Thesis for the degree
Doctor of Philosophy

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

עבודת גמר (תזה) לתואר
דוקטור לפילוסופיה

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

By
Maria Gabriele Pszolla

מאת
מריה גבריאל פשולה

תכנון חישובי של קישור בנוגדנים קונבנציונליים ובנוגדנים בעלי שרשרת כבדה בלבד

Computational Design of Novel Interactions in Conventional and Heavy Chain Antibodies

Advisor:
Sarel-Jacob Fleishman

מנחה:
שראל-יעקב פליישמן

July 2017

אב תשע״ז

'But I don't want to go among mad people,' Alice remarked.
'Oh, you can't help that,' said the Cat: 'we're all mad here. I'm mad. You're mad.'
'How do you know I'm mad?' said Alice.
'You must be,' said the Cat, 'or you wouldn't have come here.'

- Lewis Carroll, "Alice's Adventures in Wonderland"

Declaration

This thesis is a presentation of my original research work. As development of the algorithm was a joined effort, contributions of others are shown throughout the thesis, and every effort is made to indicate these clearly. Briefly, writing of new code required for installment of new design principles was largely executed by Gideon Lapidoth; the final algorithm was published [42], Gideon was first author of this publication. Cloning and YSD testing of designs against insulin was performed by Dror Baran; Dror and myself are equally contributing author of the publication encompassing the part of this thesis concerned with conventional antibodies [Baran et al., PNAS 2017 *in press*]. Identification of the factors contributing to the shortcomings of each design round were identified largely by Dror Baran and myself, and in discussion with Sarel Fleishman, we worked out possible approaches for overcoming them. Soluble expression of Fab fragments for characterization of the designed antibodies in solution was done at The ISrael STructural Proteomics Center at the Weizmann Institute.

Abstract

Natural proteins must meet two criteria in order to fulfil their specific tasks within the biological context: First, they have to fold into a stable conformation, and second, they have to exert their molecular function. In computational design, both issues have been addressed separately to date: On the one hand, stable and atomically accurate proteins have been produced by using ideal folds, rich in secondary structure and almost devoid of loops. On the other hand, molecular function, such as binding and catalysis, has been designed by placing appropriate residues on the surface of existing, stable scaffolds, usually consisting of ideal folds. In natural proteins, however, molecular function generally demands non-ideal features, including large loops and buried polar interaction networks, which have remained inaccessible to fold design. Through five design/experiment cycles, we established an algorithm that succeeded in designing stable and functional antibody variable fragments (Fv). In each cycle, thousands of models were generated and the best selected for experimental testing. Selected designs were converted into single-chain variable fragments (scFvs), and as such tested for expression and ligand-binding in yeast -surface display (YSD). The experimental results from each cycle were used to improve the algorithm before starting another design cycle. functional designs were then expressed solubly as Fabs for further characterization. I succeeded in generating two binding designs against ACP, and another antibody recognizing its designed target insulin was generated in the lab, using the same algorithm. After in-vitro evolution using error-prone PCR and fluorescence-activated cell sorting (FACS), the Fab constructs of all designs showed affinities in the mid-nanomolar range and stabilities comparable to natural antibodies.

For the insulin binder as well as another non-binding insulin design, structures of the unbound antibodies could be determined by crystallography and demonstrated atomic accuracy throughout the framework and in all CDRs for the non-binding design and in four out of six in the insulin binder. Also the rigid-body orientation between the light and heavy chain were modeled accurately. I further adapted the algorithm to producing expressible and stable single-domain antibodies (v_HH)and attempted to design these for function by active-site matching of a phenylalanine binding site. Out of 34 designs tested in total, 24 were designed for stability only while ten were designed for phenylalanine binding. 32 of the designs expressed well, but none recognized the target, suggesting that either the scaffolds were not ideal, or that biotin/streptavidin is not a suitable method for detection of small molecule binding.

Acknowledgements

I want to express my deep gratitude to my supervisor, Sarel Fleishman, for his unfaltering support in the years leading up to this day, for nurturing my curiosity and allowing me to follow my ideas, regardless of how little promise it might have held at the time. I'm also very grateful to the entire Fleishman lab for their advice and support inside, and their friendship outside the lab.

I want to thank Emmanuel Levy and Ron Diskin for helping me work out the kinks throughout my research and the reviewers for their time and effort in critically reading this thesis.

I am also very happy and proud to have gained new friends, many of whom have become my family over the last years, who have filled my life with joy, and helped me build a second home away from home - I will always hold these memories deep in my heart.

Last, but certainly not least, I want to use this opportunity to thank my family, who have always had my back, without whom none of this would have been possible, and whose latest addition, my niece Clara Marie, inspires me every day to be a good example, and ensure her generation will inherit the best world it possibly can.

Contents

Chapter 1

Introduction

1.1 Ideal and Non-Ideal Features in Proteins

1.1.1 Overall architecture of proteins is determined by secondary-structural elements

For almost seven decades already, the propensity of certain peptide sequences to assume secondary structures, α-helices and β-sheets, has been known [57]. By formation of regular hydrogen bonds between the backbones of adjacent residues, these conformations are stabilized within themselves, and by formation of further interactions, such elements can be stabilized against each other to form tertiary interactions. Arrangement of the secondary-structural elements within a protein happens according to two principles: Hydrophobic residues are packed against each other tightly to exclude the surrounding water, eventually forming the core of the protein, while polar and charged residues remain solvated on the surface [3].

The term 'ideal protein' has been coined in the design literature of recent years, meaning a protein that incorporates all the features we know help stabilize it [23,46,56]: ideal bond lengths and angles, high content in secondary structures and very short loops to connect these elements; this results in overall globular shapes with tightly packed cores and no cavities. In an 'ideal' protein, the sequence is designed to optimally

meet the requirements of each specific position in terms of propensity to form secondary structures, solvation, and hydrophobic packing. In all natural proteins, some of these features occur, eventually determining its overall architecture or 'fold'.

1.1.2 In nature, function is usually encoded in non-ideal sections

Loops fine-tune the protein surface

While the tight bonding networks within secondary structures provide stability, the lack of backbone conformational freedom also restricts the attainable overall shape, yielding mostly flat or convex surfaces [32,35]. While such surfaces in proteins may be capable of performing some functions, they pose a disadvantage to others: Particularly interactions involving small-molecule binding, such as catalysis, are generally carried out on concave surface areas, sometimes with very deep substrate binding pockets [66]. This provides two obvious advantages over ligand binding on a flat surface: The area of the binding interface is increased, which is crucial in particular for small-molecule binding, where binding energies are low. Further, burial of the active site in a deep groove or pocket allows not only better ligand or substrate selectivity, but also helps shield the site from the solvent as well as stripping solvent molecules off the ligand during binding [4, 62]. On natural proteins, such concave surfaces are usually located in areas of the protein largely devoid of secondary structures; the advantage of such non-ideal features is their inherent flexibility. Unlike in α-helices and β-sheets, the backbone is not locked down in a rigid pattern with fixed dihedral angles, but can form polar and non-polar interactions with the backbone and side chains of other residues. It can thus not only be adapted to specifically recognize its ligand, but also easily undergo conformational changed (induced-fit) upon binding [63].

