

AbDesign: An algorithm for combinatorial backbone design guided by natural conformations and sequences

Gideon D. Lapidoth,¹ Dror Baran,¹ Gabriele M. Pszolla,¹ Christoffer Norn,¹ Assaf Alon,¹ Michael D. Tyka,² and Sarel J. Fleishman^{1*}

¹ Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

² Google Inc., 1600 Amphitheatre Pkwy, Mountain View, California 94043

ABSTRACT

Computational design of protein function has made substantial progress, generating new enzymes, binders, inhibitors, and nanomaterials not previously seen in nature. However, the ability to design new protein backbones for function—essential to exert control over all polypeptide degrees of freedom—remains a critical challenge. Most previous attempts to design new backbones computed the mainchain from scratch. Here, instead, we describe a combinatorial backbone and sequence optimization algorithm called *AbDesign*, which leverages the large number of sequences and experimentally determined molecular structures of antibodies to construct new antibody models, dock them against target surfaces and optimize their sequence and backbone conformation for high stability and binding affinity. We used the algorithm to produce antibody designs that target the same molecular surfaces as nine natural, high-affinity antibodies; in five cases interface sequence identity is above 30%, and in four of those the backbone conformation at the core of the antibody binding surface is within 1 Å root-mean square deviation from the natural antibodies. Designs recapitulate polar interaction networks observed in natural complexes, and amino acid sidechain rigidity at the designed binding surface, which is likely important for affinity and specificity, is high compared to previous design studies. In designed anti-lysozyme antibodies, complementarity-determining regions (CDRs) at the periphery of the interface, such as L1 and H2, show greater backbone conformation diversity than the CDRs at the core of the interface, and increase the binding surface area compared to the natural antibody, potentially enhancing affinity and specificity.

Proteins 2015; 83:1385–1406.
© 2015 Wiley Periodicals, Inc.

Key words: CDRs; V(D)J recombination; computational protein design; fuzzy-logic design; Rosetta; canonical conformations; modular segments; conformation-sequence optimization.

INTRODUCTION

Molecular recognition underlies many central biological processes. The ability to design novel protein interactions is a stringent test of our understanding of the physico-chemical principles that govern molecular recognition and holds promise for creating specific and sensitive molecules for use as therapeutics, diagnostics, and research probes. Recent strategies in protein-binder design used naturally occurring proteins as scaffolds on which binding surfaces were designed.^{1–4} These strategies relied either on a small number of protein scaffolds^{2,3} or several hundred different scaffolds^{4–7} to achieve the structural characteristics required for binding. In all cases the designed scaffolds were treated as rigid structures with minimal perturbation of their backbone degrees of freedom. These strategies

resulted in the experimentally validated design of homooligomers,^{8–12} inhibitors,^{4,6} and a protein purification reagent.⁵ Several generalizations have been made about successfully designed binding surfaces: (1) they comprise surfaces rich in secondary structure (α -helices and

Additional Supporting Information may be found in the online version of this article.

Grant sponsors: Israel Science Foundation; Center of Research Excellence (I-CORE) in Structural Cell Biology, the Human Frontier Science Program; a Marie Curie Reintegration Grant; a European Research Council Starter's Grant; an Alon Fellowship; the Yeda-Sela Center; the Geffen Fund; the Minerva Foundation; Sam Switzer; Martha S. Sagon Career Development Chair; Boehringer Ingelheim Fonds. *Correspondence to: Sarel Fleishman, Department of Biological Chemistry Ullman Building Room 301c Weizmann Institute of Science Rehovot 76100, Israel. E-mail: sarel@weizmann.ac.il

Received 12 November 2014; Revised 13 January 2015; Accepted 26 January 2015
Published online 9 February 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24779

β -sheets); (2) Interactions with the ligand are largely mediated by hydrophobic amino acid sidechains; and (3) The buried surface area upon binding is at or smaller than the average for naturally occurring protein–protein interactions (1600 Å²).¹³ The design of large and polar surfaces, essential to make computational binder design general, remains an unmet challenge.^{14–17}

We reasoned that a key to solving the challenge of designing large and polar binding surfaces lies with the design of the protein backbone, since the backbone provides many additional conformation degrees of freedom that have so far been untapped by binder-design strategies. Designing backbones for function, however, is an unsolved problem due to uncertainty in assessing the contributions to free energy from polar groups and due to the large conformation space open to the protein backbone.¹⁸ As a step to address the challenge of designing backbones in binders we suggest an algorithm that uses conformation and sequence information from naturally occurring proteins belonging to the same fold family in order to constrain backbone design and amino acid sequence choices, thereby limiting modeling uncertainty while exposing a large space of conformations to computational design. We test this approach by computing antibody models that target pre-chosen protein epitopes and assess sequence and backbone conformation recovery compared to natural antibodies in complex with the same epitopes.

Antibody structure, function, and engineering

The key challenge in the design of backbones for function is that the designed surface needs both to bind its target and be conformationally stable; the design of antibodies for function therefore brings into focus the need to develop methods that simultaneously optimize both binding and stability, challenges which have hitherto been approached by computational design separately.^{7,19,20} Natural antibodies are built of sequence blocks that alternate conserved with highly variable segments.^{21–23} The molecular structures of antibodies show that the conserved segments belong to a structurally homologous and rigid structure known as the framework, which confers stability to the antibody, whereas the variable segments cluster at the ligand-binding surface, and are therefore termed the complementarity-determining regions (CDRs). Despite their tremendous binding-surface diversity, all CDRs except H3 fall into a handful of discrete conformations termed “canonical conformations.”^{22,24–26} For instance, in hundreds of antibody molecular structures only seven conformation variants are observed for L2.²⁴ Each canonical conformation is characterized by key conserved residue identities, which are important for maintaining the backbone conformation.^{21,22,24,25,27} Some other fold families, such

as ankyrin-repeat proteins and the Rossman fold, which similar to antibodies are associated with many molecular functions, are likewise modular, with a clear separation between a structurally conserved region and a variable region, where function is typically encoded.²⁸ The ability to design diverse backbones within fold families could therefore have many applications.

A key attraction for protein engineering lies in antibodies’ modular architecture, suggesting that a large combinatorial complexity of well-folded backbones could be tapped. As early as the 1980s, observations on the structural modularity of antibodies made by Lesk and Chothia²⁹ proposed that synthetic antibodies could be constructed by combining fragments of naturally occurring antibodies. From this insight, Winter and co-workers devised a method for antibody humanization, in which CDRs from a mouse antibody were grafted onto a human antibody framework to generate a humanized functional antibody,^{30,31} opening the way to safe therapeutic antibody engineering. These early advances raised excitement that the complete design of antibodies from first principles is achievable,³² but until recently, computational tools for protein design had not matured sufficiently to realize this objective.

An important advantage of computational design over conventional protein-engineering methods has been its ability to generate binders of specific sites of interest on target molecules with atomic accuracy. Site-specific targeting has been essential to the design of broad-specificity influenza inhibitors, a pH-sensitive binder, and an enzyme inhibitor.^{4–6} In contrast to this ability to target specific molecular surfaces, conventional antibody-engineering methods, such as animal immunization and repertoire selection,³³ are capable of isolating binders to target molecules, but there is no general method to target binders to specific epitopes,^{34,35} hampering efforts to generate specific binders, inhibitors, and allosteric effectors.^{36–38} To be sure, in certain systems selection methods were developed to isolate antibodies that bind specific epitopes on target molecules^{39–43}; yet these capabilities are challenging and rely on specific properties of the target molecule, such as naturally high antigenic variability outside the target site.⁴⁴ Computational antibody design may in future complement existing antibody-engineering techniques, and open new avenues for generating antibodies that target specific sites and even rare conformations on target receptors and enzymes.

Recent work on computational antibody design aimed to increase binding affinity,^{45–47} identify favorable positions for experimental random mutagenesis,⁴⁸ modify binding specificity⁴⁹ and increase thermo-resistance.⁵⁰ An antibody design strategy was suggested by Pantazes et al.^{51–53} that capitalizes on observations that antibody CDRs exhibit canonical conformations. In this method a representative set of antibody CDRs is designed from canonical conformations, and then docked and designed

to bind the target epitope. The resulting output from this procedure is a CDR library that can be grafted on a selected antibody framework. The *AbDesign* algorithm reported below differs in three important respects from this strategy: first, rather than segmenting the CDRs individually *AbDesign* segments the antibody backbone using junctions of high structure conservation, generating structurally compatible framework-CDR interactions; second, *AbDesign* derives sequence information from natural antibodies to constrain sequence optimization to amino acid identities that are important for the stability of the modeled conformation; and third, *AbDesign* conducts combinatorial backbone design, sampling backbones from all the natural antibodies in the structure database, including highly homologous ones, to improve binding affinity and antibody stability. The procedure is general and can be adapted, in principle, to any modular protein fold family.

MATERIAL AND METHODS

Source code and structure models availability

The methods have been implemented within the Rosetta macromolecular modeling software suite⁵⁴ and are available through the Rosetta Commons agreement. All of the methods have been implemented through RosettaScripts,⁵⁵ and all scripts are available as Supporting Information. The ten top-ranked structure models targeting each of the epitopes studied in this article are provided in the Supporting Information. These models were automatically generated, filtered, and ranked using the methods presented below; we note that designs chosen for experimental testing are typically selected from a larger pool, visually inspected for flaws and manually corrected prior to testing.^{1,2,6,7,56}

Binding mode criteria

Following the Critical Assessment of PRredicted Interactions (CAPRI) convention we use interface-root mean square deviation (I_RMS) with a cutoff of 4 Å to define which designs fail to recapitulate the natural binding mode.⁵⁷ This measure computes the C α RMSD on all ligand residues with atoms within 10 Å of the antibody in a structure in which the natural and designed antibody structure are aligned.

Energy and structure filters

Shape complementarity (*Sc*) was computed using the algorithm described in Ref. 58 implemented in Rosetta.⁵⁴ *Sc* ranges from 0 (no shape complementarity) to 1 (perfect complementarity). Protein packing quality at the antibody core and antibody–ligand interface were

calculated using RosettaHoles (Packstat).⁵⁹ Antibody designs with *Sc* or Packstat values <0.57 were rejected.

The binding energy is defined as the difference between the total system energy in the bound and unbound states. In each state, interface residues are allowed to repack. For numerical stability, binding-energy calculations were repeated three times, and the average was taken.

Antibody stability is defined as the Rosetta all-atom system energy of the antibody monomer when the ligand is eliminated from the system.

All-atom energies were calculated using the default Rosetta energy (score12), which is dominated by contributions from van der Waals packing, solvation, and hydrogen bonding.⁵⁴

Docking antibody scaffolds to the target epitope

Each initial antibody scaffold was aligned to the natural antibody framework in the experimentally determined molecular structure using a customized PyMol script,⁶⁰ and the ligand coordinates were combined with the antibody scaffold model to produce a single coordinate file used as input for design simulations. The resulting binding mode was perturbed with RosettaDock⁶¹ using low-resolution docking (centroid mode).

Boltzmann conformational probabilities of interface side chains

Boltzmann conformational probabilities were calculated as described in Ref. 62. For each partner in the complex and for each residue that contributes >1 Rosetta energy units (R.e.u) to the predicted binding energy we iterate, in the unbound state, over all the backbone-dependent rotamers in the Dunbrack library defined within the Rosetta software. For each rotamer, all residues within a 6 Å shell are repacked and minimized. The energy *E* of each such state is then evaluated using the Rosetta all-atom energy function (score12).⁶³ The probability of the conformation of residue *i*, *P_i*, is then computed assuming a Boltzmann distribution:

$$P_i = \frac{e^{\frac{-E_i}{K_B T}}}{\sum_s e^{\frac{-E_s}{K_B T}}} \quad (1)$$

Where *s* is the rotameric state, *K_B* is the Boltzmann constant, and *T* is the absolute temperature. *K_BT* was set to 0.8 R.e.u. *E_i* is the energy of the unbound state.

Backbone-segment clustering and sequence profiles

The antibody structures were aligned separately to the variable heavy and variable light domains of antibody 4m5.3, (PDB entry 1X9Q).⁴⁶ We then extracted the

Table 1
Comparison of CDR Definitions

| CDR | Chothia | CDR definitions used for sequence constraints ^a | Segment definitions used for PSSMs and backbone modeling |
|-----------------|---------|--|--|
| VL ^b | L1 | L24-L34 | L23-L35 |
| | L2 | L50-L56 | L46-L55 |
| L3 | | L89-L97 | L88-L98 |
| VH ^b | H1 | H26-H32 | H26-H37 |
| | H2 | H50-H58 | H45-H58 |
| H3 | | H95-H102 | H93-H103 |

^aSee Results, section b.

^bIncluding CDRs 1 and 2 and intervening framework positions.

coordinates of the CDRs according to VL, L3, VH, and H3 definitions (Table 1) and clustered them according to length. For L3 and H3, which show high conformational variability,^{22,24,64,65} we performed additional clustering using BCL::cluster⁶⁶ with a 2.0 Å C_α RMSD radius. The resulting clusters were visually inspected for common sequence motifs, and clusters that contained several different sequence motifs were manually split; conversely, conformation clusters that shared sequence motifs were merged. Clustering results in 207 H3 bins, and the top 50 clusters (by size) were used to generate the conformation representatives (algorithm, section d).

For each backbone conformation cluster we generated a Position Specific Scoring Matrix (PSSM) of unique sequences that belong to this cluster using the PSI-BLAST suite⁶⁷ with default parameters. In the case of singleton backbone conformations (in H3) the BLOSUM62 scoring matrix is used to provide a statistical model for tolerance to amino acid substitutions.

Code-integrity tests

The integrity of the Rosetta source code is maintained through a set of integration tests. Antibody conformation sampling and sequence design are maintained with three tests: the splice_out integration test ensures that the algorithm can properly extract backbone segments from the source antibody and create a new torsion database; the splice_in integration test checks that the algorithm can read the torsion database and impose a new backbone conformation onto the template antibody; and the splice_seq_constraints integration test checks that the algorithm can add sequence constraints to an antibody structure.

