Thesis for the degree
Master of Science

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

עבודת גמר (תזה) לתואר
מוסמך למדעים

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

By
Dror Baran

מאת
דרור ברן

עיצוב ממוחשב של נוגדנים הקושרים אינסולין

Computational design of de-novo insulin binding antibodies

Advisor:
Sarel Fleishman

מנחה:
שראל פליישמן

September 2014

כסלו תשע״ה

Abstract

Antibodies are the most versatile system for molecular recognition known, able to bind a bewildering range of target molecules with high affinity and specificity. Although vast, antibody repertoires are highly biased and redundant due to evolutionary selection for tolerance to self-epitopes, thus limiting the antibody response towards many drugs, diagnostic targets, and antigens. Due to these features, antibodies are desirable targets for protein design. To date, methodologies for protein engineering of antibodies are restricted to improvement of binding affinity or stabilization of existing antibodies and the ability to generate *de novo* antibodies for a specific target still remains elusive. In this research, I describe a novel strategy, combining iterative computational *de novo* design algorithm development and high throughput experimental screening. Implementing such a `learning loop` strategy enabled us to raise a hypothesis, test it experimentally by high throughput screening, and gain insight that was later employed in further development of the algorithm, enabling us to develop an innovative algorithm for designing antibodies from first principle. Here, I present the status quo of the ongoing algorithm development, the generation of *de novo* design of antibodies targeting insulin, the experimental evaluation and *in vitro* evolution of these antibodies, yielding a high affinity insulin binder. These experimental results demonstrate the potential of our algorithm to generate *de novo* antibodies against, in principle, any target molecule, potentially revolutionizing the antibody engineering discipline and yielding novel diagnostics, therapeutics, and molecular probes.

Table of Contents

List of Abbreviations

VH -	Variable heavy domain
VL -	Variable light domain
CDR -	Complementarity-determining region
YSD -	Yeast surface display
PDB -	Protein data bank
ScFv -	Single-chain variable fragment
Fab -	Fragment antigen-binding
PSSM -	Position-specific scoring matrix
APBS -	Adaptive Poisson-Boltzmann solver
PCR -	Polymerase chain reaction
W.T. -	Wild type
epPCR -	Error prone polymerase chain reaction
K_d -	Equilibrium dissociation constant
HA -	Influenza hemagglutinin
ACP -	Apo *Mycobacterium tuberculosis* acyl carrier protein 2
IGF1 -	Insulin growth factor1
IGF2 -	Insulin growth factor2
R.e.u -	Rosetta energy units
Sasa -	Solvent-accessible surface area

Introduction

Molecular recognition is the most fundamental to many biological processes [1] which is mastered by antibodies, earning them the title "magic bullets" by Paul Ehrlich over 100 years ago [2] due to their vast repertoire, high affinity and superb binding specificity. The justification for such a reputation is attributed, among others, to their vast variability, exceeding in theory an astronomical number of $>2x10^{13}$ unique antibodies in humans [3]. Each antibody can potentially undergo affinity maturation, reaching high affinity and specificity in light of the immune system requirement following antigen encounter. The central piece in antibody repertoire relies in somatic gene recombination of their tandemly arranged germline gene segments known as V(D)J recombination, a process that shuffles the modular antigen-binding region of both heavy chain (VH) and light chain. Each chain is composed of a conserved β-sheet framework, providing structural stability and support to the 3 highly variable complementarity-determining regions (CDRs) that form the antibody binding interface [4]. Despite their tremendous binding surface diversity, each CDR except H3 assumes one of only a handful of discrete conformations termed canonical conformations [5–7]. Over the last decades, antibody engineering has evolved and attained many goals, ranging from natural CDR grafting, framework humanization, and bispecificity to advanced combinatorial library screening[8–12].

The early observation of antibodies' modular structure and the growing number of structure-solved antibodies has led to an increased understanding of antibodies sequence-structure relationships. Such vital knowledge along with advances in computational protein design, has enabled successes in computational methods for antibody re-engineering including higher affinity, specificity, and humanization [13–15]. While protein re-engineering by computational means was at first templated on naturally occurring complexes, recent years have seen the first reports of novel protein-protein interactions in non-cognate proteins [16–19]. Yet, these achievements mainly rely on pre-existing protein designed scaffolds that were treated as rigid structural elements with minimal perturbation of their backbone degrees of freedom. Thus, computational *de novo* design of protein backbone for function remains an unmet challenge due to the complication

inherent in correctly balancing the contributions to free energy from polar groups as well as the large conformation space open to the protein backbone [20].

Natural antibody repertoires, in spite of their theoretically vast size and potential, are dramatically decreased in size by the adaptive immune system acting as two arms of a sensitive scale. On one hand, the necessity to produce large diversity of antibodies to dispose of pathogens, hazardous chemicals, and various other antigen, on the other hand, antibody self-reactivity towards one out of thousands of the organism's natural entities can result in severe pathologies as autoimmunity diseases [21]. The resulting functional repertoire is thus limited and redundant [22,23], suppressing many antibodies with therapeutic and diagnostic potential.

Realizing that the natural antibody repertoire is thus highly redundant and biased, Winter and Milstein [8] described the need for experimental high throughput strategies capable of screening and isolating antibody candidates from libraries of immunized animals' repertoires. Over time, multiple methods for high throughput screening were developed, relying on the generation of a diversified antibody repertoire, coupling antibodies genotype and phenotype, applying selective pressure towards a desired trait and the ability to screen and isolate specific clones[24]. Nowadays, ribosomal, phage, bacteria, and yeast surface display (YSD) have turned into mainstream tools, allowing the engineering of full length antibodies and variable fragments such as scFvs or Fabs, optimizing the antibodies' size to a minimum, which resulted in increased expression and simplified DNA cloning strategies. Transcending the low avidity and difficult screening process of phage display and exhibiting superior protein folding-efficiency in relation to bacterial and ribosomal display, YSD has become a preferred tool in antibody combinatorial screening and isolation of antibodies [9,25]. In the YSD system, $5x10^4$ scFv copies on average are fused into the yeast cell, enabling quantitative screening through the use of fluorescence-activated cell sorting, monitoring simultaneously scFv expression using c-Myc epitope tag flanking the scFv and scFv binding to a fluorophore conjugated antigen [26]. As screening and antibody isolation become simpler and the need for antibody based therapeutics increases, rational design of combinatorial libraries holds the promise of increasing the probability of obtaining more specific and higher affinity antibodies [10,27] .

Research Aims:

In this study, as a part of our laboratory research in developing computational methods of *de novo* protein design, I will focus on insulin as a specific antibody design target. In this research I aimed to:

1. **Computationally design de-novo anti-insulin antibodies:** My prime goal in this research is to design anti-insulin antibodies with high affinity and specificity, using our laboratory computational design method.

2. **Screening and identification of insulin binding antibodies.** Utilizing yeast surface display platform to robustly screen a large number of designed antibodies (~100) and isolating insulin binding antibodies.

3. **Affinity maturation of insulin binding antibodies.** Conduct *in vitro* evolution rounds using the binding antibodies, generating new libraries of mutants from which clones with higher affinity, specificity and stability will be isolated.

Material and Methods

Computational algorithm development

The computational methods have been implemented within the Rosetta macromolecular modeling software suite[28] and are available through the Rosetta Commons agreement. All of the methods have been implemented through RosettaScripts[29], and all scripts are elaborated in depth in the AbDesign paper (Appendix, AbDesign paper, methods section).

Generation of yeast surface display plasmids by homologous recombination

ScFv recombination cassette was designed according to YSD manual[26], in which each scFv is tandemly synthesized as $(Gly_4Ser)_3$ peptide linker followed by the VH, an additional $(Gly_4Ser)_3$ peptide linker, the VL and a c-Myc epitope tag (except for the initial designs, in which the VL proceeded the VH and the interconnecting peptide linker was taken from the original 4M5.3 framework). Candidate ScFv recombination cassettes were reverse translated to DNA sequence codon optimized for expression[30] in *Saccharomyces cerevisiae* except for the interconnecting $(Gly_4Ser)_3$ peptide linker sequence that was modified in order to prevent its DNA homology to the upstream linker.

ScFv dsDNA cassettes were synthesized (GeneBits, Gen9) with flanking homology regions according to YSD pETCON sequence (~50 bp) and amplified by PCR. 50 µl of scFv amplified PCR products were cloned with a 100 ng of NdeI and XhoI (NEB) double digested pETCON plasmid by yeast LiAc homologous recombination into EBY100 YSD strain according to protocol[31], subjecting the cells to 20 min heat shock and recovery on SC-Trp selective plates for 48 hours at 30°C The transformants' plasmids were recovered (ZymoprepII, Zymoresearch) and transformed by electroporation into E. coli Clooni strain (Lucigen) for high copy production. Plasmids were recovered from isolated colonies, verified by DNA sequencing and retransformed to EBY100 by yeast LiAc homologous recombination.

Yeast surface display screening and flow cytometry

ScFv expression and binding trials were performed by YSD according to manual[26] with the following modification: 1. Prior to surface display induction, overnight cultures were diluted in SD to OD_{600}=0.1 and re-grown for 5 hours reaching OD_{600}=0.5 for cells induction during exponential growth phase. 2.2 g/l dextrose and 20 g/l raffinose were added to the SGCAA induction medium to improve cell surface expression.

Surface display primary labeling was performed for 40 min at 4°C in 50 µl labeling reaction containing 1:50 primary mouse anti-myc antibody (9E10, Santa Cruze biotech) and varied concentration of biotinylated antigen of interest in PBSF (IBT; insulin, IGF1, IGF2: BioINS100, BIO-IGF1-10, BIO-IGF2-10 respectively). Secondary labeling was performed for 20min at 4°C in 50 µl labeling mix containing 1:50 streptavidin-APC (7100011M, SouthernBiotech) and 1:500 Alexa-Fluor 488 conjugated goat anti-mouse secondary antibody (A-21121, Molecular Probes) in PBSF. Fluorescence intensity was measured using a BD Accury C6 flow cytometer.

Generation of binder variants by homologous recombination

All point mutation variant were generated by site directed mutagenesis PCR as follows: Two overlapping primers were designed to insert the desired mutations and were used in combination with the primers for general scFv amplification using the original plasmid as template. PCR reactions were then treated with DpnI and inserted to an NdeI and XhoI (NEB) double digested pETCON plasmid by yeast LiAc homologous recombination.

Sequence analysis and germline comparison

Sequence analysis and germline conservation evaluation of the scFv designs were performed using the *IgBLAST* sequence analysis tool[32]. Germline antibody sequences were retrieved from the *IMGT/GENE-DB* database[33].

Creation of random mutagenesis libraries and FACS sorting

Low mutation rate (1-2 mutations per clone) epPCR libraries were generated using the GeneMorph II random mutagenesis kit (Agilent) according to manual using 1µg of Ins16 plasmid for 17 thermal cycles. The epPCR product was treated with DpnI, reamplified by PCR, and transformed by yeast electroporation according to Benatuil *et al.*'s protocol[34].

Library expression and binding trials were conducted in 500 µl YSD labeling reaction and were sorted using FACSAriaIII (BD bioscience). Sorting rounds were executed by collection of ~5% top binding population and subsequent 48 hours recovery in SD-Trp selective medium. After 4 sorting rounds, a sample from the library was plated, single colonies were chosen for sequencing and individual YSD assessment.

Results

A novel strategy combining computational de-novo design of antibodies and high-throughput experimental screening

We have developed a unique approach combining a novel algorithm for antibody design with high-throughput experimental screening. This approach enabled us to rapidly iterate between hypothesis, design-algorithm development, and experimental testing (Figure 1). Each iteration yielded valuable experimental information and insights that were implemented and evaluated in the subsequent step, resulting in a continuous design 'learning loop' (Figure 2). Our antibody design algorithm has therefore been subject to major change as we gained more insight into antibody design principles. The main elements of an advanced version of the antibody design algorithm are presented in the appendix; in the following, I present how key elements in the design method were conceived and validated experimentally. I chose human insulin as our initial target since it is a self-antigen, antibody-bound structures are not known, and insulin is available for purchase at high purity. Thus, insulin is a prime example for a broad class of biomedically important targets where traditional methods have failed to produce clinically useful binders and a biochemically convenient target.

A. Initial design algorithm

Natural antibodies repertoire generation relies mainly on V(D)J recombination, a process in which antibodies gene rearrangement creates a very diverse set of antibody scaffolds [3]. Our key idea was to create a very large backbone diversity by combining fragments taken from structure solved antibodies and place them on a conserved antibody framework.

The computational process started by generating a diverse set of backbones for each of the 6 antibody complementarity-determining regions (CDRs) modeled to be structurally close to the CDRs observed in >1,000 high-resolution antibody crystal structures (Appendix: *AbDesign* paper, results section B) while residue identities were preserved only for glycines and prolines. The CDRs are then recombined and grafted onto the framework of 4m5.3, generating a library of scFv backbone chimeras (Appendix: *AbDesign* paper, results section D, except in this initial algorithm, the framework residue

identities of 4m5.3 were not altered, all 6 CDRs were grafted with no relation to one another, and all the glycines and prolines in the CDRs were retained). Each chimera was selectively docked, allowing only binding modes in which the CDRs are buried by the binding interface and insulin N-terminals are exposed to ensure a natural antibody-like binding mode while preventing clashes by the un-modeled biotin label of the insulin. Docked models were then subjected to CDR sequence design and refinement to optimize binding. This process yielded hundreds of candidate scFvs from which top scoring candidates were picked by computed binding energy [35] and shape complementarity [36].

17 scFv designs were picked and visually inspected. At this initial stage of design-algorithm development we noticed that combinatorial backbone design produces many flaws, including voids in the core of the antibody and at the ligand-binding surface and inappropriately positioned charged residues at the antibody core. To address these defects I manually introduced, on average, 6 substitutions per design. The designs were synthesized as dsDNA oligonucleotide containing the designed VL and VH interconnected by 4m5.3's original linker and cloned into the yeast surface display vector by homologous recombination [26,37].Yeast cell surface expression and binding of the scFv clones were assessed by flow cytometry. Out of the 17 designs, 14 designs exhibited no scFv expression while 3 designs presented low expression (positive expression was defined by detection of >4.6% of scFv-displaying yeast population, supplementary information, table S1). Examination of the expressing designs in comparison to the non-expressing scFvs and natural antibody structures revealed multiple cavities within the design cores that were suggested to cause instability due to lack of backbone rigidity, exposure of the hydrophobic core to the solvent and scFv misfolding (Figure 3A). In addition, as opposed to natural antibody structures, I noticed that conserved residue identities at key-positions, which may be crucial for maintaining the backbone conformation of the CDRs were replaced or misplaced (Figure 3B). This setback highlights the complexity of combinatorial backbone design, which, unlike fixed backbone design, demands careful consideration of the stability and foldability of the designed protein.

Furthermore, to increase the overall scFv expression, our experimental platform was optimized by switching the VH/VL domain order and replacing the charged connecting linker (originally taken from 4m5.3 framework) with a shorter, more flexible $(Gly_4Ser)_3$ linker according to the general yeast surface display strategy [26]. In order to test these experimental changes, the three low expressing designs in addition to 4m5.3 were recloned and retested for expression. Though still modest, all three designs exhibited an increase in expression levels, ranging from 30% to 130% (Supplementary information, table S1).

B. Identity restriction at consensus positions and cavity filtering

Core cavities destabilize proteins by reduction of hydrophobic stabilization and the high energetic cost of their formation in the folded state[38]. In order to address the low stability of the initial designs, we added an additional packing filter to evaluate cavities and packing flaws in our designs (Figure 4A). The principle of the packing filter is generating a set of void-filling balls covering all caged voids within the protein model, clustering them into contiguous cavities and report the likelihood of these to resemble the cavities observed in high-resolution crystal structures [39]. The filter threshold was set to 0.6 according to visual examination of a packing-score sorted database of native antibodies and our design models, thus allowing native-like small cavities to continue the design process. In fact, the packing filtering results suggested that ~30% of the initial designs that did not express contained numerous cavities likely explaining their poor expression. In addition, to address the loss of conserved residue identities within the CDRs, residue identities at positions L33, H27 and H29 (Kabat numbering) were restricted to a single consensus or few identities present in many natural antibodies, which have a crucial role in stabilizing the CDR backbone.

30 additional designs were selected using the improved algorithm, manually fixed by 6 mutations on average, and were tested for binding and expression. In contrast to our initial poor expression results, out of all designs tested, one design (3%) exhibited high expression, 11 designs (36%) exhibited medium expression, 16 design (~53%) exhibited low expression and only 2 designs (~7%) failed to express, yet insulin-binding was not detected in any design (Supplementary information, table S2). These results suggested

that retention of only low-cavity models eliminates unstable designs that are unlikely to express. Moreover, fixation of consensus residues improves design stability and such conservation of consensus identities should be widely implemented throughout the CDRs. In light of these results, further analysis of our models revealed additional positions of conserved identities, not only of positive design nature but also negative design nature such as negative design in beta sheets and charged peripheral residues eliminating VH-VL domain swapping[40]. In addition, although many of the unstable large-cavity containing designs were eliminated by the packing filter, we noticed that many small cavities persisted and may be minimized by improving the shape complementarity between the VH-VL design interfaces (Figure 4B). Insertions of charged residues at the designed CDR stems facing the scFv cores also occurred in many of our scFv models, destabilizing the designs' hydrophobic core, resulting in impaired stability. This we intended to address by restricting residue identity of buried position (Figure 4C). These results highlight again the complexities of backbone design compared to previous design of function studies; combinatorial backbone design must include explicit elements to optimize both, affinity as well as protein stability.

C. VH-VL shape complementarity, broader use of identity conservation and core hydrophobicity

The association between the VH and VL domains forms the antibody variable domain, and it has been shown that improving the shape complementarity between these domains increases binding affinity toward the ligand [41]. Considering the low VH-VL shape complementarity of our previous designs, we lessened the backbone minimization constraints, thus improving the interface shape complementarity and reducing the VH-VL interface packing flaws. In an effort to further pursue the successful positive and negative consensus-identity retention strategy, an additional 16 positions were restricted to specific identity or a limited few (Supplementary information, table S3). As a concrete example for the importance of these identity restriction for stabilization by negative design elements, position L70 has a conserved surface pointing aspartate residue protecting from β aggregation and domain swapping, a common strategy in Ig-fold

families [40]. Finally, residue identities located at the buried CDR stems were confined to uncharged residues increasing the core and VH-VL hydrophobicity and packing quality.

In order to assess these algorithm modifications, an additional experimental iteration was preformed testing 48 designed scFvs (Supplementary information, table 3A). Although no insulin binding was observed, expression levels continued to increase: 13 designs (~27%) exhibited high expression, 16 designs (~33%) exhibited medium expression, 12 design (25%) exhibited low expression and 7 designs (~14%) failed to express. Analysis of the expression data in comparison to our design models revealed that all of the non-expressing designs had relatively long H3 loops (Figure 5A). Consequently, insufficient backbone support by hydrogen bonds and loose packing as a result of large distance from other loops and framework might explain the low stability of those models reflected in the expression results. Moreover, we observed that most of our scFv binding interfaces had a high preference to contain charged residues, independent from the insulin binding mode (Figure 5B). Such preference may be explained by the fact that the design objective function exclusively aims at binding, disregarding overall stability. While surface exposed charged residues are known to increase protein solubility and stability [11], neglecting interactions stabilizing the CDRs at their expense may decrease the overall design stability. In addition, while the identity restriction, both for positive and negative design had proven to increase the designs' stability, a general strategy for identifying such positions and restricting such to position specific identities might improve sequence-structure relationship, yielding scFvs binders with higher stability.

These observations and experimental results served as impetus to develop a general strategy for encoding sequence-structure rules in antibodies by using sequence constraints inferred from large sequence alignments of homologues elaborated fully in Appendix *AbDesign* paper, results, section c and detailed briefly below.

D. Multistate binding function, novel antibody segmentation, and PSSMs

My modeling and experimental results led to two major conclusions regarding deficiencies in our original concept of combinatorial backbone design: 1. Natural sequence-structure rules in the core of the protein that are essential for protein stability

and foldability are not preserved in combinatorial backbone design; and 2. Optimizing designed binders solely for binding the target molecule compromises the stability of the designed protein. To address these challenges, we developed a strategy that incorporates the following three additional elements:

1.Antibodies' binding sites are mainly composed of their CDRs, a fact that has led to extensive characterization of CDRs conformation, length, and specific boundaries[5,7,42]. Therefore, antibody engineering and computational design mainly rely on CDR-based rational library construction, redesigning the antibody binding site with minimal variation to the antibody framework[43–45]. In an attempt to improve the stabilizing interaction between the CDRs and the antibody framework, we decided to mimic the natural V(D)J recombination process and use the disulfide-linked cysteines in each of the variable fragments as stems for a single segment comprising CDR1 and CDR2, while the second disulfide-linked cysteine and a conserved position at the end of CDR3 as the stems for the CDR3 segments (position 100 in the variable κ domain and position 103 in the variable heavy domain, according to Kabat numbering. Appendix: *AbDesign* paper, results section B).

2. Antibodies are composed of a well-conserved rigid framework which serves as support for the highly variable antigen binding CDRs. Despite their diversity, all of the CDRs except for H3 assume one of a limited number of discrete conformations known as canonical structures [5]. Each canonical structure contains key conserved residue identities with crucial roles in stabilization of the CDR backbone conformation by positive and negative design elements. To constrain the design algorithm to these identities, position-specific scoring matrices (PSSM) were generated, encoding the identity frequencies at each position according to its conformation cluster in natural antibodies. Specific conservation cutoffs were applied depending on the position, allowing high diversity in the binding interface while retaining core identity towards the sequence consensus (Appendix: *AbDesign* paper, results section C).

3. Protein stability and binding affinity are two major traits often competing, resulting in a complex tradeoff that improves fitness[46–48]. In order to recapitulate such tradeoffs in computational design, each trait must be explicitly encoded, correctly quantified and

formulated to an objective function. To encode a multistate objective function that will encode both binding affinity and stability, we formulated binding energy as the bound-state fractional occupancy and stability as the system energy of the monomer in its bound conformation but dissociated from the ligand. The two terms were then summed as Boolean operators according to fuzzy-logic principles and used as the algorithm objective function [49].

In order to test these implementations, another 19 designs were selected and experimentally tested (Supplementary information, table S4). Experimental data revealed that all designs exhibited superior expression, exceeding the levels of natural antibodies and affinity-evolved scFvs. Additionally; two designs exhibited low-affinity insulin binding (5Ins_16 and 5Ins_20, Figure 6A-D). The high expression results achieved with these designs emphasizes the importance of the antibody segmentation, the PSSMs and the multistate-design strategies in our algorithm. In addition, optimization of the PSSMs strategy by alleviating the conservation cutoff and the multistate-design strategy by diversion of the multistate objective function towards binding might increase the ability to isolate additional binders at a minuscule expense to expression. Furthermore, the manual mutation of the designed scFvs were mainly restricted for optimization of long-range electrostatic interactions, and may be accounted for by the inaccurate score function of the Rosetta or the inability to grasp the significance of such.

Sequence characterization of the isolated binders

Taking into account the sequence constrains according to solved antibody structures, the use of lengthened fragments in comparison to the CDR based design strategy, and the originally murine framework, I hypothesized that my designs would exhibit high sequence similarity to germline antibodies. *IgBLAST*[32,33] Sequence alignment analysis of the 4m5.3 antibody used as the initial design, concurred that 4m5.3 has high conservation to mouse germline, sharing 96% and 91.8% identity to germline VL and VH respectively. In contrast, aligning both binders to mouse germline revealed that 5Ins_16 shares 78.8% and 70.4% identity with germline VL and VH, and 5Ins_20 shares 77.8% and 70.1% with different mouse germline VL and VH (Figure 7A-B). The observed low germline conservation exhibited by these design resembles tremendous somatic mutation

frequencies seen in broadly neutralizing HIV antibodies that undergo extensive nucleotide somatic mutation of 32% in the VH and 20% in the VL[50]. Unlike secondary immune response antibodies, that exhibits high germline conservation rarely surpassing 10% difference[50–52], the low conservation of my designs may suggest that our design algorithm generates models which are less biased to germline sequences than natural antibodies are. To evaluate the new fragment strategy, utilizing CDR1 and CDR2 as a single fragment during the design process, sequence analysis was conducted on the interconnecting framework regions between the two CDRs of both chains (Figure 7C). While in 4m5.3 the VL interconnecting region was entirely conserved and the VH interconnecting showed 75% germline conservation, 5Ins_16 retain only 50% identity and 70% in the VH and VL interconnecting regions, and 5Ins_20 retained 60% and 62.5% in the VL and VH interconnecting region, respectively. The low conservation seen in the CDR1 and CDR2 interconnecting region along with the designs' high stability may be explained by improved packing of the scFv core as guided by improved global sequence structure relationship between the framework and the segments.

Experimental characterization of the isolated insulin binders

In order to examine whether the actual binding mode of the insulin binding designs are in agreement with the model, five point mutation variants were generated for each binder: 4 variants, in which a disruptive mutation was inserted at the binding interface to abolish binding; and 1 mutation with no predicted effect inserted in the periphery and was modeled to not affect binding (Figure 8A).

Experimental results confirmed the modeled binding mode insofar as all 4 disruptive mutations of 5Ins_16 binder greatly impaired binding by the introduction of substantial steric clashes into the binding interface. Moreover, the increased binding observed for the T183K variant which was predicted not to affect binding, may be explained by improved long-range electrostatic interaction with insulin's negatively charged surface which was not taken into consideration by Rosetta's scoring function (Figure 8B). In contrast to 5Ins_16, three of the predicted disruptive variants of 5Ins_20 binder exhibited increased binding while binding was entirely abolished by introducing the fourth predicted disruptive as well as the indifferent mutation. The discrepancies between experimental

results and model in 5Ins_20 may be explained by an alternative binding mode. Therefore, we decided to continue characterization solely on the 5Ins_16 binder. In light of these results, one can appreciate the capability of our experimental screening strategy to quickly provide supporting data on binding mode.

Development of long-range electrostatic algorithm for design improvements

Electrostatic interactions are known for their long-range vital influence in protein-protein interactions, unbound to any atomic contacts, which makes their computational prediction and simulation extremely difficult [53,54]. Insulin is a negatively charged protein with large negative surface patches and a net charge of -2. The improved binding observed for the T183K 5Ins_16 variant, which was predicted not to affect binding, stressed the tremendous advantage and potential in maximizing long-range electrostatic interactions to increase affinity. To that end, we used an adaptive Poisson-Boltzmann solver (APBS) algorithm [55] to help identify additional long-range electrostatic mutations that will improve binding. The algorithm first identifies all peripheral exposed residues and samples the positively charged side chains, measuring the difference in binding energy (restriction to positively charged residues only was due to the negatively charged surface of insulin). The APBS algorithm returned a noticeable improvement in binding energy for the 5Ins_16 T183K variant, and another 7 positive charge substitutions were tested for binding. The experimental results revealed an increased binding for all clones tested, emphasizing the substantial contribution of long-range electrostatic interactions to binding energy (Figure 9A-B). Out of these 7 enhanced variants, only 3 of the variants were predicted by the APBS algorithm to increase binding. These results may be explained by lack of sensitivity of our algorithm and by the high negative net charge of insulin, yet serve to emphasize the potential of improving Rosetta's ability to notice long-range electrostatics for robust improvement of binding[11].

Binding affinity improvement by in-vitro evolution

In order to improve the initial binding (estimated K_d of 470nM) affinity seen by 5Ins_16 binder, a library of $2*10^7$ variants was made by low error rate (1-2 mutations per scFv on average) epPCR. Library sorting for improved binders yielded clones with higher affinity

than the original binder, which were separately tested and evaluated (Figure 10A). While most mutations were distributed throughout the design framework, peripheral surface, and core, no mutations were found on the designed CDRs involving direct interaction with insulin (Figure 10B). These results were unexpected based on previous findings[56] in which most mutations that improved binding were either at the core of the interface or surrounding it while none were detected in the core or on the surface away from the binding interface. Nevertheless, recent studies on affinity maturation of protein binding antibodies suggest that such distant mutations can increase binding affinity by improving antibody rigidity and VH-VL shape complementarity[57]. Assessment of the isolated affinity improved variants exhibited an average decrease of 31% in expression and an average binding improvement of 41% (Figure 10C). Such results may be explained by small changes in backbone conformation and residue-packing that stabilize the binding interface while impairing overall design stability. In addition, absence of mutations in the binding interface further confirms the modeled binding-mode and indicates that the sequence designed by our algorithm for the interface is the optimal solution for the designed binding mode.

In order to generate an evolved high-affinity binder, I generated double and triple mutation variants, coupling mutations that increased the affinity to the largest extent with minimal expression reduction. This resulted in a high affinity insulin binder with an estimated K_d of 30 nM comprising the following mutations: S144Y, H38Q and Q213R.

Examination of the evolved binder revealed that both S144Y and H38Q strongly support the H3 loop by forming two hydrogen bonds each with the H3 backbone while the Q213R mutation, contributing to long-range electrostatic forces by a surface-exposed sidechain conformation and relieving strain for the H38Q to correctly pack (Figure 11A-C). To investigate why our algorithm failed to use these affinity improving mutations, we introduced them into the design model and noticed that they were rejected mostly due to high stability penalties and only negligible binding energy contribution predicted by Rosetta. In addition, as the evolved binder's mutation could only be introduced to the binder after backbone minimization, loosening these constraints during design might allow such identities to be introduced *a priori*. Furthermore, as the number of hydrogen

bonds supporting the H3 CDR increased with the high affinity mutation, a strategy emphasizing backbone hydrogen bonds might be considered.

Specificity evaluation of the W.T. and evolved binders

To determine the insulin binding designs specificity toward insulin, additional binding trials were conducted using ACP as an unrelated protein and the highly homologs IGF1 and IGF2. To examine the specificity of Ins16 binder and the evolved variant, binding trials were conducted using biotinylated insulin, IGF1, IGF2 and ACP using a ligand concentration of 500 nM. To rule out unspecific adherence of the tested targets to yeast cells or to displayed scFvs in general; specificity trials were also subjected on anti-tissue factor 5G9 scFv that demonstrated no apparent binding. According to the binding trials results, the original 5Ins_16 binder demonstrated negligible binding towards both IGFs and no ACP binding was observed in either the original binder or in the evolved variant. The evolved binder, however, bound IGF-1 and IGF-2 at 87% and 64% (measured as fractional bounded population), respectively, compared to insulin binding (Figure 12A). Structural comparison between Insulin and IGF-1 (PDB entries 1G7A and 2DSR) revealed that both share striking interface residue identities and the net charge of -2. The comparatively weaker binding of the evolved binder to IGF-1 may be explained by slight structural incompatibility: first, the S9A substitution results in the loss of a hydrogen bond to the design Trp100. Second, insulin's H10E substitution, located within the interface periphery might cause steric clashes with the design VL, increasing repulsion force (Figure 12B). Third, the E13D substitution might lead to formation of a small cavity within the binding interface. In addition, removal of the π-stacking interaction between insulin's Tyr16 and design's Trp209 by substitution to glutamine might also explain binding decrease.

The weaker binding to IGF-2 in comparison to insulin might be explained by similar substitutions in IGF-1; H10E, E13D, Y16Q and S9G, which forms a minor interface cavity in addition to loss of the original hydrogen bond. Although these substitutions were seen also in IGF-1 and might by attributed only to 13% decrease in binding accordingly, IGF-2 has a lower net charge of -1 and 4 Arginine residues, which are not present on insulin, are positioned next to the VH positively charged residues. Such

electrostatic repulsion may cause the weaker binding of IGF-2 in comparison to IGF-1 and insulin (Figure 12C). More detailed interaction studies using surface plasmon resonance are required to determine the specificity of 5Ins_16 to insulin over its homologs, but my preliminary results suggest low specificity at the concentrations tested.

The unspecific binding demonstrated by the evolved binder relative to the original 5Ins_16 binder may be explained by the increase in charge compatibility of 5Ins_16 to insulin and its homologs by the Q213R mutation. In addition, the increased H3 stability by the H38Q and S144Y mutations might have resulted in a more stable binding interface or a moderate change in the VH-VL packing able to encompass the two insulin homologues. Furthermore, the inability of both binders to bind ACP and the significant binding of insulin along with the homologs IGFs suggests that predicted and actual binding mode coincide.