The interactions stabilizing such binding sites are by their very nature irregular and strongly depend not only on the sequence of the loop itself, but also on the structure and sequence of its supporting scaffold. This makes conformation prediction

difficult and design a challenge that has to date not been met by computational means [24, 31, 64].

Exposed hydrophobic patches and buried charges are crucial for function

Life is made possible only by the fact that proteins carry out their function within a given biochemical context; at the beginning of any such function stands a more or less specific interaction with ligand or substrate - and this interaction is facilitated not only by sculpting the protein's surface to complement that of its interacting partners, but also by furnishing it with the required physicochemical properties.

While ideal proteins constitute a tightly packed, apolar core and a polar surface that energetically favor the native fold in its aqueous environment, interactions may require non-ideal features, such as exposure of hydrophobic side chains. On the surface of many natural protein binders, exposed hydrophobic patches can be found; these patches often form the core of an interface, and desolvation of the formerly exposed apolar side chains makes a sizeable contribution to binding energy [18, 36, 65]. While this strategy ensures a considerable affinity, it favors rather unspecific binding, so many proteins use conformational and sequence variability of a proximal loop for specificity [34]. Conversely, catalytic sites often require charged residues in their mechanism, and with active site placement in a more or less deep trough for many enzymes, this results in charged residues placed essentially at the core of the protein [73]. Just like exposed hydrophobic residues, also buried charges are inherently destabilizing, particularly in already conformationally fragile regions devoid of secondary structures. For millions of years, evolution has succeeded in walking this fine line between stabilization and functionality, and protein engineers are striving to catch up.

1.1.3 Protein design to date

De-novo fold design

To date, all de-novo protein design approaches have, in one way or another, aimed at fold design, either recapturing existing folds - including α-helix bundles, the Rossman fold, β/α barrels - with novel sequences or generating novel folds by arranging secondary structural elements in novel ways [27, 29, 41]. Design methods largely achieve this by following a funnel in the energy landscape of the sequence-conformation space to finding a minimum [24].

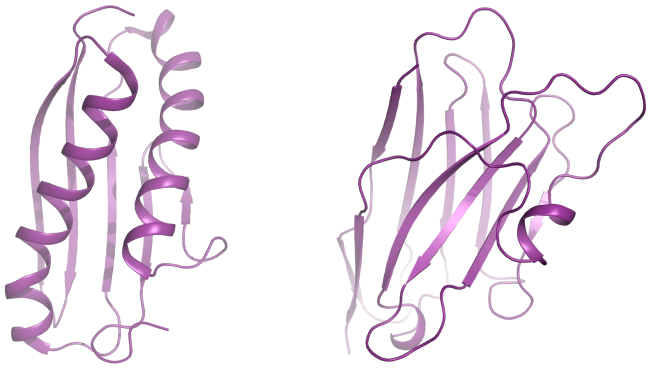

Figure 1.1: The result of the first fold design study (left, PDB access code 2mbl) and for comparison, a natural protein (right, PDB access code 1mel). While the designed protein contains almost exclusively secondary structual elements with shortest possible connecting loops, the natural protein - in this case a single-domain antibody - contains long loops between the β-strands of the framework [41].

In most cases, at this energetic minimum lies such an ideal structure with high secondary structure content and very short loops connecting them (see **Figure 1.1**), hydrophobic side chains buried in a tightly packed core, and polar residues solvated on the surface. The main goal of this approach seems achieved now, as years of engineering efforts have allowed us to understand the rules governing the arrangement and stabilization of α-helices and β-strands in such bundle-like folds fairly well [40]. In some cases, this systematic in silico optimization results in even higher stability in designed than in natural proteins [23].

Ideal protein design, however, has so far avoided one very critical aspect in their studies: functionality.

Design for function

For reasons explained above, ideal proteins possess only limited suitability for placement of active sites, and yet recent years have seen some success in design-of-function using ideal scaffolds.

The basic idea is that between tens of thousands of solved protein structures to date, virtually every architecture can be found on an existing protein, and with the overall surface sculpture in place, fine details and physicochemical requirements can be adjusted for by sequence design alone. With the superior stability of idealized proteins, it is assumed that a a small number of destabilizing features, such as single hydrophobic residues on the surface, can be introduced without endangering overall stability.

To show the common strategy for design-of-function to date, two different examples shall be examined in more depth; both make use of existing scaffolds, re-designing their surface to achieve the desired function - in one case, Kemp elimination catalysis, a model reaction for proton transfer from carbon, and in the other case, recognition (and neutralization) of a conserved antigenic region on the influenza hemagglutinin (HA) protein.

Active site matching for catalysis

Almost a decade ago already, Röthlisberger at al. [60] have succeeded in generating a novel enzyme for this catalytic reaction by computational design. Briefly, the functional enzyme was generated by molecular dynamics modelling of an ideal active site (see **Figure 1.2**), and subsequent placement of this active site onto existing scaffolds.

17 scaffolds were selected for active-site matching; all scaffolds possessed a ligand-binding pocket, and the structure was solved to high resolution to improve the accuracy of active site placement. In total, 59 designs were tested experimentally, out of which eight exhibited measurable catalytic activity. While this does represent a breakthrough, in particular with the high success rate, it has to be noted that the

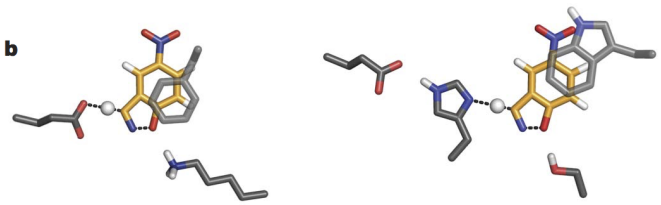

Figure 1.2: Shown are the active sites modeled for the Kemp elimination in the presented work. The design algorithm tries to place one of two modeled active sites on existing scaffolds, and subsequently design the surrounding area to stabilize the conformation of the active residues. *Image taken from [60].*

method relies on use of existing, specialized scaffolds: Most of the proteins re-designed in this study were TIM barrels, which naturally feature a very similar catalytic site, so the active site has already been stabilized through evolution.