Algorithm performance

Following precomputation of sequence and backbone-torsion databases a typical design trajectory takes ~7 h on an Intel Xeon 2.4 GHz CPU. The protocol is divided into two parts. First, the complex formed between the designed antibody scaffold (algorithm, section d) and the target molecule is subjected to docking, design, and min-

imization (algorithm, section e); this step takes only 3 minutes; the vast majority of time is spent in the downstream refinement steps (algorithm, section f). To make efficient use of computational resources *AbDesign* applies energy and structure filtering before going into refinement; on average, only 4% of all trajectories pass this filtering. Depending on the availability of computational resources and the magnitude of the design problem, filters at this step can be adjusted.

Checkpointing

We use checkpointing to ensure that if a design trajectory is prematurely terminated due to computer resource outage it can be resumed from the last backup point. A PDB-formatted file containing the coordinate information of the complex is saved to disk along with the details on the design stage, complex stability, and binding energies, whenever a sampled backbone improves the objective function (algorithm, section g). When *AbDesign* is started it automatically checks for the existence of checkpointing files; if those are found, *AbDesign* will continue from the last checkpoint. Restarting simulations from the backup point takes <30 s.

RESULTS

AbDesign: An algorithm for combinatorial backbone-sequence optimization in protein-fold families

Using Figure 1 as a visual guide we present a step-by-step description of the *AbDesign* process. *AbDesign* addresses four related challenges: (1) Leveraging knowledge from conformation and sequence databases to constrain design choices; (2) Encoding residue correlations between the variable segments, which largely lack stabilizing secondary-structure elements, and the framework, which forms a tightly packed and stable structural foundation; (3) Efficient sampling of the large backbone and sequence combinatorial space encoded in a fold family; and (4) Designing conformations and sequences that optimize both protein stability and target-molecule binding. In the following sections we describe the different elements of the algorithm in detail and how they relate to these design challenges.

a. Sequence constraints from natural antibodies guide amino acid design choices

The stability of a protein conformation relative to unfolded and misfolded states relies on both positive and negative design elements; whereas positive design elements address the target conformation and are amenable to modeling, negative design with respect to the vast conformation space open to a loop is impractical for modeling.^{18,70} A key advantage of computational

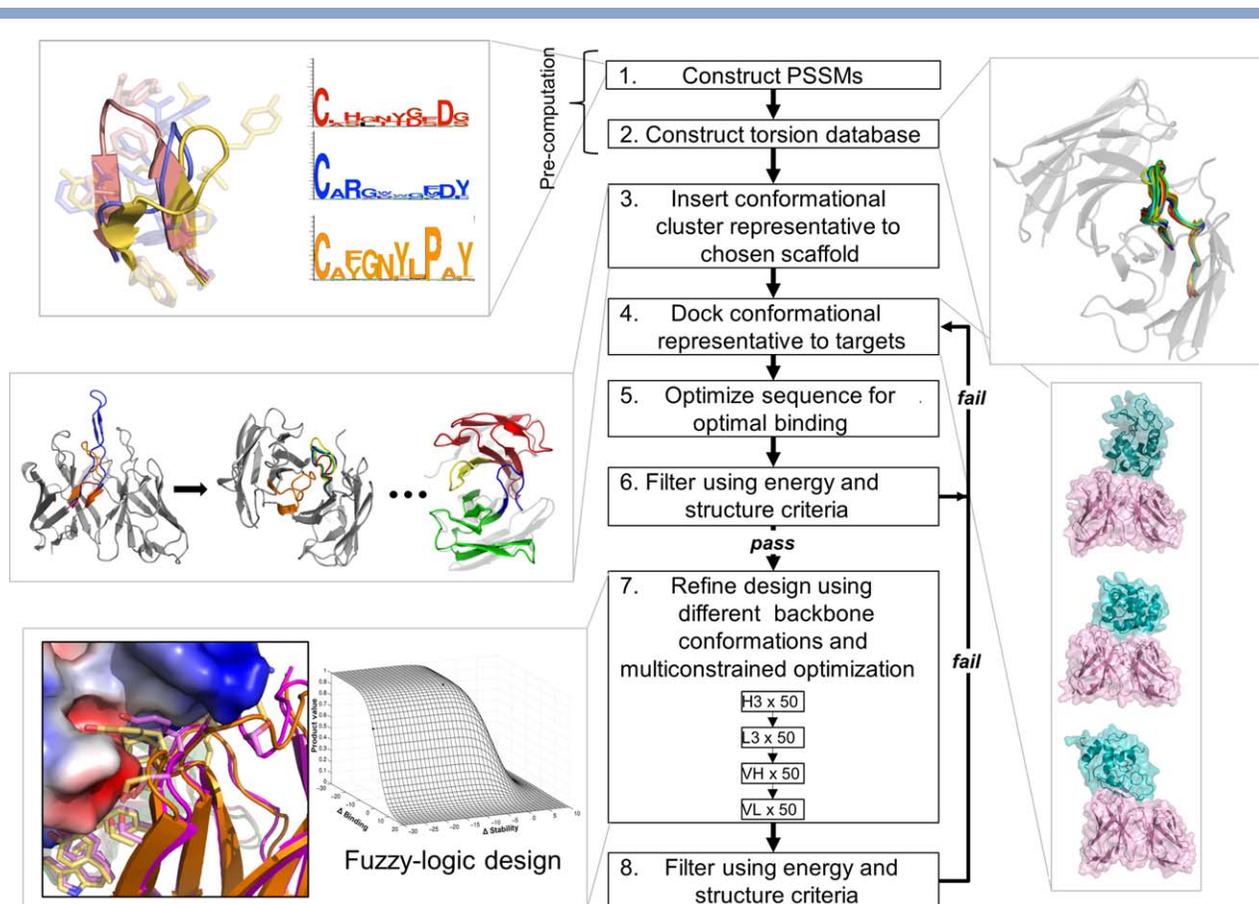


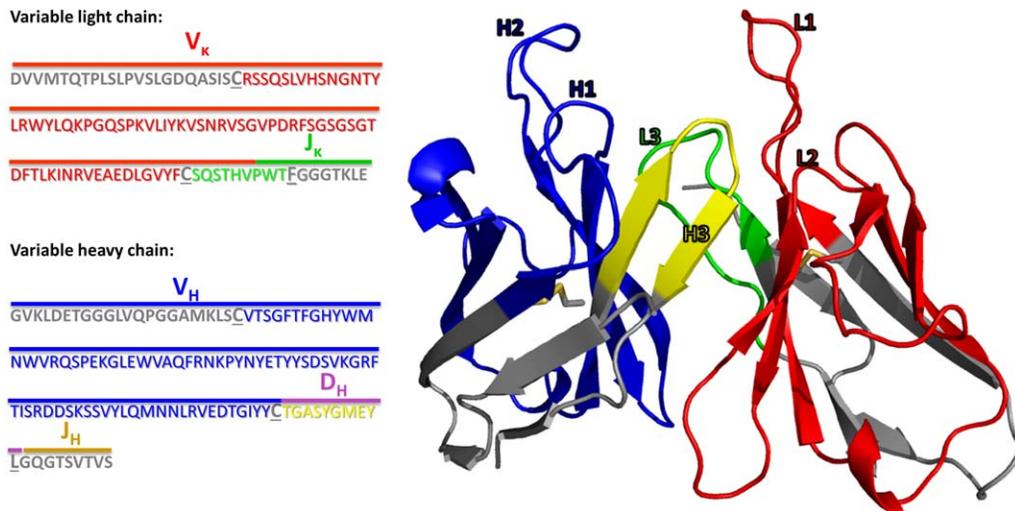
Figure 1

Overview of the design protocol workflow. Briefly, structures of naturally occurring antibodies are extracted from the Protein Data Bank (PDB)⁶⁸ and aligned to a template antibody structure. Backbone segment conformations and sequences are extracted into two correlated databases: a Position-Specific Site Matrix database (step 1) and a backbone-torsion database (step 2), where PSSMs and their respective torsion databases are linked. From the torsion database a set of antibody conformations representing all combinations of canonical conformations is generated (step 3), docked against the target surface (step 4) and designed for optimal binding affinity, subject to sequence constraints derived from the PSSMs (step 5). Antibodies passing structure and energy filters (step 6) are then subjected to a backbone and sequence refinement protocol (step 7): for each backbone segment (VL, VH, L3, H3) alternative conformations are sampled from the pre-computed torsion database and designed in the context of the modeled antibody-bound structure. The backbone conformation with the highest computed stability and affinity for the ligand is selected using fuzzy-logic design⁶⁹ and serves as input in the optimization of the next backbone segment. Finally, designs are filtered using energy and structural criteria derived from natural antibodies (step 8). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

design of proteins belonging to a diverse fold family, such as antibodies, is that we can extract statistics regarding amino acid choices on a per-position basis that encode at least some of these elements, and use these statistics to guide the design process. Moreover, by correlating natural backbone conformations and sequences we can classify sets of natural protein-segment sequences that fold into particular conformation classes (such as antibody canonical conformations), and maintain for each of these classes its own unique sequence profile.

For each segment cluster we generate a Position Specific Scoring Matrix (PSSM) using the PSI-BLAST software package.⁷¹ The sequence constraints encoded in the PSSMs are stringent in the antibody framework and

relaxed at the CDRs, giving sequence optimization room to explore different residue combinations for interacting with ligand, while maintaining the antibody's structural integrity. The PSSM is used during all design calculations in two ways. First, design sequence choices are restricted only to identities above a conservation threshold according to the PSSM. The cutoffs are determined separately for the binding site (PSSM score ≥ 0 for all antibody residues with C_{β} 's within a 10 Å distance cut-off of the ligand), CDRs (≥ 1 , Table I), and framework positions (≥ 2). Effectively, positions that are important for binding are allowed more room to vary from the family consensus than positions in the antibody framework. Second, the all-atom energy function is modified to include a term that biases the sequence toward the more

**Figure 2**

Natural V (D) J gene segmentation versus conformation segmentation used in *AbDesign* represented on the 4m5.3 (PDB entry 1X9Q) antibody. *AbDesign* segments the antibody structure at the disulfide-linked cysteines and in a structurally conserved position at the end of CDR3 (stem positions are underlined). Natural antibody recombination follows a similar, but not precisely the same, segmentation (bars above sequence and the V, D, and J labels). Sequence and structure are color-coded by conformation segments (red: CDRs L1, L2, and framework, green: L3, blue: H1, H2, and framework, yellow: H3). Gray segments are only subjected to sequence, rather than backbone optimization.

likely identities according to the PSSM. The bias toward the sequence consensus is weighted 50% more strongly away from the binding site.

b. A precomputed database of backbone conformations for each antibody segment

Backbone-conformation sampling is computationally demanding^{72–75} and despite some success⁷⁶ backbone design for function has led to conformations that deviated from the original computed models.^{19,77} By designing proteins in a conformationally highly diverse family, such as antibodies, we can make use of hundreds of naturally occurring conformation variants for each backbone segment, where the conformations are likely to be stable within the host protein fold. In a pre-computation step we extract the conformations of natural antibodies (Fig. 1 step 2) and store them in a database for use during design. 788 variable light κ chains and 785 variable heavy-chain structures are superimposed on a template antibody (throughout this manuscript, we use as template antibody 4m5.3, Protein Data Bank (PDB) entry 1X9Q, a high-expression, high-affinity anti-fluorescein antibody,⁷⁸ although the choice of template is arbitrary). λ variable light chains were not included in the current database because of the relatively small number of available structures in the PDB (1300 variable κ chains vs. 265 variable λ chains⁷⁹), although *AbDesign* can address λ chains without changes to the algorithm.

Next we identify positions on the protein backbone, which are structurally highly conserved in all antibody

molecules; due to the high homology such positions can serve as effective junctions or stems for recombining backbones from antibodies. Past structural analysis suggested to use stems corresponding to each individual CDR, for instance at the start and end of CDR1 (L24–L34, H26–H32, Chothia numbering²⁵), CDR2 (L50–L56, H52–H56), and CDR3 (L89–L97, H95–H102).^{21,22,25–27,30,47,80–84} Our preliminary *in silico* experiments using such stem choices, however, resulted in structurally unrealistic designed antibodies with poor packing between the CDR and the framework sidechains. Instead we use the disulfide-linked cysteines in each of the variable domains as stems for a segment comprising CDR1 and CDR2 and the framework region, and the second disulfide-linked cysteine and a conserved position at the end of CDR3 (position numbers 100 in the variable κ domain 103 in the variable heavy domain,^{25,26} Fig. 2) as the stems for the CDR3 segments; these stems are very well aligned in all antibodies of known structure (Fig. 2). The genomic recombination of the V and (D) J genes occurs at a variable position C-terminally to the second cysteine in each variable domain, but the genomic-recombination sites are structurally poorly aligned in a set of diverse antibodies compared to the disulfide-linked cysteines, which therefore provide more favorable junctions for joining conformation fragments.