Overall, binding affinity improvement using *in vitro* evolution methods is known to result in decreased specificity due to a solely positive selection for binding without any negative selection influence. If high specificity insulin binding were desired, affinity maturation could be undertaken in the presence of an excess of unlabeled IGF1 and IGF2 to serve as competitors during affinity maturation.

Development of reconstitution algorithm of the *in vitro* evolved mutations

In order to recapitulate the 3 in vitro evolved mutations, accounting for a 15-fold increase in affinity without compromising stability, I noticed that proper insertion to the above into the w.t model must be accompanied by backbone minimization to relieve resulted strain, similar to observations in earlier binder design studies[17]. In order to improve the design algorithm's ability to introduce these mutations, I have implemented an iterative backbone minimization algorithm. The algorithm chooses every CDR position in turn, substitutes the original identity to all residues allowed by a permissive PSSM, followed by successive rotamer and constrained backbone minimization steps.

In agreement with our experimental data, the minimization algorithm successfully recapitulated the S144Y substitution, increasing the binding energy by 3.6 R.e.u and Q213R with a more modest improvement of 0.6 R.e.u. In contrast, H38Q resulted in a

decrease in binding energy of 4 R.e.u. The inability to recapitulate this substitution may be explained by the strain introduced with the attempt to pack the glutamine. This strain can be alleviated by minor modifications to the rigid body orientation of VH and VL, however, such minimization was not performed. These results suggest that addition of an iterative, per-position backbone minimization step can identify more suitable residue identities thus increasing the initial affinities. The *AbDesign* method has recently been extended to deal with broader sampling of the docking angle between the heavy and light chains of the variable domain.

Discussion

Recent successes in computational *de novo* design of proteins with novel molecular function [18,58,59] rely mainly on existing protein scaffolds with high secondary structure content [60] or mimicking natural interaction interfaces by grafting rigid backbone segments [61] while *de novo* design of unstructured segments for function remains elusive and until now, was only recently shown *in silico* by our lab (Appendix, *AbDesign* paper). The successful isolation of an insulin binding scFv and its evaluation represent a milestone towards the goal of designing *de novo* antibodies from first principle. This and the additional success of our design algorithm to generate experimentally validated *de novo* antibodies against ACP indicate that our design algorithm can most likely produce antibodies for any given target of known structure, revolutionizing antibody generation for therapeutic, diagnostic, and research use. Nevertheless, further, more accurate experimental evaluation is needed, such as surface plasmon resonance and structure determination to accurately measure the binding affinity and specificity of my binders, and confirm the accuracy of the modeling. Given that two designs out of 20 designs tested exhibited insulin binding, and considering the ability of our algorithm to generate a virtually infinite number of designs for any given epitope, additional algorithm development is needed to increase our binding retention efficiency. Additional filters are required, which are capable of distinguishing between our successful, designed models and natural antibodies on the one hand and the designs that failed to recognize the target are needed. Moreover, numerous structure determinations of binding and non-binding designs are crucial to assess potential deviations of the actual solved structure from the models, yielding valuable information for additional algorithm improvement [62].

The low germline antibody conservation of our isolated binders emphasizes how computational antibody design could span the sequence and conformation space more broadly than natural immune systems and can yield unique antibodies that would not occur in nature.

Comparing the average scores of the recent designs to earlier design tested (Supplementary information, table S5), despite the overall improvement in stability as seen by the high expression, a noticeable decrease in the calculated total score, and

binding energy is observed while shape complementarity, packing score and sasa scores showed no obvious tendency. The total score and binding energy deterioration seen in the second and third experimental iterations (2Ins_# and 3Ins_#) may be explained by the utilization of the identity restriction strategy and the inaccuracy in Rosetta score function, failing to assess the identity contribution at these positions.

Compared to the improvement in stability that originated from the identity restriction strategy, the ability of the PSSMs to broaden such strategy to all positions and restricting to identities observed in natural antibodies, highlights the importance of such strong sequence structure relationship in combinatorial backbone design, which is also emphasized by the superb expression of the designs. Moreover, the high expression that was exhibited by my most recent designs tested, in which both the PSSMs and the multistate objective function were implemented (5Ins_#), exemplify that encoding information from large protein families such as antibodies to infer sequence-structure relationships, as well as integrating natural-like multistate objective functions are crucial evolutionary aspects that should guide computational backbone design.

Unlike the 6 CDRs design strategy that is widely used in antibody engineering, our novel segmentation backbone imitating natural V(D)J gene segments can be regarded as one of the prominent causes to the improved stability of my designs. Utilizing these natural-like 4 segments directly contain conformation and sequence correlations between the CDRs and the framework, encoding both local and global sequence-structure relationships. Extending the boundaries of the segments from the CDR positions facilitates design not only of the exterior binding surface of the antibody, but also of the antibody core and framework. The high stability of my designs emphasizes the algorithm's ability to design *de novo* intertwined protein core and exterior interface for function.

Antibodies' modularity and the large number of antibody structures have enabled us to accesses a large space of feasible backbone conformations and vast information for combinatorial backbone design. Our experimental results suggest that such a strategy for combinatorial backbone design may be generalized to other large protein families with sufficiently heterogeneous sets of solved structures, such as TIM barrels or certain classes of enzymes.

The isolated binder's ability to bind IGF1 and IGF2 in lower affinity, in contrast to ACP binding, demonstrates its relative specificity, as the IGFs and insulin are highly homologous on the interacting surface. While binding to IGF1 and IGF2 was expected since no negative design strategies were employed in the design algorithm, such can easily be formulated within the multistate objective function by aligning homologs to the antigen binding mode and introducing these interactions as negative states.[49] The two major advantages of the multistate objective function in designing specificity are the generation of designs that are rationally focused on high specificity, thus avoiding experimental directed evolution, which could result in stability deterioration.

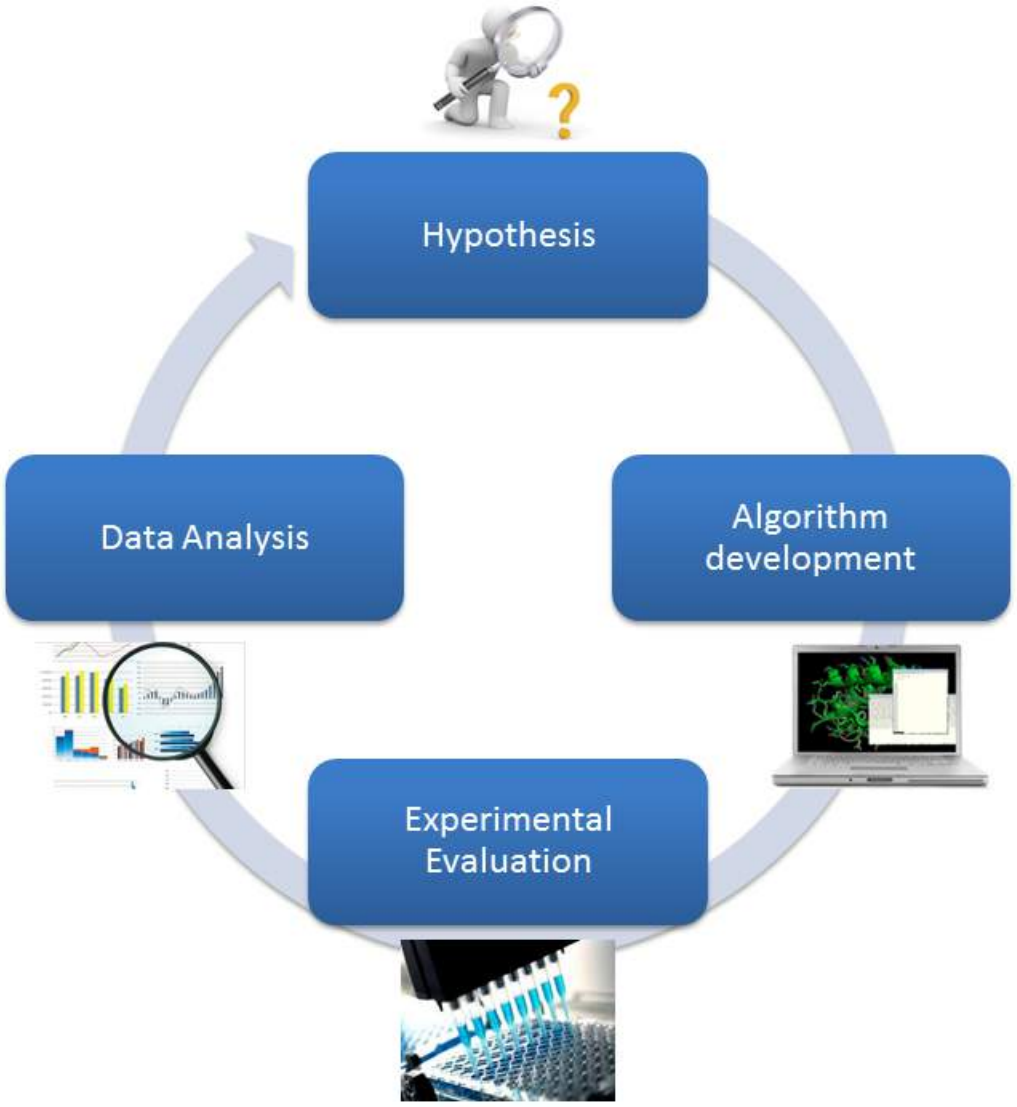
The key element that directed our algorithm development was the learning loop strategy. As seen throughout the algorithm development process, the ability to disassemble the process into defined loops directed and refined our development progress by the abundant experimental information gathered. As computational methods and experimental techniques continuously progress, the overall individual methods can only improve to some extent. [63,64] Therefore, new methodologies of combining the two, such as our learning loop strategy should be utilized.

As the demand for antibody based therapeutic constantly increases, the ability to engineer epitope specific antibodies remains elusive. To date, epitope specific antibodies are based mainly on iterative combinatorial screenings for high-affinity binding of a target protein and subsequent screening against agonist or antagonist activities, a functional trait which may not be applicable for most targets [65,66]. Similar to my insulin binding scFv, our algorithm can design epitope specific antibodies without infringing on antigen functionality and their functional integrity quickly be confirmed by experimental screening.

Currently, the most common strategies for isolation of novel antibodies are generation of antibody combinatorial repertoires [14,67] and rational designed repertoires composed of intentionally biased toward a specific antigen or a desired functionality [44,68,69]. Despite past successes of both strategies, both suffer from poor antibody stability, minor retention of binders, high dependency on massive repertoire screening and tedious directed evolution for improved affinity and stability of the isolated clones. Exceeding natural

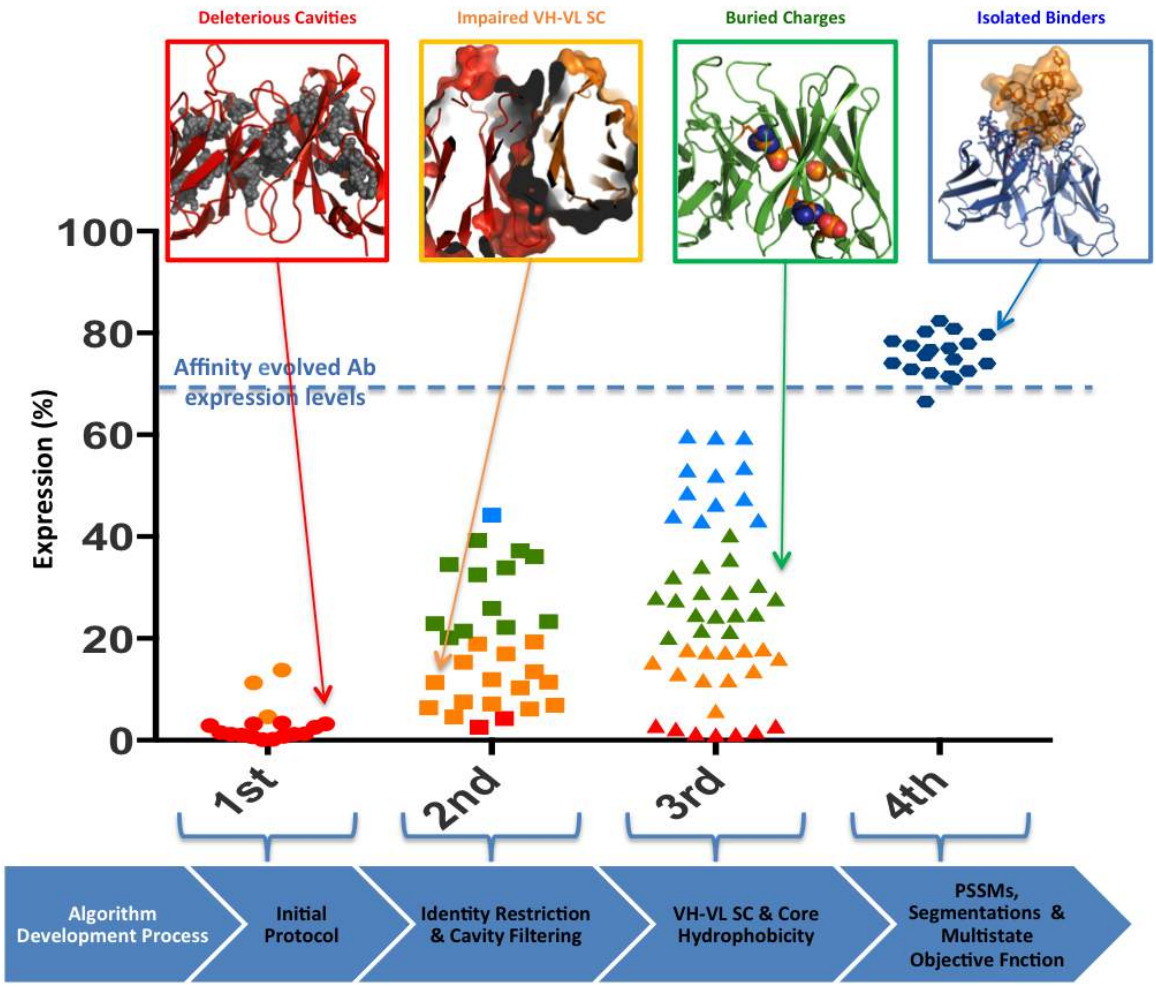
antibody expression levels for all models tested, our designs algorithm yields highly stable antibodies capable of standing random mutagenesis rounds of directed evolution without significant stability deterioration. It can potentially generate very large unbiased repertoires of highly stable antibodies, thereby transgressing the limitations of any previously known strategy.

Figures



<u>Figure 1:</u> Novel iterative strategy for antibody design.

Our research strategy combines computational de-novo design of antibodies with high-throughput experimental screening. Briefly, the hypothesis is transformed into an algorithm and implemented into our design protocol. Antibody models generated by the design algorithm are then evaluated in yeast surface display. The experimental results are evaluated, leading to further advancement of the algorithm.



<u>Figure 2:</u> Expression improvement over the course of algorithm development.

Each column (and sign) shows the experimental data obtained at a specific phase of algorithm development. Each data point marks the expression level of one design, non-expressing scFvs shown in red, low expressing scFvs shown in orange, medium expressing scFv shown in light blue and highly expressing scFvs shown in dark blue. Expression level of a previously reported, highly expressing scFv is marked by the dashed blue line. Significant expression improvement can be observed from each stage of algorithm development to the next. Designs representative of the problems addressed in each iteration are represented in the panels above, frame and dot colors signify expression levels.

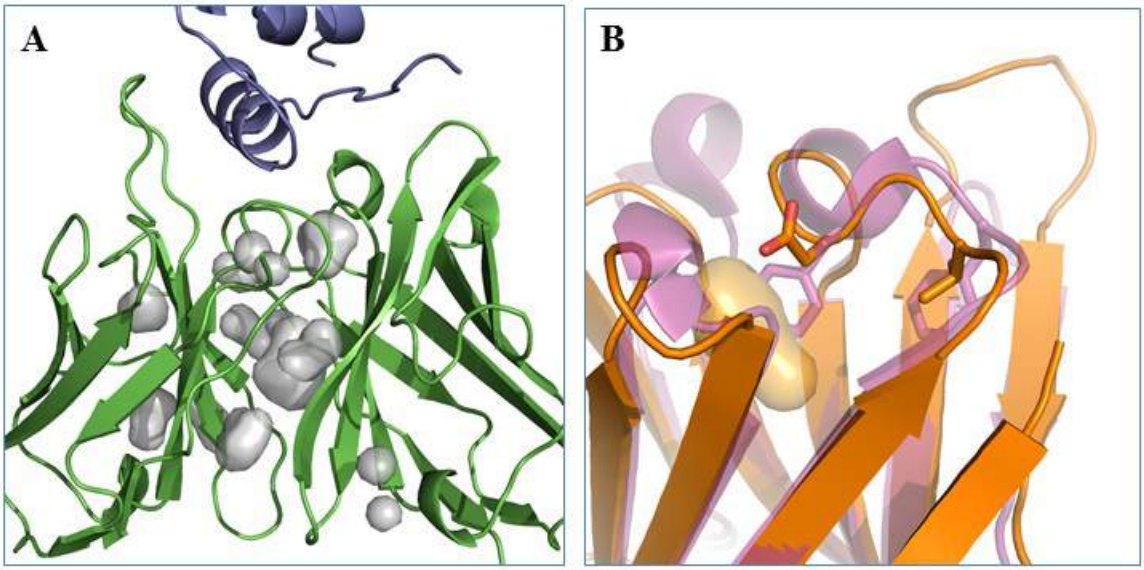


Figure 3: Initial algorithm anti-insulin scFv models.

(A) The extent of inadequate packing within the scFv (Cavities shown in gray, scFv model shown in green, insulin shown in purple). **(B)** Comparison between design model (gold) and 4m5.3 (magenta) reveals two well-conserved phenylalanine residues which were substituted by valine and aspartate, destabilizing the backbone and exposing the scFv core by forming a cavity.

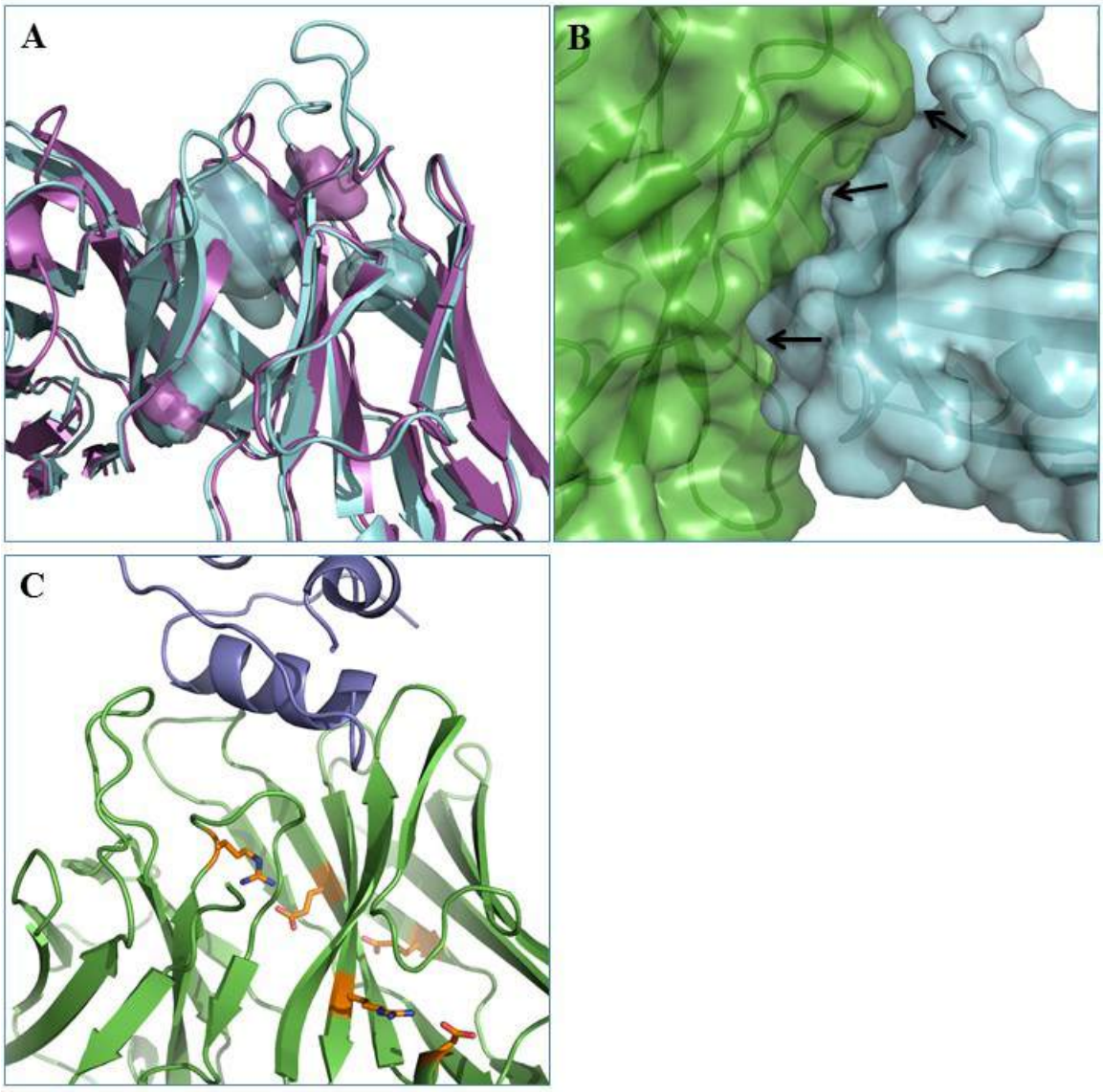


Figure 4: Cavity filtering and identity restriction at consensus positions.

(A) Comparison between high-scoring, well packed design (magenta, packstat score of 0.679) to a low-scoring design with multiple cavities (cyan, packstat score of 0.577). Cavities within the scFv VH-VL interface greatly reduce protein stability and can be identified by the packstat filter. **(B)** Shape-complementarity inaccuracy between the VL (green) and VH (cyan) results in stability deterioration, incompatibilities marked by black arrows. **(C)** Charged residues (gold) are buried within the design's (green) hydrophobic core, reducing hydrophobic forces. Insulin shown in purple.

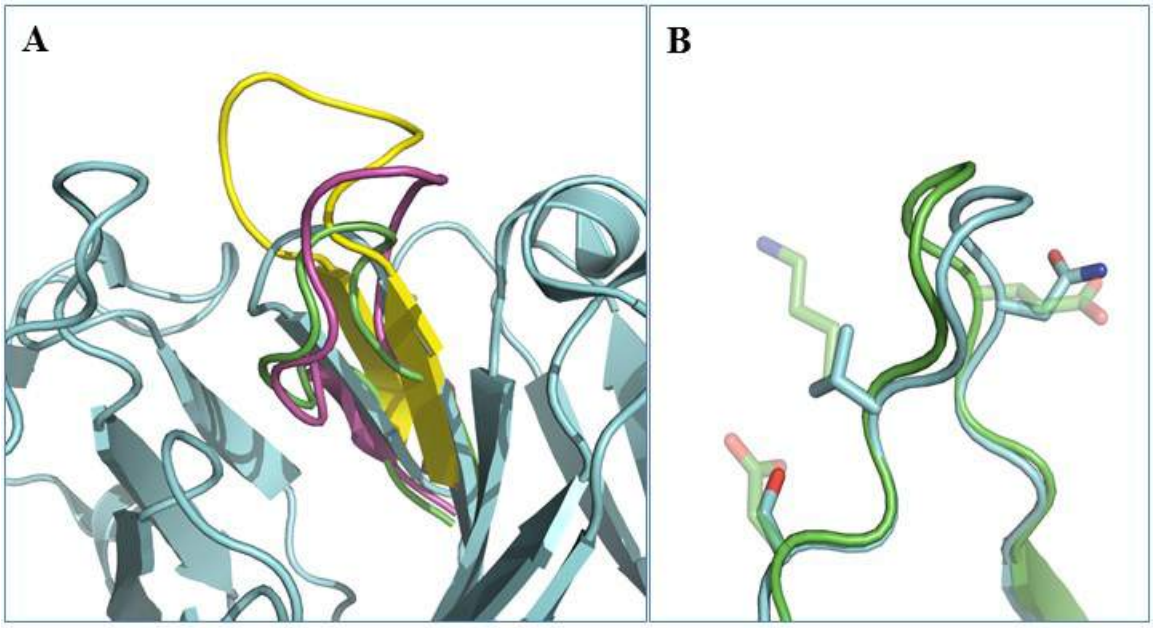


Figure 5: Long H3 CDRs and highly charged substitutions.

(A) Comparison of long H3 loops from 3 non-expressing designs (yellow, magenta and green) which impair design stabilization, with the original 4m5.3 scFv (cyan). **(B)** Multiple charged residue substitution in a designed H1 CDR (green) relative to 4m5.3 (cyan) decrease rigidity, thus destabilizing the loop.

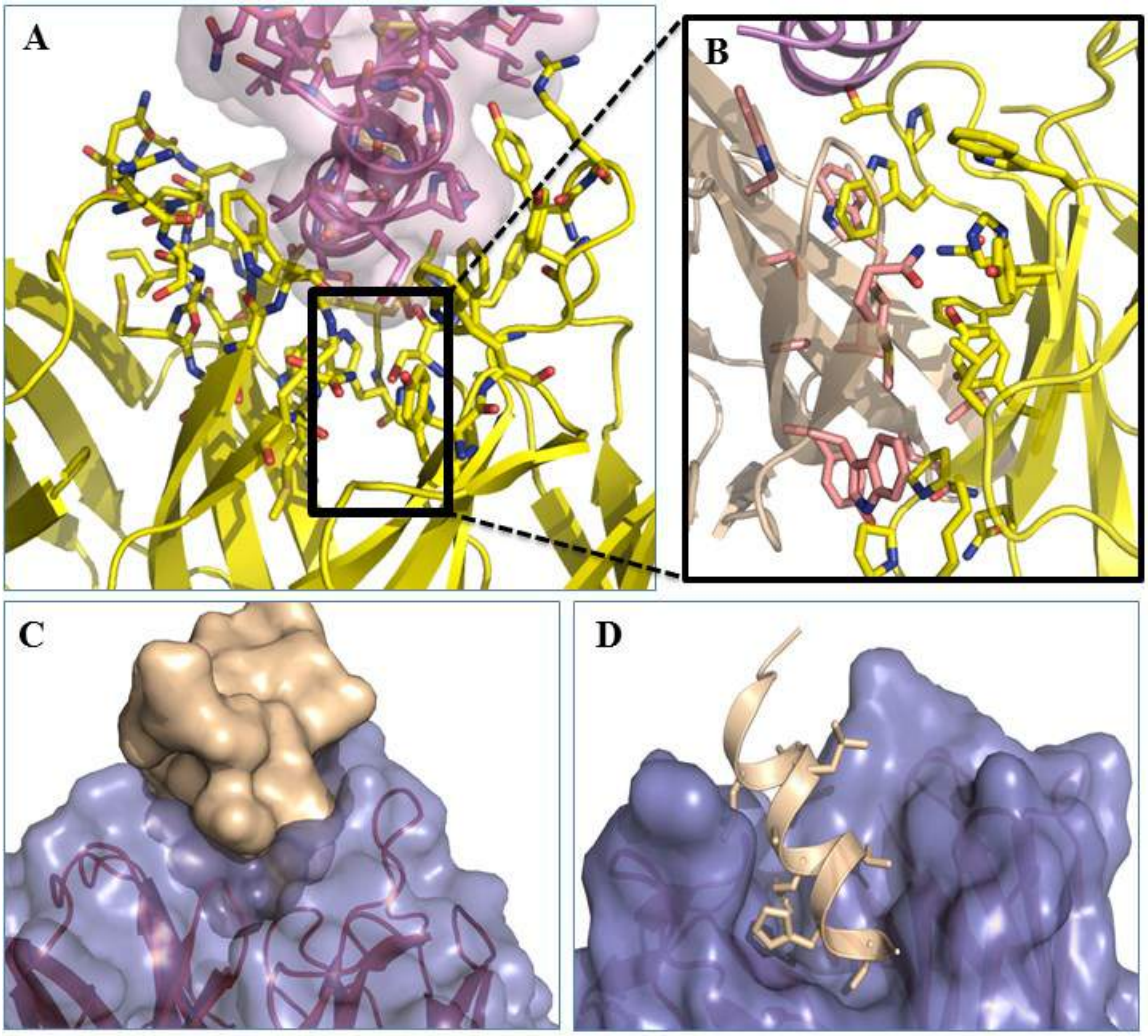


Figure 6: Isolation of insulin-binding scFvs.

(A) Interface view of the 5Ins_16 binder (yellow) and insulin (magenta) in the designed binding coordination. **(B) 5**Ins_16 tight core packing, VH (pale yellow) and VL (yellow) interface. **(C)** The large binding surface of 5Ins_16 (purple) engulfs insulin (pale yellow). **(D)** Superb shape complementarity of the binding interface.

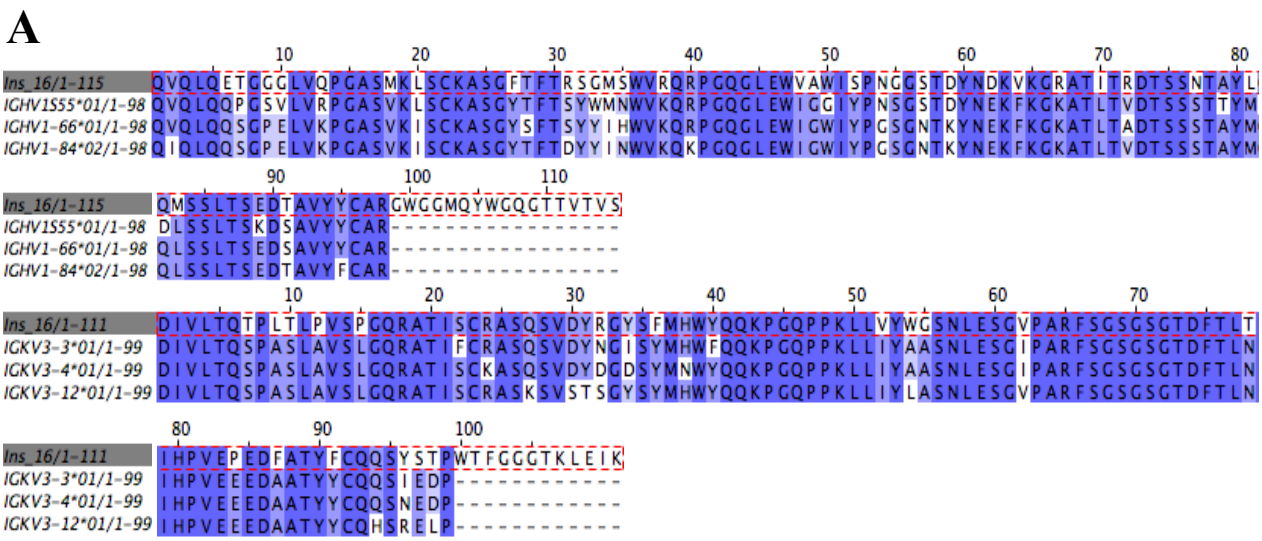


Figure 7: Sequence analysis of the isolated binders relative to germline antibodies.

(A) Multiple sequence alignments of 5Ins_16 (top row, marked in red dashes) and 3 most similar germline antibodies variable domains reveal low germline conservation of our designed binder (residue color by degree of sequence conservation). H3 and L3 were not aligned. **(B)** Many non-conserved mutations (red) are distributed throughout the 5Ins_16 (blue) framework and CDRs. H3 and L3 shown in gold. **(C)** The connecting regions (red) of CDR1 and CDR2 (both show in green) composing the entire segment as utilized in the most recent design protocol, enable design of the scFv core.