Hotspot matching for interaction

The overall strategy for HA binder design in this study [25] is very similar to that for the Kemp eliminase: side chain conformations for an interaction hotspot are modeled, and these residues then placed on potential scaffolds from the protein database (PDB). In this study, the scaffolds are selected after coarse docking and identification of surfaces with approximate shape complementarity to the target region of HA.

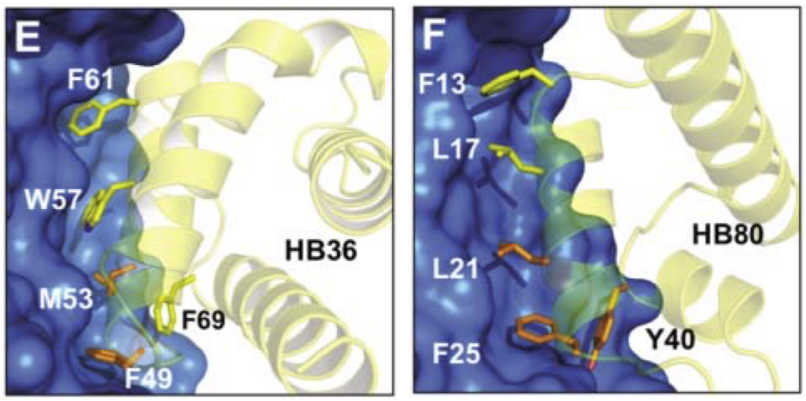

Figure 1.3: Shown are the designed hemagglutinin binders HA36 and HA80. Notably, all interacting residues are placed on secondary structural elements of the scaffolds to ensure stability in spite of the inherently destabilizing properties of the exposed hydrophobic side chains. *Images taken from [25].*

In this study, 865 high-resolution scaffolds selected from the PDB by experimental considerations were used. After placement of the hotspot, the sequence of the surrounding region is re-designed to stabilize the binding rotamers of the hotspot residues, generating a total of 88 designs on 79 different scaffolds. Out of these, 73 designs expressed and two exhibited binding.

Similar to Kemp eliminase design, also this study uses natural scaffolds that were stabilized throughout evolution, and yet introduction of an average of eleven mutations is sufficient to completely abolish expression in almost 20% of the designs; as **Figure 1.3** shows, the active site of both binding designs is entirely located on α-helices in the scaffold, further proving the energetic strain put on the scaffold by introduction of these exposed hydrophobic patches. However, the approach demonstrates success, and as the scaffolds used in designs are not pre-selected by natural function, it is much more general and thus applicable to a wider range of problems.

Like most design-of-function approaches, the two presented above make use of active site modeling and subsequent placement on an existing naturally stable or even ideal scaffold; since this method has demonstrated success, it has been claimed that it provides solutions to most protein design problems [29]. It does, however, have its limitations: Due to the rigid architecture of the protein, very little adaptations can be made, and single point mutations on the target - in case of the binder - or alterations in the substrate in case of the catalyst, require a total re-make, whereas a non-ideal structure could be adapted to a slightly altered requirement. In theory, this might not be problematic, but in practice, requirements change very rapidly, and the amount of time that needs to be invested in even minor adaptations is not sustainable. Another drawback of using ideal proteins as scaffolds is that arrangements of α-helices and β-strands typically result in globular architectures, with all its surfaces being either flat or convex. As can be seen in the HA example, this requires the epitope on the binding partner to have a certain architecture itself, and poses a challenge for binding globular proteins or even small molecules. Only recently, ideal-fold design efforts have begun to focus on generating concave shapes - although not yet for function [47] - but due to the rigid nature of secondary structures, there is a limit to the architectures that can be encoded.

A de-novo designed TIM barrel as the first step towards functional protein design?

As *de novo* protein design attempts have progressed over recent years from mere design of simple α/β folds on to larger constructs, the deficiencies in our understanding of protein stabilization became apparent: As a prime example, Huang et al. succeeded in generating a 4-fold symmetric TIM barrel [30] strictly by applying rules for the arrangement of secondary structural elements [40] without relying on previous knowledge of natural sequences.

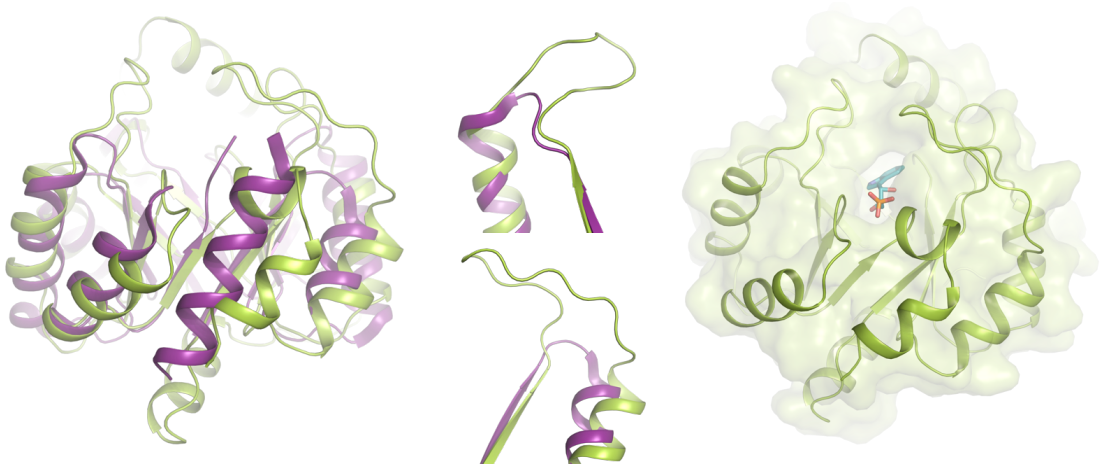

Figure 1.4: Comparison of a designed TIM barrel (purple, PDB access code 5bvl) and a natural TIM barrel (limon, PDB access code 1a53. Shown are the overall structures (left) and details of the loops (center) connecting the secondary structural elements in both proteins. The much longer loops in the natural enzyme contain the substrate binding site and bury the catalytically active residues (right; substrate analog in deepteal).

With this being the scaffold used for design of active Kemp eliminases, it seems like a major step towards de-novo enzyme design, and while the achievement of generating such a large structure is substantial, comparison of the designed protein with natural structures of the same family highlights where it falls short: The catalytically active site in the natural protein is formed by loops protruding from the barrel scaffold. These loops are up to 15 residues long and create a roof over the substrate binding site, shielding it from the surrounding solvent. In contrast, the same loops in the designed protein are only 3 residues long, resulting in an overall much smaller protein

(184 amino acids in the designed compared to 248 in the natural TIM barrel) and no sculpting or coverage of what would be the active site in an active enzyme. This stark contrast between the natural and designed protein only highlights our inability to control the non-ideal features required for its function.

