	R K G Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D72	30	0.68	2.3	Y Y Y Y S S Y Y	P F F Y Y S S S I	S I Y Y S S S S S T S	R G Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(q)	D73	70	2.0	2.9	Y Y Y Y S S Y Y	P I F S S S S I	S I Y Y S S S S S T S	R Y N Y M W Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D74	30	0.90	3.0	S Y Y Y S S Y Y	P I F S S S S I	S I Y Y S S S S S T S	R F Y K Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D75	28	0.89	3.2	S Y Y Y S S Y Y	P F L S S S S I	S I Y Y S S S S S T S	R W F P I S G W R F W Y Y Y Y Y Y Y Y
	D76	14	0.45	3.3	S S S Y Y Y S Y	L F V Y Y S S S I	S I Y Y S S S S S T S	R G G Y K E Y H Y Y Y Y Y Y Y Y Y Y Y
(r)	D77	22	1.4	6.3	S Y Y Y Y Y Y Y	P F I S S S S I	S I Y Y S S S S S T S	R Y W Q M Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D78	17	2.3	13	Y Y Y Y Y Y Y Y	P V F I S S S S I	S I Y Y S S S S S T S	R L S T Q K E T Y Y Y Y Y Y Y Y Y Y Y
	D79	15	2.2	15	S Y Y Y Y Y S Y	L I Y S S S S I	S I Y Y S S S S S T S	R L Y Y F G W A Y Y Y Y Y Y Y Y Y Y Y
	D80	20	0.85	4.3	Y Y Y Y Y Y Y Y	P V I S S S S I	S I Y Y S S S S S T S	R L A S Y Y W A Y Y Y Y Y Y Y Y Y Y Y
(s)	D81	16	1.4	9.0	S S Y Y Y S S S	P F I S S S S I	S I Y Y S S S S S T S	R V Y Y Y Y D E D S Y S I K G M Y Y Y Y
	D82	22	2.2	10	Y S Y Y Y S S S	P I F S S S S I	S I Y Y S S S S S T S	R G G W Y F T G S G S A F Y Y Y Y Y Y
	D83	5.3	0.8	15	S S S S S S S S	P F F S S S S I	S I Y Y S S S S S T S	R V M V H S Y M Y Y Y Y Y Y Y Y Y Y Y
	D84	14	0.13	0.9	Y Y Y Y Y S S L	P L L Y Y S S I	S I Y Y S S S S S T S	R Y Y L Y Y S S Y Y Y Y Y Y Y Y Y Y Y
(t)	D85	26	0.27	1.0	Y S Y Y Y S S L	P L L Y Y S S I	S I Y Y S S S S S T S	R R L S L Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D86	32	0.36	1.1	X X X X X X X X	I S S S S I	S I Y Y S S S S S T S	R E W Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D87	127	4.7	3.7	S Y Y Y Y Y P L	L Y Y S S S I	S I Y Y S S S S S T S	R A K Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D88	60	3.1	5.2	S Y Y Y Y Y P V	L Y Y S S S I	S I Y Y S S S S S T S	R Q W K T R A I Y Y Y Y Y Y Y Y Y Y Y
(u)	D89	20	1.5	7.6	H Y T T Y Y Y Y	P P F S S S S I	S I Y Y S S S S S T S	R K Y Y F Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D90	8.1	0.82	10	S Y S S S S Y Y	P F F S S S S I	S I Y Y S S S S S T S	R F T Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D91	5.9	7.5	126	S S Y S S S Y Y	L I V S S S S I	S I Y Y S S S S S T S	R Y Y Y P W Y Y Y Y Y Y Y Y Y Y Y Y
	D92	13	<0.1	<1	S Y Y Y Y Y Y Y	L V L S S S S I	S I Y Y S S S S S T S	R Y Y Y F Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(v)	D93	10	0.16	1.3	S Y Y Y Y Y Y Y	L V L S S S S I	S I Y Y S S S S S T S	R V G R N Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D94	13	0.18	1.4	Y Y Y Y S S Y Y	L F F Y Y S S S I	S I Y Y S S S S S T S	R T D R Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D95	10	0.16	1.6	Y Y Y Y S S Y Y	L F F Y Y S S S I	S I Y Y S S S S S T S	R F H G K E Y Y Y Y Y Y Y Y Y Y Y Y Y
	D96	3.6	0.06	1.7	S Y S Y Y Y Y Y	P V V Y Y S S S I	S I Y Y S S S S S T S	R V M S Y M H Y Y Y Y Y Y Y Y Y Y Y
(w)	D97	8.6	0.52	6.0	S S S Y Y Y Y Y	P L I G Y S S S I	S I Y Y S S S S S T S	R V P P R N Y Y M A W Y Y Y Y Y Y Y
	D98	1.8	0.21	12	S S S Y Y Y Y Y	L I I G Y S S S I	S I Y Y S S S S S T S	R N D R Y L S S F A M Y Y Y Y Y Y Y
	D99	3.9	0.51	13	S S S Y Y Y S Y	P I I G Y S S S I	S I Y Y S S S S S T S	R R K Y Y L L G Y Y Y Y Y Y Y Y Y Y Y
	D100	4.6	0.86	19	S S S Y Y Y S Y	L V I G Y S S S I	S I Y Y S S S S S T S	R H O O Y R Y Y Y Y Y Y Y Y Y Y Y Y
(x)	D101	2.1	<0.1	<1	S Y S S S S S L	F Y Y S S S I	S I Y Y S S S S S T S	R G G F Y Y Y P V S S G Y Y Y Y Y Y Y
	D102	9.9	<0.1	<1	S Y Y Y Y Y Y Y	L S S S S S I	S I Y Y S S S S S T S	R V V M Y Y Y P M G H S Y E T G S A I Y
	D103	17	0.16	1.0	S Y Y Y Y Y Y Y	L L L S S S S I	S I Y Y S S S S S T S	R G W Y Y Y Y R W E G G A F Y Y Y Y Y
	D104	24	1.0	4.4	S Y S S S S S Y	L F F Y Y S S S I	S I Y Y S S S S S T S	R Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
(y)	D105	22	0.73	5.8	S S Y Y Y Y Y Y	P V V Y Y S S S I	S I Y Y S S S S S T S	R Y V Y R A F G S L T Y A L Y Y Y Y Y
	D106	22	1.6	7.3	S S S S S S S Y	L V F Y Y S S S I	S I Y Y S S S S S T S	R V V Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D107	9.0	1.2	13	S S S S S S S Y	P F F Y Y S S S I	S I Y Y S S S S S T S	R G Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
	D108	5.3	1.0	19	S Y S S S S S Y	P F F Y Y S S S I	S I Y Y S S S S S T S	R V W Y Y Y Y Y Y Y Y Y Y Y Y Y Y

Figure 6 (legend on next page)

from the MCSG proteins and also the extracellular domain of prolactin receptor.³²

With our standardized, semi-automated procedures, a single researcher was able to routinely complete library selections against 12 antigens in one week. After three or four rounds of selection against each antigen, the enriched phage pool showed a significant level of antigen-specific binding, as measured by the “enrichment ratio”, which is the ratio of phage recovered from sorting against antigen over that from sorting without antigen. The frequency of antigen-binding clones, as tested by phage ELISA of individual clones, was typically greater than 80%.