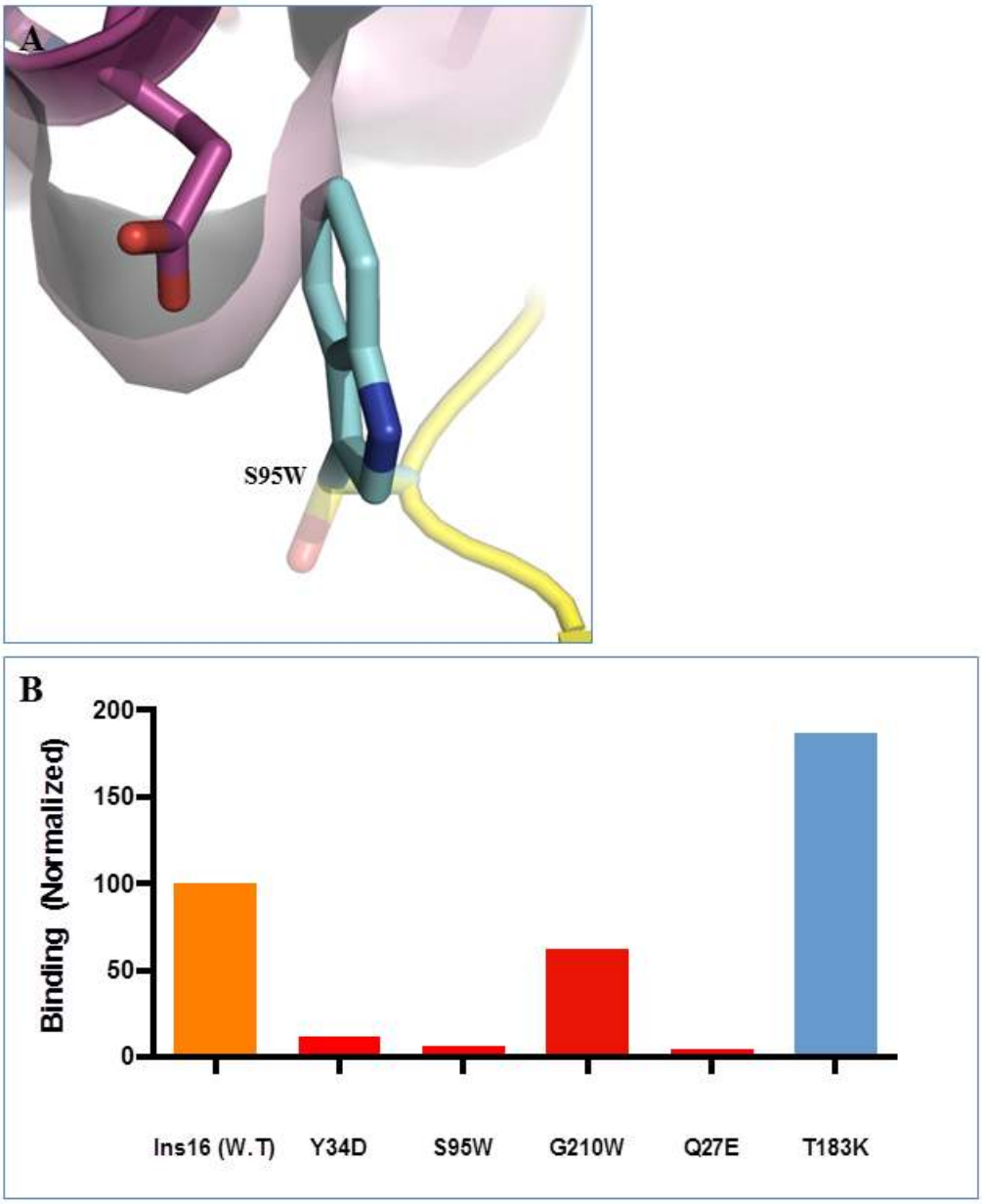


Figure 8: Mutant studies support the designed binding mode.

(A) Predicted negative mutations (cyan) cause sidechain clashes that disrupt the designed binding-mode (binder in yellow; insulin shown in magenta). **(B)** Changes in binding levels of the mutants (predicted negative mutations in red, predicted non-affecting mutation in blue, W.T. shown in orange). All negative variants support the binding model while the non-affecting mutation variant exhibits binding improvement, presumably as a result of long-range electrostatic forces.

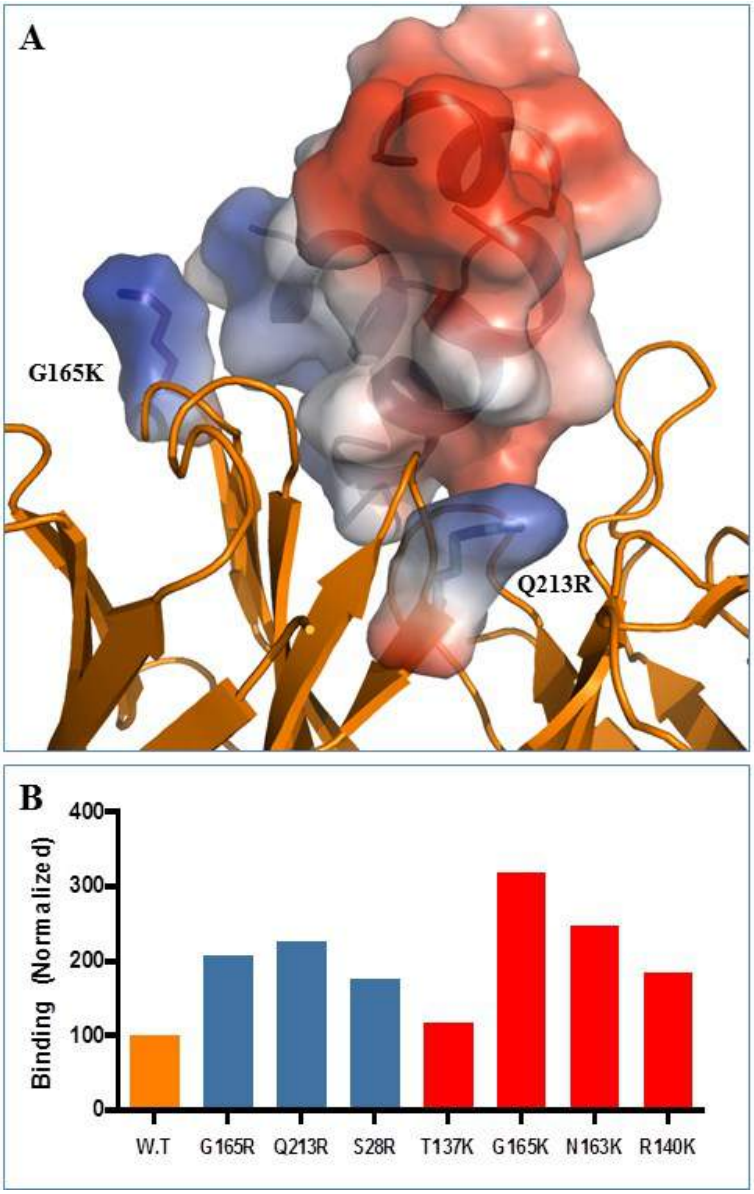


Figure 9: Long range electrostatic mutations dramatically increase binding.

(A) Electrostatic surface visualization of insulin and two tested mutations which were predicted to increase (Q213R) or decrease (G165K) the binder's (gold) binding. Q213R counters an acidic patch on insulin with a positive charge, while G165K places a positive charge opposite a basic patch. Experimental data, however, shows an increase in affinity for both, indicating an influence on long-range electrostatic interactions. **(B)** Change in affinity of the long-range electrostatic variants. Variants predicted by the APBS algorithm to increase affinity shown in blue, variants predicted to decrease affinity shown in red, original 5Ins_16 binder shown in orange.

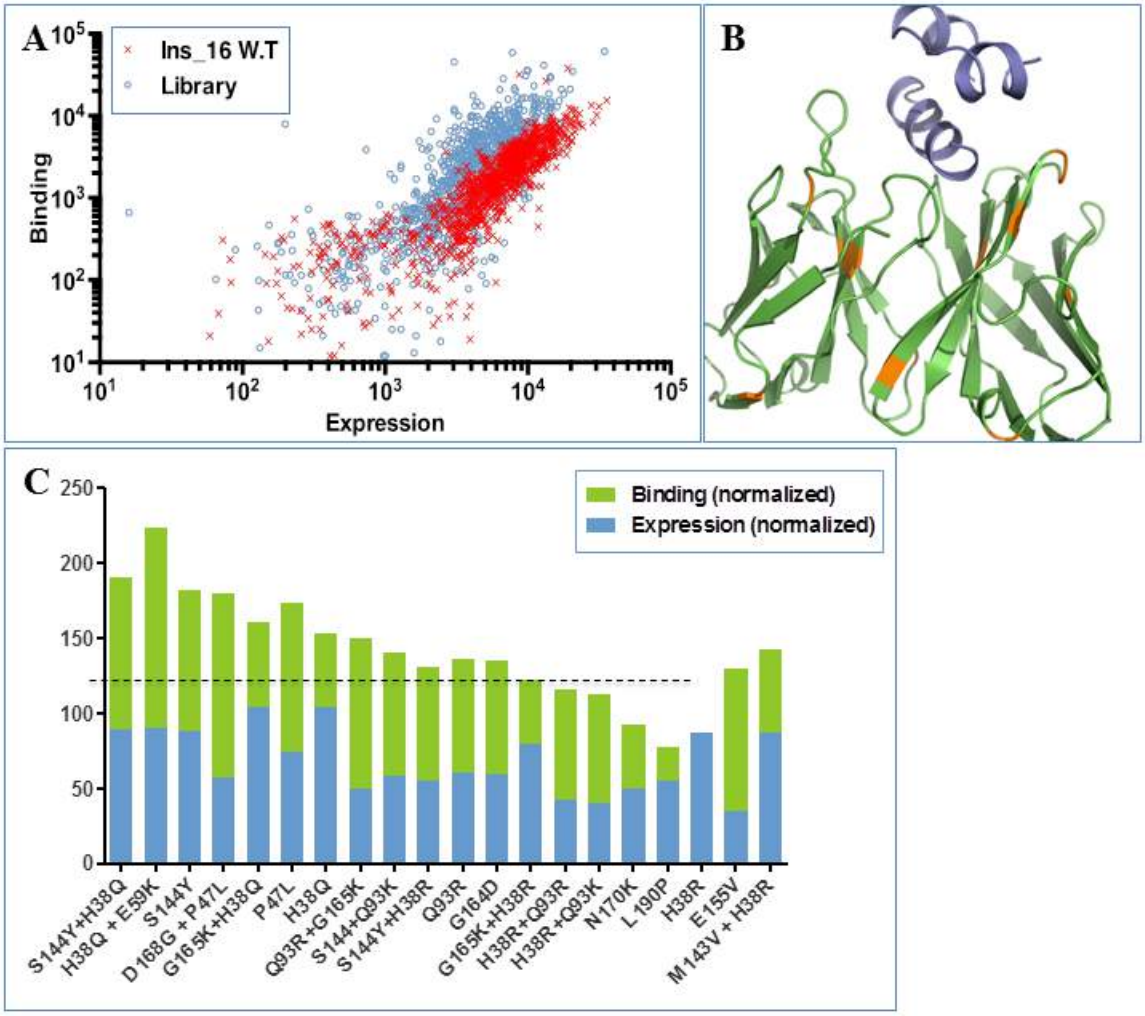


Figure 10: ***In vitro*** **evolution of 5Ins16 binder.**

(A) Yeast surface display of the epPCR library reveals variants with increased affinity. **(B)** *In vitro* evolved mutations (gold) are widely dispersed throughout the scFv framework (green) while no mutation in direct contact with insulin (magenta) was identified. **(C)** Changes in binding (green) and expression (blue) of the *in vitro* evolved isolates in comparison to the original 5Ins_16 (dashed black line) binder exhibit an increase in affinity at the expense of deteriorating stability

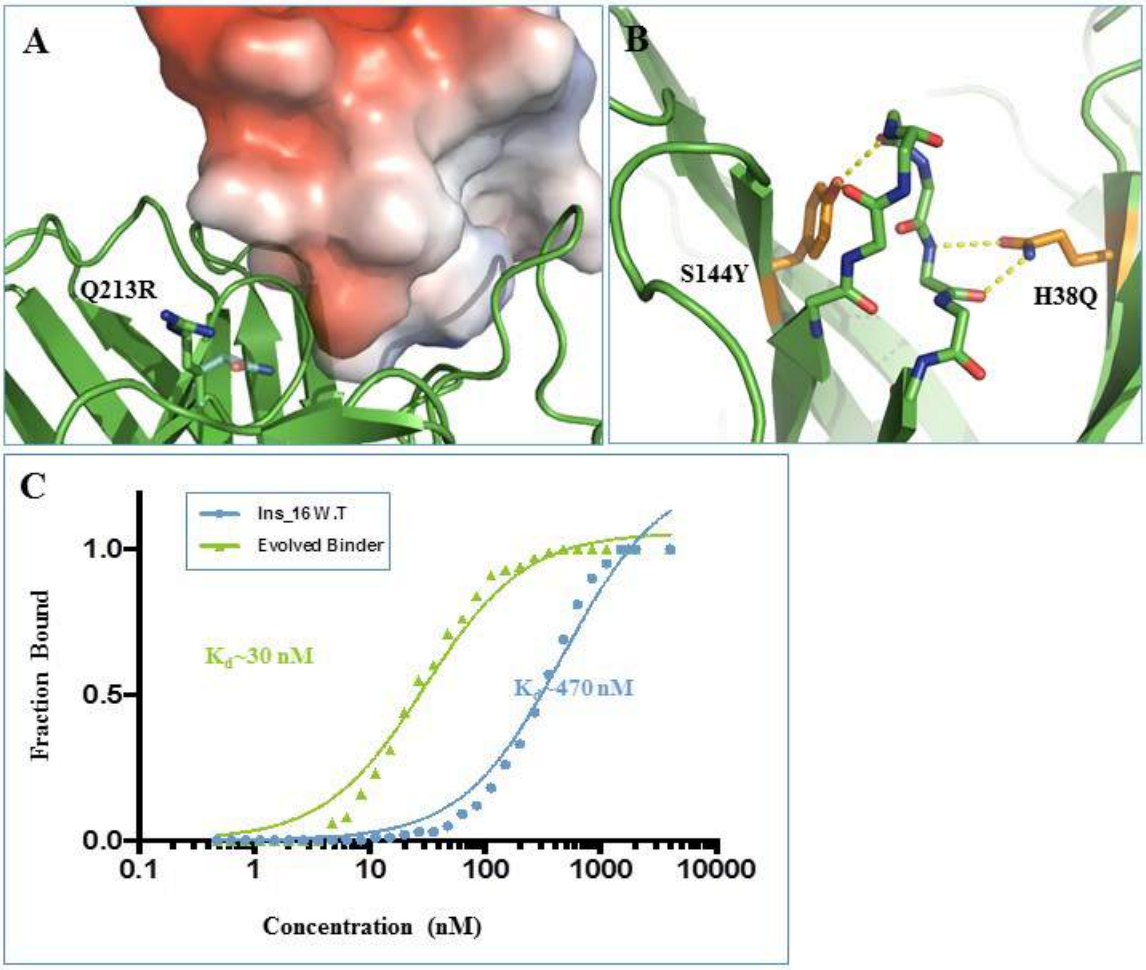


Figure 11: High affinity evolved 5Ins_16 Binder.

(A) Glutamine (cyan) to arginine mutation improves long range electrostatic interaction between the evolved binder (green) and insulin (electrostatic potential surface). **(B)** The evolved mutation (gold) stabilizes the design's (green) H3 CDR by forming multiple hydrogen bonds (yellow dashed lines). **(C)** Titration curves of 5Ins_16 W.T. binder (blue) in comparison with its evolved high affinity binder (blue) reveal a 15-fold increase in apparent affinity.

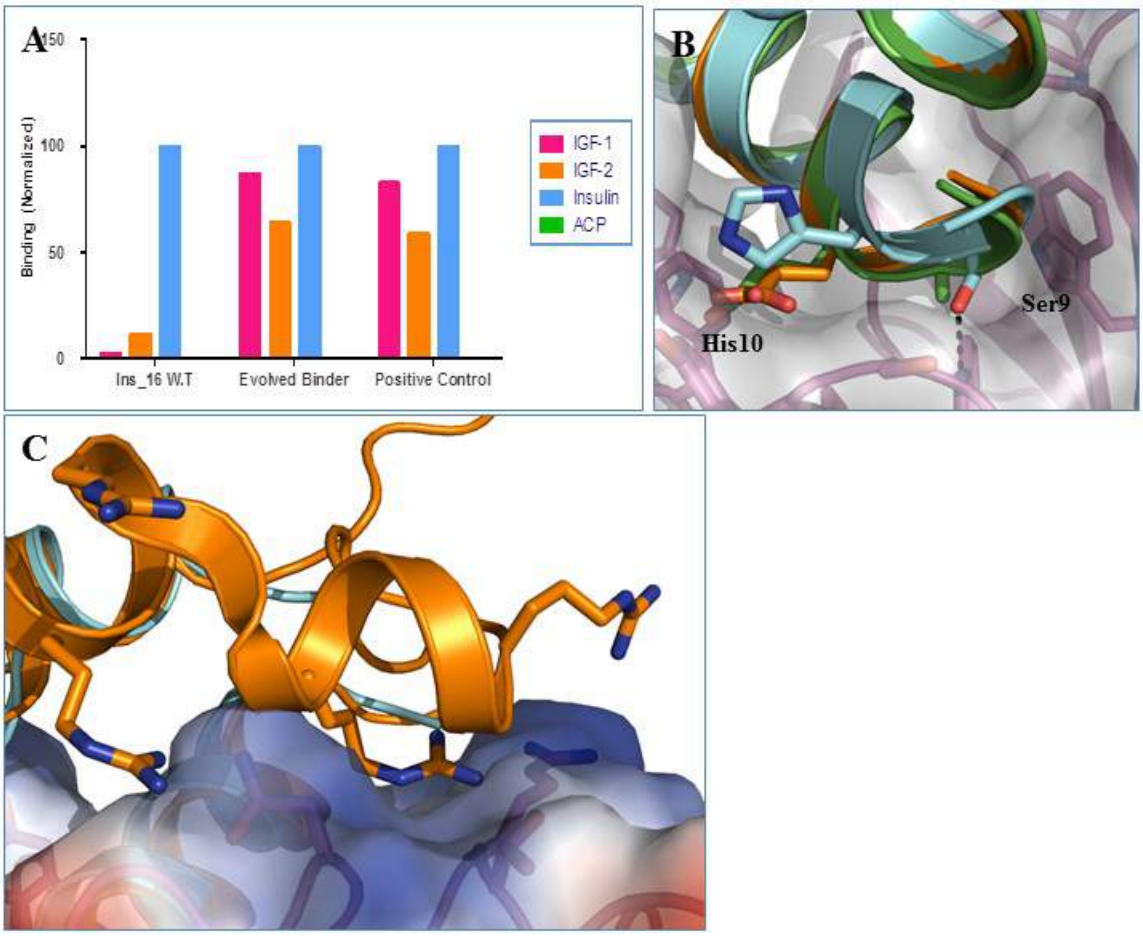


Figure 12: Specificity evaluation of the original and evolved binders.

(A) Comparison of binding affinity towards insulin and towards IGF-1, IGF-2, and ACP. 5Ins_16 W.T. exhibits higher specificity towards insulin than the evolved binder. ACP binding was not detected in either the original or the evolved design. **(B)** Alignment of IGF-1 (green) and IGF-2 (gold) to insulin (cyan) in its designed binding mode reveals minor identity changes of interface residues, reducing binding affinity (binder in magenta). **(C)** IGF-2 (gold) as opposed to insulin (cyan) possesses a positively charged surface countering the positive net-charge of both, W.T. and evolved binders.

References

1. Baron, R. & McCammon, J. A. Molecular recognition and ligand association. *Annu. Rev. Phys. Chem.* **64,** 151–75 (2013).

2. Strebhardt, K. & Ullrich, A. Paul Ehrlich's magic bullet concept: 100 years of progress. *Nat. Rev. Cancer* **8,** 473–80 (2008).

3. Georgiou, G. *et al.* The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* **32,** 158–68 (2014).

4. Vargas-Madrazo, E. & Paz-García, E. An improved model of association for VH-VL immunoglobulin domains: asymmetries between VH and VL in the packing of some interface residues. *J. Mol. Recognit.* **16,** 113–20 (2003).

5. Chothia, C. & Lesk, A. M. Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196,** 901–17 (1987).

6. North, B., Lehmann, A. & Dunbrack, R. L. A new clustering of antibody CDR loop conformations. *J. Mol. Biol.* **406,** 228–56 (2011).

7. Al-Lazikani, B., Lesk, a M. & Chothia, C. Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.* **273,** 927–48 (1997).

8. Winter, G. & Milstein, C. Man-made antibodies. *Nature* **349,** 293–9 (1991).

9. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* **15,** 553–7 (1997).

10. Finlay, W. J. J. & Almagro, J. C. Natural and man-made V-gene repertoires for antibody discovery. *Front. Immunol.* **3,** 342 (2012).

11. Miklos, A. E. *et al.* Structure-based design of supercharged, highly thermoresistant antibodies. *Chem. Biol.* **19,** 449–55 (2012).

12. Buss, N. a P. S., Henderson, S. J., McFarlane, M., Shenton, J. M. & de Haan, L. Monoclonal antibody therapeutics: history and future. *Curr. Opin. Pharmacol.* **12,** 615–22 (2012).

13. Lee, C. C., Perchiacca, J. M. & Tessier, P. M. Toward aggregation-resistant antibodies by design. *Trends Biotechnol.* 1–9 (2013). doi:10.1016/j.tibtech.2013.07.002

14. Lippow, S. M., Wittrup, K. D. & Tidor, B. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat. Biotechnol.* **25,** 1171–6 (2007).

15. Kuroda, D., Shirai, H., Jacobson, M. P. & Nakamura, H. Computer-aided antibody design. *Protein Eng. Des. Sel.* 1–15 (2012). doi:10.1093/protein/gzs024

16. Schreiber, G. & Fleishman, S. J. Computational design of protein-protein interactions. *Curr. Opin. Struct. Biol.* **23,** 903–10 (2013).

17. Fleishman, S. J. *et al.* Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332,** 816–21 (2011).

18. Jha, R. K. *et al.* Computational design of a PAK1 binding protein. *J. Mol. Biol.* **400,** 257–70 (2010).

19. Huang, P., Love, J. J. & Mayo, S. L. A de novo designed protein protein interface. *Protein Sci.* **16,** 2770–4 (2007).

20. Fleishman, S. J. & Baker, D. Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* **149,** 262–73 (2012).

21. Pan, P.-Y., Ozao, J., Zhou, Z. & Chen, S.-H. Advancements in immune tolerance. *Adv. Drug Deliv. Rev.* **60,** 91–105 (2008).

22. Starr, T. K., Jameson, S. C. & Hogquist, K. a. Positive and negative selection of T cells. *Annu. Rev. Immunol.* **21,** 139–76 (2003).

23. Palmer, E. Negative selection--clearing out the bad apples from the T-cell repertoire. *Nat. Rev. Immunol.* **3,** 383–91 (2003).

24. Hoogenboom, H. R. Selecting and screening recombinant antibody libraries. *Nat. Biotechnol.* **23,** 1105–16 (2005).

25. Boder, E. T., Midelfort, K. S. & Wittrup, K. D. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 10701–5 (2000).

26. Chao, G. *et al.* Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1,** 755–68 (2006).

27. Pandey, M. & Mahadevan, D. Monoclonal antibodies as therapeutics in human malignancies. *Futur. Oncol.* **10,** 609–636 (2014).

28. Das, R. & Baker, D. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **77,** 363–82 (2008).

29. Fleishman, S. J. *et al.* RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* **6,** e20161 (2011).

30. Hoover, D. M. & Lubkowski, J. DNAWorks : an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30,** 1–7 (2002).

31. Gietz, R. D. & Schiestl, R. H. Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2,** 1–4 (2007).

32. Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* **41,** W34–40 (2013).

33. Giudicelli, V., Chaume, D. & Lefranc, M.-P. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.* **33,** D256–61 (2005).

34. Benatuil, L., Perez, J. M., Belk, J. & Hsieh, C.-M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23,** 155–9 (2010).

35. Kortemme, T., Kim, D. E. & Baker, D. Computational alanine scanning of protein-protein interfaces. *Sci. STKE* **2004,** pl2 (2004).

36. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234,** 946–50 (1993).

37. Gietz, R. D. & Woods, R. a. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350,** 87–96 (2002).

38. Eriksson, a E. *et al.* Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255,** 178–83 (1992).

39. Sheffler, W. & Baker, D. RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci.* **18,** 229–39 (2009).

40. Richardson, J. S. & Richardson, D. C. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 2754–9 (2002).

41. Midelfort, K. S. *et al.* Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. *J. Mol. Biol.* **343,** 685–701 (2004).

42. Wu, T. T. & Kabat, E. A. 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.* **132,** 211–50 (1970).

43. Hoet, R. M. *et al.* Generation of high-affinity human antibodies by combining donor-derived and synthetic complementarity-determining-region diversity. *Nat. Biotechnol.* **23,** 344–8 (2005).

44. Larman, H. B., Xu, G. J., Pavlova, N. N. & Elledge, S. J. Construction of a rationally designed antibody platform for sequencing-assisted selection. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 18523–8 (2012).

45. Riechmann, L., Clark, M., Waldmann, H. & Winter, G. Reshaping human antibodies for therapy. *Nature* **332,** 323–7 (1988).

46. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. How protein stability and new functions trade off. *PLoS Comput. Biol.* **4,** e1000002 (2008).

47. Foit, L. *et al.* Optimizing protein stability in vivo. *Mol. Cell* **36,** 861–71 (2009).

48. Beadle, B. M. & Shoichet, B. K. Structural Bases of Stability–function Tradeoffs in Enzymes. *J. Mol. Biol.* **321,** 285–296 (2002).

49. Warszawski, S., Netzer, R., Tawfik, D. S. & Fleishman, S. J. A "fuzzy"-logic language for encoding multiple physical traits in biomolecules. *J. Mol. Biol.* (2014). doi:10.1016/j.jmb.2014.10.002

50. Kwong, P. D. & Mascola, J. R. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity* **37,** 412–25 (2012).

51. Moody, M. A. *et al.* H3N2 influenza infection elicits more cross-reactive and less clonally expanded anti-hemagglutinin antibodies than influenza vaccination. *PLoS One* **6,** e25797 (2011).

52. Wrammert, J. *et al.* Rapid cloning of high-affinity human monoclonal antibodies against influenza virus. *Nature* **453,** 667–71 (2008).

53. Ren, P. *et al.* Biomolecular electrostatics and solvation: a computational perspective. *Q. Rev. Biophys.* **45,** 427–91 (2012).

54. Gilson, M. K. Theory of electrostatic interactions in macromolecules. *Curr. Opin. Struct. Biol.* **5,** 216–223 (1995).

55. Baker, N. a, Sept, D., Joseph, S., Holst, M. J. & McCammon, J. a. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10037–41 (2001).

56. Whitehead, T. A. *et al.* Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* **30,** 543–8 (2012).

57. Acierno, J. P., Braden, B. C., Klinke, S., Goldbaum, F. a & Cauerhff, A. Affinity maturation increases the stability and plasticity of the Fv domain of anti-protein antibodies. *J. Mol. Biol.* **374,** 130–46 (2007).

58. Murphy, G. S. *et al.* Computational de novo design of a four-helix bundle protein - DND_4HB. *Protein Sci.* 1–32 (2014). doi:10.1002/pro.2577

59. Huang, P., Love, J. J. & Mayo, S. L. A de novo designed protein protein interface. *Protein Sci.* **16,** 2770–4 (2007).

60. Schreiber, G. & Fleishman, S. J. Computational design of protein-protein interactions. *Curr. Opin. Struct. Biol.* **23,** 903–10 (2013).

61. Li, T., Pantazes, R. J. & Maranas, C. D. OptMAVEn--a new framework for the de novo design of antibody variable region models targeting specific antigen epitopes. *PLoS One* **9,** e105954 (2014).

62. Blundell, T. L. *et al.* Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **361,** 413–23 (2006).

63. Davids, T., Schmidt, M., Böttcher, D. & Bornscheuer, U. T. Strategies for the discovery and engineering of enzymes for biocatalysis. *Curr. Opin. Chem. Biol.* **17,** 215–20 (2013).

64. Coelho, E. D., Arrais, J. P. & Oliveira, J. L. From protein-protein interactions to rational drug design: are computational methods up to the challenge? *Curr. Top. Med. Chem.* **13,** 602–18 (2013).

65. Zhang, H., Wilson, I. A. & Lerner, R. A. Selection of antibodies that regulate phenotype from intracellular combinatorial antibody libraries. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 15728–33 (2012).

66. Liu, T. *et al.* Rational design of CXCR4 specific antibodies with elongated CDRs. *J. Am. Chem. Soc.* **136,** 10557–60 (2014).

67. Hoet, R. M. *et al.* Generation of high-affinity human antibodies by combining donor-derived and synthetic complementarity-determining-region diversity. *Nat. Biotechnol.* **23,** 344–8 (2005).

68. Bernett, M. J. *et al.* Engineering fully human monoclonal antibodies from murine variable regions. *J. Mol. Biol.* **396,** 1474–90 (2010).

69. Almagro, J. C. Identification of differences in the specificity-determining residues of antibodies that recognize antigens of different size: implications for the rational design of antibody repertoires. *J. Mol. Recognit.* **17,** 132–43 (2004).

Supplementary information

Table S1. Scores and experimental expression of the initial design algorithm (section A)

Design Name	Total Score	Predicted Binding Energy (R.e.u)	Packing [b]	Binding Surface Area (Å) [a]	Shape complementarity [c]	Expression (%) [d]
1Ins_1	-495.849	-41.083	0.614	1708.756	0.678	13.8
1Ins_2	-456.209	-29.82	0.621	1344.616	0.629	0.2
1Ins_3	-450.818	-39.399	0.598	1767.845	0.7	1.2
1Ins_4	-491.035	-44.216	0.671	1816.745	0.725	1.5
1Ins_5	-486.069	-31.125	0.64	1484.892	0.497	1
1Ins_6	-479.888	-32.012	0.598	1634.637	0.662	4.6
1Ins_7	-505.342	-40.593	0.583	1926.641	0.615	11.3
1Ins_8	-482.806	-31.236	0.638	1562.815	0.766	3.4
1Ins_9	-438.411	-28.671	0.573	1630.722	0.639	2.6
1Ins_10	-498.366	-22.627	0.657	1378.111	0.593	2.9
1Ins_11	-504.979	-33.036	0.571	1660.087	0.647	0.1
1Ins_12	-493.492	-30.991	0.657	1326.133	0.703	1.3
1Ins_13	-482.931	-28.563	0.623	1387.389	0.617	0.7
1Ins_14	-495.185	-40.577	0.577	1684.938	0.731	0.8
1Ins_15	-487.141	-36.739	0.599	1815.692	0.696	3.2
1Ins_16	-483.471	-30.773	0.679	1712.469	0.647	3.2
1Ins_17	-500.99	-36.696	0.613	1720.974	0.619	1.2
1Ins_1 [SW]	-495.849	-41.083	0.614	1708.756	0.678	17.9
1Ins_6 [SW]	-479.888	-32.012	0.598	1634.637	0.662	6.5
1Ins_7 [SW]	-505.342	-40.593	0.583	1926.641	0.615	26.4

[a] - Binding surface area is the area excluded from solvent upon ligand binding.
[b] - A criterion that measures how well the antibody core and the binding surface are packed.
[c] - Shape complementarity between the binding surfaces of antibody and ligand.
[d] - Experimental expression observed by the fractional bounded population. Non-expressing scFvs shown in red (<4.6%), low expressing scFvs shown in orange (4.6-20.5%), medium expressing scFv shown in green (20.5-40%) and highly expressing scFvs shown in dark-blue (>40%).
[SW] - Modified designs in which the original VH-VL order was switched and the original linker replaced by a $(Gly_{(4)}Ser)_3$ linker.