Protein design has made some significant strides over the last decade, but our understanding of the basic rules for protein stabilization is still very deficient. None of the methods combine scaffold design with design of function, thus limiting their applicability, and where design of function is involved at all, only reluctantly incorporate non-ideal features, which are so crucial for function in natural proteins.

Table 1.1: The design tasks and non-ideal features addressed in the approaches presented above.

	scaffold design	active site design	exposed hydrophobics	buried charges	loop design
Kemp eliminase	✗	✓	✗	✓	✗
HA binders	✗	✓	✓	✗	✗
TIM barrel	✓	✗	✗	✗	✗
AbDesign	✓	✓	✓	✓	✓

In contrast, the method presented here - AbDesign [42] - designs antibody scaffolds with both, the framework and the active site loops. In design of the fold, charged residues form buried H-bonds and salt bridges and aromatic side chains on the surface form interaction sites with the ligands. The proteins generated with this algorithm assume the designed conformations as well as fulfil their designed function, signifying an important step in protein design.

1.2 Antibodies

The antigen-binding region of an antibody lies in the variable fragment (F_v) and consists of two immunoglobulin (Ig) domains, one from each, the light and the heavy

chain (LC and HC, respectively). The overall fold of the Ig domain and thus the F_v is very conserved with few deviations in the framework - between antibodies within one species as well as between species; this high conservation also warrants their physical stability [39, 69]. The source of antibodies' ability to recognize virtually all potential intruders into the mammalian body lies within the paratope, or complementarity-determining region (CDR). This region mediates the interaction with the antigen, and does so by assuming a wide range of different shapes and physical properties. Its variablity is owed to the fact that it is almost completely devoid of secondary structural elements, stabilizing the loops against each other without the 'ideal' regular bonding networks of other proteins.

1.2.1 Different types of natural antibodies

Beside the conventional 150 kDa antibody type, members of the family *Camelidae* possess an additional variant, consisting only of two HCs, in each of which the c_H1 domain is replaced by an elongated hinge region and lacks LCs entirely (see **Figure 1.5**), the heavy chain antibody (HCAb).

Consequently, the resulting antibody is significantly smaller, and particularly the variable domain (v_HH), which mediates interaction with the antigen, is only half the size of the variable fragment of a conventional antibody. To account for the thus exposed, typically hydrophobic v_L/v_H interface, many such antibodies incorporate large CDR3 loops, mostly covering the hydrophobic patch. These antibodies often contain disulfides, either connecting two cystein residues within the loop for internal stabilization or linking it

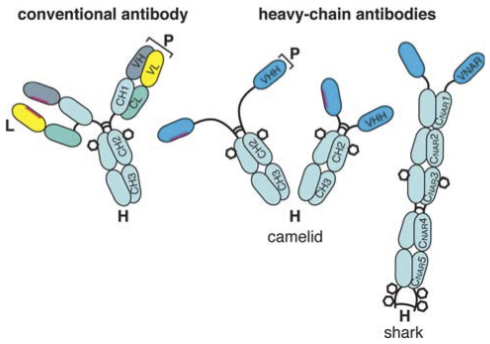

Figure 1.5: Comparison between a conventional antibody, a 150 kDa heterotetramer (left) and the camelid HCAb (middle) of only approximately half its size. *Image taken from [71].*

to either CDR1 or the framework of the v_L/v_H interface to ensure structural integrity. While typically v_HH originating from *C. dromedarius* contain such an additional disulfide, v_HH from *V. pacos* or *L. glama* generally comprise shorter CDR3 loops, which do not cover the v_L/v_H interface. In these examples, stabilization seems to be achieved by introduction of more polar residues in the formerly hydrophobic surface, thus increasing solubility. Both modes of stabilization are encoded in the sequences of the framework and CDRs 1 and 2, and it has been shown that a family of VH genes designated for v_HH exists [54]. In spite of their practical advantges, like smaller size and increased stability [53], they also seem to naturally target a different group of antigens, although it is not clear yet what the basis of this distinction is [16].

1.2.2 Antibodies in clinical use

The main concerns for protein drugs are affinity, specificity, stability, and low immunogenicity. With their diversity, antibodies have the capability of forming high-affinity interactions with nearly every ligand, and with the typically large interfaces and high shape complementarity, they have the potential of highly specific high affinity binding, although some antibodies show poor specificity. With apparent melting temperatures of around 70°C [69], they are also very stable. Immunogenicity has been a problem in the past, because antibodies originally raised in animals may elicit an immune response and lead to serums sickness. To prevent this, attempts have been made at humanizing animal antibodies, but our understanding of what causes immunogenicity is still very deficient [8, 33, 51], and attempts to tamper with the Fv's framework often result in a decrease in stability. Nonetheless, antibodies have become a very successful class of protein therapeutic; to date, dozens of antibodies are approved for clinical use, and hundreds more are currently being developed.

1.2.3 Previous methods for antibody generation

The methods used to date for generation of antibodies for research, diagnostic, or therapeutic purposes are powerful: Several antibodies generated by these means are in therapeutic use, and thousands more generated for research purposes; these methods, however, have severe limitations under more than one aspect: Immunization of animals and isolation of antibodies from the serum is time consuming, not to speak of creation of hybridoma cell lines, and specificity and affinity of the resulting antibody is uncertain. The most important drawback, however, is that the animal's self-tolerance mechanism will severely repress formation of an antibody against a protein resembling one of the animal's own repertoire, so that evolutionarily conserved proteins pose a serious problem. These, however, are the ones typically targeted in immunotherapy, e.g. for cancer [7].

With the advent of high-throughput methods such as phage display and YSD, efforts have been made to generate synthetic repertoires of antibodies to make use of this powerful scaffold without the drawbacks of natural repertoires. In such libraries, antibodies are generally expressed as a single-chain constructs and diversity is achieved by recombination or randomization of CDRs from natural sources on few fixed frameworks. Similar to immunization of animals, antibodies with initial affinity for a given target are then isolated from the library by sorting and their properties improved by random mutagenesis and subsequent sorting.

One method for generation of isolation of antibodies using such libraries is electroporation of *S. cerevisiae* and subsequent YSD [9], where the antibodies are expressed as scFvs and displayed on the surface of the cells. The protein of interest, in this case the scFv, is expressed in yeast as an Aga2 fusion protein with N- and C-terminal tags and secreted to the extracellular space, where it is disulfide-bonded to the mebmrane-bound Aga1 protein. Thus covalently linked to the outer membrane, the scFvs can be fluorescently labeled and the yeast cells then sorted in FACS. Cells expressing a

specific variant of the antibody remain viable, so several rounds of sorting can be performed, single clones characterized, and if necessary, further diversification achieved by randomized mutagenesis of the contained plasmids.