By using segment boundaries that are close to the V (D) J genomic segmentation we directly embody conformation and sequence correlations between the CDRs and the framework that were refined by natural selection, encoding both local and global sequence–structure

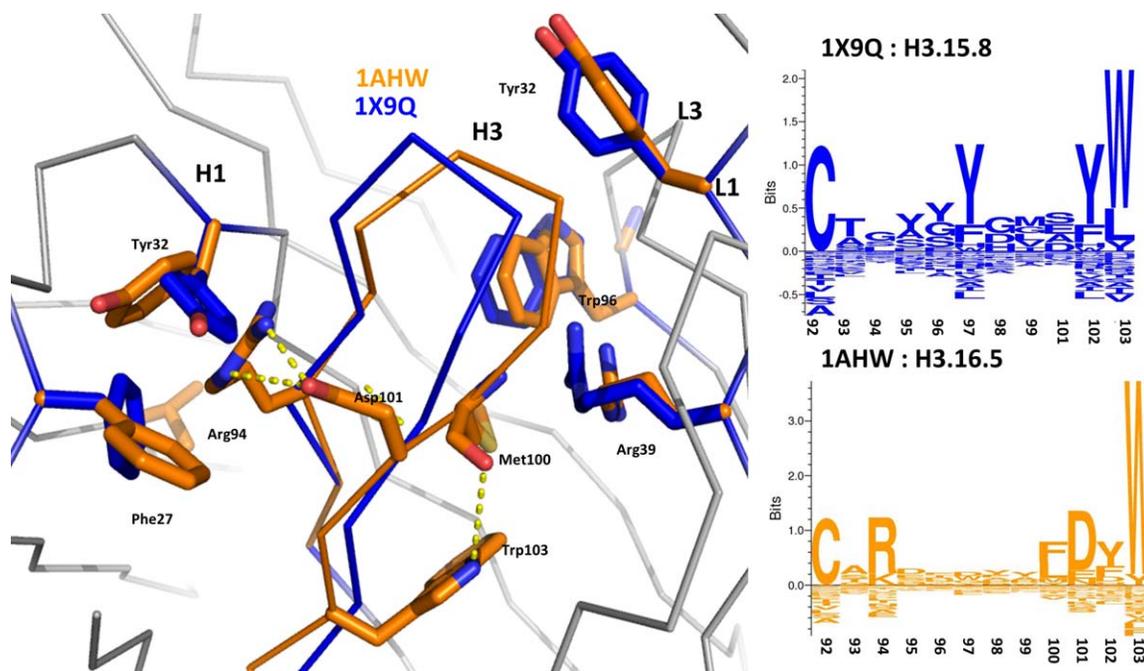


Figure 3

Sequence and conformation coupling during design. During design of a new backbone the PSSMs used to constrain sequence choices are altered. In this example, the H3 backbone segment from antibody 5G9 (PDB entry 1AHW) is modeled in the context of the 4m5.3 antibody (PDB entry 1X9Q). Web-logos for the two conformation segments are shown on the right, revealing different amino acid conservation patterns, which are important for the structural integrity of the modeled segment. For instance, the H3 backbone conformation from 1X9Q is in an extended conformation, whereas the imposed H3 backbone conformation is kinked,⁸⁵ and characterized by a hydrogen bond between the conserved stem Trp (Trp103, Chothia numbering) Ne1 atom and a carbonyl oxygen (Met100, Chothia numbering). The conserved salt bridge between Arg94 and Asp101 is similarly frequently observed in kinked conformations. Surrounding residues in a 6 Å shell around the inserted backbone segment are also designed and repacked under sequence constraints to accommodate the new backbone conformation. In this example, residues Phe27 and Tyr32 from the heavy chain and residue Tyr32 from the light chain are repacked to avoid clashes with the designed H3 conformation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

relationships. As a concrete example for the importance of the correlations between conformations and sequences, the H3 backbone cluster H3.15.8 (Supporting Information Table S1) is in an extended conformation, while the conformation from the cluster H3.16.5 is kinked⁸⁵; the two clusters' sequence profiles are correspondingly characterized by different amino acid conservation patterns (Fig. 3), which are encoded in the respective PSSMs used during design.

For each segment (VL, L3, VH, and H3) of each of the natural antibodies in our database we extract the backbone dihedral angles (Φ , Ψ , and Ω) from the source antibody and replace the segment in the template with the source segment's dihedral angles (Fig. 1, step 2), introducing a main-chain cut site in a randomly chosen position in the inserted segment. Where the inserted segment is longer than the template antibody segment, residues are added to the model using idealized bond lengths and angles. We then refine the main chain using cyclic-coordinate descent (CCD),⁷⁴ small, and shear moves, as implemented in the CCD mover in Rosetta. During refinement the standard Rosetta all-atom energy

function (score12)⁶³ is modified by the addition of an energy term that favors closing the main-chain gap, and harmonic restraints that bias the C_{α} positions and the backbone-dihedral angles of each modeled amino acid to the values observed in the source antibody to minimize deviation from the source conformation. CCD alternates backbone moves with combinatorial amino acid side-chain packing. During packing steps we also allow combinatorial stochastic sequence optimization in the entire modeled segment and in a 6 Å shell surrounding the segment subject to amino acid constraints derived from the antibody PSSM. At the end of CCD we compute the root mean square deviation (RMSD) of the modeled segment from the source segment and if it exceeds 1 Å or if the main-chain gap score is >0.5 we repeat the procedure. Segments that fail to meet the criteria above after 10 trials are discarded from further consideration. Although CCD was originally conceived as a method for loop closure,⁷⁴ here we find that, guided by coordinate and dihedral constraints from naturally occurring segments, CCD effectively refines segments up to 74 amino acids long to the RMSD and main-chain gap criteria

above within, on average, 1.2 attempts. Given the above selection criteria, the natural backbone conformations are fitted onto the template scaffold in the majority of antibody entries in our database, ranging from 74% of the VH segments to 96% of L3 segments. Each trajectory takes on average 4.6 hours on an Intel Xeon 2.4 GHz CPU. Backbone dihedral angles of successfully fitted segments are recorded in a backbone torsion database for subsequent use during design (Supporting Information Table SII).

c. Design subject to sequence constraints derived from natural antibodies

In sections (a) and (b), above, we precomputed correlated PSSM and backbone-conformation databases. During design we load the pre-computed PSSM matrices associated with the current conformation (4 PSSM segments, one for each backbone segment), and combine them to generate a single PSSM matrix for the entire antibody. Whenever a different backbone conformation is sampled *AbDesign* replaces the relevant PSSM matrix associated with the swapped segment, automatically synchronizing the sequence constraints with the backbone conformation.

For efficiency, at different phases of design different sets of residues are subjected to combinatorial sequence optimization. For instance, several initial design phases only optimize the ligand-binding surface, whereas at the final stages of design there are several steps of sequence optimization over all antibody positions. Sequence constraints (section a) considerably reduce the combinatorial design problem: in a representative case, the latter step of full design over a 230 amino acid antibody variable fragment has a total of $\sim 10^{117}$ different possible sequence combinations, equivalent to full combinatorial design of only 93 positions; increasing the PSSM cutoffs would further reduce this combinatorial space.

d. A representative set of antibody conformations

Combining the four antibody segments (VL, L3, VH, and H3) using all backbone conformations extracted in section (b) above would result in a prohibitively large library of antibody scaffolds for design. Observations made by Chothia and coworkers,^{22,24} however, highlighted that each antibody backbone segment other than H3 falls into a handful of canonical conformations. We start the design process by generating a library of representative antibody backbones that spans the space of these canonical conformations plus a set of 50 H3 backbone conformations (Fig. 1, step 3). We extract the conformation mean from each cluster and reduce the number of representative structures further by eliminating similar conformations by visual inspection. This procedure results in 5 (VL) \times 2 (L3) \times 9 (VH) \times 50 (H3) = 4500 non-redundant conformation representatives (Supporting Information Table SIII), exceeding the number of solved

antibody structures (Methods). All sequence and conformation information from the template antibody is eliminated in constructing the conformation representatives, except for the relative orientation of the disulfide-bonded cysteines in the variable light and variable heavy domains. In other protein fold families, where canonical conformations have not been characterized, automated clustering of backbone conformations can be employed to generate the reduced set of conformation representatives.²⁴

e. Low-resolution docking and sequence design

Each of the 4,500 representative conformations generated in step d is aligned using PyMol⁶⁰ to the natural antibody to obtain conformations where the representative antibodies are bound to the target molecule in approximately the same orientation as the natural antibody. In each design trajectory, this conformation is perturbed using low-resolution (centroid) RosettaDock (Fig. 1, step 4)⁸⁸ to randomize the initial binding orientation within the vicinity of the naturally observed binding mode, and the target protein–ligand surface is repacked to eliminate memory of the bound sidechain conformations. This procedure is in keeping with previous studies of binder and enzyme design,^{7,20,89} where the target site for binding was constrained to the one observed in the natural complex to avoid sampling the impractically large space of orientation and sequence open to design of function. In the context of antibody design sequence-structure space is still larger than in previous studies due to the additional backbone-conformation degrees of freedom. Indeed, where intense experimental effort was invested many different antibodies and epitopes were discovered that target a single molecule,⁹⁰ suggesting that without restricting to the natural target epitope a potentially large number of different binding modes and sequences might result. In cases where the target epitope or binding mode are not defined *a priori* docking software, such as PatchDock⁹¹ or RosettaDock,⁸⁸ can be used to generate the initial bound conformations, as was done in binder design applications^{4–6}; the *AbDesign* methodology can therefore be extended, in principle, to binder design in the absence of a known antibody-bound complex.

Following docking the antibody is designed subject to the PSSM constraints above and ligand sidechains within 10 Å of the antibody are repacked (Fig. 1, step 5). We then minimize the sidechain conformations on the ligand and antibody and assess the complex using energy and structure filters (Fig. 1, step 6).

f. Combinatorial rigid body, conformation, and sequence sampling

In step e we optimized the sequence of the representative antibodies; in this step we also sample the antibody backbone degrees of freedom from the torsion databases

computed above (step b). For each of the four antibody segments we randomly sample 50 different backbone conformations from the relevant torsion database (Fig. 1, step 7). The optimization objective function used to select the best conformation of the 50 randomly chosen conformations is specified in step g below. To improve the chances of acceptance, each sampled backbone is within a predefined sequence-length change with respect to the input conformation, ranging from ± 2 for segment types VL, VH, and L3, and ± 4 for H3. In this step *AbDesign* rigorously samples natural backbone conformations that are similar to the initial conformational representative antibody. Some of the sampled backbones vary by sub-angstrom RMSD values, thereby fine-tuning the backbone conformation.

AbDesign in effect samples combinations of naturally observed backbone conformations from a pre-computed menu of conformations, accessing an unprecedented combinatorial space of backbones for design, and addressing an important shortcoming of current design of function strategies, which have relied on a limited number of backbones (typically under 3,000).^{7,92,93} In a protein superfamily comprising m protein structures each segmented into n structural fragments, a total diversity on the order of m^n backbones could, in principle, be accessed through *AbDesign*; applied to the antibodies in our set, $m = 700$ molecular structures and $n = 4$ segments (VL, L3, VH, and H3), leading to a total space of 10^{11} different backbones. To be sure, not all resulting backbones are physically realistic, and the stability optimization of section g below tests that combinations of backbone fragments that destabilize the protein are not selected.

Changing the current segment's backbone conformation to any other conformation in the torsion database simply consists of imposing the backbone dihedral angles specified in the pre-computed database and can be done in well under a second on a standard CPU, opening the way to efficient sampling of backbone conformation space. At each backbone-sampling step we use combinatorial side-chain packing to design the sequence subject to the PSSM constraints above, and repack ligand residues within 10 Å of the antibody. We then simultaneously minimize the sidechains on the ligand-binding surface and antibody and rigid-body orientation of the antibody relative to the ligand and the antibody heavy chain relative to the light chain. We repeat this design-minimization cycle three times starting with a soft-repulsive potential and ending with the standard all-atom energy function (score12). We then use the rotamer trials-minimization procedure, whereby single sidechains are selected at random, packed, and minimized to improve sidechain packing in the antibody core and in the antibody–ligand interface.

g. Optimization of ligand binding and protein stability

A key challenge in protein design of function is that the protein needs to be both stable in its designed conforma-

tion and bind its target molecule.¹⁸ *AbDesign* implements a novel multiconstrained optimization scheme, fuzzy-logic design,⁶⁹ to select the designed backbone segments that best optimize ligand-binding energy and antibody stability. As explained in the previous step, for each of the four backbone segments (VL, L3, VH, and H3) we randomly sample 50 backbone conformations derived from that segment's torsion database (section f), compute the binding energy (E_B) and stability (E_S) of the redesigned antibody, and transform each according to the following sigmoid function:

$$f(E) = \frac{1}{1 + e^{(E-o)s}} \quad (2)$$

Where E is either the binding energy (E_B) or the energy of the unbound antibody (E_S), o is the sigmoid midpoint, where $f(E)$ assumes a value of $1/2$ and s is the steepness of the sigmoid around the midpoint. The sigmoid approaches values of 1 at low energies and 0 at high values. Before sampling conformations for each of the segments, parameter o in Eq. (1) is automatically reset to the energy value of the currently designed antibody, so both sigmoids are close to their midpoints at the start of refinement of each segment. The optimization objective function is the product of the two sigmoids: $o = f(E_S) \times f(E_B)$, resulting in values approaching 1 when both E_S and E_B are low and values approaching 0 if either one of the energy criteria is high. The effect of optimizing this objective function is to find a backbone conformation that is both sufficiently stable and high affinity. For instance, a backbone conformation that improves binding energy by 10 Rosetta energy units (R.e.u.) has a transformed sigmoid value of 0.99, and improved stability by 10 R.e.u. (transformed value of 0.97), the product ($E_S \times E_B$) equals 0.963, would be preferred to a backbone conformation that improves the binding energy by 1 R.e.u. (transformed value of 0.61) and the stability by 30 R.e.u. (transformed value 0.999, product equals 0.6) [Fig. 4(A)]. By optimizing this function during combinatorial backbone design binding energy and antibody stability improve by on average 100 R.e.u. and 5 R.e.u., respectively, relative to the starting designed scaffold antibody [Fig. 4(B,C)]; thus, combinatorial backbone sampling and fuzzy-logic design can considerably improve two of the most important parameters for design of function.

h. Structure and energy filters derived from natural antibodies

At the end of the design simulation we filter antibody structure models using four parameters: predicted binding energy, buried surface area, packing quality between the antibody's variable light and heavy domains and the bound ligand,⁵⁹ and shape complementary⁵⁸ between the antibody and bound ligand (Fig. 1, step 8). Cutoffs