Table S2. Scores and experimental expression of the second iteration (section B)

Design Name	Total Score	Predicted Binding Energy (R.e.u)	Packing [b]	Binding Surface Area (Å) [a]	Shape Complementarity [c]	Expression (%) [d]
'2Ins_2	-404.924	-34.992	0.674	1967.859	0.677	44.2
'2Ins_3	-487.809	-29.354	0.641	1558.341	0.731	39.3
'2Ins_4	-481.913	-22.099	0.662	1468.558	0.689	36.1
'2Ins_5	-463.76	-23.15	0.654	1366.986	0.681	6.9
'2Ins_6	-479.103	-25.216	0.652	1528.704	0.694	2.6
'2Ins_7	-457.469	-31.597	0.66	1682.507	0.668	4.6
'2Ins_8	-470.757	-27.166	0.68	1678.209	0.691	33.9
'2Ins_9	-481.703	-32.827	0.671	1435.371	0.702	22.2
'2Ins_10	-492.642	-37.01	0.643	1735.768	0.679	7.5
'2Ins_11	-457.471	-27.261	0.665	1453.1	0.669	21.4
'2Ins_12	-464.152	-27.627	0.668	1482.59	0.666	7.1
'2Ins_13	-485.039	-34.46	0.605	1960.685	0.647	6.4
'2Ins_14	-475.739	-28.351	0.64	1592.869	0.59	23.3
'2Ins_15	-501.462	-37.016	0.601	1688.948	0.663	11.9
2Ins_2	-454.341	-20.314	0.593	1726.89	0.623	11.32
2Ins_3	-488.13	-29.333	0.658	1718.515	0.62	19.3
2Ins_5	-472.721	-33.077	0.64	1704.294	0.679	15.3
2Ins_6	-448.091	-32.863	0.632	1834.101	0.602	13.5
2Ins_7	-505.055	-39.723	0.66	1906.876	0.71	34.5
2Ins_8	-461.941	-30.225	0.654	1431.213	0.795	32.5
2Ins_9	-465.126	-25.053	0.622	1704.818	0.669	11.4
2Ins_10	-486.629	-20.718	0.663	1511.809	0.586	4.3
2Ins_11	-451.601	-34.058	0.599	1772.358	0.654	18.9
2Ins_12	-448.455	-30.992	0.609	1656.872	0.579	10.3
2Ins_14	-507.408	-41.248	0.687	1878.534	0.673	25.9
2Ins_15	-467.095	-20.507	0.636	1419.087	0.556	37.2
2Ins_16	-507.91	-28.542	0.697	1733.699	0.685	20.1
2Ins_17	-495.849	-33.644	0.627	1708.756	0.678	17
2Ins_18	-479.888	-28.322	0.629	1634.637	0.662	6.14
2Ins_19	-505.342	-36.415	0.612	1926.641	0.615	22.9

[a] - Binding surface area is the area excluded from solvent upon ligand binding.
[b] - A criterion that measures how well the antibody core and the binding surface are packed.
[c] - Shape complementarity between the binding surfaces of the antibody and ligand.
[d] - Experimental expression observed by the fractional bounded population. non-expressing scFvs shown in red (<4.6%), low expressing scFvs shown in orange (4.6-20.5%), medium expressing scFv shown in green (20.5-40%) and highly expressing scFvs shown in dark-blue (>40%).

Table S3. Position identity conservation (section C)

Kabat Position	Original Identity	Restricted Identities
L33	Leu	L,I,V
L91	Ser	A,F,G,I,L,M,N,P,Q,S,T,V,W,Y
H101	Glu	A,F,G,I,L,M,N,P,Q,S,T,V,W,Y
H103	Leu	L,I,W,F
L71	Phe	F,Y
L24	Arg	Restricted identity
L45	Lys	Restricted identity
L70	Asp	Restricted identity
L72	Thr	Restricted identity
H23	Val	Restricted identity
H25	Ser	Restricted identity
H27	Phe	Restricted identity
H29	Phe	Restricted identity
H57	The	Restricted identity
H70	Ser	Restricted identity
H77	Ser	Restricted identity
H79	Tyr	Restricted identity
H106	Gly	Restricted identity

Residue identities and positions that were restricted in the third version of the design algorithm.

Table S3A. Scores and experimental expression of the third iteration (section C)

Design Name	Total Score	Predicted Binding Energy (R.e.u)	Packing [b]	Binding Surface Area (Å) [a]	Shape Complementarity [c]	Expression (%) [d]
3aIns_1	-479.946	-20.961	0.675	1392.02	0.717	32.08
3aIns_2	-476.196	-23.29	0.676	1634.427	0.709	27.80
3aIns_3	-498.346	-29.734	0.611	1657.709	0.696	35.53
3aIns_4	-510.11	-27.303	0.678	1602.795	0.671	11.82
3aIns_5	-502.594	-30.149	0.66	1692.129	0.679	24.67
3aIns_6	-490.4	-22.758	0.652	1381.043	0.667	28.99
3aIns_7	-499.014	-25.781	0.691	1508.079	0.715	53.10
3aIns_9	-494.526	-24.765	0.668	1444.505	0.66	5.83
3aIns_10	-501.115	-24.062	0.653	1327.362	0.7	40.30
3aIns_11	-503.711	-23.287	0.625	1366.629	0.68	24.40
3aIns_12	-450.465	-21.579	0.65	1515.44	0.727	44.04
3aIns_13	-456.229	-31.041	0.638	1667.313	0.767	2.90
3aIns_14	-442.635	-20.566	0.661	1164.092	0.742	30.38
3aIns_15	-454.385	-35.03	0.653	1757.234	0.701	13.05
3aIns_16	-457.727	-14.474	0.682	1343.542	0.773	1.74
3aIns_17	-439.968	-25.958	0.607	1718.874	0.685	16.07
3aIns_18	-451.601	-22.373	0.612	1604.79	0.678	47.46
3aIns_19	-458.308	-18.364	0.651	1223.923	0.692	34.17
3aIns_20	-459.035	-23.443	0.688	1375.404	0.725	59.55
3aIns_21	-446.828	-22.339	0.658	1489.794	0.659	53.61
3aIns_22	-461.91	-23.653	0.655	1425.075	0.68	17.66
3aIns_23	-459.626	-21.693	0.691	1286.388	0.684	1.13
3aIns_24	-477.156	-17.296	0.683	1217.687	0.709	11.86
3aIns_25	-469.216	-23.592	0.655	1589.617	0.673	20.24
3aIns_26	-462.968	-18.044	0.655	1613.808	0.681	15.33
3aIns_28	-450.921	-20.834	0.672	1300.851	0.759	2.79
3aIns_29	-435.394	-18.159	0.633	1643.127	0.691	1.15
3aIns_30	-457.95	-18.525	0.62	1336.739	0.673	21.57
3aIns_31	-465.686	-20.976	0.645	1393.843	0.648	59.48
3aIns_32	-437.942	-21.855	0.687	1466.796	0.783	43.12
3aIns_33	-457.751	-19.858	0.666	1285.558	0.655	24.59
3aIns_34	-459.961	-28.783	0.654	1721.786	0.677	17.41
3aIns_35	-443.483	-17.016	0.595	1312.353	0.711	29.00
3bIns_1	-450.117	-23.159	0.679	1368.836	0.72	17.27
3bIns_2	-481.787	-25.261	0.691	1471.31	0.73	59.76
3bIns_3	-421.179	-24.908	0.697	1304.183	0.766	43.25
3bIns_4	-452.18	-30.708	0.658	1649.552	0.721	1.32
3bIns_5	-452.018	-24.434	0.667	1506.064	0.714	46.31
3bIns_6	-470.404	-27.421	0.67	1421.592	0.765	17.98

3bIns_7	-450.041	-30.165	0.673	1586.748	0.733	28.03
3bIns_8	-484.087	-29.113	0.655	1731.922	0.704	17.81
3bIns_10	-490.268	-34.381	0.652	1814.713	0.683	2.21
3bIns_11	-481.662	-30.781	0.662	1674.093	0.733	13.66
3bIns_12	-466.148	-25.425	0.642	1458.171	0.578	21.32
3bIns_13	-446.379	-23.895	0.633	1642.005	0.651	24.79
3bIns_14	-461.691	-32.541	0.679	1565.8	0.812	48.64
3bIns_15	-443.86	-21.628	0.65	1532.266	0.792	27.53
3bIns_16	-439.547	-31.125	0.669	1482.089	0.708	52.04

[a] - Binding surface area is the area excluded from solvent upon ligand binding.
[b] - A criterion that measures how well the antibody core and the binding surface are packed.
[c] - Shape complementarity between the binding surfaces of the antibody and ligand.
[d] - Experimental expression observed by the fractional bounded population. non-expressing scFvs shown in red (<4.6%), low expressing scFvs shown in orange (4.6-20.5%), medium expressing scFv shown in green (20.5-40%) and highly expressing scFvs shown in dark-blue (>40%).

Table S4. Scores and experimental expression of the fourth iteration (section D)

Design Name	Total Score	Predicted Binding Energy (R.e.u)	Packing [b]	Binding Surface Area (Å) [a]	Shape Complementarity [c]	Expression (%) [d]
5Ins_2	-424.944	-23.535	0.653	1546.039	0.666	75.60
5Ins_3	-424.883	-29.443	0.626	1587.874	0.708	76.70
5Ins_4	-420.759	-28.159	0.666	1684.834	0.707	77.50
5Ins_5	-402.578	-20.867	0.689	1519.306	0.72	71.50
5Ins_6	-415.838	-24.57	0.697	1553.412	0.759	82.40
5Ins_7	-387.708	-21.031	0.671	1613.86	0.657	72.90
5Ins_8	-398.704	-30.837	0.661	1703.929	0.693	77.90
5Ins_9	-416.417	-17.309	0.655	1888.964	0.614	74.10
5Ins_10	-429.493	-23.384	0.648	1473.735	0.653	72.10
5Ins_11	-353.668	-22.194	0.638	1529.724	0.571	79.70
5Ins_12	-421.206	-24.976	0.69	1742.764	0.715	74.00
5Ins_13	-402.412	-29.179	0.634	1551.148	0.729	78.40
5Ins_14	-445.334	-25.288	0.716	1694.903	0.678	80.30
5Ins_15	-364.749	-21.697	0.632	1623.501	0.628	66.50
5Ins_16	-398.237	-18.15	0.657	1469.701	0.628	72.50
5Ins_17	-403.915	-23.44	0.657	1378.826	0.671	77.00
5Ins_18	-395.924	-17.094	0.665	1521.381	0.641	80.80
5Ins_19	-411.3	-24.333	0.654	1455.946	0.616	74.80
5Ins_20	-404.817	-25.808	0.65	1671.377	0.631	70.90

[a] - Binding surface area is the area excluded from solvent upon ligand binding.
[b] - A criterion that measures how well the antibody core and the binding surface are packed.
[c] - Shape complementarity between the binding surfaces of the antibody and ligand.
[d] - Experimental expression observed by the fractional bounded population. non-expressing scFvs shown in red (<4.6%), low expressing scFvs shown in orange (4.6-20.5%), medium expressing scFv shown in green (20.5-40%) and highly expressing scFvs shown in dark-blue (>40%).

Table S5. Averaged scores and experimental expression of all algorithm designs iteration

Design Name	Total Score	Predicted Binding Energy (R.e.u)	Packing [b]	Binding Surface Area (Å) [a]	Shape Complementarity [c]	Expression (%) [d]
1Ins_#	-484.29±19.01	-34.01±5.76	0.618±0.03	1621.4±180.6	0.657±0.06	3.118±3.79
2Ins_#	-474.98±22.47	-30.11±5.63	0.644±0.02	1662.3±170.9	0.661±0.04	18.932±11.95
3Ins_#	-464.68±21.09	-24.43±4.8	0.658±0.02	1493.1±161.9	0.706±0.04	26.557±17.34
5Ins_#	-406.47±21.75	-23.75±3.95	0.661±0.02	1590.1±120.8	0.668±0.04	75.558±4.01

[a] - Binding surface area is the area excluded from solvent upon ligand binding.
[b] - A criterion that measures how well the antibody core and the binding surface are packed.
[c] - Shape complementarity between the binding surfaces of the antibody and ligand.
[d] - Experimental expression observed by the fractional bounded population. non-expressing scFvs shown in red (<4.6%), low expressing scFvs shown in orange (4.6-20.5%), medium expressing scFv shown in green (20.5-40%) and highly expressing scFvs shown in dark-blue (>40%).

Table S6. Protein sequences of all designs tested

>1Ins_1
SDVVMTQTPLSLPVSLGDQASISCRSSQSAVLRNGLTFLFWYLQKPGQSPKVLIYKVSNRVSGVPDR
FSGSGFGTDFTLKINRVEAEDLGVYFCAADSWMANENRFGGGTKLEIKSSADDAKKDAAKKDDA
KKDDAKKDGGVKLDETGGGLVQPGGAMKLSCVTLGFNFMEYAMNWVRQSPEKGLEWVAAFYL
NNSAQYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTADARQNQQGNAAVKSGQGTSV
TVS
>1Ins_2
SDVVMTQTPLSLPVSLGDQASISCISADAEAYYSFVMFQLQKPGQSPKVLIYKVSNRVSGVPDRFYG
TGMGILYFLKINRVEAEDLGVYFCGVNWSKSIKFGQGTKLEIKSSADDAKKDAAKKDDAKKDDAK
KDGGVKLDETGGGLVQPGGAMKLSCVTVGFVFAFYLMNWVRQSPEKGLEWVAAFDPWSDSTEY
ADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCEASESKFNSSSQAAGQGTSVTVS
>1Ins_3
SDVVMTQTPLSLPVSLGDQASISCIAAKMAISWDGDVFDRWYLQKPGQSPKVLIYKVSNRVSGVPD
RFEGTGRQRMYYLKINRVEAEDLGVYFCGSFGPFGEETSYGFGTKLEIKSSADDAKKDAAKKDDA
KKDDAKKDGGVKLDETGGGLVQPGGAMKLSCTGSGFNVTGYMMAWVRQSPEKGLEWVGSYAY
MEWMVAWADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCRATANGLFTDAQMGWGTSVT
VS
>1Ins_4
SDVVMTQTPLSLPVSLGDQASISCRSSQSLWMLNGVIALAWYLQKPGQSPKVLIYKVSNRVSGVPD
RFSGSGNGNDFTLKINRVEAEDLGVYFCAGAARGYQDFGPGTKLEIKSSADDAKKDAAKKDDAKK
DDAKKDGGVKLDETGGGLVQPGGAMKLSCVTEGFVFGDTIMNWVRQSPEKGLEWVAGFYFNQD
ASYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCRSSSMYSKNYAGRGTSVTVS
>1Ins_5
SDVVMTQTPLSLPVSLGDQASISCRSSQYLVMSDGETRLRWYLQKPGQSPKVLIYKVSNRVSGVPD
RFSGSGSGTDFTLKINRVEAEDLGVYFCAAQDSEANRYSFGWGTKLEIKSSADDAKKDAAKKDDA
KKDDAKKDGGVKLDETGGGLVQPGGAMKLSCVTAGFYFGKYLMNWVRQSPEKGLEWVAFWAY
ANAAYYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCEAAAYAYDNAYRTAAAYAGAGT
SVTVS
>1Ins_6
SDVVMTQTPLSLPVSLGDQASISCKSKKTSFKDTFWYLQKPGQSPKVLIYKVSNRVSGVPDRFYGT
GEGREYKLKINRVEAEDLGVYFCSQFFSYEVSFGYGTKLEIKSSADDAKKDAAKKDDAKKDDAKK
DGGVKLDETGGGLVQPGGAMKLSCTASGLDSSRAFYMWVRQSPEKGLEWVATYTSGNLAYYAD
SVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAIRQSKLNSAMAVMGQGTSVTVS
>1Ins_7
SDVVMTQTPLSLPVSLGDQASISCFAKFMVYDYWKNNSHVAWYLQKPGQSPKVLIYKVSNRVSGV
PDRFYGTGSGRFFRLKINRVEAEDLGVYFCAQRASIPWAATAGGGTKLEIKSSADDAKKDAAKKD
DAKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCRAEGVDDSEMTFEWVRQSPEKGLEWVAAFT
DNNSAAYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCEAASENISNTAFIYIGDGTSVTVS
>1Ins_8
SDVVMTQTPLSLPVSLGDQASISCRSSLSLKDSSGMTLTAWYLQKPGQSPKVLIYKVSNRVSGVPD
RFKGEGEGKDFVLKINRVEAEDLGVYFCAQFSIAPFTFGFGTKLEIKSSADDAKKDAAKKDDAKKD
DAKKDGGVKLDETGGGLVQPGGAMKLSCVTTGFNFGDTIMNWVRQSPEKGLEWVASFWSGNLT
RYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCAAAKTGLAWWMGDGTSVTVS
>1Ins_9
SDVVMTQTPLSLPVSLGDQASISCFSESSILDNSGSTSSAWYLQKPGQSPKVLIYKVSNRVSGVPDRF
YGSGSGKTFRLKINRVEAEDLGVYFCSDDQSRKKTFGFGTKLEIKSSADDAKKDAAKKDDAKKDD
AKKDGGVKLDETGGGLVQPGGAMKLSCDSQGFWFWDYSFSWVRQSPEKGLEWVATFFSNNLSA
YADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCAADTIAKAFSGDGTSVTVS
>1Ins_10
SDVVMTQTPLSLPVSLGDQASISCKSSKYVYQKNFNESKVRWYLQKPGQSPKVLIYKVSNRVSGVP
DRFEGTGSGREYKLKINRVEAEDLGVYFCAAWESDQGRSKFGGGTKLEIKSSADDAKKDAAKKDD

AKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCDTQGFTARKSVFMWVRQSPEKGLEWVAAFW
NDNLAFYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCESASGSTAWVGLYTGLGTSVTV
S
>1Ins_11
SDVVMTQTPLSLPVSLGDQASISCWASTTLIDSSGFTGLAWYLQKPGQSPKVLIYKVSNRVSGVPDR
FSGSGAGTFYTLKINRVEAEDLGVYFCAQNMYAPRTFGGGTKLEIKSSADDAKKDAAKKDDAKKD
DAKKDGGVKLDETGGGLVQPGGAMKLSCVTSGFRFFYYIMNWVRQSPEKGLEWVAQYEDLTNN
YRAMYSDSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCVSDAGAFTGAAWIGKGTSVTVS
>1Ins_12
SDVVMTQTPLSLPVSLGDQASISCRSSGSLMHSSGWTSTSWYLQKPGQSPKVLIYKVSNRVSGVPD
RFSGSGSQTDFTLKINRVEAEDLGVYFCAQSFQSRWAFGGGTKLEIKSSADDAKKDAAKKDDAKK
DDAKKDGGVKLDETGGGLVQPGGAMKLSCVTEGFTFRNSFMNWVRQSPEKGLEWVAEFRNLPTN
YNTAYSDSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTAASDFSESDWYSGQGTSVTVS
>1Ins_13
SDVVMTQTPLSLPVSLGDQASISCRSSFSLTSADGFTRAVWYLQKPGQSPKVLIYEVSNRVSGVPDR
FSGSGSGNEFTLKINRVEAEDLGVYFCGAASSDYWSYGRGTKLEIKSSADDAKKDAAKKDDAKKD
DAKKDGGVKLDETGGGLVQPGGAMKLSCVGRGFEFKTYTMNWVRQSPEKGLEWVAGFKNNNW
WRYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTSSTSRQEMYDAMALWMGQGTSVTV
S
>1Ins_14
SDVVMTQTPLSLPVSLGDQASISCKAKFQLYYESNNEARAVWMLQKPGQSPKVLIYKVSNRVSGV
PDRFYGEGSGAEFRLKINRVEAEDLGVYFCAAVSDRSNKSSYSFGGGTKLEIKSSADDAKKDAAKK
DDAKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCYAYGTKARDLIFSWVRQSPEKGLEWVASF
MWNNAEFYADSVEGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCSAASARFSYVIWYSGDGTSVT
VS
>1Ins_15
SDVVMTQTPLSLPVSLGDQASISCFSSDSLINNETNLAITIWYLQKPGQSPKVLIYMVSNRVSGVPDR
FYGEGSGKSYRLKINRVEAEDLGVYFCGGSSNYAARFGGGTKLEIKSSADDAKKDAAKKDDAKKD
DAKKDGGVKLDETGGGLVQPGGAMKLSCYMAGFWFQGVVFMWVRQSPEKGLEWVAAFYWNN
EAAYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCFASIANVTTYVGDGTSVTVS
>1Ins_16
SDVVMTQTPLSLPVSLGDQASISCRSQENAKSNIAWYLQKPGQSPKVLIYKVSNRVSGVPDRYSGS
GRGKDFTLKINRVEAEDLGVYFCAAYSWRSYKRRFGGGTKLEIKSSADDAKKDAAKKDDAKKDD
AKKDGGVKLDETGGGLVQPGGAMKLSCVTTGFLFYWYQMNWVRQSPEKGLEWVAGFDPTNGFV
MYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMASWSGDTFYEGQGTSVTVS
>1Ins_17
SDVVMTQTPLSLPVSLGDQASISCKAAFDLWNFWGEIFLTWFLQKPGQSPKVLIYKVSNRVSGVPD
RFYGTGSGQKFKLKINRVEAEDLGVYFCAQDAWYPSKTQFGPGTKLEIKSSADDAKKDAAKKDDA
KKDDAKKDGGVKLDETGGGLVQPGGAMKLSCYSYGFSSSAYIYTWVRQSPEKGLEWVAQFTSNP
FNYWTYYSDSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAAKFGLSGSLWAGSGTSVTVS

>2Ins_2
GGGGSGGGGSEVKLDETGGGLVQPGGAMKLSCVTSGFTFSQYIMFWVRQSPEKGLEWVAAFSGT
NTSTYYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMGSGSGGDSSTYSGDGTSVTVSSG
GGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSQQSVSQLQWYLQKPGQSPKVLISEVSN
RVSGVPDRFSGTGSGTDYTLKINRVEAEDLGVYFCAQEATVPWTFGGGTKLEIK
>2Ins_3
EVKLDETGGGLVQPGGAMKLSCVASGFTFSKYYMGWVRQSPEKGLEWVAIFQFANTTYYSDSVK
GRFTISRDDSKSSVYLQMNNLRVEDTGIYYCWGSSDNASMAFFYNGSGTSVTVSSGGGGSGGGGS
GGGGSDVVMTQTPLSLPVSLGDQASISCRSSKSLVTSSGNTYLVWFLQKPGQSPKVLIYDVSKRVS
GVPDRFSGTGSGTDFTLKINRVEAEDLGVYFCAQFTNSTTFFTFGGGTKLEIK
>2Ins_5

EVKLDETGGGLVQPGGAMKLSCVTSGFRFATYMMMWVRQSPEKGLEWVANFDSDNYTNYADSV
KGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCSGPKTNSTSSTYLGQGTSVTVSSGGGGSGGGGSG
GGGSDVVMTQTPLSLPVSLGDQASISCRSSSSLSKYLEWYLQKPGQSPKVLIYNTSNRVSGVPDRFS
GSGSGTDFTLKINRVEAEDLGVYFCGQWSYWPSTFGGGTKLEIK
>2Ins_6
EVKLDETGGGLVQPGGAMKLSCVTSGFTFSYFVMVWVRQSPEKGLEWVAAFYSGNSTYYADSVK
GRFTISADASKSSVYLQMNNLRVEDTGIYYCIGTAGTTYLGQGTSVTVSSGGGGSGGGGSGGGGSD
VVMTQTPLSLPVSLGDQASISCRSSQSLYESLYWYLQKPGQSPKVLIYKVSNRVSGVPDRFYGSGSG
TDYTLKINRVEAEDLGVYFCAAYSSESRAITFGGGTKLEIK
>2Ins_7
EVKLDETGGGLVQPGGAMKLSCVASGFTFGHYEMVWVRQSPEKGLEWVANFSQSNNTNYADSV
KGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAAGSDGAWWAVSTMIYTGQGTSVTVSSGGGG
SGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSKSLVWSSGWTLLSWYLQKPGQSPKVLIYK
VSNRVSGVPDRFSGTGSGTDFTLKINRVEAEDLGVYFCSQSSYVPWTFGGGTKLEIK
>2Ins_8
EVKLDETGGGLVQPGGAMKLSCVTSGFTFGKTVMGWVRQSPEKGLEWVAAYYWSNITFYADSV
KGRFTISRDESKSSVYLQMNNLRVEDTGIYYCGGSGGFGEFVFRYLGNGTSVTVSSGGGGSGGGGS
GGGGSDVVMTQTPLSLPVSLGDQASISCRSSQSLVDSDGNTRLAWYLQKPGQSPKVLIYNVSNRVS
GVPDRFSGTGSGTDFTLKINRVEAEDLGVYFCAQSTDAPWTSGGGTKLEIK
>2Ins_9
EVKLDETGGGLVQPGGAMKLSCVTSGFTFGSYKMSWVRQSPEKGLEWVAAFSPTVNVTSYADSV
KGRFTISSDGSKSSVYLQMNNLRVEDTGIYYCMAPNPVLMSSLGQGTSVTVSSGGGGSGGGGSGG
GGSDVVMTQTPLSLPVSLGDQASISCRSSQVVSNYLGWYLQKPGQSPKVLIYKVSNRVSGVPDRFY
GTGSGTDFTLKINRVEAEDLGVYFCMADFADSYSFGGGTKLEIK
>2Ins_10
EVKLDETGGGLVQPGGAMKLSCVTSGFTFGKYTMSWVRQSPEKGLEWVAAFYWVNLTVYADSV
KGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCVYSSSWSSSMTSNWGQGTSVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRSSLPLVSRDGSPNLVWFLQKPGQSPKMIIYNVSNRVS
GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAAYSSDLDAFWFFGGTKLEIK
>2Ins_11
EVKLDETGGGLVQPGGAMKLSCVTSGFTFGWYLMSWIRQSPEKGLEWVAIFAYNNQTRYSDSVK
GRFTISRDDSKSSVYLQMNNLRVEDTGIYYCGGLFWIYLGQGASVTVSSGGGGSGGGGSGGGGSD
VVMTQTPLSLPVSLGDQASISCRSSMNVTDDLYWMLQKPGQSPKVLIYKVSQRVSGVPDRFTGSG
QGTDFTLKINRVEAEDLGVYFCAQSAWWPSTFGGGTKLEIK
>2Ins_13
EVKLDETGGGLVQPGGAMKLSCVTSGFDFSNYSMMWVRQMPEKGLEWVAAFYNNNSTTYSDSV
KGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMGSSYTSSSTYLGQGTSVTVSSGGGGSGGGGSGG
GGSDVVMTQTPLSLPVSLGDQASISCRGDSTNVSYTTLSWYLQKPGQSPKVLIYNGSNRVSGVPDR
FSGSGSSTDYTLKINRVEAEDLGVYFCMTFWSEFITFGPGTKLEIK
>2Ins_14
EVKLDETGGGLVQPGGAMKLSCVTSGFTFGHYSMLWVRQSPEKGLEWVAMFAPADYSTYYADD
VKGRFTISSDPSKSSVYLQMNNLRVEDTGIYYCWASSYTPNFSSSTADLYSGQGTSVTVSSGGGGSG
GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCSMASDIQPYSALTWYLQKPGQSPKVLIYKVSLRV
SGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCASWNETFYYRTFGGGTKLEIK
>2Ins_15
EVKLDETGGGLVQPGGAMKLSCVTSGFTFGNYIMNWVRQMPEKGLEWVASFTPKPDNNSTRYSD
SVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTASSSTSSSLMTYWGQGTSVTVSSGGGGSGGG
GSGGGGSDVVMTQTPLSLPVSLGDQAQISCRSSSPLSNFLAWYLQKPGQSPKVLIYQVSNRVSGVP
DRFSGSGNGTDFTLKINRVEAEDLGVYFCSQFSVNPPTFGGGTKLEIK
>2Ins_16
EVKLDETGGGLVQPGGAMKLSCVTSGFTFSKYDMNWMRQSPEKGLEWVASFYDDNDTSYADSV
KGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTTSSQDETSSSDFSWYLGQGTSVTVSSGGGGSGG
GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSHSLTSSSGHTTLAWFLQKPGQSPKVLIYKVSNR
VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQSWFSPFTFGPGTKLEIK

>2Ins_17
EVKLDETGGGLVQPGGAMKLSCVTLGFNFMEYAMNWVRQSPEKGLEWVAAFYLNNSAQYADSV
KGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTADARQNQQGNAAVKSGQGTSVTVSSGGGGSGG
GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSQSAVLRNGLTFLFWYLQKPGQSPKVLIYKVSN
RVSGVPDRFSGSGFGTDFTLKINRVEAEDLGVYFCAADSWMANENRFGGGTKLEIK
>2Ins_18
EVKLDETGGGLVQPGGAMKLSCTASGLDSSRAFYMWVRQSPEKGLEWVATYTSGNLAYYADSV
KGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAIRQSKLNSAMAVMGQGTSVTVSSGGGGSGGG
GSGGGGSDVVMTQTPLSLPVSLGDQASISCKSKKTSFKDTFWYLQKPGQSPKVLIYKVSNRVSGVP
DRFYGTGEGREYKLKINRVEAEDLGVYFCSQFFSYEVSFGYGTKLEIK
>2Ins_19
EVKLDETGGGLVQPGGAMKLSCRAEGVDDSEMTFEWVRQSPEKGLEWVAAFTDNNSAAYADSV
KGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCEAASENISNTAFIYIGDGTSVTVSSGGGGSGGGGS
GGGGSDVVMTQTPLSLPVSLGDQASISCFAKFMVYDYWKNNSHVAWYLQKPGQSPKVLIYKVSN
RVSGVPDRFYGTGSGRFFRLKINRVEAEDLGVYFCAQRASIPWAATAGGGTKLEIK
>'2Ins_2
EVKLDETGGGLVQPGGAMKLSCVTTGFDSSKYEMNWVRQSPEKGLEWVAATSPLNYYYSYADSV
KGRFTISKDNSKSSVYLQMNNLRVEDTGIYYCMASSYGSTAWTWTGNGTSVTVSSGGGGSGGGGS
GGGGSDVVMTQTPLSLPVSLGDQASISCRALTEKAGNRTVTWYLQKPGQSPKVLIYMVSNRVSGV
PDRFSGSGSGSDYTLKINRVEAEDLGVYFCATYSSDNFKQFFGGGTK
>'2Ins_3
EVKLDETGGGLVQPGGAMKLSCVTSGFDFSKAVMNWVRQSPEKGLEWVAAFVPVKNMTFYADS
VKGRFTISADTSKSSVYLQMNNLRVEDTGIYYCMGAVWTDNDELDYWVMYTGQGTSVTVSSGGG
GSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCYAAVDTSRYVAWYLQKPGQSPKVLIYKVSN
RVSGVPDRFSGSGQGRVYTLKINRVEAEDLGVYFCAQYAFWPWSWGGGTKLEIK
>'2Ins_4
EVKLDETGGGLVQPGGAMKLSCVTTGFNFGKSYMNWIRQSPEKGLEWVASFWTLNWTYYADSV
KGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCAASESASTSTMWIGDGTSVTVSSGGGGSGGGGSG
GGGSDVVMTQTPLSLPVSLGDQASISCKSSSSLVNEKGEIYTSWYLQKPGQSPKILIYKVSNRVSGV
PDRFSGSGSGDTFTLKINRVEAEDLGVYFCAQDSEDPSTWGGGTKLEIK
>'2Ins_5
EVKLDETGGGLVQPGGAMKLSCVTTGFTFGVYFMNWVRQSPEKGLEWVASFTSNPNNNLTYYSD
SVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAAFSQRSDNWYVGDGTSVTVSSGGGGSGGG
GSGGGGSDVVMTQTPLSLPVSLGDQASISCKANDLVWTKGFASLSWYLQKPGQSPKVLIYQVSSRT
SGVPDRFSGSGSGLVFTLKINRVEAEDLGVYFCATSYKYPATFGGGTKLEIK
>'2Ins_6
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKFFMNWVRQSPEKGLEWVANVWADNRQTYYSDS
VAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMSALAQKAWWTGDATVVTVSSGGGGSGGGGS
GGGGSDVVMTQTPLSLPVSLGDQASISCRALYKVAWIVWYLQLPGQSPKVLVYFIYIRVSGVPDRF
SGSGSGTDFTLKINRVEAEDLGVYFCAQFFYWPATFSTGTKLEIK
>'2Ins_7
EVKLDETGGGLVQPGGAMKLSCVQSGFDFGQATMNWVRQSPEKGLEWVAAMSSDDVSTSYAAS
VRGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCAASDKDAGWKYMGDATVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRAWWRAWNQVIWYLQLPGQSPRALIVNTNDRVSGVP
DRFSGSGSGTDFTLKINRVEAEDLGVYFCAQYAASATFSTGTKLEIK
>'2Ins_8
EVKLDETGGGLVQPGGAMKLSCVTSGFVFGLTVMNWVRQSPEKGLEWVAAVITWTNDTMYSDS
VAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAFTTSLTSRSQAFWGQSTLVTVSSGGGGSGGG
GSGGGGSDVVMTQTPLSLPVSLGDQASISCAALRQAAQAYIYWYLQLPGQSPKVLVFFTRYRVSG
VPDRFSGSGSGTDFTLKINRVEAEDLGVYFCQYFVSAPALWAEGTKLEIK
>'2Ins_9
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGRTNMNWVRQSPEKGLEWVASARDSSNNKDTRYSD
SVRGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCWAAARAFASSGQATVVTVSSGGGGSGGGGSG