A decided drawback of any such undirected approach is that any surface of the antigen can be targeted by the antibody, and epitope testing is time consuming and in some cases very difficult. By modeling the interaction computationally, we can generate antibodies that specifically target one epitope; this allows either direction of the antibody toward a specific site, eg to block a binding site and prevent a natural interaction, or away from specific surfaces of the ligand, to allow other interactions to still occur. This advantage makes our method more powerful than previous methods, and more apt to generating antibodies meeting the stringent requirements placed on therapeutic antibodies.

Chapter 2

Materials and Methods

2.1 Computational design procedures

2.1.1 A database of natural antibody conformations and sequences

A set of 924 antibodies from the SAbDab [20] database were aligned to the Fv of antibody 4M5.3 (PDB entry: 1x9q [49]). The dihedral angles (ϕ, ψ, ω) from each aligned segment were extracted, modeled on PDB entry 1x9q, and relaxed using the methods described in [42]. In the current study, only κ antibodies were built.

During the course of algorithm development, we changed the conformational segments' boundaries. Specifically, in design cycles 1-4, each of the six CDRs constituted an independent conformation segment, and the framework was based on the coordinates in PDB entry 1x9q. The segment boundaries in cycles 1-4 were (using Kabat numbering [72]), L1:23-34, L2:45-54, L3:87-100, H1:27-35, H2:45-57, H3:91-104. Boundaries in cycle 5: vL:1-88, L3: 89-105, vH: 7-92, H3:93-112.

In cycles 4-5, backbone conformations were clustered by root-mean-square deviation (rmsd) with a cutoff of 1.0 Å, and for each conformation class, a sequence alignment and Position-specific Scoring Matrix (PSSM) was computed, essentially as described [42].

2.1.2 Sequence constraints during design

Similar to previous binder and enzyme design studies [25, 60], in design cycles 1-3 we disallowed sequence design at positions in which the natural segments had Gly, Pro, or Cys. In cycle 2, we further disallowed Gly, Pro, Cys, or His to be introduced during design. In cycle 3, we added specific sequence constraints that mainly targeted positions in the antibody framework, where we noticed persistent design flaws, such as underpacking or buried but unsatisfied polar groups. For instance, the following restrictions were implemented (numbering in accordance with PDB entry 1x9q): Phe/Tyr (position 90), Leu (91), Val (174), Phe (178, 180), Tyr (233). In this design cycle, we additionally restricted core positions to apolar residue identities. For each position, we calculated the number of neighboring C_α atoms ($< 8\,\text{Å}$), excluding sequence neighbors (10 positions N- or C-terminally). Any position for which there were > 5 such neighbors was considered core, and design was restricted in these positions to apolar residues.

With the implementation of conformation-specific sequence constraints using PSSMs in design cycle 4, all the above sequence constraints were lifted. Instead, identities with PSSM cutoffs of -2 and 0 were used in the CDRs and framework, respectively.

2.1.3 A set of conformationally representative designed antibodies

Using the conformation databases, above, a set of representative antibody structures was designed (2,000-8,000 structures, depending on the design cycle). For each segment, a single representative conformation was arbitrarily selected from the conformation databases, and all representative conformations were recombined with one another to assemble backbone conformations of the complete Fv. The sequence of each antibody was then optimized using RosettaDesign.

2.1.4 Docking and fixed-backbone design

We used as antigens human insulin and *M. tuberculosis* ACP2 (PDB entries: 2a3g and 2cgq, respectively). The structure of 2a3g was truncated on the termini of both chains after initial design runs strongly favored interactions with the terminal regions, and interaction with those would experimentally be hindered by the N-terminal biotinylation of the ligand. The structures were docked against the set of representative antibody backbones using the feature-matching algorithm PatchDock [61]. Briefly, the parameters were set to prediction of antibody-ligand interaction, and poses that did not bind the antigen through HCDR3 and LCDR3 were eliminated.

For each representative antibody, the first 2,000 PatchDock entries were used as seeds for Rosetta simulations, in which all-atom docking, sequence design, backbone minimization, and combinatorial sidechain packing were carried out. At the end of this trajectory, the designed complexes were automatically filtered according to the following criteria: shape complementarity (Sc > 0.5) [44], binding energy $<$ -10 Rosetta energy units (R.e.u.), and binding surface area $> 1{,}100\,\text{Å}^2$.

2.1.5 Stability and binding optimization through modular backbone design

Designed antibodies that passed the previous filters were optimized through modular backbone design, essentially as described [42]. Briefly, for each starting design, 150 moves were conducted, in each of which, a backbone segment was chosen at random, and replaced with another segment from the conformation databases. Following segment replacement, the sequence of the segment and every residue within 8 Å was designed, and iterations of backbone and sidechain minimization, sidechain packing, and rigid-body minimization at the binding interface were conducted.

Designs in cycles 1 and 2 showed poor packing between Fv light and heavy chains. Starting with design cycle 3, we therefore added rigid-body minimization between

these two chains (in conjunction with rigid-body minimization of the Fv relative to the antigen).

In design cycles 3-5, at the end of each design move comprising backbone exchange, sequence design, packing, and minimization, the stability and binding energies were computed and evaluated through the "fuzzy"-logic design framework with an objective function $O = B(Ab, L)S(Ab)$, where B and S are computed energies of binding and stability, respectively, Ab is the antibody, and L is the ligand [70]. The new pose was accepted if it passed the Metropolis criterion with a gradually decreasing temperature, thereby effecting simulated-annealing Monte Carlo optimization [42].

At the end of the optimization trajectory, the structures were filtered by structure and energy criteria. Insulin-binding designs were filtered by: shape complementarity ($\mathrm{Sc} > 0.6$) [44], binding energy <-15 R.e.u., and binding surface area $> 1{,}300\,\text{Å}^2$. ACP-binding designs were filtered by stricter criteria reflecting the larger size of ACP: shape complementarity ($\mathrm{Sc} > 0.6$), binding energy $<$ -18 R.e.u., and binding surface area $> 1{,}400\,\text{Å}^2$.

2.1.6 Manual selection and design improvement

Designs were ranked according to shape complementarity for the antigen (Sc) [44], and visualized using PyMOL. Designs with unpaired charges or cavities at the binding interface or within the antibody core were eliminated, and the remaining designs were manually optimized using FoldIt [13]. Briefly, in rounds 1-3, mutations were mainly introduced to optimize interactions between the framework and the CDRs, and structures of natural antibodies of similar conformation were often reviewed in making these design decisions. With the introduction of PSSM constraints in design cycle 4, such manually introduced mutations were no longer necessary. Additional mutations were introduced to improve charge complementarity with the antigen through long-range electrostatics.