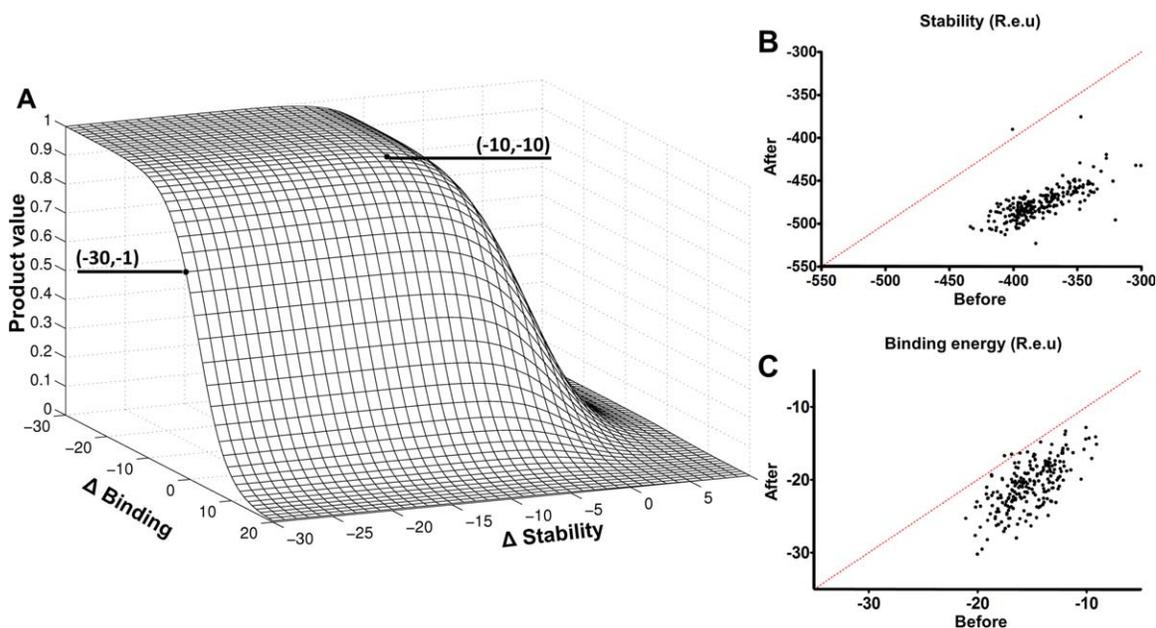


Figure 4

Antibody affinity and stability optimization using fuzzy-logic design.⁶⁹ (A) Plot of the fuzzy-logic objective function, which is the product of the stability and binding sigmoids [Eq. (2)]. A transformed value of a -10 R.e.u change in binding and stability is preferred to a -30 R.e.u change in stability and a -1 R.e.u change in binding. The product of the two transformations gauges the effect of the incorporated segment on antibody stability and target affinity relative to the baseline score (of the best-scoring antibody structure so far). (B and C). Comparison between the stability and binding energy of a set of 222 designed antibodies before and after refinement (algorithm, section f). The x axis is the calculated energy (R.e.u) of the antibody-target complex after sequence optimization (algorithm, section e) and before refinement. y axis is the designed antibody energy (R.e.u) after the backbone refinement phase (algorithm, section f). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

for each of these parameters are derived from a set of 303 natural antibody–protein complexes (Supporting Information Table SIV) extracted from the PDB⁶⁸ using the Structural Antibody Database (SabDab)⁷⁹ (Fig. 5).

Structural characteristics of designed antibodies

To test *AbDesign*'s performance and highlight areas for future improvement, we selected a set of nine high-affinity ($K_d < 20$ nM), medium-to-high crystallographic resolution (≤ 2.5 Å), protein-binding antibodies from SABDab⁷⁹ (Table III) as targets for design. For each natural antibody–protein complex we retain only the natural binding orientation, and eliminate all antibody sequence and backbone information; sidechains on the target molecule binding site are allowed to repack and minimize, in keeping with previous design of function studies.^{7,20} The natural antibody set comprises human antibodies Fab40, D5 neutralizing mAb, and BO2C11 (PDB entries: 3K2U⁹⁴ 2CMR,⁹⁵ 1IQD,⁹⁶ respectively), murine antibodies E8, D1.3 mAb, F10.6.6, JEL42, and 5E1 Fab (PDB entries: 1WEJ,⁹⁷ 1VFB,⁹⁸ 1P2C,⁹⁹ 2JEL,¹⁰⁰ 3MXW¹⁰¹), and the humanized murine antibody D3H44 (PDB entry 1JPS¹⁰²). The ligand targets comprise convex (2JEL, 1IQD), flat (1P2C), and concave (3MXW) surfaces, con-

taining helical (2CMR), sheet (1JPS), and loop (1P2C, 3K2U, 1IQD) secondary-structural elements. The conformation propensities in our dataset mirror those of protein-binding κ antibodies in the PDB (Supporting Information Table SV): eight antibodies in the set use the H1.14_H2.15 backbone conformation in the VH backbone segment (compared to 80% out of all protein-binding κ antibodies in our database), D1.3 mAb belongs to H1.14_H2.14 (14%). Antibodies E8, D1.3 mAb, 5E1, D5, D3H44, Fab40, and F10.6.6 use L1.11_L2.8 (82%), antibody JEL42 uses L1.16_L2.8 (2%) and antibody BO2C11 uses L1.12_L2.8 (3%). All antibodies in the set use the L3.10.1 conformation, which dominates natural κ light chains (86%). The H3 lengths in the set range from 7 (F10.6.6, Kabat and Chothia length^{25,26}) to 10 (Fab40 and D5) (7%, 8%, 22%, and 13%, respectively). A broader study would be needed to cover the full extent of antibody conformation diversity, particularly for H3.

We filter resulting designs using metrics developed to assess structure models, including binding energy, antibody stability, shape complementarity, packing statistics, and buried surface area. To rank the final design models we use only computed binding energy, and contrast the highest-affinity design with the target natural antibody, bound to the same epitope, according to the following structure and energy criteria: sequence identity, C α

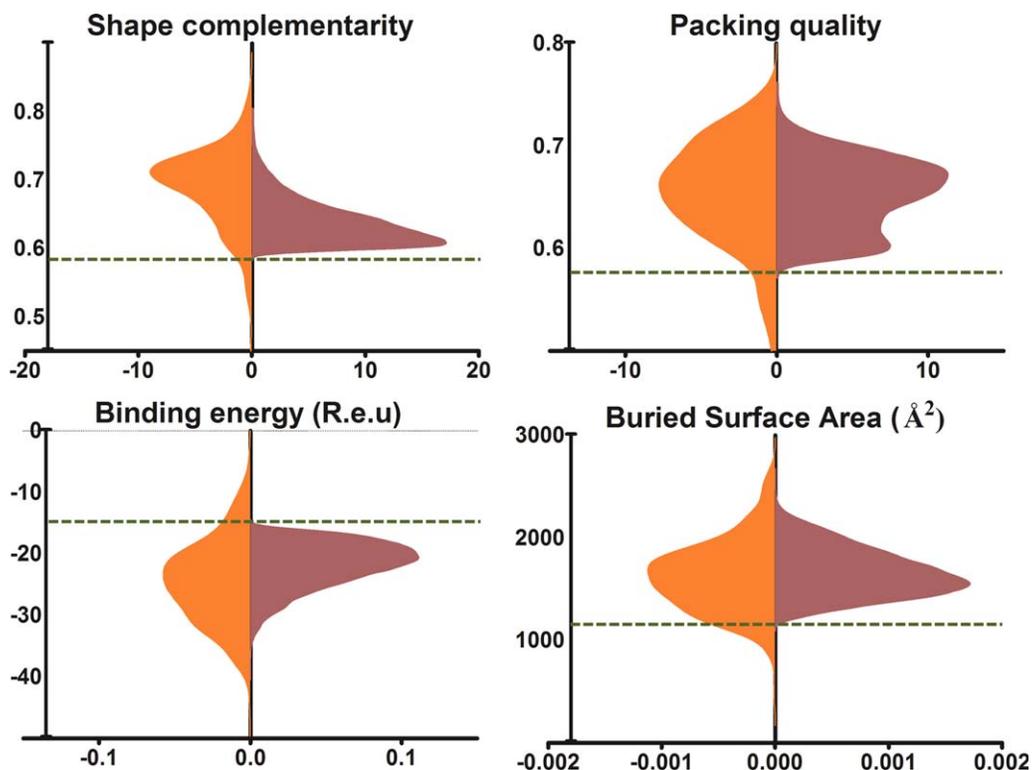


Figure 5

Energy and structure criteria used to filter designed antibody structures. In the final step of *AbDesign* we filter the designed antibodies according to four parameters: predicted binding energy, buried surface area, shape complementarity (Sc^{58}) between antibody structure and ligand, and packing quality (using RosettaHoles⁵⁹) between the variable light and heavy domains and the ligand (see Methods). Cutoffs (green dashed lines) were derived from a set of 303 natural protein-binding antibodies (Supporting Information Table SIV). Antibody designs (purple) that passed all filters are compared to the natural protein-binding antibodies (gold). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

RMSD, interface shape complementarity (Sc),⁵⁸ packing statistics,⁵⁹ buried surface area, binding energy (Fig. 5), and backbone-conformation clustering. The distributions of natural and filtered designs along the four parameters are similar, except in shape complementarity, which tends to be lower in designed complexes than in the natural ones. This is a general trend for designed complexes, which reflects the challenge of achieving the subtle steric complementarities seen in natural binders. Even though the design trajectories start from a binding orientation similar to the experimental bound complex, during design some antibodies migrate and bind at other sites. For the purposes of recapitulation analysis we eliminate designed antibodies with interface RMSD⁵⁷ values $>4 \text{ \AA}$.

Subject to the above selection criteria, six out of the nine natural antibodies in our study set select H3 and L3 CDRs of the same length as the natural antibody targets; of those six, five select fragments belonging to the same conformation clusters as those of the target natural antibody and four are at the top 10% ranking in terms of computed binding energy (Table III). Generally, designs tend to high buried surface area ($>1800 \text{ \AA}^2$), suggesting that *AbDesign* is biased toward large interfaces; with

larger binding surfaces computed binding energy rises and the number of conformations and sequences that are compatible with forming favorable inter-chain contacts drops, thereby increasing the probability of recovering the natural conformation.

To focus on the atomic details of designed antibodies we consider antibodies that target the same surface as the humanized anti-tissue factor antibody D3H44 (PDB entry 1JPS) and the anti-sonic hedgehog protein 5E1 (PDB entry 3MXW). In preliminary simulations we noticed that the tissue-factor targeting designs used backbone conformations for H3 that were derived from naturally occurring anti-tissue factor antibodies; to eliminate bias, we removed anti-tissue-factor entries from the backbone conformation dataset, and repeated the analysis. All backbone conformation segments comprising the designed antibodies belong to the same backbone conformation clusters as the experimentally determined structure of 1JPS (L1.11_L2.8, L3.10.1, H1.14_H2.15, H3.16.5) and 3MXW (H1.14_H2.15, H3.18.17, L1.11_L2.8, L3.10.1) (Fig. 6). The designs' backbone conformations show a high level of agreement with the natural antibodies (CDR $C\alpha$ RMSD between design and natural antibody is $<1 \text{ \AA}$ for all six CDRs; Table III). Previous

Table II
Structure and Energy Parameters of Natural and Designed Antibody Complexes

| PDB entry | Ligand | K_d^a (nM) | Natural antibodies | | | | | Designed antibodies | | | | | |
|-------------|------------------------------------|--------------|---|---------------------------------------|----------------------------|------------------------------------|---|---------------------------------------|----------------------------|------------------------------------|---------------------------------------|---------------------------|--|
| | | | Predicted binding energy (R.e.u) ^b | Buried surface area (Å ²) | Packing score ^b | Shape complementarity ^b | Predicted binding energy (R.e.u) ^b | Buried surface area (Å ²) | Packing score ^b | Shape complementarity ^b | Buried surface area (Å ²) | Ligand interface RMSD (Å) | Predicted binding energy rank ^c |
| 1JPS | Tissue factor | 0.1 | -25 | 1950 | 0.66 | 0.70 | -34.4 | 2171 | 0.62 | 0.59 | 2171 | 2.00 | 17 (5212) |
| 1WEJ | Cytochrome C | 15.8 | -16 | 1220 | 0.70 | 0.75 | -24.5 | 1535 | 0.67 | 0.62 | 1535 | 2.00 | 3 (93) |
| 2CMR | Transmembrane glycoprotein | 0.05 | -22 | 2110 | 0.58 | 0.72 | -26.3 | 2162 | 0.57 | 0.60 | 2162 | 1.40 | 11 (297) |
| 3MXW | Sonic hedgehog protein | 0.7 | -21 | 1882 | 0.70 | 0.51 | -32.2 | 2011 | 0.69 | 0.58 | 2011 | 2.72 | 24 (1274) |
| 1VFB | Lysozyme | 3.7 | -22 | 1405 | 0.67 | 0.69 | -24.3 | 1493 | 0.64 | 0.60 | 1493 | 3.20 | 24 (250) |
| 2JEL | Phosphocarrier protein HPr | 3.7 | -17 | 1549 | 0.66 | 0.58 | -20.4 | 1353 | 0.62 | 0.60 | 1353 | 2.70 | 9 (50) |
| 3K2U | Hepatocyte growth factor activator | 0.16 | -29.2 | 1982 | 0.62 | 0.68 | -26.6 | 1695 | 0.58 | 0.62 | 1695 | 3.20 | 51 (112) |
| 1P2C | Lysozyme | 0.098 | -17 | 1467 | 0.68 | 0.67 | -22.2 | 1566 | 0.68 | 0.60 | 1566 | 3.90 | 138 (659) |
| 1I0D | Coagulation factor VIII | 0.014 | -32 | 2134 | 0.70 | 0.78 | -24.7 | 1632 | 0.66 | 0.67 | 1632 | 2.80 | 762 (2802) |

^aObtained from the "SabDab" database.

^bSee methods for details.

^cValues in parentheses signify how many designs with $L_{RMS} < 4$ Å were computed (Supporting Information Table SVI).