GGGSDVVMTQTPLSLPVSLGDQASISCRALYAVNNRVVWYLQLPGQSPKVLIYNVSNRVSGVPDR
FSGSGTGTDFTLKINRVEAEDLGVYFCAQDSSEPATFSTGTKLEIK
>'2Ins_10
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGTTIMNWVRQSPEKGLEWVAAIKMSALNRQTMYSD
SVAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCWAAARANASSGTSTVVTVSSGGGGSGGGGSG
GGGSDVVMTQTPLSLPVSLGDQASISCRAKFVVRDYVVWYLQLPGQSPKALIFRIRYRVSGVPDRF
SGSGSGTDFTLKINRVEAEDLGVYFCSQFAIWPASFSTGTKLEIK
>'2Ins_11
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGVDAMNWVRQSPEKGLEWVTRVDSWDNRTEYSDS
VAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCYSDSNWWTGSATVVTVSSGGGGSGGGGSGGG
GSDVVMTQTPLSLPVSLGDQASISCRALYSANNSVTWYLQLPGQSPKVLIWKVSNRVSGVPDRFSG
SGSGQDFTLKINRVEAEDLGVYFCAQDSEKSSYSNGTKLEIK
>'2Ins_12
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKSNMNWVRQSPEKGLEWVAFASIFNYATYYAASV
RGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCWASNNVDYAMMLTGNSTVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFYMNTFAVFWYLQLPGQSPKVLFSKVFNRVSGVP
DRFSGSGSGTDFTLKINRVEAEDLGVYFCAVMSNSSSWSTGTKLEIK
>'2Ins_13
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYAMVWVRQSPEKGLEWVVSAVTWSNQTSYSDS
VAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCWAQYAGKATTVAGSATVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFVLTDQFNNKTWVTWYLQLPGQSPKVLIYKVSN
RVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAATSWYLSTSYSTGTKLEIK
>'2Ins_14
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYFMMWVRQSPEKGLEWVAYTDATNTSTRYSDS
VAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCLAFAADSEWKTKKTGNATVVTVSSGGGGSGGG
GSGGGGSDVVMTQTPLSLPVSLGDQASISCRARYATADFVYWYLQLPGQSPKVLIYKVSNRVSGVP
DRFSGSGRGTDFTLKINRVEAEDLGVYFCTTMTANYYSFSTGTKLEIK
>'2Ins_15
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGQYTMNWVRQSPEKGLEWVAAASDNSKMTFYSDS
VRGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCAAALYNWIWYMGQATVVTVSSGGGGSGGGGS
GGGGSDVVMTQTPLSLPVSLGDQASISCRAAYSLENAGTAYVLWYLQLPGQSPTLLVYNTNQRVS
GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQASTYPITYSTGTKLEIK
>'2Ins_16
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGLYSMVWYRQSPEKGLEWVAWASWWNLQTEYSDS
VRGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCSAWTDWLVVVWAWTVTGNATVVTVSSGGGGS
GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASRYLRSEDVMWYLQLPGQSPKLLVFLTYFR
VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQTWRRPAVYSTGTKLEIK
>'2Ins_17
EVKLDETGGGLVQPGGAMKLSCVTDGFTFGATVMNWVRQSPEKGLEWVASFWANNYTYYADSV
KGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCWAAVANSMRLTGTGTSVTVSSGGGGSGGGGSGG
GGSDVVMTQTPLSLPVSLGDQASISCRAAYTVEKNLLWYLQKPGQSPKVLIYKVSNRVSGVPDRFS
GSGSGTVYDLKINRVEAEDLGVYFCAQSSTDPSTFGGGTKLEIK
>'2Ins_18
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYSMNWVRQSPEKGLEWVAAMDYNDNRTYYSDS
VAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMASSDGTNAFTWTGDATVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYEVKAMVVWYLQLPGQSPKVLIYKVSNRVSGVP
DRFSGSGSGNDFTLKINRVEAEDLGVYFCAQTATFPWSFSNGTKLEIK
>'2Ins_19
EVKLDETGGGLVQPGGAMKLSCVTSGFDFSTSWMNWVRQSPEKGLEWVAEVAWSNLITLYSDSV
RGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMASSYSSSSKFTGNATVVTVSSGGGGSGGGGSGG
GGSDVVMTQTPLSLPVSLGDQASISCRSKYATTYSIVWYLQLPGQSPKVLIYNIYVRVSGVPDRFSG
SGSGTDFTLKINRVEAEDLGVYFCAQSSSWWRKVTYSTGTKLEIK

>3aIns_1
EVKLDETGGGLVQPGGAMKLSCVTSGFSFGTTIMNWVRQSPEKGLEWVASITSAEYNFATYYSDS
VRGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCMASYNAYTAIGNGTVVTVSSGGGGSGGGGSGG
GGSDVVMTQTPLSLPVSLGDQASISCRAAYVTNQTVYWYLQLPGQSPKVLVSRTSLRVSGVPDRFS
GSGSGTDFTLKINRVEAEDLGVYFCAQNSQWPQSYSTGTKLEIK
>3aIns_2
EVKLDETGGGLVQPGGAMKLSCVTSGFNFGLYYMNWVRQSPEKGLEWVAAKTPDNNDSYYSDS
VRGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAASSKSKTVSLWGSGTVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRASRQLYIQNLNSSYTVWYLQLPGQSPKVLVMNVNFR
VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFAWWPWTYSTGTKLEIK
>3aIns_3
EVKLDETGGGLVQPGGAMKLSCVTSGFNFGATIMNWVRQSPEKGLEWVAATDPENNRTWYSDSV
AGRFTISYDISKSSVYLQMNNLRVEDTGIYYCMAAAYLSAAFIGNGTLVTVSSGGGGSGGGGSGGG
GSDVVMTQTPLSLPVSLGDQASISCRAAYSLKVDQDNTAVISWYLQLPGQSPKTLVFSTDKRVSGV
PDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFAIWPATFSTGTKLEIK
>3aIns_4
EVKLDETGGGLVQPGGAMKLSCVTSGFNFGKTIMNWVRQSPEKGLEWVAAATNADNNKLTYYS
DSVRGRFTISIDESKSSVYLQMNNLRVEDTGIYYCMAAWQKFAAYIGTGTVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRASRRLVSSSGNIYIAWYLQLPGQSPKLLVFNVRYRVS
GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFAYWPWTWSTGTKLEIK
>3aIns_5
EVKLDETGGGLVQPGGAMKLSCVTSGFFFGFYDMNWVRQSPEKGLEWVASVSKTNDYTYYSDSV
AGRFTISIDISKSSVYLQMNNLRVEDTGIYYCMASSWAKTTTVIGSGTLVTVSSGGGGSGGGGSGG
GGSDVVMTQTPLSLPVSLGDQASISCRALLSLKDSNNNEIWIVWYLQLPGQSPKVLVSNNNWRVSG
VPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFAWWPFSYSTGTKLEIK
>3aIns_6
EVKLDETGGGLVQPGGAMKLSCVTSGFNFGNAWMNWVRQSPEKGLEWVAAVRPEQNLTWYSDS
VRGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCAAAMRKEAATITLGTGTVVTVSSGGGGSGGGGS
GGGGSDVVMTQTPLSLPVSLGDQASISCRAAYSLKDDNNNKIYVTWYLQLPGQSPSTLVYNVRYR
VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCSQFWAWPATYAEGTKLEIK
>3aIns_7
EVKLDETGGGLVQPGGAMKLSCVTSGFVFGWSIMNWVRQSPEKGLEWVAAADSDNSRTWYSDS
VAGRFTISEDRSKSSVYLQMNNLRVEDTGIYYCWSSSNDRTSASWLGSGTLVTVSSGGGGSGGGGS
GGGGSDVVMTQTPLSLPVSLGDQASISCRASRALVSRYGTSNVSWYLQLPGQSPKTLVTNVSYRVS
GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFTYYPFTFSTGTKLEIK
>3aIns_9
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGNFQMNWVRQSPEKGLEWVAAASDNNASTYYSDS
VAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCAAISWGLNSYSYWGNGTLVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRSSTDLKSSEGKINVLWYLQLPGQSPKTLVFFVSARVT
GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQYWREPYSFSNGTKLEIK
>3aIns_10
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGNYIMNWVRQSPEKGLEWVAAIKANENKTWYSDSV
AGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCAAMSWGSSSLSWWGNGTVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCAALYTAAWWYIMWYLQLPGQSPKELVFATSARTSGV
PDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQSSQWPSTFSTGTKLEIK
>3aIns_11
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGQAIMNWVRQSPEKGLEWVASVKTAKYNYETRYSD
SVRGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCAASFWNSSWLIGTGTVVTVSSGGGGSGGGGSG
GGGSDVVMTQTPLSLPVSLGDQASISCRASRDLKSSNGRYWVVWYLQLPGQSPKTLVFWVSARVS
GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFTYKATYSTGTKLEIK
>3aIns_12
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKTIMNWVRQSPEKGLEWVAAVTSAAYNWLTYYS
DSVAGRFTISWDKSKSSVYLQMNNLRVEDTGIYYCWASARSFASWGSGTAVTVSSGGGGSGGGGS

GGGGSDVVMTQTPLSLPVSLGDQASISCRSRSSSDNRVAWYLQLPGQSPKALVVNTYIRVSGVPDR
FSGSGSGTDFTLKINRVEAEDLGVYFCAQFSRWPISWSTGTKLEIK
>3aIns_13
EVKLDETGGGLVQPGGAMKLSCVASGFDFGQTVMNWVRQSPEKGLEWVAAVDTTNNRTYYSDS
VAGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCWASASLYRSTAAFGTGTVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRSRSSSNNIVTWYLQLPGQSPKTLVWFTYWRVSGVPD
RFSGSGSGTDFTLKINRVEAEDLGVYFCSQNSQNPQTFSNGTKLEIK
>3aIns_14
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGRTIMNWVRQSPEKGLEWVAATSRDNAETRYSDSV
AGRFTISQDRSKSSVYLQMNNLRVEDTGIYYCMYLAYEASLEQYRYVKYIGNATVVTVSSGGGGS
GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFVADKTFIVWYLQLPGQSPKTLVVWTYIRV
SGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQYSQWPATFSNGTKLEIK
>3aIns_15
EVKLDETGGGLVQPGGAMKLSCVTSGFHFGAFIMAWVRQSPEKGLEWVAAAVPWNNLTYYSDSV
AGRFTISFDISKSSVYLQMNNLRVEDTGIYYCMMFWEHVSVVLGTGTVVTVSSGGGGSGGGGSGG
GGSDVVMTQTPLSLPVSLGDQASISCRAKYVISRDVMWYLQLPGQSPKTLVYNVSLRVSGVPDRFS
GSGSGTDFTLKINRVEAEDLGVYFCAQSYIWPITYSTGTKLEIK
>3aIns_16
EVKLDETGGGLVQPGGAMKLSCVTSGFRFGLFWMSWVRQSPEKGLEWVATVWMYNWATYYSD
SVRGRFTISRDVSKSSVYLQMNNLRVEDTGIYYCMAYSIYWYWYSVAVWGHATVVTVSSGGGGS
GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALYKTNYTVYWYLQLPGQSPKVLVFLVMFRV
SGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCASWATVSNSMSFSTGTKLEIK
>3aIns_17
EVKLDETGGGLVQPGGAMKLSCVSSGFSFGTTWMNWVRQSPEKGLEWVASVWAEQFNKLTYYS
DSVRGRFTISMDISKSSVYLQMNNLRVEDTGIYYCAASTAKASTYLGNATVVTVSSGGGGSGGGGS
GGGGSDVVMTQTPLSLPVSLGDQASISCRAKFDSSRFVAWYLQLPGQSPKTLVFSVYYRVSGVPDR
FSGSGSGTDFTLKINRVEAEDLGVYFCATYASFPFSYSTGTKLEIK
>3aIns_18
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKFIMNWVRQSPEKGLEWVASYVPQYNLTYYSDSV
RGRFTISADISKSSVYLSMNNLRVEDTGIYYCAAAAAYERALSYKIGNGTAVTVSSGGGGSGGGGS
GGGGSDVVMTQTPLSLPVSLGDQASISCRALFDTSNYVAWYLQLPGQSPKALVVSTWIRVSGVPDR
FSGSGSGTDFTLKINRVEAEDLGVYFCAQSSSFPSTFSTGTKLEIK
>3aIns_19
EVKLDETGGGLVQPGGAMKLSCVTSGFAFGYTWMNWVRQSPEKGLEWVAAVWTWDNATYYSD
SVAGRFTISVDFSKSSVYLQMNNLRVEDTGIYYCMAMFATATNLSMAAIKYFGNATVVTVSSGGG
GSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALYATSNYVIWYLQLPGQSPKTLVFLTNIR
VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQVSLFPFSYSTGTKLEIK
>3aIns_20
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKTIMAWVRQSPEKGLEWTAAVDPNNNRTWYSDSV
RGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCTASTSSYSSSWGDGTAVTVSSGGGGSGGGGSGGG
GSDVVMTQTPLSLPVSLGDQASISCRALLDVSSNVSWHLQLPGQSPKTLVSHAKTRVSGVPDRFSG
SGSGTDFTLKINRVEAEDLGVYFCSQSYFWPWSFSNGTKLEIK
>3aIns_21
EVKLDETGGGLVQPGGAMKLSCVTSGFFFGKTIMSWVRQSPEKGLEWTASIKPEENQTWYSDSVR
GRFTISVDVSKSSVYLQMNNLRVEDTGIYYCSFFAYDNSSTFWGNATQVTVSSGGGGSGGGGSGG
GGSDVVMTQTPLSLPVSLGDQASISCRAKYATNNHVTWYLQLPGQSPKALVTDTDRRVSGVPDRF
SGSGSGTDFTLKINRVEAEDLGVYFCAQFSHWPASFSTGTKLEIK
>3aIns_22
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGSHVMTWVRQSPEKGLEWTAAVTSQLWNRLTYYS
DSVRGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCAASSWYASAYLGNGTVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYVVSISVAWYLQLPGQSPKTLVVRVYIRVSGVPD
RFSGSGSGTDFTLKINRVEAEDLGVYFCAQFTTFPWTFSTGTKLEIK
>3aIns_23

EVKLDETGGGLVQPGGAMKLSCVTSGFNFGEYIMMWVRQSPEKGLEWVAAITSTNVTFYSDSVA
GRFTISIDVSKSSVYLQMNNLRVEDTGIYYCASSSRSRAVMYFGNGTQVTVSSGGGGSGGGGSGGG
GSDVVMTQTPLSLPVSLGDQASISCAASRKLWIWAYLAAFVVWYLQLPGQSPKALVVLVNVRVSG
VPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFAYWPASYATGTKLEIK
>3aIns_24
EVKLDETGGGLVQPGGAMKLSCVTSGFLFGLYVMAWVRQSPEKGLEWVAAIAWDRSSTYYAASV
RGRFTISEDRSKSSVYLQMNNLRVEDTGIYYCWAWLLWWNSAMWWGSGTVVTVSSGGGGSGGG
GSGGGGSDVVMTQTPLSLPVSLGDQASISCRASHSLVDDTRNKAFIVWYLQLPGQSPKVLVWATR
YRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQYARMPATFSTGTKLEIK
>3aIns_25
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYIMSWVRQSPEKGLEWTASVTDQLSNWKTYYSD
SVRGRFTISWDKSKSSVYLQMNNLRVEDTGIYYCTGARAKYTMYWGDGTQVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCAAAYSTWVWQYMTAFVVWTLQLPGQSPKTLVHTVN
DRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQMTTFPWTFSTGTKLEIK
>3aIns_26
EVKLDETGGGLVQPGGAMKLSCVTSGFYFGRFVMMWVRQSPEKGLEWVASAKSNNKSTYYAAS
VAGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCVAVYWALSLWGSGTVVTVSSGGGGSGGGGSGG
GGSDVVMTQTPLSLPVSLGDQASISCAASVDLKWATGEIWVSWYLQLPGQSPKELVYNVRLRVSG
VPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAAVAVISNSVSFATGTKLEIK
>3aIns_28
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGRYWMSWIRQSPEKGLEWTATVDSTSNVTWYAASV
AGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCAAEQSYNSSTLAIGDGTVVTVSSGGGGSGGGGSG
GGGSDVVMTQTPLSLPVSLGDQASISCRALLSASNSLAWYLQLPGQSPKVLVYDVNRRVSGVPDRF
SGSGSGTDFTLKINRVEAEDLGVYFCATYNYWPHSFSTGTKLEIK
>3aIns_29
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYWMTWVRQSPEKGLEWTAMVDSTSNSTAYAAS
VAGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCMAAFAYTSAAFGTGTVVTVSSGGGGSGGGGSG
GGGSDVVMTQTPLSLPVSLGDQASISCRSSYDVDSVAWYLQLPGQSPKALVLDAKNRVSGVPDRF
SGSGSGTDFTLKINRVEAEDLGVYFCAQFKYFPFSWATGTKLEIK
>3aIns_30
EVKLDETGGGLVQPGGAMKLSCVTSGFIFGLSVMNWVRQSPEKGLEWVAAIEPVLNLTYYSDSVA
GRFTISIDVSKSSVYLQMNNLRVEDTGIYYCAATVEQYFSSYSSAAIIGQGTVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRALFSSSAAVAWYLQLPGQSPKTLVFKVSIRVSGVPDR
FSGSGDGTDFTLKINRVEAEDLGVYFCASFLRFPFTFSTGTKLEIK
>3aIns_31
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKTQMAWIRQSPEKGLEATAIIDSTNNRTWYAASVA
GRFTISVDISKSSVYLQMNNLRVEDTGIYYCSFFAYDNSSTFWGNGTQVTVSSGGGGSGGGGSGGG
GSDVVMTQTPLSLPVSLGDQASISCRAAYATNDNVTWYLQLPGQSPKTLVTNTNQRVSGVPDRFS
GSGSGTDYTLKINRVEAEDLGVYFCAAYSDMAAFFSTGTKLEIK
>3aIns_32
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYVMSWVRQSPEKGLEWTASVIPQRNQTYYSDSV
RGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCAAALAHFMWYIGNGTVVTVSSGGGGSGGGGSGG
GGSDVVMTQTPLSLPVSLGDQASISCAASANTDYVTWYLQLPGQSPKVLVWWIYYRVSGVPDRFS
GSGSGTDFTLKINRVEAEDLGVYFCAQTTKYPWTFSTGTKLEIK
>3aIns_33
EVKLDETGGGLVQPGGAMKLSCVASGFTFGLASMNWVRQSPEKGLEWVASTTPFLNLTWYAASV
AGRFTISVDLSKSSVYLQMNNLRVEDTGIYYCAASSSWTSSNASVAAWSTGTQVTVSSGGGGSGG
GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALFSASMVVFWYLQLPGQSPKLLVFWVRYRVSG
VPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQASRFPVTFSTGTKLEIK
>3aIns_34
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGLMTMAWVRQSPEKGLEFVAAAWPARNLTYYSDS
VRGRFTISMDISKSSVYLQMNNLRVEDTGIYYCTYSQFIADSWHDSTIVYWGNGTVVTVSSGGGGS
GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASHDSKRDVIWTLQLPGQSPKTLVYNVYYRV
SGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCASTLSHPFTFSTGTKLEIK

>3aIns_35
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGMFIMNWVRQSPEKGLEWVAAYVTVANLTFYSDSV
RGRFTISADISKSSVYLQMNNLRVEDTGIYYCAASKSWWSDSWYIGNATVVTVSSGGGGSGGGGS
GGGGSDVVMTQTPLSLPVSLGDQASISCRASKSVNSVAWYLQLPGQSPKVLVYEVYKRVSGVPDR
FSGSGSGTDFTLKINRVEAEDLGVYFCAAWAVDSLSASFSTGTKLEIK
>3aIns_36
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKTIMNWVRQSPEKGLEWVAAVIARYSATYYSDSV
RGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCMASSSTSDTYWSTTSTIGSGTVVTVSSGGGGSGG
GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASQNTNFVVWYLQLPGQSPKVLISNIWFRVSGVP
DRFSGSGSGTDFTLKINRVEAEDLGVYYCAQTWSFPYTYSTGTKLEIK
>3aIns_37
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKFVMAWVRQSPEKGLEWTAAVTSADNNFKTYYS
DSVRGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCIAWYNGASSAYWGSGTQVTVSSGGGGSGGG
GSGGGGSDVVMTQTPLSLPVSLGDQASISCRALLTAKDRVAWALQLPGQSPKTLVSLVSIRVSGVP
DRFSGSGSGTDFTLKINRVEAEDLGVYFCYATVSWPWSFSTGTKLEIK
>3bIns_1
EVKLDETGGGLVQPGGAMKLSCVTSGFNFGKSTMTWVRQSPEKGLEWTAAVSSQDDNWKTYYS
DSVAGRFTISWDKSKSSVYLQMNNLRVEDTGIYYCASSTSSYSSSWGDGTAVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFEVKTSVAWTLQLPGQSPKTLVSHVSWRVSGVPD
RFSGSGSGTDFTLKINRVEAEDLGVYFCSQSYQFPWTFSTGTKLEIK
>3bIns_2
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKFVMMWVRQSPEKGLEWTAAVTNDNNNWATYY
SDSVAGRFTISWDKSKSSVYLQMNNLRVEDTGIYYCLASSHNTTIYIGNGTVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRALLATNDNVTWWLQLPGQSPKALVVWTSIRVSGVP
DRFSGSGSGTDYTLKINRVEAEDLGVYFCAQFTYWPASFSTGTKLEIK
>3bIns_3
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGSSIMAWVRQSPEKGLEWVAAVMSDLNLTYYSDSV
AGRFTISMDISKSSVYLQMNNLRVEDTGIYYCIAWAYMNVSVSWGNGTVVTVSSGGGGSGGGGSG
GGGSDVVMTQTPLSLPVSLGDQASISCRSTHVLNQHVAWTLQLPGQSPKTLVYMNYYRTSGVPDR
FSGSGSGTDFTLKINRVEAEDLGVYFCSQYEIMPASYSTGTKLEIK
>3bIns_4
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGQYAMAWIRQSPEKGLEWVATITTLQYSTVYAASV
KGRFTISMDISKSSVYLQMNNLRVEDTGIYYCMATSDLFNFSLFWWGHGTVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFVTSLHVVWFLQLPGQSPKTLVYDTYYRVSGVPD
RFSGSGSGTDFTLKINRVEAEDLGVYFCAQTFIVPWSFSTGTKLEIK
>3bIns_5
EVKLDETGGGLVQPGGAMKLSCVTSGFWFGSTKMAWTRQSPEKGLEWVASADPNSNSTYYSDSV
RGRFTISWDESKSSVYLQMNNLRVEDTGIYYCMAFSDNAALTIGTGTVVTVSSGGGGSGGGGSGG
GGSDVVMTQTPLSLPVSLGDQASISCRSSHDSKRIVMWWLQLPGQSPKTLVYEVRKRTSGVPDRFS
GSGSGTDFTLKINRVEAEDLGVYFCAQFTTYPWTYATGTKLEIK
>3bIns_6
EVKLDETGGGLVQPGGAMKLSCVSSGFDFGYYWASWVRQSPEKGLEWVASTHSSNMWTLYAAS
VRGRFTISADKSKSSVYLQMNNLRVEDTGIYYCSFHSKTSMTTTVWGNGTQVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRALYDTWWIVTWYLQLPGQSPKTLVYNTSIRVSGVPD
RFSGSGSGTDYTLKINRVEAEDLGVYFCAQYSRWPASFSTGTKLEIK
>3bIns_7
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYFMTWIRQSPEKGLEWVATTASYNMWTHYAAS
VRGRFTISADVSKSSVYLQMNNLRVEDTGIYYCAADMSVYNDSLAIGTGTVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRALYSADRTVSWMLQLPGQSPKVLVYHVRERVSGVP
DRFSGSGSGTDFTLKINRVEAEDLGVYFCSAVRGDKISFGGGTKLEIK
>3bIns_8
EVKLDETGGGLVQPGGAMKLSCVASGFIFGLYIMAWVRQSPEKGLEWTAAVSSTSNLTWYAASV
AGRFTISADVSKSSVYLQMNNLRVEDTGIYYCTAWYYWVATIWGDGTVVTVSSGGGGSGGGGSG

GGGSDVVMTQTPLSLPVSLGDQASISCRAAYVLRFKTSDLSIIAWTLQLPGQSPKTLVYNTRARVSG
VPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQYSIMPATYSTGTKLEIK
>3bIns_10
EVKLDETGGGLVQPGGAMKLSCVTSGFSFGNSWMAWVRQSPEKGLEWVASARNASHNHKTFYS
DSVRGRFTISMDESKSSVYLQMNNLRVEDTGIYYCAAAVAQAITRIGTGTVVTVSSGGGGSGGGGS
GGGGSDVVMTQTPLSLPVSLGDQASISCAASMDLWITSHNSAYVIWYLQLPGQSPKVLVFNTYYR
VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSYYPWTYSTGTKLEIK
>3bIns_11
EVKLDETGGGLVQPGGAMKLSCVTSGFHFGAFIMSWIRQSPEKGLEWVASASRYSVYTWYSDSVR
GRFTISQDISKSSVYLQMNNLRVEDTGIYYCVASWDNTIMLIGNGTVVTVSSGGGGSGGGGSGGGG
SDVVMTQTPLSLPVSLGDQASISCRAAYVLYLSDGSEHLTWWLQLPGQSPKLLVYNTRLRVSGVPD
RFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSYWPWTWSTGTKLEIK
>3bIns_12
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGKHIMLWTRQSPEKGLEWTAAVASALWNKQTWYS
DSVAGRFTISIDISKSSVYLQMNNLRVEDTGIYYCWAFFYTADNLFFIGNGTVVTVSSGGGGSGGGG
SGGGGSDVVMTQTPLSLPVSLGDQASISCRASRMLISSDGFVIVWWYLQLPGQSPKVLVAYVRWR
VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSYWPATYSTGTKLEIK
>3bIns_13
EVKLDETGGGLVQPGGAMKLSCVTSGFDFGSTIMLWIRQSPEKGLEWTAAVASMRNQTWYAASV
AGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCWSSIATYTGTGTLVTVSSGGGGSGGGGSGGGGSD
VVMTQTPLSLPVSLGDQASISCRAKFKTRFVVWYLQLPGQSPKVLVYYIYYRVSGVPDRFSGSGSG
TDYTLKINRVEAEDLGVYFCAQYSIWPATFSTGTKLEIK
>3bIns_14
EVKLDETGGGLVQPGGAMKLSCVTSGFHFGAFQMSWTRQSPEKGLEWTATVDPTNNRTWYSDSV
RGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCTAQWRSLSFYGSGTVVTVSSGGGGSGGGGSGGG
GSDVVMTQTPLSLPVSLGDQASISCRAAYVTSHDVTWHLQLPGQSPKTLVYSTRLRDSGVPDRFSG
SGSGTDFTLKINRVEAEDLGVYFCSQSYISPTTYSTGTKLEIK
>3bIns_15
EVKLDETGGGLVQPGGAMKLSCVTSGFNFGAYIMAWVRQSPEKGLEWTAAVDPSTNTTWYSDSV
RGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCTAHLIRYDLYIGTGTVVTVSSGGGGSGGGGSGGG
GSDVVMTQTPLSLPVSLGDQASISCRASHDVLAVAWHLQLPGQSPKTLVYWIYWRVSGVPDRFSG
SGSGTDFTLKINRVEAEDLGVYFCAQYSRWPATYSTGTKLEIK
>3bIns_16
EVKLDETGGGLVQPGGAMKLSCVTSGFWFGKSIMQWVRQSPEKGLEWVATARSEDWNKLTFYSD
SVRGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCTASSSSYASMWGDGTQVTVSSGGGGSGGGGSG
GGGSDVVMTQTPLSLPVSLGDQASISCRAKFDTSSEVSWTLQLPGQSPKTLVMVTLTRVSGVPDRF
SGSGSGTDYTLKINRVEAEDLGVYFCSQNTTDPATFSTGTKLEIK

>5dIns_2
QVQLQETGGGLVQPGASMKLSCKASGFNFTSSGMAWVRQRPGQGLEWVAWISPGGGTTHYNDK
VKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCAADPWNSYGFNYWGQGTLVTVSGGGGSGGGG
SGGGGSDIVLTQTPLTLPVSPGQRATISCRASQSVDNSGNSYMHWYQQKPGQPPKLLVYRASNLES
GVPARFSGSGSGTDFTLTIHPVEPEDFAMYYCLQGYSYPFTFGQGTKLEIK
>5dIns_3
QVQLQETGGGLVQPGASMKLSCKASGFTFTDYSMSWVRQRPGQGMEWVANISPNGGSRSYNDKV
KGRATITRDTSSNTAYLQMSSLTSEDTAVYYCARLDSSGDSGFAYWGQGTLVTVSGGGGSGGGGS
GGGGSDIVMTQTPLTLPASPGDRVTISCKASQDVGKYVAWFQQKPGGSPKLLIYYASNLYSGVPSR
FSGSGSGTDFTFTVSNVQPEDFGTYYCVQGSYAPPTFGQGTKVEIK
>5dIns_4
QVQLQESGGGLVQPGASMKLSCKASGFTFTRYAMHWVRQRPGQGLEWVAGISPDGGETHYNDK
VKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCAREYWHSLDYWGQGTTVTVSGGGGSGGGGSG
GGGSDIVLTQTPLTLPVSVGQRATISCRSSQSLVDSNGNTYLNWYLQKPGQSPKLLIYQVSNRFSGV
PDRFSGSGSGTDFTLKISRVEAEDFGVYFCQQGYRHPLTFGQGTKLEIK