2.1.7 High-performance computing using Google Exacycle

Design cycles 1-3 produced few models that passed the structure and energy filters described above. With the introduction of conformation-dependent PSSM constraints in cycle 4 and the improved backbone segmentation of cycle 5, the productivity of the design algorithm improved dramatically. To address the low productivity of the algorithm in initial design cycles, we deployed the algorithm on the Google Exacycle infrastructure. Briefly, during algorithm development, the modified Rosetta source code was regularly committed to the Rosetta development community's common code repository, and the conformation and sequence databases as well as the PatchDock output files were computed on the local computer cluster at the Weizmann Institute of Science. The resulting intermediate data files, as well as RosettaScripts [25] protocols detailing the design trajectory were electronically transferred to Google, where the updated Rosetta source code was compiled and deployed on the Exacycle platform.

Exacycle was a highly parallel infrastructure using hundreds of thousands of idle cores at Google. The high likelihood of preemption of the low-priority design jobs required runtimes to be kept $< 40\,\mathrm{min}$. Because the strict filtering following design cycles 1-3 yielded very few good candidates but executed relatively quickly, this computation was a perfect fit for the Exacycle platform. We were able to cast an extremely wide net, rapidly exploring a huge number of possible candidates using very low-cost computing resources. The vast majority of these candidates were quickly rejected but the remaining ones (in the thousands) were of high quality and were further refined and optimized, now requiring only moderate computing resources.

Designs computed on Exacycle were saved in compact Rosetta-formatted binary silent files and electronically transferred to the Weizmann Institute for further analysis.

2.1.8 Rosetta energy function

Throughout all modeling calculations, the Rosetta score12 energy function was used. This energy function is dominated by van der Waals packing, hydrogen bonding, and an implicit solvation model [17].

2.1.9 RosettaScripts and commandline instructions

```
<dock_design>
 <SCOREFXNS>
  <score12_coordcst weights=score12_w_corrections>
   <Reweight scoretype=coordinate_constraint weight=0.06/ >
  < /score12_coordcst>
  <soft_rep_res_type_cst weights=soft_rep>
   <Reweight scoretype="res_type_constraint" weight=0.2/ >
  < /soft_rep_res_type_cst>
   <sc12_w_correction weights=score12_w_corrections>
  <Reweight scoretype="res_type_constrain" weight=0.2/ >
   < /sc12_w_correction>
   <soft_rep_coordcst weights=soft_rep>
  <Reweight scoretype=coordinate_constraint weight=0.06/ >
   < /soft_rep_coordcst>
 < /SCOREFXNS>


 <TASKOPERATIONS>
  <SeqprofConsensus name=seqprofcons min_aa_probability=2 conservation_cut
off_protein_interface_design=0 conservation_cutoff_aligned_segments=1 probab
ility_larger_than_current=0 ignore_pose_profile_length_mismatch=1>
   <ProteinInterfaceDesign design_chain1=1 design_chain2=0/ >
   <RestrictToAlignedSegments chain=1>
```

```
    <L1 source_pdb="%%template%%" start_res=24 stop_res=41/ >
    <L2 source_pdb="%%template%%" start_res=51 stop_res=60/ >
    <L3 source_pdb="%%template%%" start_res=94 stop_res=105/ >
    <H1 source_pdb="%%template%%" start_res=135 stop_res=147/ >
    <H2 source_pdb="%%template%%" start_res=155 stop_res=171/ >
    <H3 source_pdb="%%template%%" start_res=209 stop_res=220/ >
   </RestrictToAlignedSegments>
  </SeqprofConsensus>

  <RestrictToAlignedSegments name=CDRs chain=1>
   <L1 source_pdb="%%template%%" start_res=24 stop_res=41/ >
   <L2 source_pdb="%%template%%" start_res=51 stop_res=60/ >
   <L3 source_pdb="%%template%%" start_res=94 stop_res=105/ >
   <H1 source_pdb="%%template%%" start_res=135 stop_res=147/ >
   <H2 source_pdb="%%template%%" start_res=155 stop_res=171/ >
   <H3 source_pdb="%%template%%" start_res=209 stop_res=220/ >
  </RestrictToAlignedSegments>
  <RestrictAbsentCanonicalAAS name=no_cys keep_aas="ADEFGHIKLMNPQRSTVWY"/ >
  <RestrictToRepacking name=rtr/ >
  <InitializeFromCommandline name=init/ >
  <ProteinInterfaceDesign name=pido design_chain1=1 design_chain2=0 interfa
ce_distance_cutoff=10.0/ >
  <ProteinInterfaceDesign name=rtr_chain2_interface design_chain1=1 design_c
hain2=0 interface_distance_cutoff=10.0 modify_before_jump=0/ >
  <PreventChainFromRepacking name=prevent_chain2 chain=2/ >
 </TASKOPERATIONS>

 <MOVERS> //design
  <TaskAwareCsts name=Ab_constraints task_operations=prevent_chain2/ >
```

<PackRotamersMover name=design_score12 scorefxn=sc12_w_correction task_operations=CDRs,seqprofcons,init,pido,no_cys/ >

<PackRotamersMover name=design_softrep scorefxn=soft_rep_res_type_cst ta
sk_operations=CDRs,seqprofcons,init,pido,no_cys/ >

<PackRotamersMover name=design_score12_fullCDRs scorefxn=sc12_w_correction task_operations=CDRs,seqprofcons,init,no_cys/ >

<PackRotamersMover name=design_softrep_fullCDRs scorefxn=soft_rep_res_type_cst task_operations=CDRs,seqprofcons,init,no_cys/ >

<TaskAwareMinMover name=soft_min scorefxn=soft_rep bb=0 jump=1 chi=1 ta
sk_operations=rtr/ >

<TaskAwareMinMover name=soft_min_all scorefxn=soft_rep_coordcst bb=1 ju
mp=1 chi=1 task_operations=rtr_chain2_interface/ >

<TaskAwareMinMover name=hard_min scorefxn=sc12_w_correction bb=0 jump=1 chi=1 task_operations=rtr/ >

<TaskAwareMinMover name=hard_min_all scorefxn=score12_coordcst bb=1 jump=1 chi=1 task_operations=rtr_chain2_interface/ >

<RotamerTrialsMinMover name=rtmin scorefxn=score12 task_operations=rtr,CDRs,init,pido/ >

<RotamerTrialsMinMover name=rtmin_fullCDRs scorefxn=score12 task_operations=rtr,CDRs,init/ >

<DockingProtocol name=dock_lowres low_res_protocol_only=1 docking_score_low=score_docking_low/ >