The enriched pool of phage-displayed antibodies was then reformatted to soluble proteins as an ensemble, and individual clones were randomly chosen for small-scale expression and semi-automated purification. The binding parameters for each Fab protein were evaluated by surface plasmon resonance. The affinity and amino acid sequences for individual Fab proteins revealed that the library robustly produced high-affinity Fabs for all the targets (Figure 6). Remarkably, the majority (83%) of clones had a K_d value below 20 nM, and a substantial proportion (10%) had a K_d lower than 1 nM (Figure 7). Also, the CDR-H3 lengths and sequences for these clones were highly diverse, indicating that our selection procedures rapidly enrich high-affinity clones without concentrating a few highest-affinity clones that could otherwise dominate the enriched pools.

Sequence characteristics of selected Fabs

The large number of unique Fabs selected from library D for a diverse panel of antigens, together with the accompanying affinity parameters, constitute a unique database for analyzing contributions of CDR residues to antigen recognition. The use of a single framework eliminates complications associated with the occurrence of multiple frameworks and framework mutations in the natural repertoire.^{33,34} As references, we determined the sequence compositions of the naïve library and those of the library after selection for binding to protein A, which we refer to as the “structure-selected” population, since protein A binds to a discontinuous epitope on the Fab framework and selects for correctly folded molecules.³⁵

In CDR-H1, -H2 and -L3, few positions among those diversified with restricted amino acid combinations showed significant enrichment or depletion (Figure 8), indicating that our structure-based design strategy was effective for choosing diversity that was both structurally well-tolerated and useful

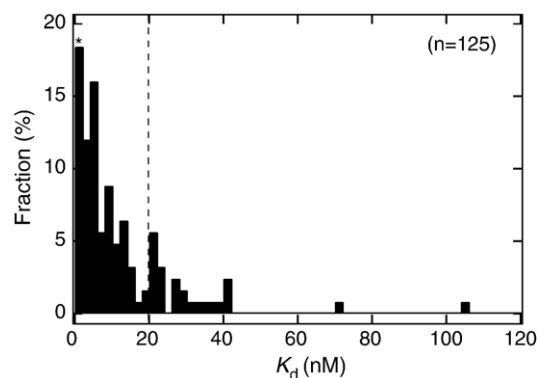


Figure 7. A histogram of the K_d values for 125 Fab clones isolated from library D. The K_d values were binned into 2 nM intervals. The lowest bin, indicated by an asterisks (*), includes the K_d values that were estimated to be lower than 1 nM but not determined accurately. The data were derived from Fabs D1-D109 (Figures 2(d) and 6), plus 16 unique Fabs that bound specifically to antigens shown in Figure 6 but were not included in the alignment due to poor sequence quality.

for antigen recognition. However, biases were evident at some positions. In CDR-H2, positions 50 and 53 were enriched with Ser or Tyr, respectively, in the structure-selected clones and even more in the antigen-specific clones, suggesting differing structural and functional requirements at these positions. Interestingly, at position 32 in CDR-H1, Tyr is enriched or depleted in the structure-selected or antigen-specific clones, respectively, suggesting a mismatch between structural and functional requirements. In CDR-L3, position-specific Tyr enrichment depends on the loop length, suggesting the importance of structural context in Tyr presentation. We also analyzed the distributions of the three different loop lengths that were allowed in CDR-L3 but found no significant changes following selection for binding to protein A or antigen (Figure 9(a)).

For CDR-H3, we allowed 12 different lengths, and thus, the numbers of analyzed clones were not sufficient to detect statistically significant deviations amongst frequencies of the different lengths. However, a qualitative assessment reveals that all of the designed lengths were present in both the naïve library and the structure-selected population, indicating that all of the loop lengths were tolerated structurally (Figure 9(b)). The antigen-binding clones used all of the designed CDR-H3 lengths, except the longest loop length. We also isolated antigen-binding CDR-H3 sequences (eight, nine or ten residues long) that were shorter than the designed range, and in particular, ten-residue sequences were highly abundant. These short

Figure 6. Sequences and binding parameters of Fabs selected against 14 protein antigens. The Fab sequences are shown in the same manner as in Figure 2. Targets used were: (a) APC22356; (b) APC23269; (c) APC23371; (d) APC35945; (e) APC24651; (f) APC24734; (g) APC24755; (h) APC24947; (i) APC35734; (j) APC35946; (k) APC36018; (l) APC23257; (m) APC24892; and (n), prolactin receptor extracellular domain. Binding parameters (k_{on} , k_{off} , K_d) were determined from kinetic analysis of Fabs binding to immobilized antigen measured by surface plasmon resonance. The origins and amino acid sequences of the MCSG proteins are available [at <http://www.mcsg.anl.gov>].

sequences likely arose from mutagenic oligonucleotides produced by incomplete coupling during the DNA synthesis reaction, and this is consistent with

the properties of syntheses involving trinucleotide building blocks, in which each incomplete coupling gives rise to a codon deletion.¹⁸ These short sequences were extremely rare in the naïve library, as would be expected for side products of DNA synthesis, and thus, the enrichment of short, undesigned CDR-H3 sequences amongst the antigen-binding clones suggests that library performance might be further enhanced by inclusion of shorter loops in the CDR-H3 design. The need for highly diverse CDR-H3 loop lengths is underscored by an analysis of the number of different antigens recognized by each length class (Figure 9(c)). While Fabs with a 17 residue CDR-H3 loop were able to bind to the largest number of targets, the length-dependence was not pronounced, suggesting that no particular length is dominant or superior to the others.