>5dIns_5
QVQLQESGGGLVQPGASMRLSCKTSGFTFTSYAISWVKQAPGQGFEWIGSINPGSGNTHYNDKFKG
RATITADTSSNTAYLQLNNLTSEDTAVYYCAASDYRSHGALDYWGQGTQVTVSGGGGSGGGGSG
GGGSDIVMTQTPLTLPASPGDRVTISCKASQNIYKHIIWYQQKPGQAPKLLVYYASNLYSGVPSRFS
GSGSGTDYTLTISNVQPEDFGTYYCAQHYSHPLTFGQGTKVEIK
>5dIns_6
QVQLQETGGGLVQPGASMKLSCKASGFTFTDYGMIWVRQRPGQGLEWVAYISPGGGYTYYNDKV
KGRATITRDTSSNTAYLQMSSLTSEDTAVYYCARDDGYGAMDYWGQGTTVTVSGGGGSGGGGSG
GGGSDIVMTQTPLTLPVSPGQRATISCRASQSVDSNGYSHVNWYQQKPGQPPKLLVYLASNLESGV
PARFSGSGSGTDFTLTIHPVEPEDFATYFCMQGYKWPYTFGQGTKLEIK
>5dIns_7
QVQLQETGGGLVQPGASMKLSCKASGFTFTRYGMSWVRQRPGQGLEWVASISPDGGETDYNDKV
KGRATITRDTSSNTAYLQMSSLTSEDTAVYYCASSHFDYWGQGTTVTVSGGGGSGGGGSGGGGSD
IVLTQTPLTLPVSPGQRATISCRASQSVDYYGYSFMHWYQQKPGQPPKILVYRGSNLYSGVPARFSG
SGSGTDFTLTIHPVEPEDFATYYCQQSYAVPPTFGQGTKLEIK
>5dIns_8
QAQLQETGGGLVQPGASMKLSCKASGYTFTDYGMSWVRQRPGQGLEWVASISPDGGTTYYNDKV
KGRATITRDTSSNTAYLQMSSLTSEDTAVYYCARMDFTSYQDSAVDYWGQGTSVTVSGGGGSGG
GGSGGGGSDIVMTQTPLTLPASPGDRVTISCKASRDVGNYVAWYQQKPGQSPQLLIYSASNLHSGV
PSRFSGSGSGTDFTLTISSVQPEDFGDYYCMQVSHSPYTFGQGTKVEIK
>5dIns_9
QVQLQETGGGLVQPGASMKLSCKASGFTFTDYGMSWVRQRPGQGLEWVAWISPNGGRTHYNDK
VKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCAMTGDYGYAFTLDYWGQGTTVTVSGGGGSGG
GGSGGGGSDIVMTQTPLTLPASPGDRVTISCRASQKVSESVAWYQQKPGQSPKLLIYSASNLYDGV
PSRFSGSGSGTDFTFTISSVQPEDFATYYCMQSSSYPYTFGQGTKVEIK
>5dIns_10
QVQLQETGGGLVQPGASMKLSCKASGFTFTRYSMNWVRQRPGQGLEWVASISPGGGSTHYNDKV
KGRATITRDTSSNTAYLQMSSLTSEDTAVYYCAASDRGYYGALGYWGQGTLVTVSGGGGSGGGG
SGGGGSDIVLTQTPLTLPVSPGQRATISCRASQSVDDNGYSYMAWYQQKPGQPPKLLVYAGSNLES
GVPARFSGSGSGTDFTLTIHPVEPEDFATYYCMQMYSTPLTFGSGTKLEIK
>5dIns_11
EVQLQETGGGLVQPSQSMSLTCTVSGDSFSNYAWSWFRQTPGNRLEYVGSISGGGSTYYHDSVKG
RFTISRDTSKNQVFLQMNSLTTEDTAVYYCARSNHAAFDYWGQGTTVTVSGGGGSGGGGSGGGG
SDIVLTQTPLTLPVSPGQRATISCRASQSVDSYGYSFMDWYQQKPGQPPKLLVYLASNLYSGVPARF
SGSGSGTDFTLTIHPVEPEDFATYFCQQHSEFPYTFGQGTKLEIK
>5dIns_12
EVQLQESGGGLVQPGASMKLSCKTSGYTFTDYWMHWVRQRPGQGLEWVGWIDPGNGSTNYNNK
FKGRATITADTSSNTAYMQMSSLTSEDTAVYYCARAGGGYRVPYGMDYWGQGTTVTVSGGGGS
GGGGSGGGGSDIVLTQTPLTLPVSPGQRATISCRASESVDSNGYSFMMWYQQKPGQPPKLLVYLAS
NLYSGVPARFSGSGSGTDFTLTIHPVEPEDFATYFCMQGHSSPWTFGQGTKLEIK
>5dIns_13
QVQLQETGGGLVQPGASMRLSCKASGFTITDHGMSWVRQRPGQGFEWIGGIYPGGGNTHYNDKF
KGKATITVDTSSNTAYMQLNSMTSEDTAVYYCARNFWAALDDWGQGTTVTVSGGGGSGGGGSG
GGGSDIVMTQTPLTLPASPGDRVTISCKASQKVGKHVAWFQQKPGQSPKLLIYNASNLHSGVPSRF
SGSGSGTDFTLTISSVQPEDFATYFCVQTHGWPYTFGQGTKVEIK
>5dIns_14
QVQLQETGGGLVQPGGSMKISCVTSGFNFSHAWMSWVRQSPGKGLEWVAEIRNKSDGYTTYYAP
SVKGRFTVSRDDSQNMVYLQMNNLRAEDTGIYYCAQSSPYSRAMDYWGQGTTVTVSGGGGSGG
GGSGGGGSDIVLTQTPLTLPVSPGQRATISCRASQSVDDNGHSFMNWYQQKPGQPPKLLVHAASY
VKSGVPARFSGSGSGTDFTLTIHPVEPEDFATYYCQQGYSHPWTFGGGTKLEIK
>5dIns_15
EVQLQESGGGLVQPGASMKLSCKTSGFTFTDHYMHWVKQRPGQGLEWVAGIDPGNGSTQYNNKF
KGKATFTADTSSNTAYMQLSSLTSEDTAVYYCARSYDAAMDDWGQGTTVTVSGGGGSGGGGSG

GGGSDIVLTQTPLTLPMSVGEKVTITCKSSRSLFSSGTQKNYMTWYQQKPGQPPKLLVYWASTRES
GVPDRFTGSGSGTDFTLTVSSVQAEDIAVYFCMQSHSSPFTFGQGTKVEIK
>5dIns_16
QVQLQETGGGLVQPGASMKLSCKASGFTFTRSGMSWVRQRPGQGLEWVAWISPNGGSTDYNDKV
KGRATITRDTSSNTAYLQMSSLTSEDTAVYYCARGWGGMQYWGQGTTVTVSGGGGSGGGGSGG
GGSDIVLTQTPLTLPVSPGQRATISCRASQSVDYRGYSFMHWYQQKPGQPPKLLVYWGSNLESGVP
ARFSGSGSGTDFTLTIHPVEPEDFATYFCQQSYSTPWTFGGGTKLEIK
>5dIns_17
QVQLQESGGGLVQPGASMKLSCKASGFTFTRYAMHWVRQRPGQGLEWVAWISPNGSNTRYNSKV
KGRATITRDKSSNTAYLQMSSLTSEDTAVYYCARDWGASMDYWGQGTTVTVSGGGGSGGGGSG
GGGSDIVMTQTPLTLPASPGDRVTISCKASQDIGTSVAWYQQKPGQSPKLLIYSASQLYDGVPSRFS
GSGSGTDFTFTISSVQPEDFATYFCMQSYTTPYTFGQGTKVEIK
>5dIns_18
QVQLQETGGGLVQPGASMRLSCVTSGFDITSSAMHWVRQRPGQGLEWVGGISPYGGETHYAPKFK
GKATITVDTSSNTAYMQMSSLTSEDTAVYYCARRENYAEAGFSYWGQGTTVTVSGGGGSGGGGS
GGGGSDIVMTQTPLTLPASPGDRVTISCRASQDVGKYVAWYQQKPGQSPKLLIYYASRRAPGVPSR
FSGSGSGTDFTLTISNVQPEDFGTYFCQQGYSWPYTFGQGTKVEIK
>5dIns_19
QVQLQETGGGLVQPGASMKLSCKASGFTFTEYAMSWVRQRPGQGMEWVAWISPGGGNTHYNDK
VKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCAAMGRHSTGAMDYWGQGTQVTVSGGGGSGG
GGSGGGGSDIVLTQTPLTLPMSVGEKITITCKSSQSLFNSRSQKNYLAWYQQKPGQSPKLLIYWAST
RESGVPDRFTGSGSGTDFTFTVSSVQAEDIAVYFCQQGYHHPYTFGQGTKVEIK
>5dIns_20
QVQLQETGGGLVQPGASMRLSCKASGFTITDYTMHWVRQRPGQGLEWVGYINPGSGNSYFNNKF
KGRATFSVDNSSNTAYMQMSSLTSEDTAVYYCAAKSNGRIGAFGYWGQGTQVTVSGGGGSGGGG
SGGGGSDIVLTQTPLTLPVSPGQRATISCRASQSVDYYGYSFMQWYQQKPGQPPKLLVYLGSNLES
GVPARFSGSGSGTDFTLTIHPVEPEDFAMYYCQQSHAHPPTFGQGTKLEIK

AbDesign: an algorithm for combinatorial backbone design guided by natural conformations and sequences and a computational benchmark for antibody design

Gideon D. Lapidoth[1], Dror Baran[1], Gabriele M. Pszolla[1], Christoffer Norn[1], Assaf Alon[1], Michael D. Tyka[2], Sarel J. Fleishman[1,*]

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

[2]Google Inc., 1600 Amphitheatre Pkwy, Mountain View, CA 94043

[*]To whom correspondence should be addressed: sarel@weizmann.ac.il

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 conformational degrees of freedom – remains a critical challenge. Most previous attempts to design new backbones relied on computing the mainchain from scratch. As a first step to addressing the challenge of backbone design, here, instead, we develop a combinatorial backbone and sequence optimization algorithm that exploits the fact that many classes of proteins belong to large families comprising modular folds, where fragments from structurally homologous proteins can be recombined to generate new backbones. The method starts by identifying positions of high structure conservation within the fold family to serve as junctions, where backbone fragments from different family members may be joined to generate new backbones. The design procedure then recombines backbone fragments observed in natural proteins belonging to the family and designs the best-energy sequence for each backbone conformation, subject to sequence constraints derived from multiple-sequence alignments of the fold family members. We assess this algorithm's ability to generate realistic sequences and backbones by applying it to the long-standing problem of designing antibodies that target a predetermined site on a protein of interest. In a benchmark of nine high-affinity antibody-bound complexes we find that, restricted to a predetermined naturally observed binding orientation, *AbDesign* is able to design and select antibody models comprising backbone fragments and sequence features observed in benchmark natural antibody set, especially when the natural binding surface is large. Predicted sidechain rigidity in model antibodies is as high as in natural antibodies, which may be important for binding affinity and specificity. By contrast, *AbDesign* favors models with large buried surface area, even when it samples the backbone conformations of natural high-affinity antibodies in our benchmark,

highlighting the need for improved energy functions and filters based on structural features. The benchmark, which is the first to test features of both protein stability and molecular function, may advance *de novo* design of function abilities.

Introduction

Molecular recognition underlies many central biological processes. The ability to design novel protein interactions is a stringent test of our understanding of the physicochemical 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 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 $Å^2$)[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 the complication inherent in correctly balancing the contributions to free energy from polar groups and due to the large conformation space open to the protein backbone[18]. As a first 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 designed backbones and amino acid sequence choices, thereby addressing some of the challenges facing backbone design. We test this approach by generating antibody designs against pre-chosen protein epitopes and assess sequence and backbone conformation recovery compared to natural antibodies in complex with 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 – problems 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 were 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 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 Chothia [29] 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], opening the way to safe therapeutic antibody engineering. These early advances raised excitement that the complete design of antibodies from first principles is achievable[31], 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[32], are capable of isolating binders to target molecules, but there is no general method to select binders of specific epitopes[33,34], hampering efforts to generate specific binders, inhibitors, and allosteric effectors[35–37]. To be sure, in certain systems selection methods were developed to isolate antibodies that bind specific epitopes on target molecules[38–42];

yet these capabilities are challenging and rely on specific properties of the target molecule, such as naturally high antigenic variability outside the target site[43]. 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[44–46], identify favorable positions for experimental random mutagenesis[47], modify binding specificity[48] and increase thermo-resistance[49]. An antibody design strategy was suggested by Pantazes *et al.* [50–52] 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 framework-CDR interactions that are guaranteed to be structurally compatible; second, *AbDesign* derives sequence constraints from natural antibodies to constrain sequence optimization to amino acid identities that are likely to confer stability to 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 protein fold family, where the protein fold has regions of high sequence and structure conservation.

Results

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

We present a step-by-step description of the *AbDesign* algorithm using Figure 1 as a visual guide. The core algorithmic elements were written in C++ within the Rosetta software suite of macromolecular modeling[53], and the design protocols were written using RosettaScripts[54] enabling users to change parameters and control the execution flow and even to extend the algorithm to protein families other than antibodies. RosettaScripts used in this paper are available in the supplement.

AbDesign addresses four related challenges: 1. Incorporating 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 current design challenges.

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

The stability of a protein conformation relies on both positive and negative design elements[18,55]. A key advantage of computational 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 (Figure 1, step 1). Moreover, by correlating natural backbone conformations and sequences we can classify sets of natural protein-segment sequences that fold into particular conformation classes, 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[56]. Typically the sequence constraints encoded in the PSSMs are stringent in the antibody core and relaxed at the CDRs, giving sequence optimization room to explore different residue combinations for interacting with ligand, while maintaining the antibody core. The PSSM is used during all design calculations in two distinct 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 3, methods), 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 core. Second, the all-atom energy function used in sequence design (Rosetta score12)[57], which is dominated by contributions from van der Waals packing, hydrogen bonding, and solvation, is modified to include a term that biases the sequence towards the more likely identities according to the PSSM. The bias towards the sequence consensus is weighted 50% more strongly away from the binding site.

b. A pre-computed database of backbone conformations for each antibody segment

Backbone-conformation sampling is computationally demanding[58–61] and despite some success [62] backbone design for function has led to conformations that deviated from the original computed models[19,63]. 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. To make computationally efficient use of the richness of backbone conformations observed in natural antibodies we implement a pre-computation step, in which we extract the conformations of natural

antibodies (Figure 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[64], although the choice of template is arbitrary). λ variable light chains were not included in the construction of the current database because of the relatively small number of available structures in the PDB (1300 variable κ chains *versus* 265 variable λ chains [65]), 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 homologous proteins belonging to the immunoglobulin fold. 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 numbering[25]), CDR2 (L50 -- L56, H52 -- H56), and CDR3(L89 -- L97, H95 -- H102)[21,22,25–27,30,46,66–70]. Our preliminary *in silico* experiments using such stem choices, however, resulted in structurally unrealistic designed antibodies with poor packing between the CDR sidechains and the framework. Instead we chose to 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 number 100 in the variable κ domain and position number 103 in the variable heavy domain, Kabat and Chothia numbering [25,26], Figure 2) as the stems for the CDR3 segments; these stems are very well aligned in all antibodies of known structure. In effect this scheme matches closely the boundaries of the V, D, and J genes as they are recombined in natural B-cell development (Figure 2); the disulfide-linked cysteines are not, however, precisely located at the sites of genomic recombination, but have the advantage for modeling of being more structurally conserved than the positions of genomic recombination.

By using segment boundaries that are close to the VDJ genomic segments 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 relationships. As a concrete example for the importance of the correlations between conformations and sequences, the H3 backbone cluster H3.15.8 (Table S1) is in an extended conformation, while the conformation from the cluster H3.16.5 is kinked [71]; the sequence profiles for the two conformation clusters are correspondingly characterized by different amino acid conservation patterns (Figure 3).

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 (Figure 1, step 2), introducing a main-chain cut site in a randomly chosen position in the inserted segment. Where the sequence length of the inserted segment needs to be increased relative to the template antibody segment, additional residues are added to the template segment using idealized bond lengths and angles. Although the stem positions are well aligned in all antibodies, small differences in their positions imply that simply imposing their dihedral degrees of freedom on the template antibody would fail to generate a chemically sound chain; instead we refine the main chain using cyclic-coordinate descent (CCD) [60], small, and shear moves, as implemented in the CCD mover in Rosetta[72]. During refinement the standard Rosetta all-atom energy function (score12)[57] 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. CCD alternates backbone moves with combinatorial amino acid sidechain 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 large (>=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[60], here we find that, guided by the coordinate and dihedral constraints, CCD 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 successfully fitted onto the template scaffold in the majority of antibody entries in our database, ranging from 66% of the H3 segments to 85% of L3 segments. Each trajectory takes on average 4.6 hours on an Intel Xeon 2.4GHz CPU. Backbone dihedral angles of successfully fitted segments are recorded in a backbone torsion database for subsequent use during design (Table S4).

c. Design subject to sequence constraints derived from natural antibodies

In sections (a) and (b), above, we pre-computed databases of amino acid identity propensities (encoded in PSSMs) correlated with backbone-conformation clusters observed in natural antibodies. *AbDesign* uses these databases by associating a unique PSSM to each segment conformation cluster, in this way encoding sequence-conformation rules that are likely important for appropriate folding. During the design trajectory *AbDesign* loads the pre-computed PSSM matrix of the current conformation variant for each of the backbone segments. The related PSSMs are then combined to generate a single PSSM matrix for the entire protein. Whenever a different

backbone conformation is sampled *AbDesign* replaces the relevant PSSM matrix associated with the swapped segment, synchronizing the sequence constraints with the backbone conformation.

For efficiency, at different phases of the algorithm different sets of residues are subjected to design. For instance, several initial design phases only optimize the ligand-binding surface, whereas at the end of the design protocol there are several iterations of full sequence optimization of 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 ~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 others[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 (Figure 1, step 3). We extract the cluster 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) x 2 (L3) x 9 (VH) x 50 (H3) = 4500 non-redundant conformation representatives (Table S2), 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

In the benchmark below, each of the 4,500 representative conformations generated in step d is aligned using PyMol[73] 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 (Figure 1, step 5)[74], 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

benchmarks of binder and enzyme design [7,20,75], 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 benchmarks 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[76], 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 unknown docking software (such as PatchDock[77] or RosettaDock[74] can be used to generate the initial bound conformations, as was done in experimental binder design applications[4–6].

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

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 (Figure 1, step 7). Additional sampling is likely to further optimize the designed antibody, but at the cost of longer computation. 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. For example, if the representative conformation undergoing design has an H3 of length *x* amino acids, refinement samples H3 backbone segments of length *x*±4. The allowed length change depends on the segment currently being designed. In the benchmark the allowed length-change parameter was set to ±2 for segment types VL, VH, and L3, and ±4 for H3. Restricting segment-length sampling reduces the bias for longer segments, which are likely to have more favorable energies. 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 and sequence for binding the target.

AbDesign in effect samples combinations of naturally observed backbone conformations from a precomputed 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)[7,78,79]. 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 precomputed 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 sidechain packing to design the sequence subject to the PSSM constraints above, and to repack the ligand residues within 10Å 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. This iterative packing and minimization procedure results in improvements to sidechain packing in the antibody core and in the antibody-ligand interface.

g. Selecting the backbone conformation and sequence that best optimize both 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 conformation and bind its target molecule[18]. *AbDesign* implements a novel multiconstrained optimization scheme to select the designed backbone segments that best optimize ligand-binding energy and antibody stability (Warszawski et al.). 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:

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

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 ½ 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 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) (Figure 4A). An example of the change in binding energy and stability before and after the segment optimization is shown in Figure 4 B&C.

h. Filtering final designs using structure and energy thresholds derived from natural antibodies

Final filtering of antibody design models is done using four parameters: predicted binding energy, buried surface area, packing quality between the antibody's variable light and heavy domains and the bound ligand[80], and shape complementary[81] between the antibody and bound ligand (Figure 1, step 8). Cutoffs for each of these parameters are derived from a set of 303 natural antibody-protein complexes (Table S3) extracted from the PDB[82] using the Structural Antibody Database (SabDab) [65](Figure 5).

A recapitulation benchmark for antibody-bound backbone conformations and sequences

To test *AbDesign*'s ability to select backbone-conformation fragments and sequences similar to naturally occurring antibodies, which target a predetermined binding site, we selected a diverse benchmark of nine high-affinity (K_d < 20 nM), medium-to-high crystallographic resolution (<=2.5 Å), protein-binding antibodies from SAbDab [65] (Table 1) as the benchmark targets. 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 benchmarks[7,20]. The natural antibody set comprises human antibodies Fab40,D5 neutralizing mAb, and BO2C11 (PDB entries: 3K2U[83] 2CMR[84], 1IQD[85] respectively), murine antibodies E8, D1.3 mAb,F10.6.6, JEL42, and 5E1 Fab (PDB entries: 1WEJ[86], 1VFB[87], 1P2C[88], 2JEL[89], 3MXW[90]), and the humanized murine antibody D3H44 (PDB entry: 1JPS[91]). The ligand targets comprise convex (2JEL, 1IQD), flat (1P2C), and concave (3MXW) surfaces, containing helical (2CMR), sheet (1JPS), and loop (1P2C, 3K2U) secondary-structural elements.

A variety of measures is used to filter designed antibodies, including binding energy, antibody stability, shape complementarity, complex packing statistics, and buried surface area. To rank the final design models we use only computed binding energy, and contrast the best design with the the target natural antibody, bound to the same epitope, according to the following criteria: sequence identity, RMSD, interface shape complementarity (*Sc*)[81], packing statistics[80], buried surface area, binding energy (Figure 5), and backbone-conformation clustering. Even though the design trajectories start from a binding orientation similar to the natural bound complex, during design some antibodies migrate and bind at other sites. For the purposes of the recapitulation analysis we eliminate designed antibodies with interface RMSD values [92] greater than 4 Å. We consider success as designed antibodies comprising backbone-conformation segments belonging to the same backbone clusters as the natural antibody targeting the same epitope; that is, *AbDesign* is free to choose from all possible backbone fragments of antibody conformation available in the PDB, and we test whether it selects a specific combination of backbone-conformation fragments similar to the natural one.

Subject to the above selection criteria, five out of nine antibodies in our benchmark set select the backbone fragments belonging to the same conformation clusters as those of the target natural antibody; of these five, four are at the top 10% ranking in terms of computed binding energy (Table 1). Bound conformations with large buried surface area (>1800 $Å^2$) are designed successfully more consistently than those with smaller buried surface area, suggesting that *AbDesign* is biased towards 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 prediction accuracy.

As representative examples of successfully designed antibodies we consider antibodies that target the same surface as the humanized anti-tissue factor antibody D3H44 (PDB entry 1JPS) and the anti-transmembrane glycoprotein D5 neutralizing mAb (PDB entry 2CMR). 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 2CMR (H1.14_H2.15, H3.18.7, L1.11_L2.8, L3.10.1). The designs' backbone conformations show a high level of agreement with the natural antibodies (Cα RMSD between design and natural antibody: 1.23 Å and 1.15 Å, for 1JPS and 2CMR, respectively; Figure 6). Previous 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 of whether the all-atom energy function used in design appropriately 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 (Figure 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 (Figure 8). The results demonstrate that when confined to choosing from naturally existing backbones and subject to sequence constraints that bias optimization towards the family consensus the all-atom energy function is capable of correctly generating and ranking polar binding surfaces and the protein core that provides structural stability to these surfaces.

Although segments from the nine antibodies comprising the benchmark 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 (Table 2); in fact all but one of the five successfully designed antibodies' segments originate from germline genes different than those of the benchmark set natural antibodies (Table 2); furthermore, the amino acid sequence identities between the germline genes that gave rise to the designed antibodies and the natural antibodies are lower than 80%, except for two segments in the target antibody D1.3 (PDB entry: 1VFB), which shares 90% sequence identity with the natural antibody's V_H germline gene and 83% sequence identity with the natural J_K gene (Table 2). We find that the sequence identities between the actual segments comprising the designed antibody and the target antibody segments are below 75% (expect for the VH segment of the target antibody D1.3, PDB entry: 1VFB). By contrast, natural antibody segments derived from the same germline typically show sequence identities higher than 80%[93]. The ability to construct antibodies using backbone-conformation segments from sources other than those used in natural binders suggests that the conformation data on natural antibody backbone conformations encoded in the PDB provide some redundancy, which may be important for fine tuning the backbone to the target site.

For some targets *AbDesign* selects models with backbone conformations similar to the natural antibody, but ranks these models poorly in comparison to other models; such cases provide valuable data on potential biases in the all-atom energy function, the design protocol, or their combination. In the case of the anti-lysozyme antibody F10.6.6 (PDB entry 1P2C) the natural antibody buries a relatively small surface area and is in the 20th percentile of the overall predicted binding-energy ranking (Table 1). Most of the top-ranked designs that target the same lysozyme epitope bury larger surfaces (>1600 Å^2) by using longer L1 and L3 segments (Figure 9a). 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 1980Å^2, while the best-ranked similar-conformation design buries only 1700 Å^2 (Table 1). This difference in buried surface area is due to a difference in the packing angle between the light and heavy variable domains of the natural and designed antibodies (Figure 9b); more extensive sampling of the orientation of the two antibody variable domains than done here may be necessary to address such inaccuracies. The benchmark and the results of applying *AbDesign* to it 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

The sequence-recapitulation rates in successful designs are in the range of past design benchmarks. The values are not directly comparable, however, since past design work dealt with either functional-site design[7,20,75] or the protein core[94], whereas the antibody-design benchmark deals with both, and since here we constrain sequence and backbone-conformation choices using experimental data, whereas past design benchmarks used all-atom energy functions and modeled backbones without additional restraints. Sequence within the antibody core is recapitulated to within roughly 60-80% identity, which is higher than the previous benchmark attempting to recapitulate native identities in the protein core (51%[94]), and the binding surface sequence identity is approximately 30%, similar to the previous protein-binding benchmark (interface residue sequence identity between 10-40%) [7]. In some cases residues at the interface and the antibody core also conserve side-chain conformations at atomic accuracy (Figure 10).

Amino acid conformational plasticity has the potential to reduce binding specificity and affinity[18,95] 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[79]. 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[95]. Designed binders using existing strategies[95] 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*[79]. We hypothesized that the sequence-structure rules encoded in the backbone-conformation library and 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 (Table S3) 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 (Figure 11). The proportion of low-probability sidechain conformations (<5% probability), which are unlikely to be in their intended conformation in the unbound state, is less than 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 [95,96] 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 towards the antibody sequence consensus *AbDesign* may encode some elements that are necessary for lock-and-key molecular recognition[97–100]. Two examples, the anti-tissue factor designed antibody and an anti-sonic hedgehog protein designed antibody, demonstrate how interface sidechain rigidity is encoded by the backbone conformation and sequence constraint rules (Figure 11).

Discussion

Despite breakthroughs in the design of new molecular function in regions high in secondary-structural elements[2,4–6,101], successful design of function in loop segments has been elusive [14,17,18,62,102]. *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 robust design of function [18]. Although ultimate proof lies in experimental validation of designed antibodies, the natural-antibody recapitulation benchmark offers promising signs that some of the current limitations in computational design may be addressed by this approach[16]: designed surfaces comprise more polar interaction networks, loops, and larger binding regions than in previous design benchmarks. Additionally, designed sidechains are predicted to be more rigid than in natural antibodies, whereas previous studies noted lower sidechain conformation probabilities than natural sets [7,95]; 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[103]. The *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[70]. 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 stability 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 elements that comprise the *AbDesign* algorithm are 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 Rossman 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 encoded[28]; 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 benchmark results 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 solutions to binding the target epitope; indeed, different natural antibodies are known to bind the same epitope[104]. 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 (Figure 9A).

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 analyses noted biases towards 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 towards designs with large binding surfaces (> 1800 Å^2, Table 1), reflecting the correlation between buried surface area and binding affinity in natural protein-protein interactions[105]. The antibodies with small binding surfaces represented in our benchmark 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 Å^2 and K_D of 16nM[86]. Despite the natural antibody's high affinity for its target, *AbDesign* prefers antibodies that bury 1500 Å^2 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[4,5,101]and the benchmark results reported here relied on structural features, such as the intermolecular shape compementarity[81]and packing defects[80]. 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 benchmark and the results of applying *AbDesign* to it can provide a useful framework 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* benchmark 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[106,107] and backbone modeling and design algorithms.

Methods

Source code availability

The methods have been implemented within the Rosetta macromolecular modeling software suite [53] and are available through the Rosetta Commons agreement. All of the methods have been implemented through RosettaScripts[54], and all scripts are available as Supplemental Data.

Binding mode criteria

Deciding which designs recapitulate the native binding mode was based on the CAPRI challenge criteria [92]. Specifically, I_RMS which measures the RMSD of the ligand interface residues (all residues with atoms within a 10 Å radius of the antibody) between design and native structure after both antibody structures are aligned. The interface cutoff was set to 4 Å.

CDR definitions

The CDR definitions used in this work are in general agreement with previous definitions[24–26]. For clarity we present CDR definitions used in this work by Chothia numbering. The design protocol uses two different CDR definitions. The first, closely matching the V(D)J gene segments, treats CDRs 1 and 2 as one unit . This definition is used during the construction of the PSSMs (algorithm section a) and during backbone sampling (section e). The second definition is similar to conventional CDR definitions and treats each CDR (1, 2, and 3) as separate units to determine the level of the sequence constraint thresholds (algorithm, section a).

Table 3. Comparison between CDR definition

CDR		Kabat	Chothia	CDR definitions used for design	Segment definitions used for PSSMs and backbone modeling
VL	L1	L24-L34	L24-L34	L23-L35	L23-L88
	L2	L50-L56	L50-L56	L46-L55	
L3		L89-L97	L89-L97	L88-L98	L88-98
VH	H1	H26-H35	H26-H32	H26-H37	H25-H92
	H2	H50-H65	H50-H58	H45-H58	
H3		H95-H102	H95-H102	H93-H103	H93-H103

Identifying structurally conserved stem positions

To identify structurally conserved stem positions within the immunoglobulin fold we superimposed separately the variable light domains (1300 structures) and the variable heavy domains (1253 structures) to the relevant domain in the template antibody 4M5.3 (PDB entry: 1X9Q) using PyMOL[73]. For each aligned set we ran *StemFinder,* implemented in Rosetta[53] using the RosettaScripts interface[54]. The following parameters were used: position RMSD, the maximum allowed RMSD between a position in the template structure to all possible positions in all of the aligned input structures, was set to 1Å, pair distance, the maximum distance allowed between any two atoms on two candidate positions, was set to 4Å, and neighbor separation, the minimal number of amino-acids (in the primary sequence) between pairs of amino acids, was set to 10 amino acids. Triplets of positions that passed all above criteria were output and manually screened for a triplet that spans all three CDRs; only the two triplets reported in the Results passed these filters.

Shape complementarity

Shape complementarity (*Sc*) was computed using the algorithm described in ref. [81] implemented in Rosetta[53]. *Sc* ranges from 0 (no shape complementarity) to 1 (perfect shape complementarity). In the current benchmark antibody designs with Sc values less than 0.6 were rejected.

Docking of the antibody scaffolds to the target epitope

Each of the 4,500 antibody scaffolds was initially aligned to the natural antibody framework in the complex structure. The ligand coordinates were then added to the antibody scaffold structure. The binding mode was then perturbed with RosettaDock[108] using low resolution docking (centroid mode).

Binding-energy calculations

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 calculations

The stability energy is defined as the system energy of the antibody monomer. To assess the stability energy of the antibody the ligand is removed and the antibody total energy score (score12) is calculated.

Packing-quality assessment

Protein packing quality at the antibody core and antibody-ligand interface were calculated using "RosettaHoles" (Packstat)[80] implemented in Rosetta[53]. In the current benchmark, antibody designs with Packstat values less than 0.57 were rejected.

Boltzmann conformational probabilities of interface side chains

Boltzmann conformational probabilities were calculated as described in ref. [95].

For each partner in the complex and for each residue that contributes more than 1 R.e.u to the predicted binding energy we iterate, in the unbound state, over all the backbone-dependent rotamers defined in the Dunbrack library. For each rotamer, all residues within a 6Å shell are repacked and minimized. The energy *E* of each such state is then evaluated using the Rosetta all-atom energy function (score12) [57]. The probability of the conformation of residue *i*, p_i, is then computed assuming a Boltzmann distribution:

$$P_i = \frac{e^{\frac{-E_i}{K_B T}}}{\sum_s e^{\frac{-Es}{K_B T}}} \qquad (2)$$

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. in all simulations. E_i is the energy of the unbound state.

Backbone segment clustering

The antibody structures in our database were aligned separately to the variable heavy and variable light domains of antibody 4m5.3, (PDB entry 1X9Q)[45]. We then extracted the coordinates of the CDRs according to VL, L3, VH, and H3 definitions (Table 3) and clustered them according to length. For L3 and H3 we preformed additional conformational clustering using BCL::cluster[109]. This additional clustering was needed due to the higher conformational diversity of L3 and H3 compared to the other segments[22,24,110,111]. Backbone conformations were clustered with a 2.0 Å RMSD radius (measured between C_α atoms). The resulting clusters were inspected manually for common sequence motifs. Clusters that contained several different sequence motifs were manually split; conversely, conformation clusters that shared sequence motifs were merged. This clustering resulting in 207 H3 bins, and the top 50 clusters (by size) were used to generate the conformation representatives (algorithm, section d).