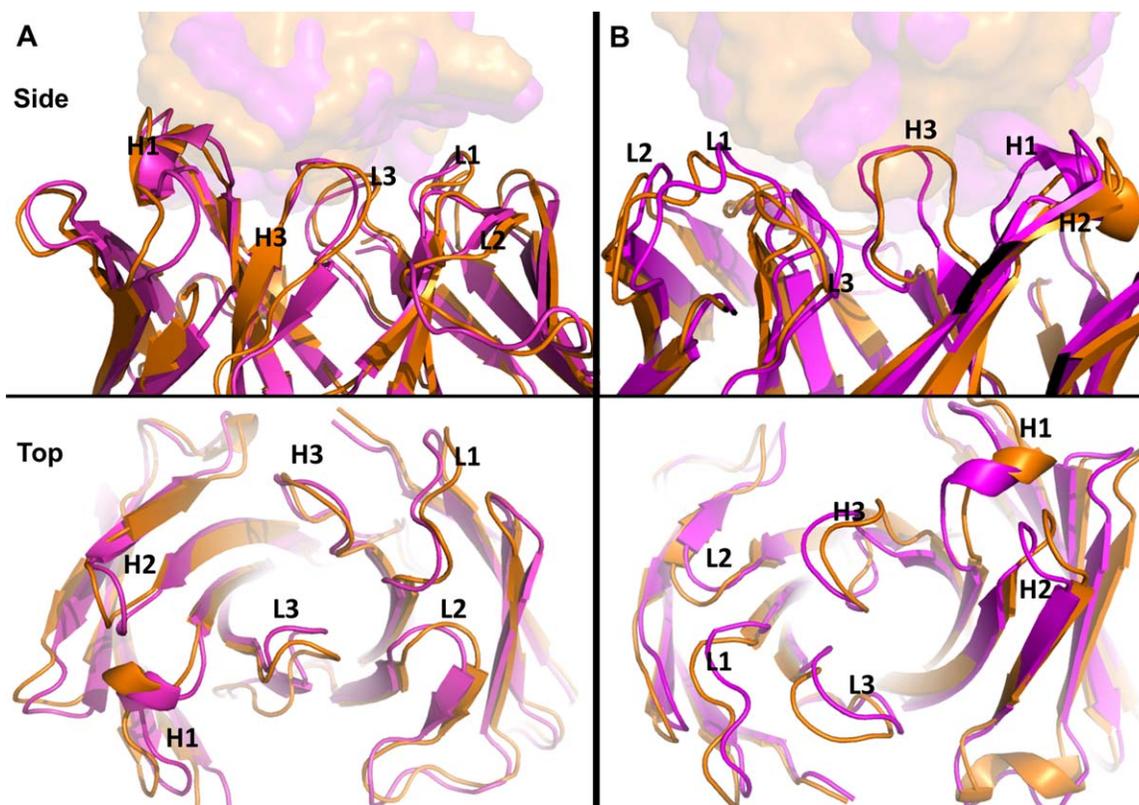


Figure 6

Antibody designs have similar backbone conformations to natural antibodies that target the same surface. Comparison between the backbone conformation of designed (magenta) and natural (orange) antibodies targeting to the same surface. (A). The anti-tissue factor protein (D3H44, PDB entry 1JPS). C_{α} RMSD between design and natural antibody is 1.1 Å and ligand interface RMSD is 2 Å. (B). The anti-sonic hedgehog protein (5E1, PDB entry 3MXW). C_{α} RMSD between design and natural antibody is 1.1 Å and ligand interface RMSD is 2.7 Å. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

studies noted that successfully *de novo* designed binding surfaces tended to be apolar and use regions high in secondary-structure content,¹⁶ raising the question whether the all-atom energy function correctly balances contributions from hydrogen bonding, solvation, and electrostatics that are crucial for designing polar surfaces.¹⁸ It is therefore encouraging that in some cases designed antibodies capture the extensive hydrogen bonding across the interface as seen for example in the highest predicted binding energy designed anti-tissue factor antibody (Fig. 7). Designed long-range interactions within the core of the variable domain, between the framework and the hypervariable CDRs, show the same characteristic hydrogen bonding, van der Waals, and aromatic stacking interactions observed in the natural antibodies from which the segment was extracted (Fig. 8). The results demonstrate that when confined to choosing from naturally existing backbones and subject to sequence constraints the all-atom energy function is capable of designing polar binding surfaces and the protein core, which provides structural stability to these surfaces.

Although segments from the nine natural antibodies were included in the conformation databases used during

design the natural backbone conformations were not selected by *AbDesign* in the majority of final high-scoring models. The designed segments are quite far from the germline compared to naturally occurring segments, ranging from 50 to 85% in designs (Table III), compared to 80–100% in natural antibodies, suggesting that design could sample parts of sequence space that are inaccessible to natural antibody diversification processes. The ability to design antibody variants using backbone-conformation segments from germline genes other than those used in natural binders suggests that the conformation data encoded in the PDB are redundant, which may be important for fine tuning the backbone to the target site.

The ability to sample and design many realistic backbone conformations can be used to highlight where design may be useful to engineering, and areas for improvement in design methodology. For some targets *AbDesign* selects models with backbone conformations similar to the natural antibody, but ranks these models poorly in comparison to others. In the case of the anti-lysozyme antibody F10.6.6 (PDB entry 1P2C) the natural

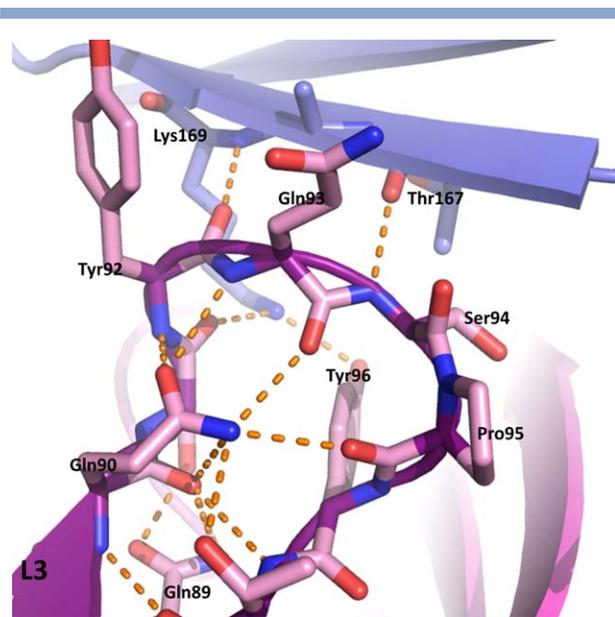


Figure 7

Designed antibody-backbone atoms form polar contacts with the ligand and supporting polar interactions within the antibody. The best predicted binding affinity design (magenta) of an anti tissue-factor antibody is shown with the target ligand (blue). Two polar contacts (dashed orange lines) are formed between the L3 Ser94 amide nitrogen and the carbonyl group of Thr167 from tissue factor and between the Tyr92 carbonyl and the amide nitrogen of Lys169 from tissue factor. The hydroxyl group of Tyr 96 forms an additional hydrogen bond with the ϵ -amino group of Lys169. In addition the conserved Gln90 forms multiple hydrogen bonds with the backbone atoms of the L3 loop that stabilize the conformation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

antibody buries a relatively small surface area and is in the 20th percentile of the overall predicted binding-energy ranking (Table III). Most of the top-ranked designs that target the same lysozyme epitope bury larger surfaces ($>1600 \text{ \AA}^2$) by using longer L1 and L3 segments [Fig. 9(a)]. These results highlight the modularity of the antibody scaffold, and a potentially useful strategy to refine existing antibodies by diversifying CDRs at the periphery of the binding site; such diversification could increase affinity and specificity for the target or increase antibody stability. Ideally, however, a design algorithm should be able to consistently predict conformations that are known to form high-affinity binding surfaces, and better methods for ranking the designed proteins should be developed to correctly identify experimentally verified binders. The anti-hepatocyte growth factor activator antibody (PDB entry 3K2U) has a binding surface area of 1980 \AA^2 , while the best-ranked similar-conformation design buries only 1700 \AA^2 (Table III). This difference in buried surface area is due to a change in the packing angle between the light and heavy variable domains of the natural and designed antibodies [Fig. 9(b)]; more

extensive sampling of the orientation of the two antibody variable domains than done here may be necessary to address such inaccuracies. The design examples studied here and provided in the Supporting Information can serve as a reference point for testing improvements in all-atom energy functions, backbone and rigid-body sampling strategies, and ranking of resulting designs.

AbDesign sequence recapitulation and interface side chain rigidity

Sequence recapitulation rates are in the range of past design studies; the values are not directly comparable, however, since past design work dealt with either functional-site design^{7,20,89} or the protein core,¹⁰⁴ whereas *AbDesign* deals with both, and since here we constrain sequence and backbone-conformation choices using data from natural antibodies, whereas past design studies used all-atom energy functions and modeled backbones without additional restraints. Sequence within the antibody core is recapitulated to within roughly 60–80% identity (Table III), which is higher than a previous study attempting to recapitulate native identities in the protein core (51%¹⁰⁴), and the binding surface sequence identity is $\sim 30\%$ (Table III), similar to a previous protein-binding study (interface residue sequence identity between 10 and 40%).⁷ Sequence identities in the CDRs (excluding interface residues) range between 50 and 90%, similar to sequence identities in the protein core; the higher recapitulation rate at noninterface CDR positions is due to the 50% higher weight on the sequence constraints imposed on noninterface CDR positions (algorithm, section a), demonstrating how *AbDesign* can be used to fine-tune the levels of divergence from the sequence consensus based on the necessities of design of function. In some cases residues at the interface and the antibody core encouragingly conserve side-chain conformations at atomic accuracy (Fig. 10).

Amino acid conformational plasticity has the potential to reduce binding specificity and affinity^{18,62} and design algorithms that rigidify sidechains at the binding surface were successful in generating the first designed protein inhibitors^{4–7} and small-molecule binders.⁹³ A computational metric to assess sidechain rigidity was suggested which computes the Boltzmann weight of the bound sidechain conformation in the ensemble of all sidechain conformations when the binder is dissociated from its target.⁶² Designed binders using existing strategies⁶² typically show lower sidechain-conformation Boltzmann weights, and presumably lower rigidity, than natural binders. Previous design attempts, which incorporated sidechain rigidity into the design scheme, have either explicitly accounted for it during design^{6,7} or have used this metric as an additional filter for evaluating designs *posteriori*.⁹³ We hypothesized that the sequence-structure rules encoded in the backbone-conformation library and

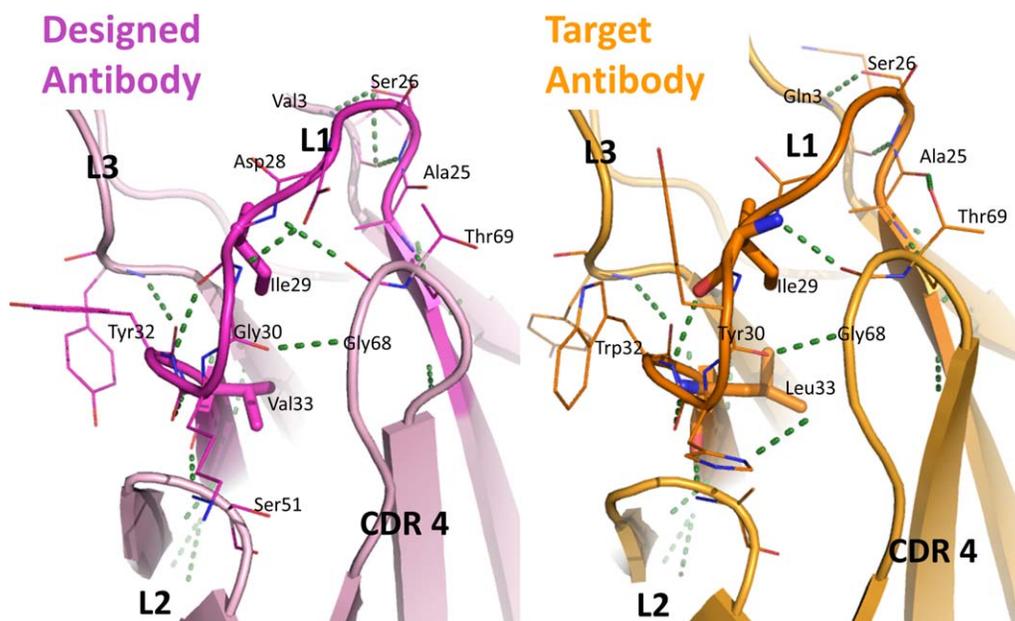


Figure 8

Designed backbone fragments conserve the stabilizing interactions observed in the natural target antibody. The natural VL segment from the target anti-gp4 antibody, D5 mAb (PDB entry 2CMR; orange), encodes long-range stabilizing interactions between CDR L1 and the framework, for instance, using hydrogen bonds (dashed green lines), and hydrophobic-packing interactions. Though the VL segment used in the design targeting gp4 originates from a different natural antibody (the anti-osteopontin antibody 23C3, PDB entry 3CXD, magenta) and has a different sequence than that of the target fragment (right), the same types of stabilizing interactions are made in the designed fragment. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the related PSSMs implicitly constrain residues in the designed antibody binding surfaces to more rigid choices. A comparison of the sidechain conformational plasticity at the binding surfaces of 303 natural high-affinity antibodies (Supporting Information Table SIV) with the designed antibodies encouragingly shows that designed aromatic residues at the binding surface that contribute more than 1 R.e.u to the predicted binding energy have conformation-probability densities somewhat higher than natural antibodies (Fig. 11). The proportion of low-probability sidechain conformations (<5% probability), which are unlikely to be in their intended conformation in the unbound state, is <10%, and more than half of the designed interface residues have sidechain-conformation probabilities above 15%, a higher fraction than in the set of natural antibodies.

The Boltzmann weight of the bound sidechain conformation is a computed metric based on sidechain-conformation libraries^{62,105} and so these results must be treated with caution in the absence of experimental structures of bound and unbound designs. Still, the high computed sidechain rigidity values suggest that by optimizing antibody stability and by biasing sequence optimization toward the antibody sequence consensus *AbDesign* may encode some elements that are necessary for lock-and-key molecular recognition.^{106–109} Two examples, the anti-tissue factor designed antibody and an

anti-sonic hedgehog protein designed antibody, demonstrate how interface sidechain rigidity is encoded by contacts between the designed aromatic sidechains and neighboring sidechain and mainchain atoms (Fig. 11).