We also analyzed the amino acid composition of CDR-H3 in an aggregate manner. The mutagenic oligonucleotides used for CDR-H3 diversification were synthesized using trinucleotides encoding for 19 amino acid residues in the following molar ratios: Tyr (20%), Ser and Gly (15% each), and 16 other amino acid residues (3.1% each). Sequencing of the naïve library revealed that the actual distribution was close to the theoretical distribution, although there were some deviations that likely arose from differing reaction rates for the different trinucleotides in the DNA synthesis reaction. In particular, Tyr was over-represented at the expense of Ser and Gly (Figure 10). In the structure-selected population, only three amino acid types (Glu, Ile and Val) are significantly depleted, suggesting the robustness of our scaffold in accommodating diverse CDR-H3 sequences. Ile and Val may restrict CDR-H3 in an unstable conformation and/or negatively affect the folding, because they are both β -branched, hydrophobic and known to promote aggregation.^{36,37} Ile was also depleted amongst the antigen-binding sequences, and we speculate that the steric demands of the bulky, β -branched side-chain of this residue may reduce conformational flexibility in a manner that compromises both protein folding and antigen-binding. In contrast, Trp, Arg and Gly were significantly enriched amongst the antigen-binding sequences, and these trends are consistent with

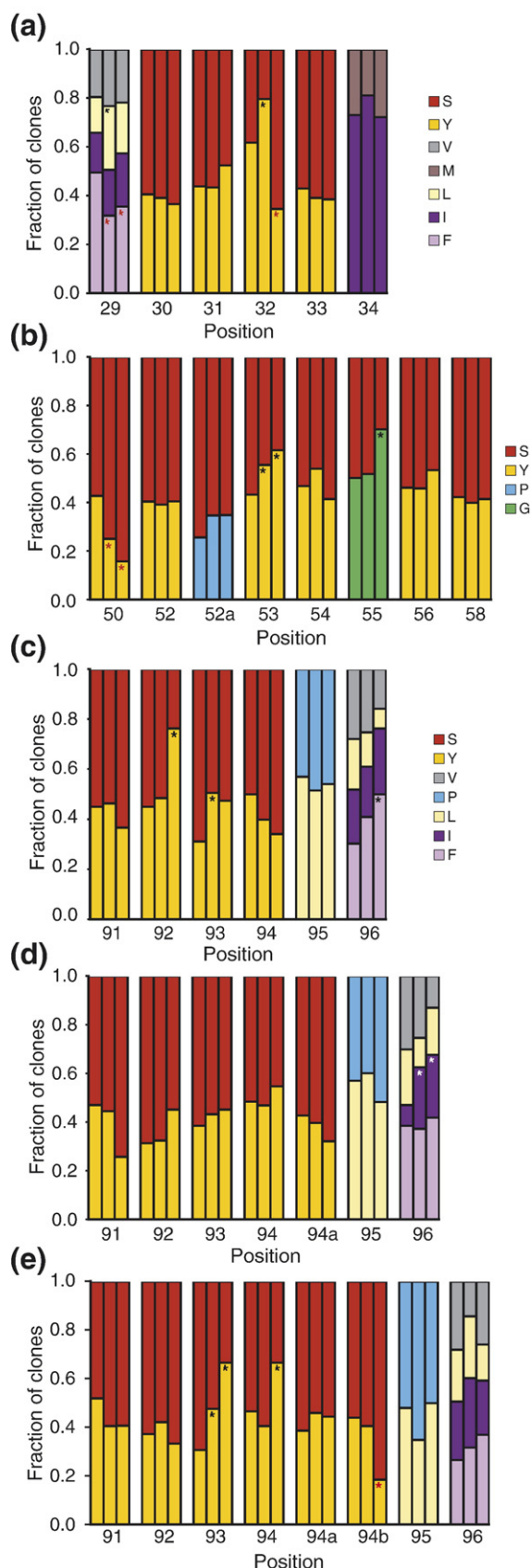


Figure 8. Amino acid compositions of antibodies from library D in (a) CDR-H1, (b) CDR-H2 and CDR-L3 loops of (c) short, (d) medium or (e) long length. The bar graphs show the composition at each position for the following populations: naïve (left bar), structure-selected (center bar), and antigen-binding (right bar). An amino acid whose composition exhibits statistically significant deviation (an unadjusted p -value < 0.05) from the corresponding value for the naïve library is marked with an asterisk (*); black or white asterisks indicate an increase in amino acid usage while a red asterisk indicates a decrease in usage. The data were derived from the following number of unique clones in each population: naïve, 229; structure-selected, 247; antigen-binding, 109 (Fabs D1-D109 in Figures 2(d) and 6).

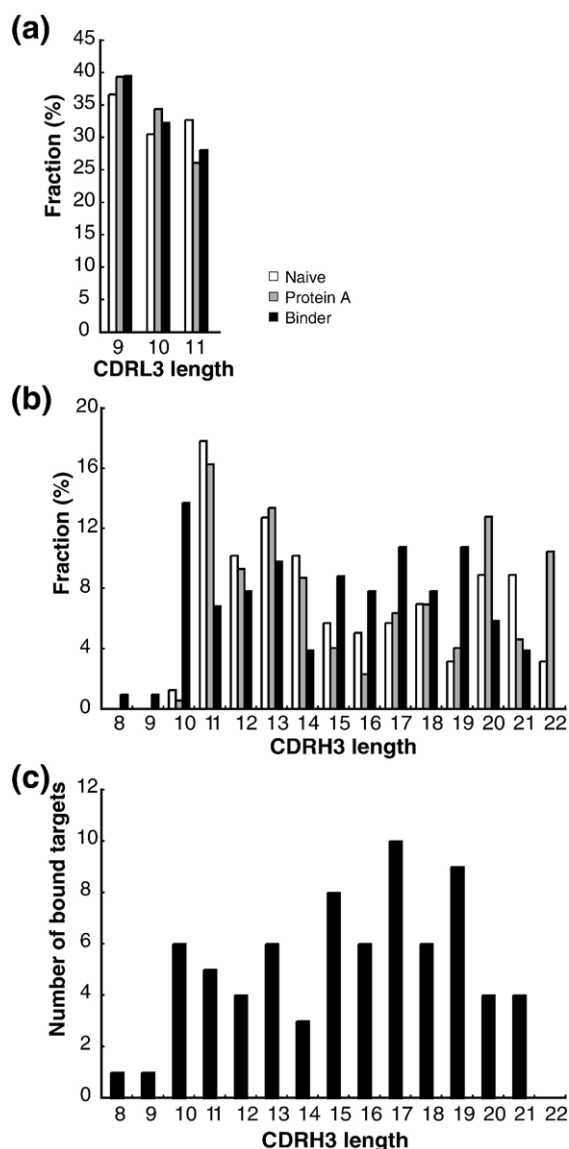


Figure 9. The length distributions of CDR loops that were varied in library D. (a) The length distribution of CDR-L3 in the naïve library (white bars), the structure-selected population (gray bars) and the antigen-binding population (black bars). (b) The length distribution of CDR-H3 shown in the same manner as in (a). (c) The number of different targets bound by Fabs plotted as a function of the length of CDR-H3. The loop lengths were calculated on the basis of the CDR boundaries defined by Kabat *et al.*¹² (positions 89–97 or 95–102 for CDR-L3 or CDR-H3, respectively). The data were derived from unique clones as described for Figure 8.