The clustering information used for the design process is stored in the Rosetta database and is under version control:

~/Rosetta/main/database/protocol_data/splice/antibodies/pdb_profile_match/

Generating sequence profiles

For each backbone conformation cluster we generated a Position Specific Scoring Matrix (PSSM). We extracted the amino-acid sequence from each structure to first generate a multiple sequence alignment from which we removed 100% sequence redundancy (every sequence in the alignment has at least a single amino acid difference from all other sequences in the alignment). The PSSMs were generated using the PSI-BLAST suite[112] with default parameters and the multiple sequence alignment as input.

The PSSMs are stored in the Rosetta database and are under version control

~/Rosetta/main/database/protocol_data/splice/antibodies/pssm/

Tests ensuring source code stability

To ensure code stability we have added to the Rosetta suite three different integration tests that check different functionalities of the algorithm. 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 tests checks that the algorithm can properly read the torsion database and impose a new backbone conformation onto the template antibody. The splice_seq_constraints integration test checks that the algorithm can add sequence constraints to an antibody structure.

Algorithm performance

A typical trajectory takes about 7 hours from submission to successful completion on a standard single CPU. The protocol is divided to two parts. First, the complex formed between the designed antibody scaffold (algorithm, section d) and the target molecule is subjected to docking, design, and minimization (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 implement a checkpointing policy that ensures that if a design trajectory is prematurely terminated it can be resumed from the last backup point. The checkpointing policy is enforced from the start of the backbone optimization step (algorithm, section d). A PDB-formatted file containing the coordinate information of the complex is saved to disk along with the precise design stage, complex stability, and binding energies, whenever a sampled backbone improves the objective function (algorithm, section g). When *AbDesign* is initiated it automatically checks for the existence of checkpointing files. If checkpointing files are found, *AbDesign* will continue from the same point it was last stopped. Resuming from the backup point takes less than 30 seconds to return to the last point of execution.

Code flow and modularity

The design protocol is implemented using RosettaScripts [54] which provides a convenient user interface to all major Rosetta functionalities. This form of implementation allows the non-expert user with no previous coding knowledge complete control over all aspects of the design protocol. The protocol is intentionally modular so prospective users can add, change or remove different elements in the protocol as they see fit.

Acknowledgments

We thank all members of the Fleishman lab who read and commented on this manuscript, and Eva-Maria Strauch, Dan Tawfik, and Meir Wilchek for helpful comments. Research in the Fleishman lab is supported by the Israel Science Foundation through an individual research grant and through the 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, and a charitable donation from Sam Switzer. SJF is the incumbent of the Martha S Sagon Career Development Chair. CN is supported by a doctoral fellowship from the Boehringer Ingelheim Fonds.

References

1. Huang, P.-S., Love, J. J. & Mayo, S. L. A de novo designed protein protein interface. *Protein Sci.* **16,** 2770–4 (2007).
2. Jha, R. K. *et al.* Computational design of a PAK1 binding protein. *J. Mol. Biol.* **400,** 257–70 (2010).
3. Karanicolas, J. *et al.* A de novo protein binding pair by computational design and directed evolution. *Mol. Cell* **42,** 250–60 (2011).
4. Fleishman, S. J. *et al.* Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* **332,** 816–21 (2011).

5. Strauch, E.-M., Fleishman, S. J. & Baker, D. Computational design of a pH-sensitive IgG binding protein. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 675–80 (2014).
6. Procko, E. *et al.* Computational design of a protein-based enzyme inhibitor. *J. Mol. Biol.* **425,** 3563–75 (2013).
7. Fleishman, S. J. *et al.* Hotspot-centric de novo design of protein binders. *J. Mol. Biol.* **413,** 1047–62 (2011).
8. King, N. P. *et al.* Computational design of self-assembling protein nanomaterials with atomic level accuracy. *Science* **336,** 1171–4 (2012).
9. Gradišar, H. *et al.* Design of a single-chain polypeptide tetrahedron assembled from coiled-coil segments. *Nat. Chem. Biol.* **9,** 362–6 (2013).
10. Fletcher, J. M. *et al.* Self-assembling cages from coiled-coil peptide modules. *Science* **340,** 595–9 (2013).
11. Stranges, P. B., Machius, M., Miley, M. J., Tripathy, A. & Kuhlman, B. Computational design of a symmetric homodimer using β-strand assembly. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 20562–7 (2011).
12. Der, B. S. *et al.* Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J. Am. Chem. Soc.* **134,** 375–85 (2012).
13. Lo Conte, L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285,** 2177–98 (1999).
14. Fleishman, S. J. *et al.* Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *J. Mol. Biol.* **414,** 289–302 (2011).
15. Moretti, R. *et al.* Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions. *Proteins* **81,** 1980–7 (2013).
16. Stranges, P. B. & Kuhlman, B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci.* **22,** 74–82 (2013).
17. Khare, S. D. & Fleishman, S. J. Emerging themes in the computational design of novel enzymes and protein-protein interfaces. *FEBS Lett.* **587,** 1147–54 (2013).
18. Fleishman, S. J. & Baker, D. Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* **149,** 262–73 (2012).
19. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302,** 1364–8 (2003).
20. Zanghellini, A. *et al.* New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **15,** 2785–94 (2006).
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.* 18–27 (1970). at <http://jem.rupress.org/content/132/2/211.abstract>
22. Chothia, C. & Lesk, a M. Canonical structures for the hypervariable regions of immunoglobulins. *J. Mol. Biol.* **196,** 901–17 (1987).
23. Strausbauch, P. H., 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* **10,** 4342–8 (1971).
24. North, B., Lehmann, A. & Dunbrack, R. L. A new clustering of antibody CDR loop conformations. *J. Mol. Biol.* **406,** 228–56 (2011).
25. Al-Lazikani, B., Lesk, a M. & Chothia, C. Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.* **273,** 927–48 (1997).
26. E.A. Kabat, T.T. Wu, H. Bilofsky, M. Reid-Miller, H. P. *Sequence of Proteins of Immunological Interest*. (National Institutes of Health, Bethesda, 1983).
27. Kabat, E. a, Wu, T. T. & 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.* **252,** 6609–16 (1977).

28. Dellus-Gur, E., Toth-Petroczy, A., Elias, M. & Tawfik, D. S. What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. *J. Mol. Biol.* **425,** 2609–21 (2013).
29. Lesk, A. & Chothia, C. Evolution of proteins formed by β-sheets: II. The core of the immunoglobulin domains. *J. Mol. Biol.* 325–342 (1982). at <http://www.sciencedirect.com/science/article/pii/0022283682901796>
30. Riechmann, L., Clark, M., Waldmann, H. & Winter, G. Reshaping human antibodies for therapy. *Nature* **332,** 323–7 (1988).
31. Winter, G. & Milstein, C. Man-made antibodies. *Nature* **349,** 293–9 (1991).
32. Michnick, S. W. & Sidhu, S. S. Submitting antibodies to binding arbitration. *Nat. Chem. Biol.* **4,** 326–9 (2008).
33. Beck, A., Wurch, T., Bailly, C. & Corvaia, N. Strategies and challenges for the next generation of therapeutic antibodies. *Nat. Rev. Immunol.* **10,** 345–52 (2010).
34. Filpula, D. Antibody engineering and modification technologies. *Biomol. Eng.* **24,** 201–15 (2007).
35. Scott, A. M., Wolchok, J. D. & Old, L. J. Antibody therapy of cancer. *Nat. Rev. Cancer* **12,** 278–87 (2012).
36. Ekiert, D. C. & Wilson, I. a. Broadly neutralizing antibodies against influenza virus and prospects for universal therapies. *Curr. Opin. Virol.* **2,** 134–41 (2012).
37. Glennie, M. J. & Johnson, P. W. Clinical trials of antibody therapy. *Immunol. Today* **21,** 403–10 (2000).
38. O'Nuallain, B. & Wetzel, R. Conformational Abs recognizing a generic amyloid fibril epitope. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 1485–90 (2002).
39. Perchiacca, J. M., Ladiwala, A. R. a, Bhattacharya, M. & Tessier, P. M. Structure-based design of conformation- and sequence-specific antibodies against amyloid β. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 84–9 (2012).
40. Schwarz, M. *et al.* Single-chain antibodies for the conformation-specific blockade of activated platelet integrin alphaIIbbeta3 designed by subtractive selection from naive human phage libraries. *FASEB J.* **18,** 1704–6 (2004).
41. Ofek, G. *et al.* Elicitation of structure-specific antibodies by epitope scaffolds. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 17880–7 (2010).
42. McLellan, J. S. *et al.* Structure of RSV fusion glycoprotein trimer bound to a prefusion-specific neutralizing antibody. *Science* **340,** 1113–7 (2013).
43. Throsby, M. *et al.* Heterosubtypic neutralizing monoclonal antibodies cross-protective against H5N1 and H1N1 recovered from human IgM+ memory B cells. *PLoS One* **3,** e3942 (2008).
44. Clark, L. A. *et al.* Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design. 949–960 (2006). doi:10.1110/ps.052030506.tional
45. Lippow, S. M., Wittrup, K. D. & Tidor, B. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat. Biotechnol.* **25,** 1171–6 (2007).
46. Clark, L. a *et al.* An antibody loop replacement design feasibility study and a loop-swapped dimer structure. *Protein Eng. Des. Sel.* **22,** 93–101 (2009).
47. Barderas, R., Desmet, J., Timmerman, P., Meloen, R. & Casal, J. I. Affinity maturation of antibodies assisted by in silico modeling. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 9029–34 (2008).
48. Farady, C. J., Sellers, B. D., Jacobson, M. P. & Craik, C. S. Improving the species cross-reactivity of an antibody using computational design. *Bioorg. Med. Chem. Lett.* **19,** 3744–7 (2009).
49. Miklos, A. E. *et al.* Structure-based design of supercharged, highly thermoresistant antibodies. *Chem. Biol.* **19,** 449–55 (2012).
50. Pantazes, R. J. & Maranas, C. D. OptCDR: a general computational method for the design of antibody complementarity determining regions for targeted epitope binding. *Protein Eng. Des. Sel.* **23,** 849–858 (2010).

51. Pantazes, R. J. & Maranas, C. D. MAPs: a database of modular antibody parts for predicting tertiary structures and designing affinity matured antibodies. *BMC Bioinformatics* **14,** 168 (2013).
52. Li, T., Pantazes, R. J. & Maranas, C. D. OptMAVEn - A New Framework for the de novo Design of Antibody Variable Region Models Targeting Specific Antigen Epitopes. *PLoS One* **9,** e105954 (2014).
53. Das, R. & Baker, D. Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **77,** 363–82 (2008).
54. Fleishman, S. J. *et al.* RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite. *PLoS One* **6,** e20161 (2011).
55. Berezovsky, I. N., Zeldovich, K. B. & Shakhnovich, E. I. Positive and negative design in stability and thermal adaptation of natural proteins. *PLoS Comput. Biol.* **3,** e52 (2007).
56. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421 (2009).
57. Kortemme, T. & Baker, D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 14116–21 (2002).
58. Mandell, D., Coutsias, E. & Kortemme, T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods* **6,** 551–552 (2009).
59. Smith, C. a & Kortemme, T. Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *J. Mol. Biol.* **380,** 742–56 (2008).
60. Canutescu, A. a, Jr, R. L. D. & Dunbrack, R. L. Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Sci.* **12,** 963–72 (2003).
61. Tyka, M. D. *et al.* Alternate States of Proteins Revealed by Detailed Energy Landscape Mapping. *J. Mol. Biol.* **405,** 607–618 (2011).
62. Hu, X., Wang, H., Ke, H. & Kuhlman, B. High-resolution design of a protein loop. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 17668–73 (2007).
63. Richter, F. *et al.* Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *J. Am. Chem. Soc.* **134,** 16197–206 (2012).
64. Midelfort, K. S. *et al.* Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. *J. Mol. Biol.* **343,** 685–701 (2004).
65. Dunbar, J. *et al.* SAbDab: the structural antibody database. *Nucleic Acids Res.* **42,** D1140–6 (2014).
66. Padlan, E. A. Structural basis for the specificity of antibody–antigen reactions and structural mechanisms for the diversification of antigen-binding specificities. *Q. Rev. Biophys.* **10,** 35–65 (1977).
67. Singer, I. *et al.* Optimal Humanization. **150,** 2844–2857 (2000).
68. Padlan, E. A possible procedure for reducing the immunogenicity of antibody variable domains while preserving their ligand-binding properties. *Mol. Immunol.* **28,** 489–498 (1991).
69. Tramontano, a, Chothia, C. & Lesk, a M. 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.* **215,** 175–82 (1990).
70. Foote, J. & Winter, G. Antibody framework residues affecting the conformation of the hypervariable loops. *J. Mol. Biol.* **224,** 487–99 (1992).
71. Shirai, H., Kidera, a & Nakamura, H. Structural classification of CDR-H3 in antibodies. *FEBS Lett.* **399,** 1–8 (1996).
72. Leaver-Fay, A. *et al.* in *Comput. Methods, Part C* (Enzymology, M. L. J. and L. B. B. T.-M. in) **Volume 487,** 545–574 (Academic Press, 2011).
73. Schrödinger, L. *The PyMOL Molecular Graphics Development Component, Version~1.0*. (2010).

74. Chaudhury, S. *et al.* Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLoS One* **6,** e22477 (2011).
75. Allison, B. *et al.* Computational design of protein-small molecule interfaces. *J. Struct. Biol.* (2013). doi:10.1016/j.jsb.2013.08.003
76. Burton, D. R., Poignard, P., Stanfield, R. L. & Wilson, I. A. Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses. *Science* **337,** 183–6 (2012).
77. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **33,** W363–7 (2005).
78. Siegel, J. B. *et al.* Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* **329,** 309–13 (2010).
79. Tinberg, C. E. *et al.* Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* **501,** 212–6 (2013).
80. Sheffler, W. & Baker, D. RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci.* **18,** 229–39 (2009).
81. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234,** 946–50 (1993).
82. F.C.Bernstein, T.F.Koetzle, G.J.Williams, E.E.Meyer Jr., M.D.Brice, J.R.Rodgers, O.Kennard, T.Shimanouchi, M. T. The Protein Data Bank: A Computer-based Archival File For Macromolecular Structures. *J. of. Mol. Biol., 112* 535 (1977).
83. Ganesan, R. *et al.* Unraveling the allosteric mechanism of serine protease inhibition by an antibody. *Structure* **17,** 1614–24 (2009).
84. Luftig, M. A. *et al.* Structural basis for HIV-1 neutralization by a gp41 fusion intermediate-directed antibody. *Nat Struct Mol Biol* **13,** 740–747 (2006).
85. Spiegel, P. C. Structure of a factor VIII C2 domain-immunoglobulin G4kappa Fab complex: identification of an inhibitory antibody epitope on the surface of factor VIII. *Blood* **98,** 13–19 (2001).
86. Mylvaganam, S. E., Paterson, Y. & Getzoff, E. D. 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.* **281,** 301–22 (1998).
87. Bhat, T. N. *et al.* Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci.* **91 ,** 1089–1093 (1994).
88. Cauerhff, A., Goldbaum, F. a & Braden, B. C. Structural mechanism for affinity maturation of an anti-lysozyme antibody. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 3539–44 (2004).
89. Prasad, L., Waygood, E. B., Lee, J. S. & Delbaere, L. T. The 2.5 A resolution structure of the jel42 Fab fragment/HPr complex. *J. Mol. Biol.* **280,** 829–45 (1998).
90. Maun, H. R. *et al.* Hedgehog pathway antagonist 5E1 binds hedgehog at the pseudo-active site. *J. Biol. Chem.* **285,** 26570–80 (2010).
91. Faelber, K., Kirchhofer, D., Presta, L., Kelley, R. F. & Muller, Y. A. 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.* **313,** 83–97 (2001).
92. Méndez, R., Leplae, R., De Maria, L. & Wodak, S. J. Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins* **52,** 51–67 (2003).
93. Tomlinson, I. M. *et al.* The imprint of somatic hypermutation on the repertoire of human germline V genes. *J. Mol. Biol.* **256,** 813–17 (1996).
94. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 10383–8 (2000).
95. Fleishman, S. J., Khare, S. D., Koga, N. & Baker, D. Restricted sidechain plasticity in the structures of native proteins and complexes. *Protein Sci.* **20,** 753–7 (2011).

96. Dunbrack, R. L. & Karplus, M. Conformational analysis of the backbone-dependent rotamer preferences of protein sidechains. *Nat Struct Mol Biol* **1,** 334–340 (1994).
97. Yin, J. *et al.* A comparative analysis of the immunological evolution of antibody 28B4. *Biochemistry* **40,** 10764–73 (2001).
98. 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.* **39,** 801–808 (2003).
99. Manivel, V., Sahoo, N. C., Salunke, D. M. & Rao, K. V. Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site. *Immunity* **13,** 611–20 (2000).
100. Yin, J., Beuscher, A. E., Andryski, S. E., Stevens, R. C. & Schultz, P. G. Structural Plasticity and the Evolution of Antibody Affinity and Specificity. *J. Mol. Biol.* **330,** 651–656 (2003).
101. Procko, E. *et al.* A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells. *Cell* **157,** 1644–56 (2014).
102. Mandell, D. J. & Kortemme, T. Computer-aided design of functional protein interactions. *Nat. Chem. Biol.* **5,** 797–807 (2009).
103. Kettleborough, C. a, Saldanha, J., Heath, V. J., Morrison, C. J. & Bendig, M. M. Humanization of a mouse monoclonal antibody by CDR-grafting: the importance of framework residues on loop conformation. *Protein Eng.* **4,** 773–83 (1991).
104. Pons, J., Stratton, J. R. & Kirsch, J. F. How do two unrelated antibodies, HyHEL-10 and F9.13.7, recognize the same epitope of hen egg-white lysozyme? *Protein Sci.* **11,** 2308–15 (2002).
105. Chen, J., Sawyer, N. & Regan, L. Protein-protein interactions: general trends in the relationship between binding affinity and interfacial buried surface area. *Protein Sci.* **22,** 510–5 (2013).
106. Leaver-Fay, A. *et al. Scientific benchmarks for guiding macromolecular energy function improvement. Methods Enzymol.* **523,** 109–43 (2013).
107. Song, Y., Tyka, M., Leaver-Fay, A., Thompson, J. & Baker, D. Structure-guided forcefield optimization. *Proteins* **79,** 1898–909 (2011).
108. Gray, J. J. *et al.* Protein–Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations. *J. Mol. Biol.* **331,** 281–299 (2003).
109. Alexander, N., Woetzel, N. & Meiler, J. Bcl:: Cluster: A method for clustering biological molecules coupled with visualization in the Pymol Molecular Graphics System. in *Comput. Adv. Bio Med. Sci. (ICCABS), 2011 IEEE 1st Int. Conf.* 13–18 (2011).
110. Shirai, H., Kidera, A. & Nakamura, H. H3-rules: identification of CDR-H3 structures in antibodies. *FEBS Lett.* **455,** 188–97 (1999).
111. Chothia, C., Lesk, A. & Tramontano, A. Conformations of immunoglobulin hypervariable regions. *Nature* (1989). at <http://www.researchgate.net/publication/20467932_Conformations_of_immunoglobulin_hypervariable_regions/file/32bfe5101479e55cfc.pdf>
112. Biegert, A. & Söding, J. Sequence context-specific profiles for homology searching. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 3770–5 (2009).
113. Giudicelli, V., Chaume, D. & Lefranc, M.-P. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. *Nucleic Acids Res.* **33,** D256–61 (2005).

Table 1. Bound antibody complexes used for recapitulation benchmark

PDB entry	Natural antibodies				Ligand	K_d[d] (nM)	Designed antibodies															
	buried surface area (Å)[a]	Packing score[b]	Shape complementarity[c]	Predicted binding energy (R.e.u)			Predicted binding energy rank[e]	Predicted binding energy (R.e.u)	buried surface area (Å)[a]	Antibody backbone RMSD (Å)[g]	Ligand interface RMSD (Å)[h]	Packing score[b]	Shape complementarity[c]	Same as natural segment[i]				Overall sequence identity (%)	Interface sequence identity (%)[j]	CDR sequence identity (%)[k]		Core sequence identity (%)[l]
														VL	L3	VH	H3			VH	VL	
1JPS	1950	0.66	0.70	-25	Tissue factor	0.1	1/1809	-38.8	2063	1.23	2.20	0.66	0.62	v	v	v	v	68	31	67	54	89
1WEJ	1220	0.70	0.75	-16	Cytochrome C	15.8	3/93	-24.5	1535	1.51	2.00	0.67	0.62	v	v	v	0	64	37	61	43	82
2CMR	2110	0.58	0.72	-22	Transmembrane glycoprotein	0.005	11/297	-26.3	2162	1.15	1.40	0.57	0.60	v	v	v	v	64	26	61	68	75
3MXW	1882	0.70	0.51	-21	Sonic hedgehog protein	0.7	24/1274	-32.2	2011	1.52	2.72	0.69	0.58	v	v	v	v	58	53	42	54	60
1VFB	1405	0.67	0.69	-22	Lysozyme	3.7	24/250	-24.3	1493	1.15	3.20	0.64	0.60	v	v	v	v	63	27	60	56	80
2JEL	1549	0.66	0.58	-17	Phosphocarrier protein HPr	3.7	9/50	-20.4	1353	1.30	2.70	0.62	0.60	v	v	v	-1	68	13	71	80	77
3K2U	1982	0.62	0.68	-29.2	Hepatocyte growth factor activator	0.16	51/112	-26.6	1695	1.31	3.20	0.58	0.62	v	v	v	-1	70	40	62	86	77
1P2C	1467	0.68	0.67	-17	Lysozyme	0.0098	138/659	-22.2	1566	1.30	3.90	0.68	0.60	v	v	-1	1	52	34	52	59	64
1IQD	2134	0.70	0.78	-32	Coagulation factor VIII	0.0014	762/2802	-24.7	1632	1.24	2.80	0.66	0.67	v	v	v	v	62	26	66	60	71

[a] Binding surface area is the area excluded from water upon ligand binding.
[b] A criterion that measures how well the antibody core and the binding surface are packed[80]
[c] Shape complementarity between the binding surface of the antibody and ligand [81]
[d] K_d values of the natural antibody taken from the "SabDab" database.
[e] All designed antibodies that are within a 4 Å distance of the target epitope are included in the ranking (see table S5 for number of total designs produced).
[f] Similar conformations are all designed antibodies that are comprised of segments of the same conformational clusters as the natural antibody
[g] RMSD was calculated over all C_α atoms
[h] Ligand interface is calculated over all ligand C_α atoms within a 10 Å radius of the antibody.
[i] A "v" signifies that design has a segment from the same conformational cluster as the natural antibody. If not, the number states the amino acid length change in the design segment relative to the natural antibody.
[j] Interface residues are all antibody residues within a 10 A distance of the ligand
[k] CDR residues do not include interface residues
[l] Core residues are all the antibody residues that are not part of the CDR and are not solvent exposed

Table 2. Analysis of germline and sequence identity of top ranked designs

PDB entry[a]	Source PDB names[b]				Sequence identity[c]				Germline gene[d]					
									VL		L3		VH	
	VL	L3	VH	H3	VL	L3	VH	H3	Germline Genes names	Seq. id[e]	Germline Genes	Seq. id	Germline Genes	Seq. id
1JPS	3IDI	1S3K	2A6I	1FGN	71%	50%	49%	72%	IGKV12-46*01(human)/ IGKV1-39*01(human)	71%	IGKJ5*01(human)/ IGKJ1*01(human)	50%	IGHV1-69*02(murine)/ IGHV3-66*01(human)	49%
2CMR	3CXD	3IFL	3GI8	2DBL	71%	41%	57%	23%	IGKV1-13*02(murine)/ IGKV1-5*03(human)	71%	IGKJ2*01(murine)/ IGKJ4*01(human)	42%	IGHV1-58*01 murine)/ IGHV1-69*01(human)	61%
3MXW	1T3F	1NBY	2I5Y	1ZTX	68%	54%	42%	53%	IGKV1-5*01(human)/ IGKV1-33*01(human)	71%	IGKJ1*02(mouse)/ IGKJ4*01(human)	75%	IGHV1-69*02(human)/ IGHV1-2*02(human)	76%
1VFB	2VDL	1CIC	3FO0	3HR5	62%	63%	85%	45%	IGKV14-100*01(mouse)/ IGKV12-41*02(mouse)	59%	IGKJ4*01(mouse)/ IGKJ1*01(mouse)	83%	IGHV2-9*01(mouse)/ IGHV2-6-7*01(mouse)	90%
1IQD	2NY7	1IGJ	1UJ3	1FL5	74%	25%	61%	36%	IGKV3-20*01(human)/ IGKV3-20*01(human)	100%	IGKJ1*01(mouse)/ IGKJ5*01(human)	75%	IGHV1-46*01(human)/ IGHV1-24*01(human)	78%

[a] Name of the natural antibody target PDB entry.
[b] Names of the antibody PDB entries from which the conformational segment used to construct the antibody design were derived.
[c] Sequence identity between the original sequence of the source antibody segment and the corresponding segment on the natural antibody.
[d] Names of the germline genes from which the segment was derived (top) and the germline gene of the natural antibody target (Bottom) (Derived from the IMGT/GENE-DB [113])
[e] Sequence identity between the germline genes of the target antibody segments and the designed antibody segments.

Figures

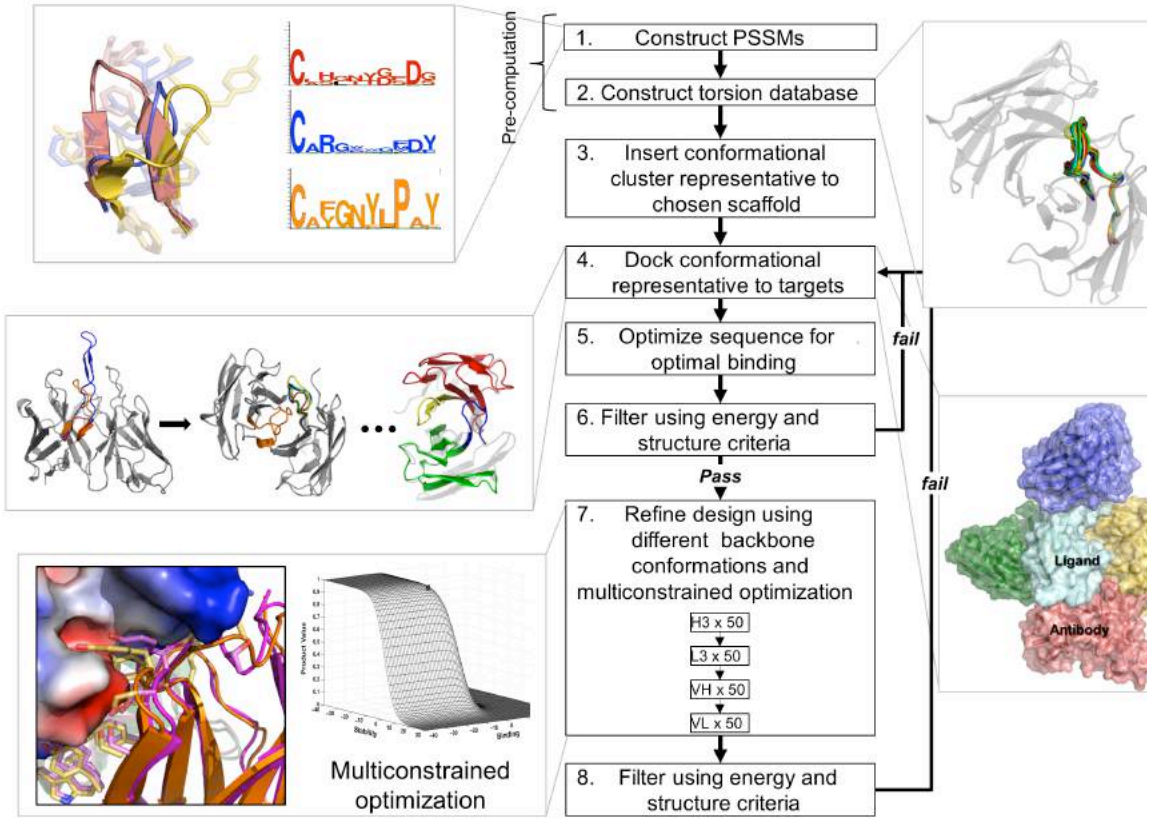


Figure 1. Overview of the design protocol workflow. Briefly, structures of naturally occurring antibodies extracted from the Protein Data Bank (PDB)[82] and aligned to a template antibody structure. Backbone segm conformations of each antibody are extracted and the segment sequences are compiled into a Position-Spec Matrix (PSSM) (step 1) to constrain amino acid residue choices during design. The backbone conformations entered into a conformation database (step 2), and a set of antibody conformations representing all combin canonical conformations is generated (step 3), docked against the target surface (step 4) and designed for c binding affinity (step 5). Antibodies passing structure and energy filters are then subjected to a backbone a sequence refinement protocol (step 7): for each backbone segment (VL, VH, L3, H3) alternative conformatic sampled from the pre-computed conformation database and designed in the context of the modeled antibo structure. The backbone conformation with the highest computed stability and affinity for the ligand is sele serves as input in the optimization of the next backbone segment. Finally, designs are filtered using energy structural criteria derived from natural antibodies (step 8).

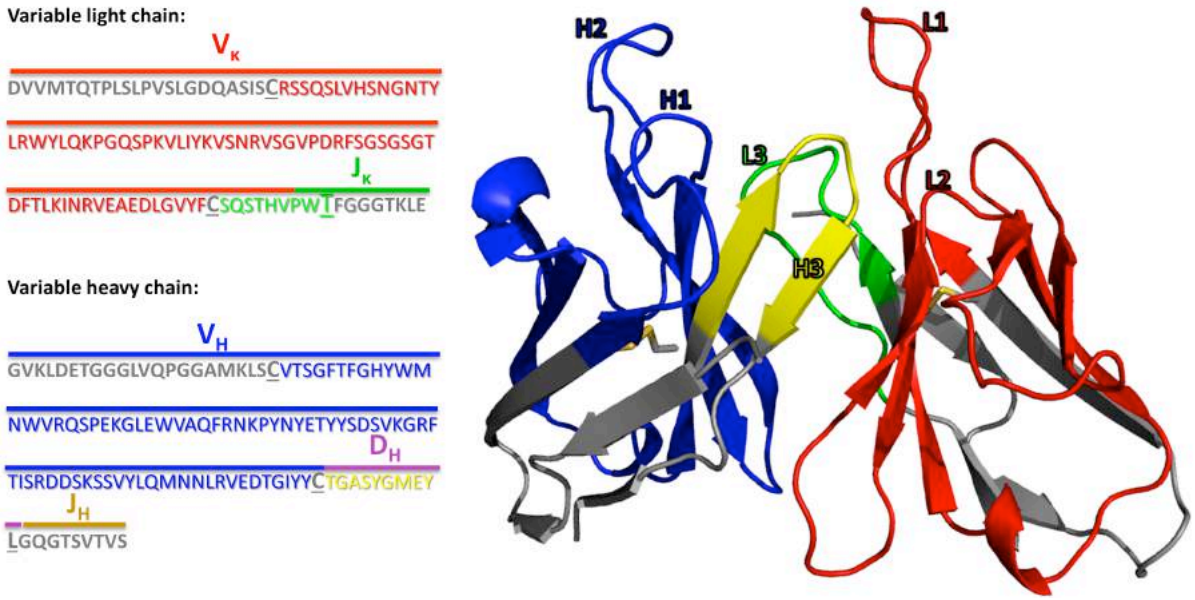


Figure 2.Comparison between V (D) J gene segmentation and conformation segmentation used in *AbDesign*. *AbDesign* segments the antibody structure in places of highest structure conservation among antibodies (the disulfide cysteines and the stem positions of CDR3) to improve the potential of different conformation segments to be joined to form artificial combinations of backbones. Natural antibody recombination follows a similar, but not precisely the same, segmentation. An antibody sequence (from antibody 4m5.3) is shown on the left, corresponding to the structure from PDB entry 1X9Q (right). Sequence and structure are color-coded by conformation segments (red: CDRs L1&L2, green: L3, blue: H1&H2, yellow: H3). Bars above the sequence denote the germline-gene origin of the antibody segments. Stem positions in the sequence are underlined. Gray segments are only subjected to sequence optimization, rather than combined backbone-sequence optimization.