<SaveAndRetrieveSidechains name=sars allsc=1 two_step=1/ >

<SwitchChainOrder name=remove_ligand scorefxn=score12 chain_order=1/ >

</MOVERS>

```
  <FILTERS>
   <PackStat name=a_packstat confidence=1/ >
   <ShapeComplementarity name=a_sc confidence=0/ >
   <Ddg name=a_ddg confidence=1 scorefxn=sc12_w_correction threshold=-12
repeats=1/ >
   <Sasa name=a_sasa threshold=1200 confidence=1/ >
   <ScoreType name=total_score score_type=total_score scorefxn=sc12_w_corre
ction threshold=100/ >
   <MoveBeforeFilter name=antibody_stability mover=remove_ligand filter=tot
al_score/ >
   <Expiry name=stopit seconds=2400 confidence=0/ >
   <CompoundStatement name=stopit_inv>
    <NOT filter_name=stopit/ >
   < /CompoundStatement>
   <Sigmoid name=sigmoid_binding filter=a_ddg offset=0 steepness=0.45 base
line_checkpoint="%%prefix%%.ddg.checkpoint"/ >
   <Sigmoid name=sigmoid_stability filter=antibody_stability offset=2 stee
pness=0.3 baseline_checkpoint="%%prefix%%.stability.checkpoint"/ >
   <Operator name=stability_and_binding operation=PRODUCT filters=sigmoid_b
inding,sigmoid_stability negate=1/ >
  < /FILTERS>

  <MOVERS> //refinement
   <Splice name=H3 profile_weight_away_from_interface=1.5 scorefxn=sc12_w_co
rrection torsion_database=H3.db dbase_iterate=1 ccd=0 delta_lengths=-1,-2,0,
1,2 design=1 template_file="%%template%%" checkpointing_file="%%prefix%%.H3
.checkpoint" design_task_operations=seqprofcons,init,no_cys segment=H3 prot
ein_family=antibodies/ >
   <Splice name=H1_H2 profile_weight_away_from_interface=1.5 scorefxn=sc12_w
```

```
_correction torsion_database=H1_H2.db dbase_iterate=1 ccd=0 delta_lengths=-1,
-2,0,1,2 design=1 template_file="%%template%%" checkpointing_file="%%prefix
%%.Vh.checkpoint" design_task_operations=seqprofcons,init,no_cys segment=H1_
H2 protein_family=antibodies/ >
  <Splice name=L3 profile_weight_away_from_interface=1.5 scorefxn=sc12_w_cor
rection torsion_database=L3.db dbase_iterate=1 ccd=0 delta_lengths=-1,-2,0,
1,2 design=1 template_file="%%template%%" checkpointing_file="%%prefix%%.L3
.checkpoint" design_task_operations=seqprofcons,init,no_cys segment=L3 prot
ein_family=antibodies/ >
  <Splice name=L1_L2 profile_weight_away_from_interface=1.5 scorefxn=sc12_w_
correction torsion_database=L1_L2.db dbase_iterate=1 ccd=0 delta_lengths=-1,
-2,0,1,2 design=1 template_file="%%template%%" checkpointing_file="%%prefix
%%.Vl.checkpoint" design_task_operations=seqprofcons,init,no_cys segment=L3
protein_family=antibodies/ >
   <AtomTree name=docking_ft docking_ft=1/ >
   <AtomTree name=two_parts_chain1 two_parts_chain1=1/ >
   <AtomTree name=simple_ft simple_ft=1/ >
   <ParsedProtocol name=post_splice_refine>
    <Add mover=design_score12/ >
    <Add mover=simple_ft/ >
    <Add mover=docking_ft/ >
    <Add mover=hard_min/ >
    <Add mover=rtmin/ >
    <Add mover=rtmin/ >
    <Add mover=simple_ft/ >
   < /ParsedProtocol>
   <ParsedProtocol name=splice_refine_L1_L2>
    <Add mover=L1_L2/ >
    <Add mover=post_splice_refine/ >
```

```
   < /ParsedProtocol>
   <ParsedProtocol name=splice_refine_L3>
    <Add mover=L3/ >
    <Add mover=post_splice_refine/ >
   < /ParsedProtocol>
   <ParsedProtocol name=splice_refine_H1_H2>
    <Add mover=H1_H2/ >
    <Add mover=post_splice_refine/ >
   < /ParsedProtocol>
   <ParsedProtocol name=splice_refine_H3>
    <Add mover=H3/ >
    <Add mover=post_splice_refine/ >
   < /ParsedProtocol>
   <GenericMonteCarlo name=mc_Vl stopping_condition=stopit_inv mover_name=s
plice_refine_L1_L2 filter_name=stability_and_binding trials=50 preapply=0
te
mperature=0 drift=0 saved_accept_file_name="%%prefix%%.checkpoint.pdb" save
d_trial_number_file="%%prefix%%.Vl.mc_trial" mover_tag="L1_L2"/ >
   <GenericMonteCarlo name=mc_L3 stopping_condition=stopit_inv mover_name=s
plice_refine_L3 filter_name=stability_and_binding trials=100 preapply=0 temp
erature=0 drift=0 saved_accept_file_name="%%prefix%%.checkpoint.pdb" saved_t
rial_number_file="%%prefix%%.L3.mc_trial" mover_tag="L3"/ >
   <GenericMonteCarlo name=mc_Vh stopping_condition=stopit_inv mover_name=s
plice_refine_H1_H2 filter_name=stability_and_binding trials=50 preapply=0
te
mperature=0 drift=0 saved_accept_file_name="%%prefix%%.checkpoint.pdb" save
d_trial_number_file="%%prefix%%.Vh.mc_trial" mover_tag="H1_H2"/ >
   <GenericMonteCarlo name=mc_H3 stopping_condition=stopit_inv mover_name=s
plice_refine_H3 filter_name=stability_and_binding trials=50 preapply=0 tempe
```

```
rature=0 drift=0 saved_accept_file_name="%%prefix%%.checkpoint.pdb" saved_tr
ial_number_file="%%prefix%%.H3.mc_trial" mover_tag="H3"/ >
   <LoadPDB name=load_from_checkpoint filename="%%prefix%%.checkpoint.pdb"/
>
   <Splice name=splice_seqconstraints scorefxn=sc12_w_correction add_sequen
ce_constraints_only=1 protein_family=antibodies segment=L3/ >
  < /MOVERS>

  <FILTERS> //checkpointing stuff
   <Sasa name=soft_dock_sasa_filter threshold=1500 confidence=1/ >
   <Sasa name=2nd_hard_min_sasa_filter threshold=1350 confidence=1/ >
   <Ddg name=ddg_final_commitment confidence=1 scorefxn=sc12_w_correction
t
hreshold=-11 repeats=1/ >
   <FileExist name=recover filename="%%prefix%%.checkpoint.pdb" ignore_zer
o_byte=1/ >
   <FileRemove name=remove_checkpointing_files delete_content_only=1 filena
mes="%%prefix%%.stability.checkpoint,%%prefix%%.ddg.checkpoint,%%prefix%%.
checkpoint.pdb,%%prefix%%.H3.checkpoint,%%prefix%%.Vh.checkpoint,%%prefix%
%.L3.checkpoint,%%prefix%%.Vl.checkpoint,%%prefix%%.H3.mc_trial,%%prefix%%.
L3.mc_trial,%%prefix%%.Vh.mc_trial,%%prefix%%.Vl.mc_trial"/ >
  < /FILTERS>

  <MOVERS>
   <If name=if_checkpoint_load filter_name=recover true_mover_name=load_from
_checkpoint/ >
   <ParsedProtocol name=recover>
    <Add filter=stopit/ >
    <Add mover=if_checkpoint_load/ >
```

```
 <Add mover=splice_seqconstraints/ >
 <Add mover=mc_H3/ >
 <Add filter=stopit/ >
 <Add mover=mc_L3/ >
 <Add filter=stopit/ >
 <Add mover=mc_Vl/ >
 <Add filter=stopit/ >
 <Add mover=mc_Vh/ >
 <Add filter=stopit/ >
 <Add mover=two_parts_chain1/ >
 <Add mover=design_softrep_fullCDRs/ >
 <Add mover=Ab_constraints/ >
 <Add mover=soft_min_all/ >
 <Add mover=design_softrep_fullCDRs/ >
 <Add mover=hard_min_all/ >
 <Add mover=design_score12_fullCDRs/ >
 <Add mover=hard_min_all/ >
 <Add mover=rtmin_fullCDRs/ >
 <Add mover=rtmin_fullCDRs/ >
 <Add mover=rtmin_fullCDRs/ >
 <Add mover=hard_min_all/ >
 <Add mover=simple_ft/ >
 <Add mover=docking_ft/ >
< /ParsedProtocol>
<SwitchResidueTypeSetMover name=fa set=fa_standard/ >
<ParsedProtocol name=start_fresh>
 <Add mover=dock_lowres/ >
 <Add mover=sars/ >
 <Add filter=soft_dock_sasa_filter/ >
```

```
    <Add mover=sars/ >
    <Add mover=design_softrep/ >
    <Add mover=soft_min/ >
    <Add mover=design_softrep/ >
    <Add mover=hard_min/ >
    <Add mover=design_score12/ >
    <Add filter=2nd_hard_min_sasa_filter/ >
    <Add mover=hard_min/ >
    <Add mover=rtmin/ >
    <Add mover=rtmin/ >
    <Add mover=rtmin/ >
    <Add mover=hard_min/ >
    <Add filter=ddg_final_commitment/ >
    <Add mover=design_score12_fullCDRs/ >
    <Add mover=hard_min/ >
    <Add mover=rtmin_fullCDRs/ >
    <Add mover=rtmin_fullCDRs/ >
    <Add mover=rtmin_fullCDRs/ >
    <Add mover=hard_min/ >
    <Add mover=recover/ >
   < /ParsedProtocol>
   <If name=checkpoint_recovery filter_name=recover true_mover_name=recover
false_mover_name=start_fresh/ >
   <AddChain name=addLig file_name="%%lig%%" new_chain=1 scorefxn=sc12_w_co
rrection random_access=0/ >
  < /MOVERS>

  <PROTOCOLS>
   <Add mover=addLig/ >
```