previously described properties of protein–protein interfaces, in which the bulky side-chains of Trp, Arg and Tyr residues are often involved in intermolecular contacts and function as hotspot residues.³⁸ Gly residues are less frequently involved in direct contacts, but instead, allow for flexibility that can be crucial for achieving a conformation favorable for binding, as observed in the CDR-H3 loop of Fab-D1 in complex with hVEGF (Figure 4). It is also

noteworthy that Tyr and Ser were maintained in high abundance amongst the antigen-binding sequences, and these results are consistent with a dominant role for these residues in mediating antigen recognition.

The small population changes of the other amino acids suggest that contributions of these amino acids are target-specific and their population changes are averaged in our aggregate analysis. It is interesting that, with the exception of Arg, none of the charged amino acid types were significantly enriched. These amino acids may be highly beneficial in making specific electrostatic interactions, but at the same time, they can cause detrimental effects when their charge is not compensated. Together, this statistical analysis clearly supports the conclusion that heavily biased amino acid diversity enriched with Tyr, Ser and Gly is a highly effective solution to designing an antibody library for protein antigens. Because this analysis minimizes influence specific to the Fab framework, we expect that this amino acid composition will be generally effective for binding interfaces with similar architectures composed of multiple loops.

Discussion

By constructing extremely large, but well-defined synthetic antibody libraries, we were able to explore the effects of conformational and chemical diversity on the capacity for antigen recognition. In particular, the incremental addition of diversity to a minimalist background allowed for the detection of statistically significant effects on antibody structure and function. In comparing the anti-hVEGF antibodies from libraries A and B, it is clear that the addition of CDR-L3 diversity does not simply produce further optimization of the same CDR-H3 loop, but rather, completely alters the nature of the optimal CDR-H3 solution (Figure 2). Thus, CDR-L3 serves to not only enhance affinity, but also, to determine the types of heavy chains that are capable of mediating antigen recognition. Because CDR-L3 is located adjacent to CDR-H3 (Figure 1), the additional sequence and length diversity of CDR-L3 alters the context in which CDR-H3 sequences are presented.

In going from library B to C, the addition of diversity at non-paratope positions allows for the use of CDR-H3 loops that differ from those obtained from either library A or B, suggesting that the additional conformational diversity within CDR-H3 afforded by these positions may enable the placement of paratope Tyr residues in just the right place for optimal binding. Thus, it appears that, rather than fine-tuning a common binding surface, the incremental addition of diversity in CDR-L3 and non-paratope residues fundamentally changes the nature of the antigen-binding site by allowing for the use of different CDR-H3 solutions in each case. The practical consequence of these effects is that library C was able to generate an high-affinity anti-hVEGF antibody (Fab-C1) with a sequence that is very

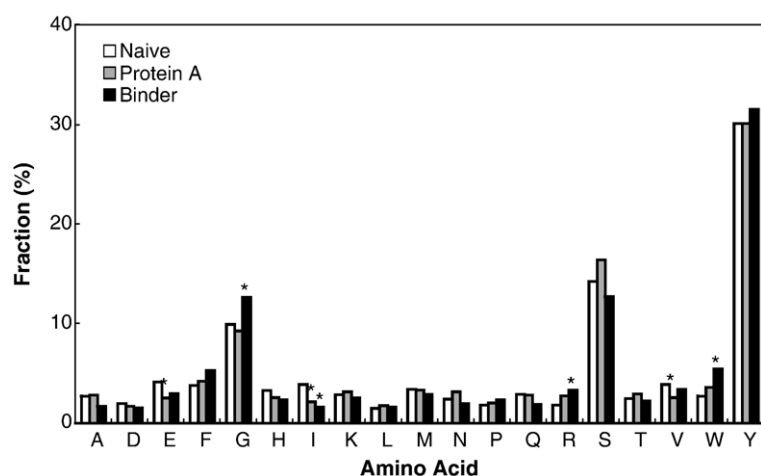


Figure 10. (a) The amino acid distribution in CDR-H3 of antibodies from library D. The following populations were analyzed: naïve (white bars), structure-selected (grey bars) and antigen-binding (black bars). Only fully diversified positions were included in the analysis. Statistically significant deviations (an unadjusted p -value < 0.05) from the naïve populations are indicated with an asterisk. The data were derived from unique clones as described for Figure 8.

different from the sequences of the numerous lower-affinity antibodies obtained from libraries A and B.

In going from library C to D, the addition of chemical diversity dramatically increased the number of clearly distinct CDR-H3 loops capable of recognizing hVEGF with high affinity. It is striking that each of the nine high-affinity anti-hVEGF antibodies (Figure 2(d)) contain CDR-H3 loops of different lengths, and there is no obvious sequence homology amongst them. Structural and functional analysis of Fab-D1 revealed a crucial role for a Gly residue in providing a CDR-H3 loop conformation suitable for high affinity binding. Each of the nine anti-hVEGF CDR-H3 loops contain Gly residues, and we speculate that many of these may be required for the particular conformations of the different loops, and thus, it appears that the addition of Gly as a third major component of CDR-H3 was crucial for affording diverse, high-affinity solutions. The highly functional nature of library D was confirmed by the generation of high-affinity antibodies against a large panel of protein antigens, and again, there was a striking level of diversity amongst the CDR-H3 loops in terms of both length and sequence (Figure 6). Gly is present in the CDR-H3 loops of the vast majority of these antibodies, often at multiple positions, suggesting that the critical role of Gly is universal. This view is supported by the high prevalence of Gly in the CDR-H3 germline genes, where it is second only to Tyr in terms of overall abundance.¹⁷

Most highly functional phage-displayed antibody libraries have been constructed by transferring natural repertoires into recombinant systems.^{39–42} Even in the design of synthetic repertoires, considerable effort has been expended on trying to replicate complex aspects of natural repertoires by combining natural CDR sequences with synthetic diversity⁴³ or using numerous combinations of light and heavy chain frameworks.¹⁸ However, our extensive analysis demonstrates that, despite simple framework and CDR design, the performance of library D compares favorably to the most successful and complex libraries reported to date.^{42–44} These findings should simplify future antibody engineering efforts, as they make clear that a single frame-

work with focused CDR diversity can generate high-affinity antibodies against diverse antigens, without the need for affinity maturation.