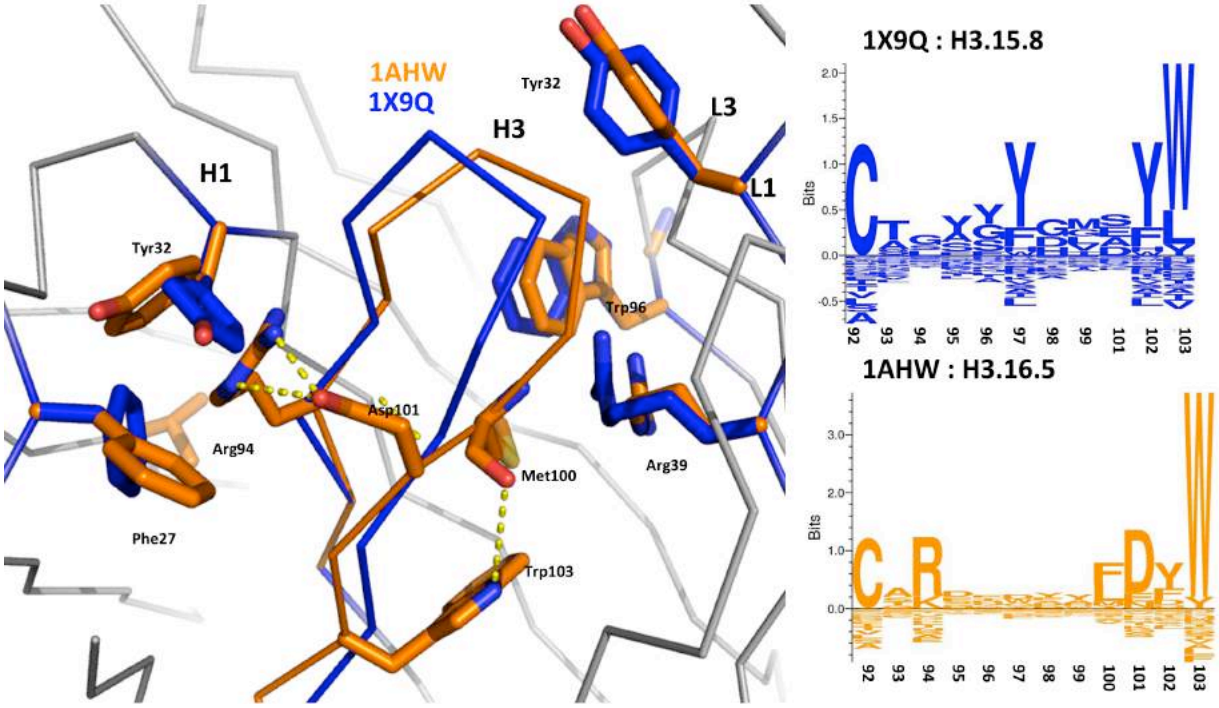


Figure 3. **Sequence and conformation coupling during design**. Imposing the H3 segment backbone conformation from the source antibody 5G9 (PDB entry 1AHW), onto the template antibody scaffold (4m5.3, PDB entry 1X9Q). Each backbone conformation cluster is associated with a different PSSM. When a backbone conformation is sampled, the PSSM of the entire antibody is re-constructed. The PSSM is used to constrain sequence design choices. For instance, The original H3 backbone conformation from 1X9Q is extended conformation whereas the imposed H3 backbone conformation is kinked [71], and characterized by a hydrogen bond between the conserved stem Trp (Trp103, Chothia numbering) Nε1 atom and a carbonyl oxygen (Met100, Chothia numbering). The conserved salt bridge between Arg94 and Asp101 is frequently observed in kinked conformations and probably serves to stabilize this conformation. 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 the example shown, residues Phe27 and Tyr32 from the heavy chain and residue Tyr32 from the light chain are repacked to avoid clashes with the new backbone conformation.

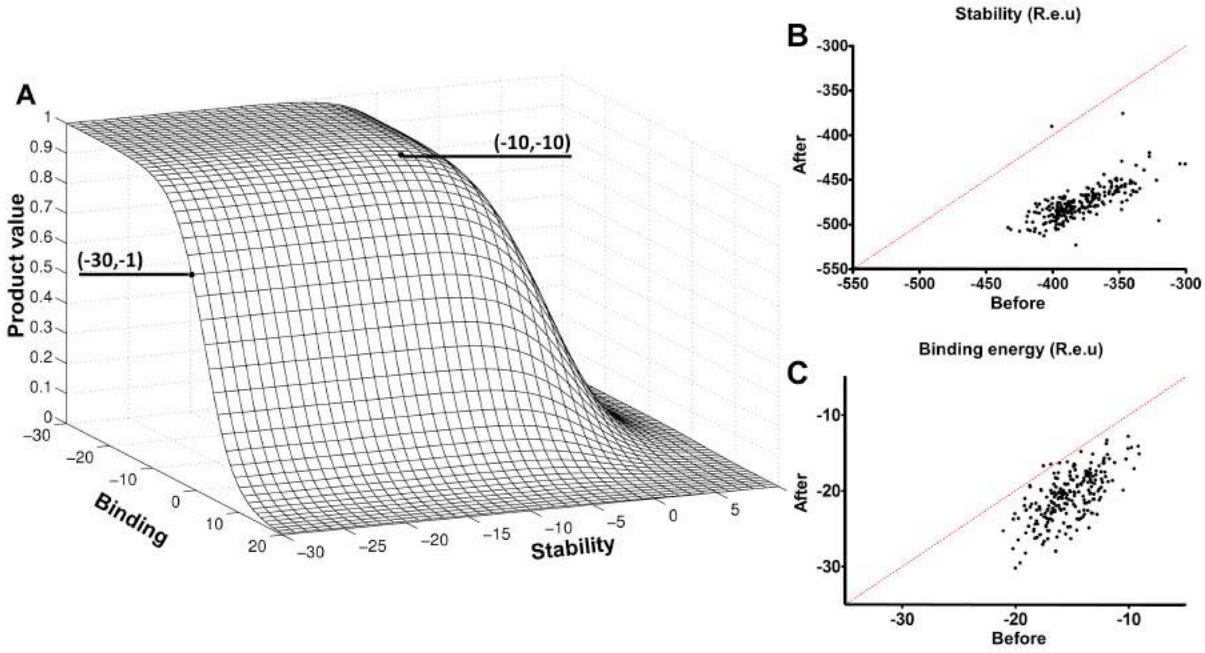


Figure 4. Designed backbone conformations are evaluated by an objective function that is constrained by both antibody stability and binding affinity. A. Plot of the objective function, the product value of the stability and binding sigmoids (Eq. 1). A transformed value of a -10 R.e.u change in binding and stability is prefered to a -30 R.e.u change in stabilty and a -1 R.e.u change in binding. The product of the two transformations gauges the effect the incorporated segment has on the antibody's stability and binding affinity for the target relative to the baseline score (the interim best scoring antibody structure so far). **B & C**. Comparison between the stability and binding energy of the 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). On average we find an improvement of 5 R.e.u for binding energy and 100 R.e.u. for antibody stability following the backbone refinement phase.

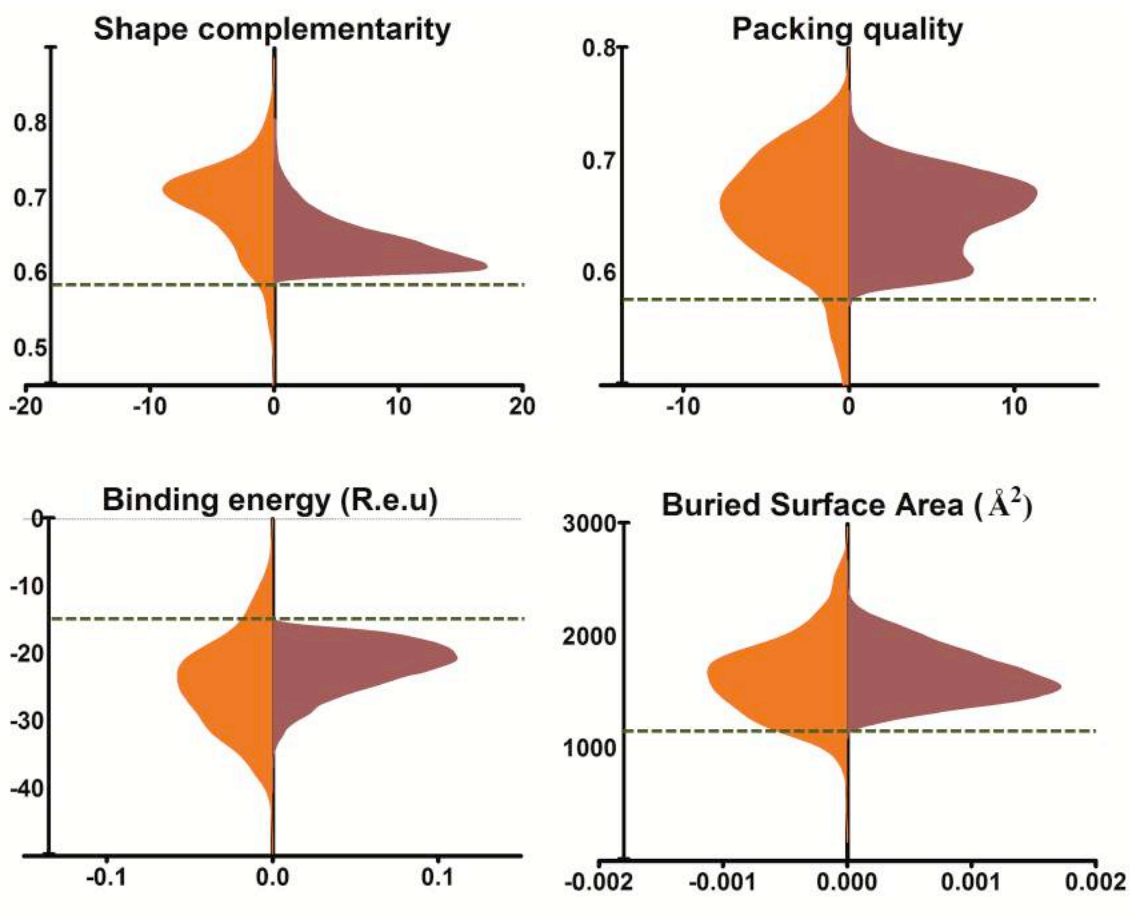


Figure 5. Energy and structure criteria used to filter designed antibody structures. In the final step of *AbDesign* we filter the designed antibodies according to four parameters: predicted binding energy, buried surface area, shape complementarity between antibody structure and ligand, and packing quality between the variable light and heavy domains and the ligand. Cutoffs (green dashed lines) were derived from a set of 303 natural protein-binding antibodies (Table S3). Antibody designs (purple) that passed all filters are compared to the natural protein-binding antibodies (gold).

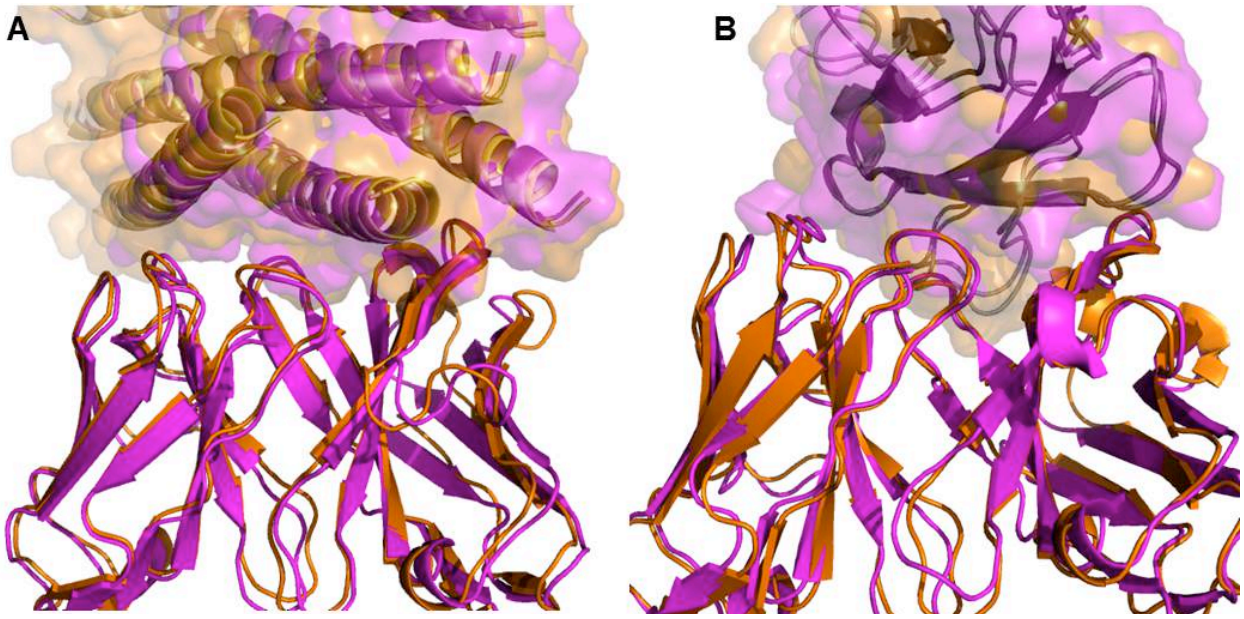


Figure 6. Antibody designs have similar backbone conformations to natural antibodies that target the same surface. Comparison between the backbone conformation of designed (magenta) and natural (orange) antibodies targeting to the same surface. (A). The anti-transmembrane glycoprotein (D5 neutralizing mAb, PDB entry 2CMR). C_α RMSD between the design and natural antibody is 1.1Å, and ligand interface RMSD is 2.7 Å. (B). The anti-tissue factor protein (D3H44, PDB entry 1JPS). C_α RMSD between design and natural antibody is 1.2 Å and ligand interface RMSD is 2.2Å.

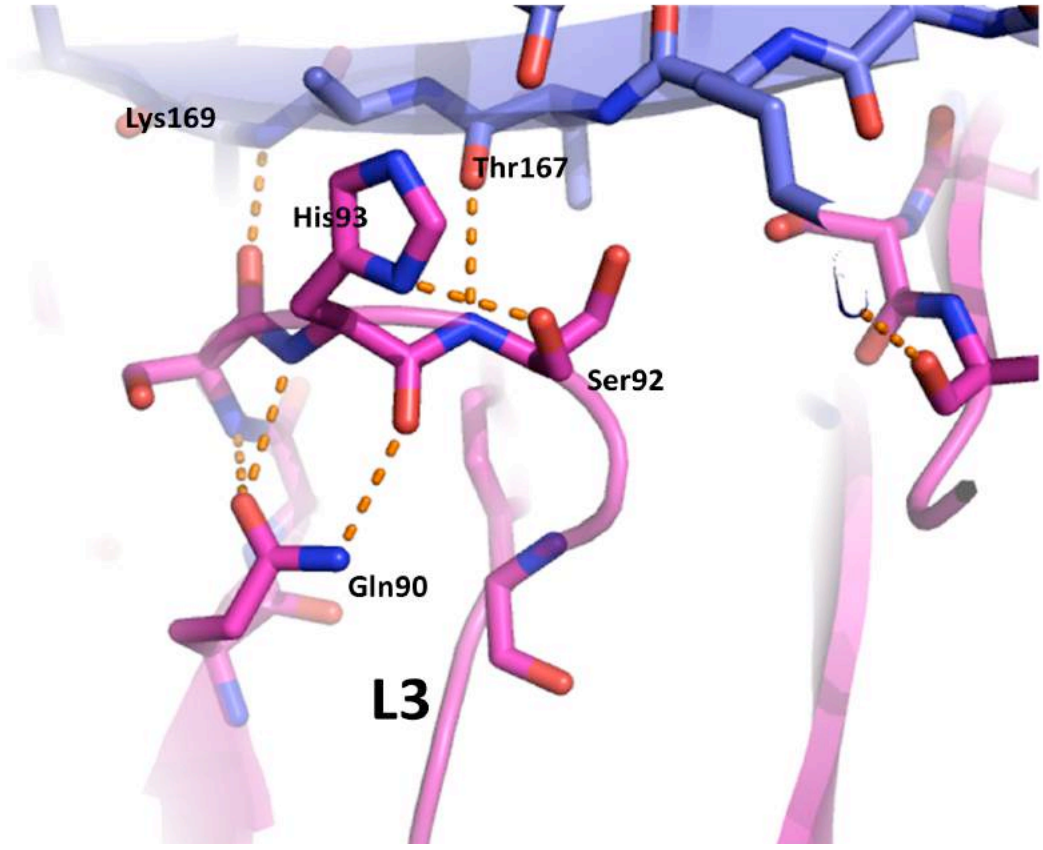


Figure 7. Designed antibody-backbone atoms form polar contacts with the ligand. The best predicted binding affinity design (magenta) of an anti tissue-factor antibody is shown with the target ligand (blue). Two polar contacts (dashed orange lines) are formed between the L3 Ser92 amide nitrogen and the carbonyl group of Thr167 from the tissue factor and between His93 carbonyl and the amide nitgoen of Lys 169 from tissue facton. In addition the conserved Glu90 forms multiple hydrogen bonds with the backbone atoms of the L3 loop that stabilze the conformation.

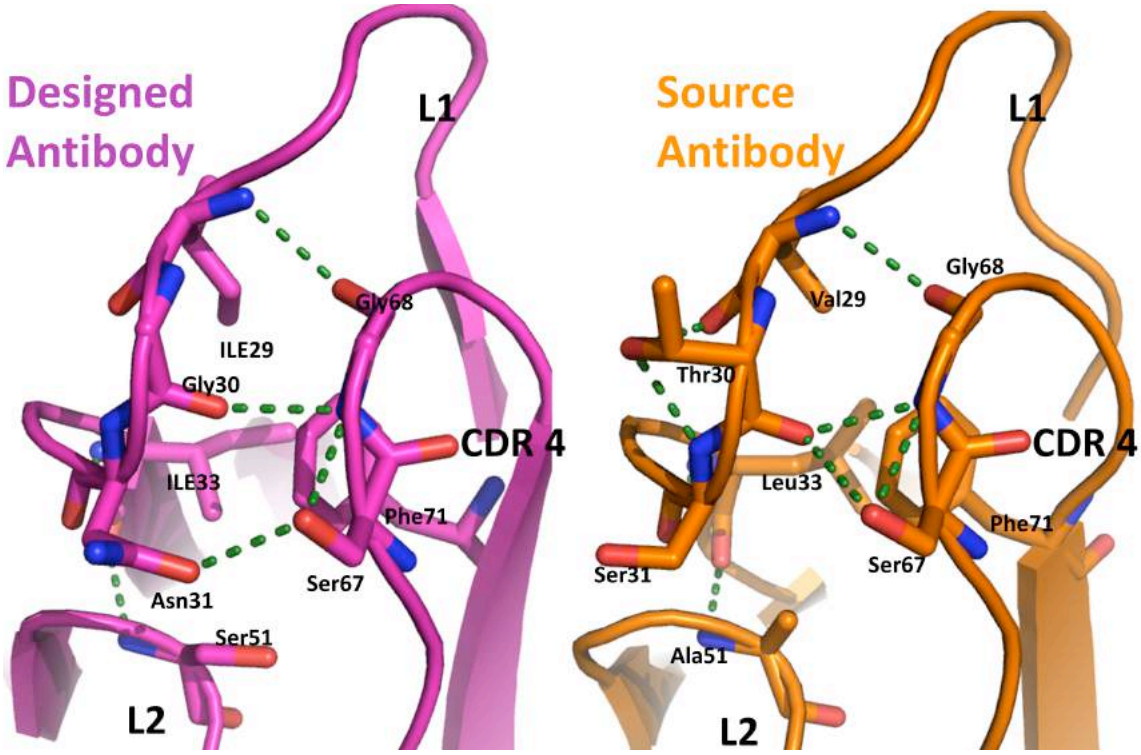


Figure 8. Designed backbone fragments conserve the stabilizing interactions observed in the natural source. The natural VL segment from PDB entry 3IDI (orange) encodes long-range stabilizing interactions between CDR L1 and the framework, for instance, using hydrogen bonds (dashed green lines) hydrophobic and aromatic-stacking interactions. Though the VL segment used in the design targeting tissue factor (target PDB entry: 1JPS, left) has a different sequence than that of the source fragment (right), the same types of stabilizing interactions are made in the designed fragment. By using sequence constraints from natural antibodies and the corresponding conformation segments that encompass both the CDRs and the framework backbone that supports the CDRs these crucial long-range interactions are maintained in design.

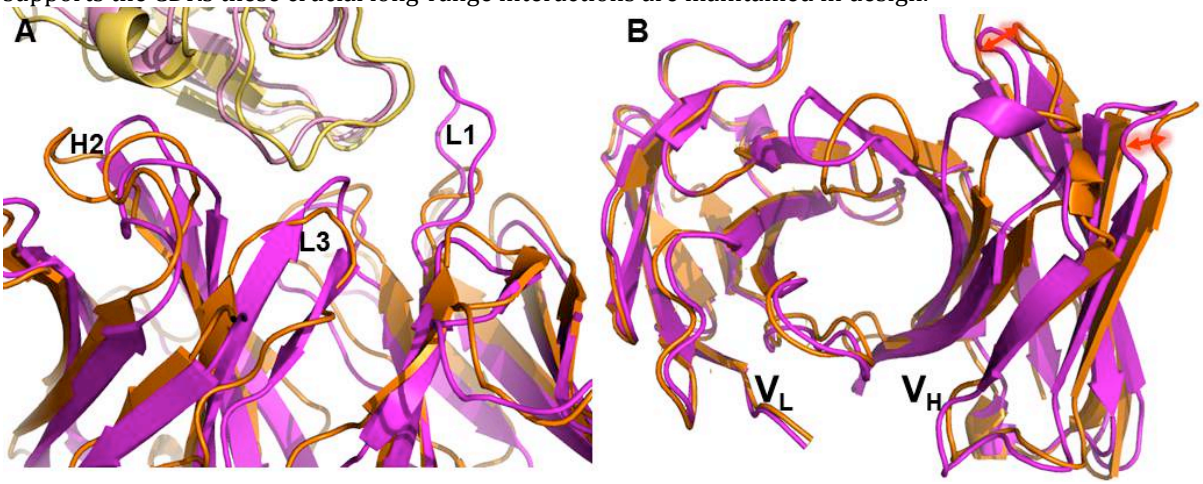


Figure 9. *AbDesign* favors larger binding surfaces. (A). Comparison between the top-ranked predicted

binding energy anti-lysozyme antibody (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 Å^2 to 1680 Å^2.(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 towards the viewer. A 10^o 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 in turn increases the buried surface area in the natural antibody.

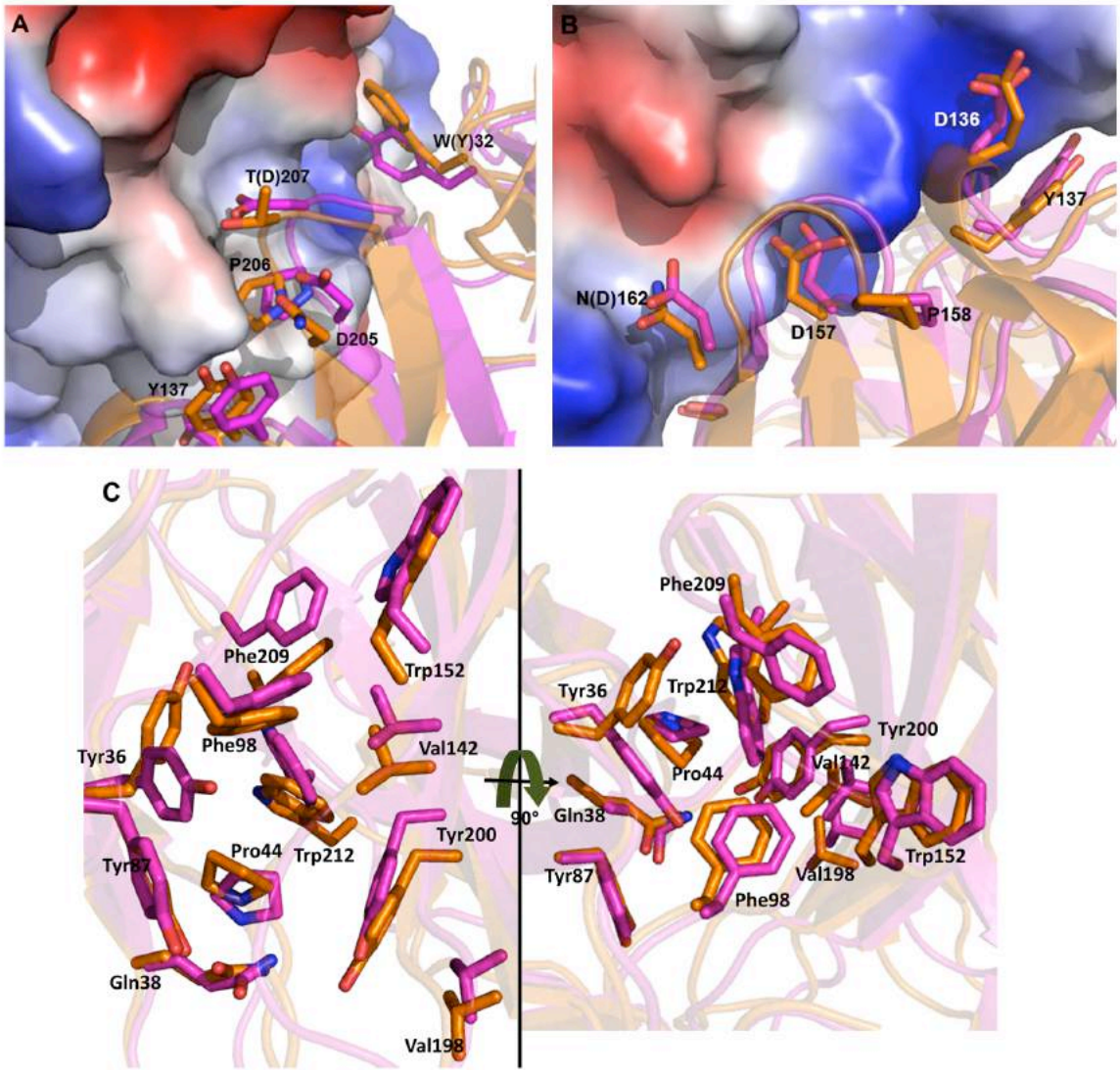


Figure 10. Designed antibodies recapitulate the identity and conformation of binding surface and core residues. Comparison between conserved interface residues on designed (magenta) and natural (gold) antibodies. (A). The anti-transmembrane glycoprotein antibody design. (B). The anti-tissue factor design. Residue designations in parentheses mark where the design residues are different from the natural residues. Ligand is shown in surface representation colored by vacuum electrostatics using Pymol (69).(C) Residues at the VL/VH interface. The natural antibody D3H44 (PDB entry: 1JPS) is colored orange.

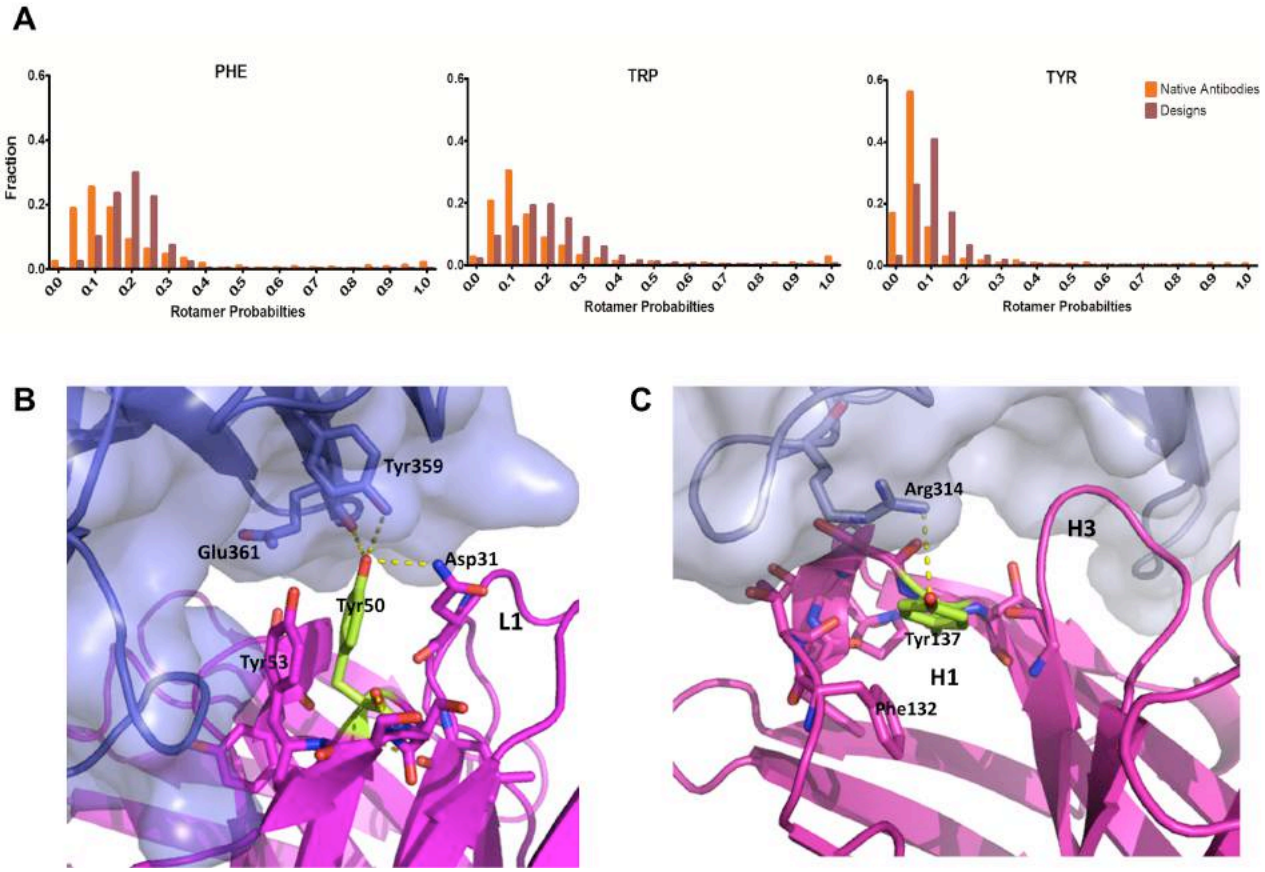


Figure 11. Designed sidechains are predicted to be rigid. (A). The sidechain conformation probabilities in the unbound state were computed using the method in Ref. [95]. *AbDesign* produces antibody complexes with a lower proportion of low-probability conformations (≤ 0.05 probability) compared to natural antibody complexes. The natural antibody complex set comprises 303 antibody-protein complexes (supplemental table S3) extracted from the PDB using the SabDab database. The designed antibody set includes all designs generated by the design protocol in the benchmark set. (B). The designed antibody against the sonic hedgehog protein. The constrained tyrosine (colored green), situated on L2, forms two hydrogen bonds with the ligand and is stabilized by packing against the backbone atoms of H3, L2 and the side chain atoms of a second tyrosine. Additional stabilization is achieved by a hydrogen bond between a glutamine side chain and the hydroxyl group of the tyrosine. According to the rigidity measure there is a 90% probability for the bound state rotamer in the unbound state. (C). The anti-tissue factor protein designed antibody. A tyrosine (green) is placed on H1 forming a hydrogen bond with an arginine on the ligand. The constrained tyrosine is stabilized by packing against the backbone atoms of H1 and H3 and the side chain of a phenylalanine. The constrained tyrosine has a 60% probability for the bound state conformation in the unbound state according to the sidechain-conformation probability measure.