DISCUSSION

Despite breakthroughs in the design of new molecular function in regions high in secondary-structural elements,^{2,4–6,110} successful design of function in loop segments has been elusive.^{14,17,18,76,111} *AbDesign* uses information encoded in large protein families, such as antibodies, to infer local and global sequence-structure relationships, within loops and between loops and spatially neighboring structural elements, and to define rules that guide the computational-design process. By sampling combinations of compatible backbone fragments, which have been refined by evolutionary selection, *AbDesign* accesses an unprecedentedly large space of feasible backbone conformations, enabling the design of fine shape and chemical complementarities needed for design of function.¹⁸ Although ultimate proof lies in experimental validation of designed antibodies, the design examples provided here offer promising signs that some of the current limitations in computational design may be addressed by this approach¹⁶: designed surfaces

Table III
Structure and Sequence Comparison of Top Ranked Designs and Natural Antibody Targets

| PDB entry | CDR C α RMSD (\AA) ^a | | | | Overall sequence identity (%) | Interface sequence identity ^b (%) | CDR sequence identity (%) ^c | | Core sequence identity (%) ^d | Source PDB entries ^e | | | | Sequence identity of design to segments' germline (%) | |
|-----------|---|------|------|------|-------------------------------|--|--|----|---|---------------------------------|------|------|------|---|----|
| | L1 | L2 | L3 | H1 | | | H2 | H3 | | VH | VL | VL | L3 | | VH |
| 1JPS | 0.52 | 0.18 | 0.4 | 0.67 | 63 | 36 | 75 | 63 | 67 | 1FVC | 1XGQ | 3MXW | 3HR5 | 76 | 56 |
| 1WEJ | 0.35 | 0.21 | 0.39 | 0.8 | 64 | 37 | 61 | 43 | 82 | 2WUB | 1RIV | 2WUC | 3HR5 | 69 | 64 |
| 2CMR | 0.26 | 0.31 | 0.42 | 2.04 | 64 | 26 | 61 | 68 | 75 | 3CXD | 3IFL | 3G18 | 2DBL | 58 | 70 |
| 3MXW | 0.33 | 0.19 | 0.45 | 0.65 | 58 | 53 | 42 | 54 | 60 | 1T3F | 1NBY | 2I5Y | 1ZTX | 73 | 55 |
| 1VFB | 0.4 | 0.19 | 0.42 | 0.32 | 63 | 27 | 60 | 56 | 80 | 2VDL | 4LVE | 3F00 | 3HR5 | 58 | 62 |
| 2JEL | 0.47 | 0.16 | 0.55 | 0.53 | 68 | 13 | 71 | 80 | 77 | 2IQ9 | 1CIC | 1I8K | 2HKF | 66 | 84 |
| 3K2U | 0.36 | 0.22 | 0.29 | 0.57 | 70 | 40 | 62 | 86 | 77 | 1U8M | 1K4D | 2CGR | 1UWE | 70 | 69 |
| 1P2C | 0.29 | 0.27 | 0.24 | 1.02 | 52 | 34 | 52 | 59 | 64 | 1NGW | 2F5A | 3F00 | 1FGN | 68 | 72 |
| 1I0D | 1.85 | 1.3 | 2.18 | 1.26 | 62 | 26 | 66 | 60 | 71 | 2NY7 | 1IGJ | 1UJ3 | 1FL5 | 62 | 60 |

^aDashes signify mismatch between the lengths of the designed and the natural antibody CDRs.^bOver all antibody residues within a 10 \AA distance of the ligand.^cOver CDR residues, excluding interface residues.^dAll buried non-CDR residues.^eSource PDBs from which the designed segments were derived.

comprise more polar interaction networks, loops, and larger binding regions than in previous design studies. Additionally, designed sidechains are predicted to be more rigid than in natural antibodies, whereas previous studies noted lower sidechain conformation probabilities than natural sets^{7,62}; higher rigidity could enhance affinity and specificity.

A key element of the *AbDesign* strategy is backbone segmentation along boundaries that are highly conserved in homologous structures (the disulfide-bonded cysteines and conserved positions at the end of CDR3). Existing strategies for CDR grafting, for instance in therapeutic antibody humanization, implant CDRs into the most homologous target framework, but these strategies often result in reduced binding affinity and specificity.¹¹² Our *in silico* results suggest that despite high sequence conservation in the framework, the specific stabilizing contacts formed between the CDRs and their natural frameworks are important for the structural integrity of the antibody, as noted by previous analysis.⁸⁴ Our strategy of using large backbone segments that contain the inter-molecular contacts between the framework and CDRs 1 and 2 generate antibody models with well-packed cores and high fidelity of the designed backbone for the one observed in the source antibody, features that are likely essential for the structural integrity of the designed segment and for its desired activity. Two additional elements of the *AbDesign* strategy are: first, direct coupling between sequence and conformation constraints to ensure that the designed sequence is compatible with its backbone; and second, selecting from among a large combination of conformations the backbones and sequences that simultaneously optimize both antibody stability and ligand binding. The *AbDesign* algorithm is general and could, in principle, be applied to any protein family with a sufficiently heterogeneous set of experimentally determined three-dimensional structures. For example, enzymes belonging to Rossmann fold and repeat proteins such as ankyrins share with antibodies the structural separation between a largely conserved scaffold that stabilizes the protein and a structurally diverse region (usually comprising loops, as in antibodies), where specific function is encoded²⁸; indeed, these fold families are unusually enriched for binding different molecules, suggesting that designing within these fold families could generate many desired molecular functions.

The design examples studied here show that *AbDesign* can in some cases retrieve backbone conformations and sequence elements observed in natural antibodies that target the same site. The design algorithm does not exclusively produce natural-like binders, however, and additional candidates, differing in backbone conformation, sequence, and binding mode, are suggested with equal and often improved computed affinity. These results might be due to inaccuracies in the forcefield or sampling method, or they could represent alternative

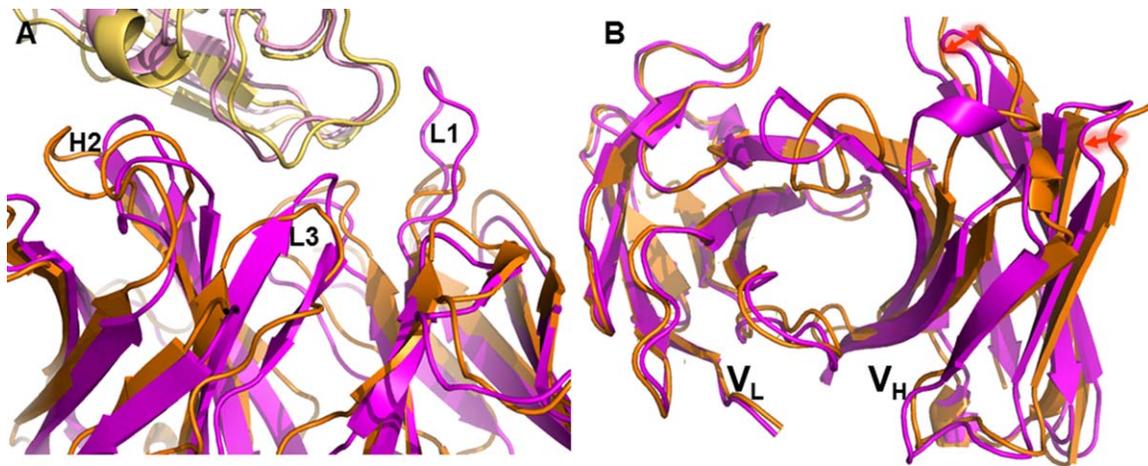


Figure 9

AbDesign favors larger binding surfaces. (A). Comparison between the top-ranked anti-lysozyme design (magenta) and the natural antibody, F10.6.6, PDB entry 1P2C (gold). The designed antibody uses a longer L1 (16 amino acid, compared to 11 in the natural antibody) and a longer L3 (11 amino acids compared to 10), increasing the buried surface area from 1470 to 1680 Å². (B) Comparison between the anti-hepatocyte growth factor activator designed antibody (magenta) and the natural antibody, Fab40, PDB entry 3K2U (orange). Structures are oriented so CDRs are pointing toward the viewer. A 10° difference in the packing angle between the variable light and heavy domains creates a gap between the CDRs of the natural antibody's variable light and heavy domains compared to the designed one (marked by red arrows). This opening in the light and heavy domain interface produces a larger binding surface in the natural antibody compared to the design. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

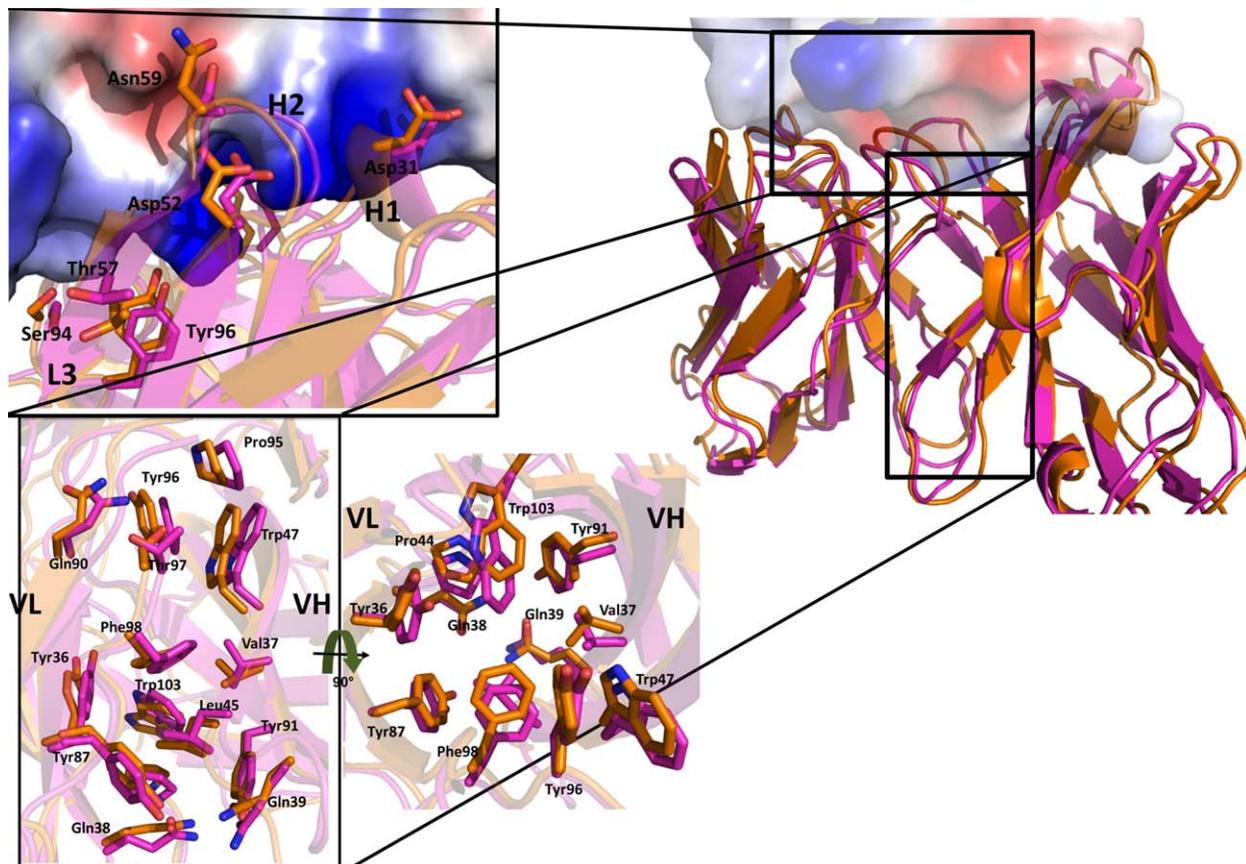
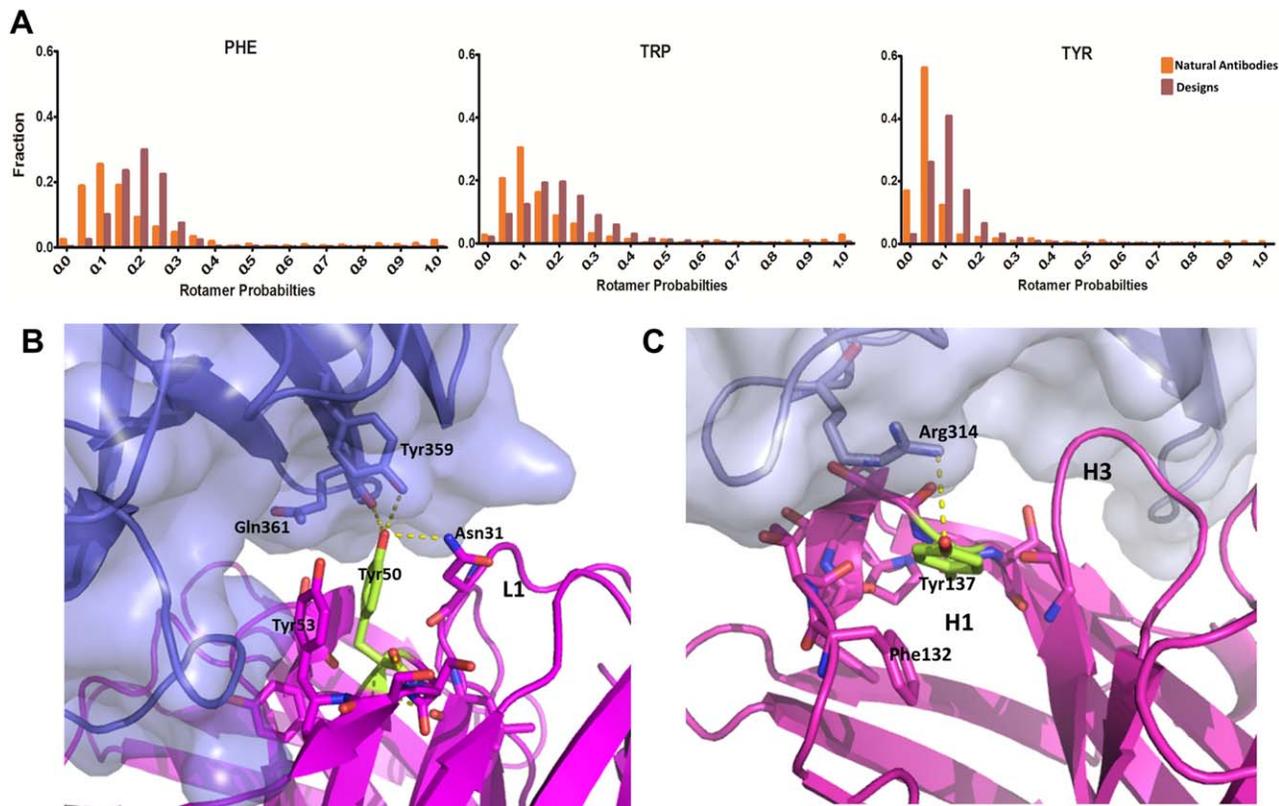