The affinities and kinetic parameters of Fabs derived from library D are suitable for most applications of antibodies as affinity reagents. Clearly, a HTP is critical for utilizing affinity reagents in proteomic and systems biology applications. As we demonstrate here, the use of a single framework greatly streamlines procedures for analyzing and reformatting Fab clones generated from phage library sorting, resulting in a higher overall throughput of the antibody generation pipeline. Furthermore, a modular design makes it possible to independently tailor the performance of the diversity design for antigen binding and the physicochemical and biological properties of the framework. Importantly, statistical analyses can detect sequence biases (Figures 8–10) that can be used to further improve library designs by incorporating features found in stable, functional antibodies.

The HTP procedures developed here (Figure 5) rapidly enrich high-affinity antibodies. The affinity of the generated antibodies was tightly clustered (Figure 7). Automation of phage sorting using biotinylated targets and solution-based selection using a magnetic bead handler have made the process highly reproducible and improved its throughput. Although chemical biotinylation could potentially result in the enrichment of antibodies specific to the modified form of a target, affinity analysis using unmodified targets (Figure 6) indicates that this is not generally the case. Our procedures should be applicable to different classes of antigens, and also, should be readily transferable to a typical research laboratory. Indeed, we have adopted them for the isolation of antibodies specific for structured nucleic acids (J. Ye, J. Piccirilli, F.A.F., S.S.S., S.K. & A.A.K., unpublished results).

While the focus of this study is on antibody engineering, our results have general implications for engineering protein–protein interaction interfaces. Development of molecular systems that produce binding proteins with high affinity and specificity is the focus of intense research activities.

Although a large number of molecular scaffolds for presenting interaction residues have been developed,⁴⁵ few studies have examined what amino acid sequence diversity should be introduced at which positions within a scaffold. Our recent demonstration of the effectiveness of the Tyr/Ser binary library on the single-domain fibronectin type III scaffold implies that, with a sufficient level of conformational diversity, a combinatorial library biased towards Tyr and Ser functions well with non-antibody scaffolds.²² Thus, we believe that our minimalist library design strategies and HTP approaches provide rational guidelines that are generally applicable to combinatorial engineering of protein binding interfaces.

Materials and Methods

Library construction

The construction of libraries A and B has been described.²⁰ For libraries C and D, a phagemid, designed to display humanized Fab-4D5 on the surface of M13 bacteriophage in a bivalent format by use of a disulfide linkage,^{21,46} was modified with four light chain mutations (D28S/N30S/T31S/F53S) and one heavy chain mutation (S93A). In addition, TAA stop codons were introduced into all three heavy-chain CDRs. The resulting phagemid (pF1359) was used as the template to construct libraries, as described.^{20,21,23,44,47} Oligonucleotide-directed mutagenesis was used to repair the TAA stop codons and replace CDR positions with degenerate codons encoding the amino acid compositions shown in Figure 1(b). With the exception of CDR-H3 in library D, CDR positions were mutagenized using oligonucleotides synthesized with standard methods. The highly diverse positions in the oligonucleotides for mutagenesis of CDR-H3 of library D were synthesized using a custom Trimer Phosphoramidite Mix (Glen Research, Sterling, VA) containing codons for 19 amino acid residues in the following molar ratios: Tyr (20%), Ser (15%), Gly (15%), 16 other natural amino acid residues excluding Cys (3.1% each).

Selection and characterization of anti-hVEGF Fabs

Phage from the libraries were cycled through rounds of binding selection with hVEGF coated on 96-well Maxisorp Immunoplates (NUNC, Rochester, NY) as the capture target, as described.^{23,47} After five rounds of selection, phage were produced from individual clones grown in a 96-well format and the culture supernatants were used in phage ELISAs to detect specific binding clones.⁴⁷ Clones that bound to hVEGF but not to bovine serum albumin (BSA; Sigma-Aldridge, St Louis, MO) were subjected to DNA sequence analysis. Approximately 100 clones were sequenced from each library and unique clones were aligned as shown in Figure 2.

A single-point competitive phage ELISA⁴⁷ was used to rapidly estimate the affinities of phage-displayed anti-hVEGF Fabs. Colonies of *Escherichia coli* XL1-blue (Stratagene, La Jolla, CA) harboring phagemids were inoculated directly into 150 μ l of 2YT broth supplemented with carbenicillin and M13-KO7 helper phage (New England Biolabs, Beverly, MA); the cultures were grown overnight

at 37 °C in a 96-well format. Culture supernatants containing Fab-phage were diluted fivefold in phosphate-buffered saline, 0.5% (w/v) BSA, 0.1% (v/v) Tween 20 (PBT buffer) either with or without hVEGF (100 nM or 1 μ M). After 1 h incubation at room temperature, the mixtures were transferred to plates coated with hVEGF and incubated for 15 min. The plates were washed with phosphate-buffered saline, 0.05% (v/v) Tween 20 and incubated 30 min with horse radish peroxidase/anti-M13 antibody conjugate (1:5000 dilution in PBT buffer). The plates were washed, developed with 3,3',5,5'-Tetramethyl-benzidine/ H_2O_2 peroxidase substrate (Kirkegaard & Perry Laboratories), quenched with 1.0 M H_3PO_4 , and read spectrophotometrically at 450 nm. For each clone, the percentage of Fab-phage uncomplexed with solution phase hVEGF was calculated by dividing the A_{450} in the presence of hVEGF by the A_{450} in the absence of hVEGF.

For kinetic analysis, high-affinity Fab proteins were purified from *E. coli* as described.⁴⁸ Binding kinetics were determined by surface plasmon resonance using a BIAcore™-3000 with hVEGF immobilized on CM5 chips at ~500 response units, as described.²¹ Serial dilutions of Fab proteins were injected, and binding responses were corrected by subtraction of responses on a blank flow cell. For kinetic analysis, a 1:1 Langmuir model of global fittings of k_{on} and k_{off} was used. The K_d values were determined from the ratios of k_{on} and k_{off} .

Shotgun scanning of anti-hVEGF Fabs

The shotgun scanning analyses²⁸ were carried out as described.²⁷ Each Fab was displayed on the surface of M13 bacteriophage by direct fusion of the heavy chain to the C-terminal domain of the M13 gene-3 minor coat protein. The light chain was expressed independently with an epitope tag (amino acid sequence: MADPNRFRGKDLGG) fused to the C terminus. A library was constructed in which the CDR-H3 loop was randomized with degenerate codons encoding for only the wild-type and a single mutation, as describe for Figure 3. The library was subjected to two independent selections for binding to either hVEGF (antigen selection) or an anti-epitope tag antibody (display selection) immobilized on Maxisorp immunoplates. Following two or three rounds of display or antigen selection, respectively, approximately 100 binding clones were sequenced and analyzed with the program SGCOUNT,²⁸ as described.²⁷ SGCOUNT aligned each unique DNA sequence against the wild-type DNA sequence and tabulated the occurrence of wild-type and mutant at each position. The F value for each mutation was derived by dividing the antigen selection wild-type/mutant ratio by the display selection wild-type/mutant ratio, to correct for biases in display levels. The F value provides a quantitative estimate of the effect of each mutation on the binding affinity of Fab for hVEGF; deleterious effects are indicated by F values greater than 1.

Crystallization, structure determination and refinement

For large-scale preparation of Fab-D1 protein, whole cell broth was obtained from a 10 l *E. coli* fermentation. The cells were lysed with a Manton–Gaulin homogenizer. The suspension was centrifuged, the supernatant was loaded on a protein A-Sepharose column (Pharmacia), and the column was eluted with 0.1 M acetic acid. The pH was

adjusted to pH 4.0 with 1.0 M Tris (pH 8.0) and the eluant was loaded on a SP-Sepharose column (Pharmacia). The column was washed with equilibration buffer (20 mM Mes (pH 5.5)) and Fab protein was eluted with a NaCl gradient in equilibration buffer.

The complex between Fab-D1 and the receptor-binding fragment of hVEGF was formed and purified, as described.⁴⁸ The complex (in PBS, 25 mM EDTA) was concentrated to an optical density of $A_{280}=10$. Hanging-drop experiments were performed using the vapor-diffusion method with 10 μ l drops consisting of a 1:1 ratio of protein solution and reservoir solution. The reservoir solution was 0.2 M ammonium sulfate, 25% (w/v) PEG 3350, 0.1 M Hepes (pH 7.5). After one to two weeks at 19 °C, plate or spindle-shaped crystals grew.

Crystals were incubated in reservoir solution supplemented with 25% (v/v) glycerol prior to flash freezing. A data set was collected from a single frozen crystal at the beam line 5.0.2 of the Advanced Light Source (Berkeley, CA). The data were processed using the programs DENZO and SCALEPACK.⁴⁹ Crystals belonged to space group $P2_1$ (cell parameters of $a=154.5$ Å, $b=90.2$ Å, $c=205.6$ Å and $\beta=90.85^\circ$), with four complexes, each composed of one hVEGF homodimer bound to two Fab-D1 molecules, in the asymmetric unit. The structure was solved by molecular replacement using the program AMoRe³⁰ and the coordinates of a previously solved Fab-hVEGF complex (PDB entry 1BJ1). The model was manually adjusted using program O⁵⁰ and refinement was performed with program REFMAC³⁰ using tight 8-fold non-crystallographic symmetry restraints.

HTP sorting of library D

Purified target proteins were generous gifts of Drs Hui Li, Frank Collart and Andrzej Joachimiak at the MCSG, the Argonne National Laboratory. Proteins that contain a free Cys were first carboxymethylated by incubation overnight with 10 mM iodoacetic acid (pH 8.0). Target proteins were biotinylated using sulfosuccinimidyl 2-(biotinamido)-ethyl-1,3-dithiopropionate (EZ-Link Sulfo-NHS-SS-Biotin; Pierce Biotechnologies). Conditions were adjusted in such a way that one to two biotin molecules were attached per protein molecule, as determined by reversed-phase chromatography. Typical reaction solution contained 1 mg/ml protein in 50 mM Mops-NaOH buffer (pH 7.0) and 0.1 mg/ml of biotinylation reagent, and the reaction was performed for 45 min at room temperature. The reaction was quenched by an addition of concentrated Tris buffer. The modified proteins were dialyzed against TBS (150 mM NaCl, 50 mM Tris-HCl (pH 7.5)).

For the first round of binding selection, biotinylated target (0.1 nmol) was adsorbed to a sufficient quantity of Streptavidin MagneSphere Paramagnetic Particles (Promega) and then free streptavidin was blocked with biotin. After washing the magnetic beads, 1 ml of phage library solution (10^{12} – 10^{13} colony-forming units (cfu)) was added. After incubating at room temperature for 15 min, the phage solution was discarded and the beads were washed twice with TBST (TBS, 0.5% Tween 20) utilizing a magnetic tube stand. The captured phages on the magnetic beads were used directly to infect *E. coli* XL1-Blue cells, and phages were amplified as described.^{23,47}

In the second round, 20 μ l of amplified phage solution was mixed with 50 nM biotinylated target in a total volume of 100 μ l. After 15 min incubation, the biotinylated target was captured using streptavidin magnetic beads.

After a total of five washing steps with 100 μ l TBST (20 s per step), the captured phages were released from the beads by cleaving the disulfide linkage in the biotinylation reagent with 100 mM DTT in 20 mM Tris (pH 8.0) for 10 min. These processes were performed using a Kingfisher instrument (Thermo Scientific). Recovered phages were then amplified.

The third round was performed in the same manner as the second round except that the target concentration was reduced to 10 nM. An additional fourth round of selection was performed to evaluate successful enrichment of binding clones. In this round, amplified phage were tested in two separate sorting experiments, one with 10 nM of a biotinylated target and the other with no target (i.e. negative control). Recovered phage were titrated and the enrichment ratio was determined as the number of phage from the target-containing selection over the number of phage from the control selection. If a significant enrichment ratio (>20) was observed, the phages amplified after the third round were used for analysis of individual clones, as described below. If the enrichment ratio was low, the output of the fourth round was used for an additional round of evaluation. In all cases, we obtained a significant enrichment ratio after a total of three or four rounds of sorting.