Figure 10

Designed antibodies recapitulate the identity and conformation of binding surface and core residues. The anti-tissue factor targeting design (magenta). Tissue factor is shown in surface representation colored by vacuum electrostatics using Pymol (69). The natural antibody D3H44 (PDB entry 1JPS) is colored orange. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

**Figure 11**

High conformational rigidity of designed sidechains. (A). The sidechain conformation probabilities in the unbound state were computed using the method in Ref. 62. *AbDesign* produces antibody complexes with a lower proportion of low-probability conformations (≤ 0.05 probability) compared to natural antibody complexes. The natural antibody complex set comprises 303 antibody–protein complexes (Supporting Information Table SIV) extracted from the SabDab database,⁷⁹ and the designed antibody set includes all designs generated and filtered by the design protocol. (B) The designed antibody against sonic hedgehog protein. The conformationally constrained tyrosine (colored green, with rotamer Boltzmann probability 90%) is stabilized by packing against surrounding backbone atoms and the side chain atoms of Tyr53 and a hydrogen bond with Asn31. (C). The anti-tissue factor protein designed antibody. Tyr137 on H1 (rotamer Boltzmann probability: 60%) is stabilized by packing against the backbone atoms of H1 and H3 and the side chain of Phe132. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

solutions to binding the target epitope; indeed, different natural antibodies are known to bind the same epitope.¹¹³ In particular, our results on the anti-lysozyme antibody suggest that *AbDesign* could propose antibodies that share large regions with natural antibodies, but that form additional interactions to those observed in the natural antibodies, highlighting the versatility of the antibody scaffold; these additional interactions could increase specificity and affinity. *AbDesign* may therefore be used to suggest variants of natural antibody binders for experimental selection of higher-affinity, higher-specificity, or higher-stability antibodies, and in the future may enable designing antibodies completely from scratch.

By sampling many different backbone combinations *AbDesign* allows us to highlight important areas for improvement in design methodology. It is encouraging that the *AbDesign* strategy is able to recapitulate sequence and structure features seen in naturally occurring polar and large binding surfaces, whereas previous design anal-

yses noted biases toward hydrophobic and small surfaces.¹⁶ A difference between previous design algorithms and the one reported here is that *AbDesign* restricts sampling to a choice between physically realistic backbone conformations and to the natural amino acid combinations that are compatible with these backbone conformations; the results suggest that when confined to such discrete choices—albeit to a very large space of such choices—the all-atom energy function can reproduce polar surfaces seen in natural binders, and often ranks them highly. *AbDesign* is generally biased toward designs with large binding surfaces ($>1800 \text{ \AA}^2$, Table III), reflecting the correlation between buried surface area and binding affinity in natural protein–protein interactions.¹¹⁴ The antibodies with small binding surfaces selected in our study nevertheless have high experimentally determined affinity for their targets, such as in the case of the anti-cytochrome c, E8 antibody (PDB entry 1WEJ) with buried surface area upon binding of 1200 \AA^2 and K_D of

16 nM.⁹⁷ Despite the natural antibody's high affinity for its target, *AbDesign* prefers antibodies that bury 1500 Å² of surface area upon binding. These results highlight the importance of developing metrics to rank designs in addition to stability and binding energy. Indeed, previous design of function applications^{4,5,110} and our study relied on structure selection criteria, such as intermolecular shape complementarity⁵⁸ and packing defects.⁵⁹ We expect that additional filters that address the geometry of hydrogen bonding and the ability of water molecules to be bound and stabilized at the binding interface may make important additional contributions to appropriate selection and ranking of design models. An advantage of design within a large family of proteins, such as antibodies, is the availability of a large set of experimentally determined structures of natural exemplars with which to test different metrics, and the current results can provide a useful reference point for testing improved metrics to accurately rank the propensity of design models to bind their intended targets. We also note that more extensive sampling of the rigid-body orientation between the antibody light and heavy domains than done here may improve the accuracy of the design calculations. The *AbDesign* method is the first, to our knowledge, to combine backbone, protein-core, and functional-site design, and could be used to test and refine molecular forcefields.^{115,116}

ACKNOWLEDGMENTS

The authors thank all members of the Fleishman laboratory who read and commented on this manuscript, and Eva-Maria Strauch, Dan Tawfik, Meir Wilchek, and an anonymous reviewer for helpful comments.

REFERENCES

- Huang PS, Love JJ, Mayo SL. A de novo designed protein protein interface. *Protein Sci* 2007;16:2770–2774.
- Jha RK, Leaver-Fay A, Yin S, Wu Y, Butterfoss GL, Szyperski T, V Dokholyan N, Kuhlman B. Computational design of a PAK1 binding protein. *J Mol Biol* 2010;400:257–270.
- Karanicolas J, Corn JE, Chen I, Joachimiak LA, Dym O, Peck SH, Albeck S, Unger T, Hu W, Liu G, Delbecq S, Montelione GT, Spiegel CP, Liu DR, Baker D. A de novo protein binding pair by computational design and directed evolution. *Mol Cell* 2011;42:250–260.
- Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch EM, Wilson IA, and Baker D. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 2011;332:816–821.
- Strauch EM, Fleishman SJ, Baker D. Computational design of a pH-sensitive IgG binding protein. *Proc Natl Acad Sci USA* 2014;111:675–680.
- Procko E, Hedman R, Hamilton K, Seetharaman J, Fleishman SJ, Su M, Aramini J, Kornhaber G, Hunt JF, Tong L, Montelione GT, Baker D. Computational design of a protein-based enzyme inhibitor. *J Mol Biol* 2013;425:3563–3575.
- Fleishman SJ, Corn JE, Strauch E-M, a Whitehead T, Karanicolas J, Baker D. Hotspot-centric de novo design of protein binders. *J Mol Biol* 2011;413:1047–1062.
- King NP, Sheffler W, Sawaya MR, Vollmar BS, Sumida JP, André I, Gonen T, Yeates TO, Baker D. Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* 2012;336:1171–1174.
- Gradišar H, Božić S, Doles T, Vengust D, Hafner-Bratkovic I, Mertelj A, Webb B, Šali A, Klavžar S, Jerala R. Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat Chem Biol* 2013;9:362–366.
- Fletcher JM, Harniman RL, Barnes FRH, Boyle AL, Collins A, Mantell J, Sharp TH, Antognozzi M, Booth PJ, Linden N, Miles MJ, Sessions RB, Verkade P, Woolfson DN. Self-assembling cages from coiled-coil peptide modules. *Science* 2013;340:595–599.
- Stranges PB, Machius M, Miley MJ, Tripathy A, Kuhlman B. Computational design of a symmetric homodimer using β -strand assembly. *Proc Natl Acad Sci USA* 2011;108:20562–20567.
- Der BS, Machius M, Miley MJ, Mills JL, Szyperski T, Kuhlman B. Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J Am Chem Soc* 2012;134:375–385.
- Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J Mol Biol* 1999;285:2177–2198.
- Fleishman SJ, Whitehead TA, Strauch E-M, Corn JE, Qin S, Zhou H-X, Mitchell JC, a Demerdash ON, Takeda-Shitaka M, Terashi G, Moal IH, Li X, a Bates P, Zacharias M, Park H, Ko J, Lee H, Seok C, Bourquard T, Bernauer J, Poupon A, Azé J, Soner S, Ovali SK, Ozbek P, Ben Tal N, Haliloglu T, Hwang H, Vreven T, Pierce BG, Weng Z, Pérez-Cano L, Pons C, Fernández-Recio J, Jiang F, Yang F, Gong X, Cao L, Xu X, Liu B, Wang P, Li C, Wang C, Robert CH, Guharoy M, Liu S, Huang Y, Li L, Guo D, Chen Y, Xiao Y, London N, Itzhaki Z, Schueler-Furman O, Inbar Y, Potapov V, Cohen M, Schreiber G, Tsuchiya Y, Kanamori E, Standley DM, Nakamura H, Kinoshita K, Driggers CM, Hall RG, Morgan JL, Hsu VL, Zhan J, Yang Y, Zhou Y, Kastiris PL, Bonvin AMJJ, Zhang W, Camacho CJ, Kilambi KP, Sircar A, Gray JJ, Ohue M, Uchikoga N, Matsuzaki Y, Ishida T, Akiyama Y, Khashan R, Bush S, Fouches D, Tropsha A, Esquivel-Rodríguez J, Kihara D, Stranges PB, Jacak R, Kuhlman B, Huang S-Y, Zou X, Wodak SJ, Janin J, Baker D. Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *J Mol Biol* 2011;414:289–302.
- Moretti R, Fleishman SJ, Agius R, Torchala M, a Bates P, Kastiris PL, Rodrigues JPGLM, Trellet M, Bonvin AMJJ, Cui M, Rooman M, Gillis D, Dehouck Y, Moal I, Romero-Durana M, Perez-Cano L, Pallara C, Jimenez B, Fernandez-Recio J, Flores S, Pacella M, Praneeth Kilambi K, Gray JJ, Popov P, Grudinin S, Esquivel-Rodríguez J, Kihara D, Zhao N, Korkin D, Zhu X, a Demerdash ON, Mitchell JC, Kanamori E, Tsuchiya Y, Nakamura H, Lee H, Park H, Seok C, Sarmiento J, Liang S, Teraguchi S, Standley DM, Shimoyama H, Terashi G, Takeda-Shitaka M, Iwadata M, Umeyama H, Beglov D, Hall DR, Kozakov D, Vajda S, Pierce BG, Hwang H, Vreven T, Weng Z, Huang Y, Li H, Yang X, Ji X, Liu S, Xiao Y, Zacharias M, Qin S, Zhou H-X, Huang S-Y, Zou X, Velankar S, Janin J, Wodak SJ, Baker D. Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions. *Proteins* 2013;81:1980–1987.
- Stranges PB, Kuhlman B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci* 2013;22:74–82.
- Khare SD, Fleishman SJ. Emerging themes in the computational design of novel enzymes and protein-protein interfaces. *FEBS Lett* 2013;587:1147–1154.
- Fleishman SJ, Baker D. Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* 2012;149:262–273.
- Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. Design of a novel globular protein fold with atomic-level accuracy. *Science* 2003;302:1364–1368.
- Zanghellini A, Jiang L, Wollacott AM, Cheng G, Meiler J, Althoff EA, Röhthlisberger D, Baker D. New algorithms and an in silico