HTP affinity characterization of Fabs

The phagemid DNA from an enriched Fab pool was purified as a mixture and used as the template for the Kunkel mutagenesis method⁵¹ to introduce a His₈ tag and a stop codon after the heavy chain constant domain. Twelve clones per target were randomly picked and used to transform *E. coli* 34B8. The transformants were grown in 0.5 ml of the CRAP media (27 mM (NH₄)₂SO₄, 2.4 mM sodium citrate-2H₂O, 14 mM KCl, 5.4 g/liter yeast extract, 5.4 g/liter HyCase SF Casein, 0.11 M Mops-NaOH buffer (pH 7.3), 0.55% (w/v) glucose and 7 mM MgSO₄) supplemented with 100 μ g/ml of ampicillin in a 96-deep well plate. The culture was lysed using B-PER (Pierce Biotechnologies) supplemented with hen egg lysozyme (0.3 mg/ml; Sigma) and deoxyribonuclease I (0.1 mg/ml; Sigma), and the Fab proteins were purified with NiNTA-magnetic beads (Dyna/Invitrogen) using a Kingfisher instrument. Briefly, the crude lysate was incubated with the beads with mixing for 40 min, washed twice with 50 mM Tris-HCl buffer (pH 8.0) containing 500 mM NaCl and 20 mM imidazole (5 min each), and the protein was eluted with 50 mM Tris-HCl buffer (pH 8.0) containing 500 mM NaCl and 500 mM imidazole. Surface plasmon resonance experiments were performed using a BIACORE 2000 instrument. The purified protein samples were diluted 20-fold to reduce the imidazole concentration. An Fab sample (20 μ l) was injected onto a Ni-NTA sensor chip and the association and dissociation kinetics of 300 nM antigen (20 μ l) was monitored using a flow rate of 5 μ l/min. Sensorgrams were analyzed in terms of the 1:1 binding reaction. The amino acid sequences of Fab clones that exhibited significant binding were then determined by DNA sequencing.

Protein Data Bank accession number

The coordinates and structure factors for the FabD1: hVEGF complex have been deposited in the RCSB Protein Data Bank with the assigned accession code 2QR0.

Acknowledgements

We thank Drs Hui Li, Frank Collart and Andrzej Joachimiak at the Argonne National Laboratory for their generous gift of protein samples. We thank Dr Jingdong Ye and Serdar Uysal for discussions. This work was supported in part by NIH grants R01-GM72688 and U54 GM74946 and by the University of Chicago Cancer Research Center. We acknowledge the use of the DNA sequencing core facilities at the University of Chicago. We thank the Genentech Fermentation and DNA sequencing groups. We also thank the staff at the Advanced Light Source (ALS), beam line 5.0.2. The ALS facility is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the US Department of Energy under contract number DE-AC02-05CH11231.

References

1. Kohler, G. & Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*, **256**, 495–497.
2. Bradbury, A. R. M. & Marks, J. D. (2004). Antibodies from phage antibody libraries. *J. Immunol. Methods*, **290**, 29–49.
3. Hoogenboom, H. R. (2002). Overview of antibody phage-display technology and its applications. *Methods Mol. Biol.* **178**, 1–37.
4. Hoogenboom, H. R. (2005). Selecting and screening recombinant antibody libraries. *Nature Biotechnol.* **23**, 1105–1116.
5. Rader, C. & Barbas, C. F. (1997). Phage display of combinatorial antibody libraries. *Curr. Opin. Biotechnol.* **8**, 503–508.
6. Winter, G., Griffiths, A. D., Hawkins, R. E. & Hoogenboom, H. R. (1994). Making antibody by phage display technology. *Annu. Rev. Immunol.* **12**, 433–455.
7. Sidhu, S. S., Fairbrother, W. J. & Deshayes, K. (2003). Exploring protein-protein interactions with phage display. *ChemBioChem.* **4**, 14–25.
8. Bradbury, A. (2003). Antibodies in proteomics I: generating antibodies. *Trends Biotechnol.* **21**, 275–281.
9. Bradbury, A., Velappan, N., Verzillo, V., Ovecka, M., Chasteen, L., Sblattero, D. *et al.* (2003). Antibodies in proteomics II: screening, high-throughput characterization and downstream applications. *Trends Biotechnol.* **21**, 312–317.
10. Fellouse, F. A. & Sidhu, S. S. (2006). Synthetic antibody libraries. In *Phage Display in Biotechnology and Drug Discovery* (Sidhu, S. S., ed), 1st edit., vol. 3, pp. 709–740, Taylor and Francis Group, Boca Raton, FL.
11. Sidhu, S. S. & Fellouse, F. A. (2006). Synthetic therapeutic antibodies. *Nature Chem. Biol.* **2**, 682–688.
12. Kabat, E. A., Wu, T. T., Redi-Miller, M., Perry, H. M. & Gottesman, K. S. (1987). *Sequences of Proteins of Immunological Interest*, 4th edit., National Institutes of Health, Bethesda, MD.
13. Padlan, E. A. (1996). X-ray crystallography of antibodies. *Advan. Protein Chem.* **49**, 57–133.
14. Davies, D. R. & Cohen, G. H. (1996). Interactions of protein antigens with antibodies. *Proc. Natl. Acad. Sci. USA*, **93**, 7–12.
15. Mian, I. S., Bradwell, A. R. & Olson, A. J. (1991). Structure, function and properties of antibody binding sites. *J. Mol. Biol.* **217**, 133–151.
16. Padlan, E. A. (1994). Anatomy of the antibody molecule. *Mol. Immunol.* **31**, 169–217.
17. Zemlin, M., Klinger, M., Link, J., Zemlin, C., Bauer, K., Engler, J. A. *et al.* (2003). Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. *J. Mol. Biol.* **334**, 733–749.
18. Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellenhofer, G. *et al.* (2000). Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J. Mol. Biol.* **296**, 57–86.
19. Fellouse, F. A., Barthelemy, P. A., Kelley, R. F. & Sidhu, S. S. (2006). Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. *J. Mol. Biol.* **357**, 100–114.
20. Fellouse, F. A., Bing, L., Compaan, D. M., Peden, A. A., Hymowitz, S. G. & Sidhu, S. S. (2005). Molecular recognition by a binary code. *J. Mol. Biol.* **348**, 1153–1162.
21. Fellouse, F. A., Wiesmann, C. & Sidhu, S. S. (2004). Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. *Proc. Natl. Acad. Sci. USA*, **101**, 12467–12472.
22. Koide, A., Gilbreth, R. N., Esaki, K., Tereshko, V. & Koide, S. (2007). High-affinity single-domain binding proteins with a binary-code interface. *Proc. Natl. Acad. Sci. USA*, **104**, 6632–6637.
23. Fellouse, F. A. & Sidhu, S. S. (2007). Making antibodies in bacteria. In *Making and Using Antibodies* (Howard, G. C. & Kaser, M. R., eds), pp. 157–180, CRC Press, Boca Raton, FL.
24. Sidhu, S. S., Lowman, H. B., Cunningham, B. C. & Wells, J. A. (2000). Phage display for selection of novel binding peptides. *Methods Enzymol.* **328**, 333–363.
25. Fuh, G., Wu, P., Liang, W.-C., Ultsch, M., Lee, C. V., Moffat, B. & Wiesmann, C. (2006). Structure-function studies of two synthetic anti-vascular endothelial growth factor Fabs and comparison with the Avastin™ Fab. *J. Biol. Chem.* **281**, 6625–6631.
26. Muller, Y. A., Chen, Y., Christinger, H. W., Li, B., Cunningham, B. C., Lowman, H. B. & de Vos, A. M. (1998). VEGF and the Fab fragment of a humanized neutralizing antibody: crystal structure of the complex at 2.4 Å resolution and mutational analysis of the interface. *Structure*, **5**, 1325–1338.
27. Vajdos, F. F., Adams, C. W., Breece, T. N., Presta, L. G., de Vos, A. M. & Sidhu, S. S. (2002). Comprehensive functional maps of the antigen-binding site of an anti-ErbB2 antibody obtained with shotgun scanning mutagenesis. *J. Mol. Biol.* **320**, 415–428.
28. Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A. & Sidhu, S. S. (2000). Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. USA*, **97**, 8950–8954.
29. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK - a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
30. CCP4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **60**, 432–438.
31. Griffiths, A. D., Williams, S. C., Hartley, O., Tomlinson, I. M., Waterhouse, P., Crosby, W. L. *et al.* (1994). Isolation of high affinity human antibodies directly from large synthetic repertoires. *EMBO J.* **13**, 3245–3260.

32. Elkins, P. A., Christinger, H. W., Sandowski, Y., Sakal, E., Gertler, A., de Vos, A. M. & Kossiakoff, A. A. (2000). Ternary complex between placental lactogen and the extracellular domain of the prolactin receptor. *Nature Struct. Biol.* **7**, 808–815.
33. Clark, L. A., Ganesan, S., Papp, S. & van Vlijmen, W. T. (2006). Trends in antibody sequence changes during the somatic hypermutation process. *J. Immunol.* **177**, 333–340.
34. Tomlinson, I. M., Walter, G., Jones, P. T., Dear, P. H., Sonnhhammer, E. L. & Winter, G. (1996). The imprint of somatic hypermutation on the repertoire of human germline V genes. *J. Mol. Biol.* **256**, 813–817.
35. Bond, C. J., Marsters, J. C., Jr & Sidhu, S. S. (2003). Contributions of CDR3 to VHH domain stability and the design of monobody scaffolds for naive antibody libraries. *J. Mol. Biol.* **332**, 643–655.
36. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature*, **424**, 805–808.
37. Lopez de la Paz, M. & Serrano, L. (2004). Sequence determinants of amyloid fibril formation. *Proc. Natl. Acad. Sci. USA*, **101**, 87–92.
38. Bogan, A. A. & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**, 1–9.
39. de Haard, H. J., van Neer, N., Reurs, A., Hufton, S. E., Roovers, R. C., Henderikx, P. *et al.* (1999). A large non-immunized human Fab fragment phage library that permits rapid isolation and kinetic analysis of high affinity antibodies. *J. Biol. Chem.* **274**, 18218–18230.
40. Sheets, M. D., Amersdorfer, P., Finnern, R., Sargent, P., Lindqvist, E., Schier, R. *et al.* (1998). Efficient construction of a large non-immune phage antibody library: the production of high-affinity human single-chain antibodies to protein antigens. *Proc. Natl. Acad. Sci. USA*, **95**, 6157–6162.
41. Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshaw, J. C. *et al.* (1996). Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nature Biotechnol.* **14**, 309–314.
42. Dobson, C. L., Minter, R. R. & Hart-Shorrock, C. P. (2005). Naive antibody libraries from natural repertoires. In *Phage Display in Biotechnology and Drug Discovery* (Sidhu, S. S., ed), 1st edit., vol. 3, pp. 659–708, Taylor and Francis Group, Boca Raton, FL.
43. Hoet, R. M., Cohen, E. H., Kent, R. B., Rookey, K., Schoonbroodt, S., Hogan, S. *et al.* (2005). Generation of high-affinity human antibodies by combining donor-derived and synthetic complementarity-determining-region diversity. *Nature Biotechnol.* **23**, 344–348.
44. Lee, C. V., Liang, W.-C., Dennis, M. S., Eigenbrot, C., Sidhu, S. S. & Fuh, G. (2004). High-affinity human antibodies from phage-displayed synthetic Fab libraries with a single framework scaffold. *J. Mol. Biol.* **340**, 1073–1093.
45. Skerra, A. (2000). Engineered protein scaffolds for molecular recognition. *J. Mol. Recognit.* **13**, 167–187.
46. Lee, C. V., Sidhu, S. S. & Fuh, G. (2004). Bivalent antibody phage display mimics natural immunoglobulin. *J. Immunol. Methods*, **284**, 119–132.
47. Sidhu, S. S., Li, B., Chen, Y., Fellouse, F. A., Eigenbrot, C. & Fuh, G. (2004). Phage-displayed antibody libraries of synthetic heavy chain complementarity determining regions. *J. Mol. Biol.* **338**, 299–310.
48. Muller, Y. A., Chen, Y., Christinger, H. W., Li, B., Cunningham, B. C., Lowman, H. B. & de Vos, A. M. (1998). VEGF and the Fab fragment of a humanized neutralizing antibody: crystal structure of the complex at 2.4 Å resolution and mutational analysis of the interface. *Structure*, **6**, 1153–1167.
49. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
50. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
51. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154**, 367–382.