- benchmark for computational enzyme design. *Protein Sci* 2006;15:2785–2794.
21. Wu, T, Te Kabat, E. An analysis of the sequences of the variable regions of bence jones proteins and myeloma light chains and their implications for antibody complementarity. *J Exp Med* 1970;132:211–250.
 22. Chothia C, Lesk AM. Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* 1987;196:901–917.
 23. Strausbauch PH, Weinstein Y, Wilchek M, Shaltiel S, Givol D. A homologous series of affinity labeling reagents and their use in the study of antibody binding sites. *Biochemistry* 1971;10:4342–4348.
 24. North B, Lehmann A, Dunbrack RL. A new clustering of antibody CDR loop conformations. *J Mol Biol* 2011;406:228–256.
 25. Al-Lazikani B, Lesk AM, Chothia C. Standard conformations for the canonical structures of immunoglobulins. *J Mol Biol* 1997;273:927–948.
 26. Kabat EA, Wu TT, Perry HM, Gottesman KS, Foeller C. Sequences of proteins of immunological interest, 5th ed. Bethesda, Maryland: National Institutes of Health; 1991.
 27. Kabat EA, Wu TT, Bilofsky H. Unusual distributions of amino acids in complementarity-determining (hypervariable) segments of heavy and light chains of immunoglobulins and their possible roles in specificity of antibody-combining sites. *J Biol Chem* 1977;252:6609–6616.
 28. Dellus-Gur E, Toth-Petroczy A, Elias M, Tawfik DS. What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. *J Mol Biol* 2013;425:2609–2621.
 29. Lesk A, Chothia C. Evolution of proteins formed by β -sheets: II. The core of the immunoglobulin domains. *J Mol Biol* 1982;160:325–342.
 30. Riechmann L, Clark M, Waldmann H, Winter G. Reshaping human antibodies for therapy. *Nature* 1988;332:323–327.
 31. Jones P, Dear P, Foote J, Neuberger M, Winter G. Replacing the complementarity-determining regions in a human antibody with those from a mouse. *Nature* 1986;321:522–525.
 32. Winter G, Milstein C. Man-made antibodies. *Nature* 1991;349:293–299.
 33. Michnick SW, Sidhu SS. Submitting antibodies to binding arbitration. *Nat Chem Biol* 2008;4:326–329.
 34. Beck A, Wurch T, Bailly C, Corvaia N. Strategies and challenges for the next generation of therapeutic antibodies. *Nat Rev Immunol* 2010;10:345–352.
 35. Filpula D. Antibody engineering and modification technologies. *Biomol Eng* 2007;24:201–215.
 36. Scott AM, Wolchok JD, Old LJ. Antibody therapy of cancer. *Nat Rev Cancer* 2012;12:278–287.
 37. Ekiert DC, Wilson IA. Broadly neutralizing antibodies against influenza virus and prospects for universal therapies. *Curr Opin Virol* 2012;2:134–141.
 38. Glennie MJ, Johnson PW. Clinical trials of antibody therapy. *Immunol Today* 2000;21:403–410.
 39. O’Nuallain B, Wetzel R. Conformational abs recognizing a generic amyloid fibril epitope. *Proc Natl Acad Sci USA* 2002;99:1485–1490.
 40. Perchiacca JM, Ladiwala ARA, Bhattacharya M, Tessier PM. Structure-based design of conformation- and sequence-specific antibodies against amyloid β . *Proc Natl Acad Sci USA* 2012;109:84–89.
 41. Schwarz M, Röttgen P, Takada Y, Le Gall F, Knackmuss S, Bassler N, Büttner C, Little M, Bode C, Peter K. Single-chain antibodies for the conformation-specific blockade of activated platelet integrin α IIb β 3 designed by subtractive selection from naive human phage libraries. *FASEB J* 2004;18:1704–1706.
 42. Ofek G, Guenaga FJ, Schief WR, Skinner J, Baker D, Wyatt R, Kwong PD. Elicitation of structure-specific antibodies by epitope scaffolds. *Proc Natl Acad Sci USA* 2010;107:17880–17887.
 43. McLellan JS, Chen M, Leung S, Graepel KW, Du X, Yang Y, Zhou T, Baxa U, Yasuda E, Beaumont T, Kumar A, Modjarrad K, Zheng Z, Zhao M, Xia N, Kwong PD, Graham BS. Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* 2013;340:1113–1117.
 44. Throsby M, van den Brink E, Jongeneelen M, Poon LLM, Alard P, Cornelissen L, Bakker A, Cox F, van Deventer E, Guan Y, Cinatl J, ter Meulen J, Lasters I, Carsetti R, Peiris M, de Kruif J, Goudsmit J. Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM+ memory B cells. *PLoS One* 2008;3:e3942.
 45. Fleishman SJ, Leaver-Fay A, Corn JE, Strauch E-M, Khare SD, Koga N, Ashworth J, Murphy P, Richter F, Lemmon G, Meiler J, Baker D. Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design. *Protein Sci* 2006;15:949–960.
 46. Lippow SM, Wittrup KD, Tidor B. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat Biotechnol* 2007;25:1171–1176.
 47. Clark L, Boriack-Sjodin PA, Day E, Eldredge J, Fitch C, Jarpe M, Miller S, Li Y, Simon K, van Vlijmen HWT. An antibody loop replacement design feasibility study and a loop-swapped dimer structure. *Protein Eng Des Sel* 2009;22:93–101.
 48. Barderas R, Desmet J, Timmerman P, Meloen R, Casal JI. Affinity maturation of antibodies assisted by in silico modeling. *Proc Natl Acad Sci USA* 2008;105:9029–9034.
 49. Farady CJ, Sellers BD, Jacobson MP, Craik CS. Improving the species cross-reactivity of an antibody using computational design. *Bioorg Med Chem Lett* 2009;19:3744–3747.
 50. Miklos AE, Kluwe C, Der BS, Pai S, Sircar A, a Hughes R, Berrondo M, Xu J, Codrea V, Buckley PE, Calm AM, Welsh HS, Warner CR, a Zacharko M, Carney JP, Gray JJ, Georgiou G, Kuhlman B, Ellington AD. Structure-based design of supercharged, highly thermoresistant antibodies. *Chem Biol* 2012;19:449–455.
 51. Pantazes RJ, Maranas CD. OptCDR: a general computational method for the design of antibody complementarity determining regions for targeted epitope binding. *Protein Eng Des Sel* 2010;23:849–858.
 52. Pantazes RJ, Maranas CD. MAPs: a database of modular antibody parts for predicting tertiary structures and designing affinity matured antibodies. *BMC Bioinform* 2013;14:168.
 53. Li T, Pantazes RJ, Maranas CD. OptMAVEN—a new framework for the de novo design of antibody variable region models targeting specific antigen epitopes. *PLoS One* 2014;9:e105954.
 54. Das R, Baker D. Macromolecular modeling with Rosetta. *Annu Rev Biochem* 2008;77:363–382.
 55. Fleishman SJ, Leaver-Fay A, Corn JE, Strauch E-M, Khare SD, Koga N, Ashworth J, Murphy P, Richter F, Lemmon G, Meiler J, Baker D. RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* 2011;6:e20161.
 56. Whitehead TA, Baker D, Fleishman SJ. Computational design of novel protein binders and experimental affinity maturation. *Methods Enzymol* 2013;523:1–19.
 57. Méndez R, Leplae R, De Maria L, Wodak SJ. Assessment of blind predictions of protein–protein interactions: current status of docking methods. *Proteins* 2003;52:51–67.
 58. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. *J Mol Biol* 1993;234:946–950.
 59. Sheffler W, Baker D. RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci* 2009;18:229–239.
 60. DeLano WL. The PyMOL Molecular Graphics System. Palo Alto, CA: DeLano Scientific LLC; 2008.
 61. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J Mol Biol* 2003;331:281–299.
 62. Fleishman SJ, Khare SD, Koga N, Baker D. Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Sci* 2011;20:753–757.
 63. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein–protein complexes. *Proc Natl Acad Sci USA* 2002;99:14116–14121.

64. Shirai H, Kidera A, Nakamura H. H3-rules: identification of CDR-H3 structures in antibodies. *FEBS Lett* 1999;455:188–197.
65. Chothia C, Lesk A, Tramontano A. Conformations of immunoglobulin hypervariable regions. *Nature* 1989.
66. Alexander N, Woetzel N, Meiler J. Bcl.: Cluster: a method for clustering biological molecules coupled with visualization in the Pymol Molecular Graphics System. *Comput Adv Biol Med Sci (ICCABS)*, 2011 IEEE 1st Int. Conf. 2011;13–18.
67. Biegert A, Söding J. Sequence context-specific profiles for homology searching. *Proc Natl Acad Sci USA* 2009;106:3770–3775.
68. Bernstein FC, Koetzle TF, Williams GJ, Meyer EE, Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi MT. The protein data bank: a computer-based archival file for macromolecular structures. *J Mol Biol* 1977;112:535.
69. Warszawski S, Netzer R, Tawfik DS, Fleishman SJ. A “fuzzy”-logic language for encoding multiple physical traits in biomolecules. *J Mol Biol* 2014;426:4125–4138.
70. Berezovsky IN, Zeldovich KB, Shakhnovich EI. Positive and negative design in stability and thermal adaptation of natural proteins. *PLoS Comput Biol* 2007;3:e52.
71. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. *BMC Bioinform* 2009;10:421.
72. Mandell D, Coutsiaris E, Kortemme T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat Methods* 2009;6:551–552.
73. Smith CA, Kortemme T. Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *J Mol Biol* 2008;380:742–756.
74. Canutescu AA, Jr, Dunbrack RL. Cyclic coordinate descent: a robotics algorithm for protein loop closure. *Protein Sci* 2003;12:963–972.
75. Tyka MD, Keedy DA, André I, DiMaio F, Song Y, Richardson DC, Richardson JS, Baker D. Alternate states of proteins revealed by detailed energy landscape mapping. *J Mol Biol* 2011;405:607–618.
76. Hu X, Wang H, Ke H, Kuhlman B. High-resolution design of a protein loop. *Proc Natl Acad Sci USA* 2007;104:17668–17673.
77. Richter F, Blomberg R, Khare SD, Kiss G, Kuzin AP, Smith AJT, Gallaher J, Pianowski Z, Helgeson RC, Grjasnow A, Xiao R, Seetharaman J, Su M, Vorobiev S, Lew S, Forouhar F, Kornhaber GJ, Hunt JE, Montelione GT, Tong L, Houk KN, Hilvert D, Baker D. Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *J Am Chem Soc* 2012;134:16197–16206.
78. Midelfort KS, Hernandez HH, Lippow SM, Tidor B, Drennan CL, Witttrup KD. Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. *J Mol Biol* 2004;343:685–701.
79. Dunbar J, Krawczyk K, Leem J, Baker T, Fuchs A, Georges G, Shi J, Deane CM. SAbDab: the structural antibody database. *Nucleic Acids Res* 2014;42:D1140–D1146.
80. Padlan EA. Structural basis for the specificity of antibody–antigen reactions and structural mechanisms for the diversification of antigen-binding specificities. *Q Rev Biophys* 1977;10:35–65.
81. Singer I, et al. Optimal humanization of 1B4, an anti-CD18 murine monoclonal antibody, is achieved by correct choice of human V-region framework sequences. *J Immunol* 1993;150:2844–2857.
82. Padlan E. A possible procedure for reducing the immunogenicity of antibody variable domains while preserving their ligand-binding properties. *Mol Immunol* 1991;28:489–498.
83. Tramontano A, Chothia C, Lesk AM. Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. *J Mol Biol* 1990;215:175–182.
84. Foote J, Winter G. Antibody framework residues affecting the conformation of the hypervariable loops. *J Mol Biol* 1992;224:487–499.
85. Shirai H, Kidera A, Nakamura H. Structural classification of CDR-H3 in antibodies. *FEBS Lett* 1996;399:1–8.
86. Chaudhury S, Berrondo M, Weitzner BD, Muthu P, Bergman H, Gray JJ. Benchmarking and analysis of protein docking performance in rosetta v3.2. *PLoS One* 2011;6:e22477.
87. Allison B, et al. Computational design of protein-small molecule interfaces. *J Struct Biol* 2014;185:193–202.
88. Burton DR, Poignard P, Stanfield RL, Wilson IA. Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* 2012;337:183–186.
89. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. Patch-Dock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res* 2005;33:W363–W367.
90. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL, Houk KN, Michael FE, Baker D. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-alder reaction. *Science* 2010;329:309–313.
91. Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL, Baker D. Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* 2013;501:212–216.
92. Ganesan R, Eigenbrot C, Wu Y, Liang W-C, Shia S, Lipari MT, Kirchhofer D. Unraveling the allosteric mechanism of serine protease inhibition by an antibody. *Structure* 2009;17:1614–1624.
93. Luftig MA, Mattu M, Di Giovine P, Geleziunas R, Hrin R, Barbato G, Bianchi E, Miller MD, Pessi A, Carfi A. Structural basis for HIV-1 neutralization by a gp41 fusion intermediate-directed antibody. *Nat Struct Mol Biol* 2006;13:740–747.
94. Spiegel PC. Structure of a factor VIII C2 domain-immunoglobulin G4kappa fab complex: identification of an inhibitory antibody epitope on the surface of factor VIII. *Blood* 2001;98:13–19.
95. Mylvaganam SE, Paterson Y, Getzoff ED. Structural basis for the binding of an anti-cytochrome c antibody to its antigen: crystal structures of FabE8–cytochrome c complex to 1.8 a resolution and FabE8 to 2.26 a resolution. *J Mol Biol* 1998;281:301–322.
96. Bhat TN, Bentley GA, Boulot G, Greene MI, Tello D, Dall’Acqua W, Souchon H, Schwarz FP, Mariuzza RA, Poljak RJ. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc Natl Acad Sci USA* 1994;91:1089–1093.
97. Cauerhff A, Goldbaum FA, Braden BC. Structural mechanism for affinity maturation of an anti-lysozyme antibody. *Proc Natl Acad Sci USA* 2004;101:3539–3544.
98. Prasad L, Waygood EB, Lee JS, Delbaere LT. The 2.5 a resolution structure of the jel42 fab fragment/HPc complex. *J Mol Biol* 1998;280:829–845.
99. Maun HR, Wen X, Lingel A, de Sauvage FJ, a Lazarus R, Scales SJ, Hymowitz SG. Hedgehog pathway antagonist 5E1 binds hedgehog at the pseudo-active site. *J Biol Chem* 2010;285:26570–26580.
100. Faelber K, Kirchhofer D, Presta L, Kelley RF, Muller YA. The 1.85 a resolution crystal structures of tissue factor in complex with humanized fab D3h44 and of free humanized fab D3h44: revisiting the solvation of antigen combining sites. *J Mol Biol* 2001;313:83–97.
101. Kuhlman B, Baker D. Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci USA* 2000;97:10383–10388.
102. Dunbrack RL, Karplus M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat Struct Mol Biol* 1994;1:334–340.
103. Yin J, et al. A comparative analysis of the immunological evolution of antibody 28B4. *Biochemistry* 2001;40:10764–10773.
104. Sagawa T, Oda M, Ishimura M, Furukawa K, Azuma T. Thermodynamic and kinetic aspects of antibody evolution during the immune response to hapten. *Mol Immunol* 2003;39:801–808.
105. Manivel V, Sahoo NC, Salunke DM, Rao KV. Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site. *Immunity* 2000;13:611–620.
106. Yin J, Beuscher AE, Andryski SE, Stevens RC, Schultz PG. Structural plasticity and the evolution of antibody affinity and specificity. *J Mol Biol* 2003;330:651–656.

107. Procko E, Berguig GY, Shen BW, Song Y, Frayo S, Convertine AJ, Margineantu D, Booth G, Correia BE, Cheng Y, Schief WR, Hockenbery DM, Press OW, Stoddard BL, Stayton PS, Baker D. A computationally designed inhibitor of an Epstein–Barr viral Bcl-2 protein induces apoptosis in infected cells. *Cell* 2014;157:1644–1656.
108. Mandell DJ, Kortemme T. Computer-aided design of functional protein interactions. *Nat Chem Biol* 2009;5:797–807.
109. Kettleborough CA, Saldanha J, Heath VJ, Morrison CJ, Bendig MM. Humanization of a mouse monoclonal antibody by CDR-grafting: The importance of framework residues on loop conformation. *Protein Eng* 1991;4:773–783.
110. Pons J, Stratton JR, Kirsch JF. How do two unrelated antibodies, HyHEL-10 and F9.13.7, recognize the same epitope of hen egg-white lysozyme? *Protein Sci* 2002;11:2308–2315.
111. Chen J, Sawyer N, Regan L. Protein–protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci* 2013;22:510–515.
112. Leaver-Fay A, O’Meara MJ, Tyka M, Jacak R, Song Y, Kellogg EH, Thompson J, Davis IW, a Pache R, Lyskov S, Gray JJ, Kortemme T, Richardson JS, Havranek JJ, Snoeyink J, Baker D, Kuhlman B. Scientific benchmarks for guiding macromolecular energy function improvement. Vol. 523; 2013. pp. 109–143.
113. Song Y, Tyka M, Leaver-Fay A, Thompson J, Baker D. Structure-guided forcefield optimization. *Proteins* 2011;79:1898–1909.