<Add mover=sars/ >

<Add filter=stopit/ >

<Add mover=splice_seqconstraints/ >

<Add mover=checkpoint_recovery/ >

<Add filter=a_sc/ >

<Add filter=a_sasa/ >

<Add filter=a_ddg/ >

Add filter=Check_ligand_rmsd_final/ >

<Add filter=a_packstat/ >

< /PROTOCOLS>

< /dock_design>

2.1.10 Command line options (wrapped in "flags" file)

- linmem_ig 10
- ex1
- ex2aro
- use_input_sc
- extrachi_cutoff 8
- ignore_unrecognized_res
- chemical:exclude_patches LowerDNA UpperDNA Cterm_amidation SpecialRotamer VirtualBB ShoveBB VirtualDNAPhosphate VirtualNTerm CTermConnect sc_orbitals pro_hydroxylated_case1 pro_hydroxylated_case2 ser_phosphorylated thr_phosphorylated tyr_phosphorylated tyr_sulfated lys_dimethylated lys_monomethylated lys_trimethylated lys_acetylated glu_carboxylated cys_acetylated tyr_diiodinated N_acetylated C_methylamidated MethylatedProteinCterm

- parser:protocol design.xml
- nstruct 10
- mute all
- pdb_comments true
- parser:script_vars lig=<ligand PDB file>

2.1.11 Command line execution

rosetta_scripts.default.linuxgccrelease @flags -parser:script_vars -s =<Input PDB after patch dock & optimization>

2.1.12 Algorithm adaptations for single-domain antibodies

The general strategy is essentially the same as the one developed for conventional antibodies with few adaptations. I used 41 source structures (PDB IDs: 1bzq, 1hcv, 1i3u,1i3v, 1ieh, 1kxq, 1mvf, 1op9, 1qd0, 1shm, 1sjv, 1sjx, 1zvh, 1zvy, 2p45, 2p4a, 2x1q, 2xa3, 3cfi, 3dwt, 3eak, 3eba, 3ezj, 3k3q, 3k7u, 3k81, 3ln9, 3ogo, 3qsk, 3qxt, 3qxu, 3qxv, 3qxw, 3rjq, 3tpk, 3k74, 3p0g, 3r0m, 3stb, 3v0a, 4dka). Due to the variability of HCDR1 in v_HHs and the small number of solved structures, I adhered to the initial strategy of maintaining a fixed backbone (I used the *L. glama* antibody against human chorionic gonadotropin deposited as PDB ID 1hcv) and recombining CDRs on it to increase variability. I also did not cluster the loops and select representatives to generate scaffolds, but used all available loops for recombination, totalling over 60,000 potential different backbones. Consequently, unlike the final design protocol for conventional antibodies, the protocol for v_HH does not contain a loop-sampling step. PSSMs are generated by multiplying each separate sequence by BLOSUM62 and used to guide sequence design as described above. Some natural v_HHs contain an addition disulfide, stabilizing its long CDR3; as our computational method introduces small backbone shifts within the loops, it cannot maintain the precise geometry required for

their formations. For this reason, I eliminated antibodies with such disulfides. Design for function, in this case binding of phenylalanine, is fully performed within Rosetta: First, suitable pockets on the surface of the scaffold are identified, then an interaction site is created within these pockets by matching. Matching is a method commonly used in enzyme design. For catalysis, a known or modeled transition state is used, and side chains capable of its stabilization are placed in favorable orientations around it. For each such side chain, all probable rotamers are modeled, and scaffolds are selected according to their ability to support a set of active site residues. After placement of the active site on the scaffold, the sequence of the surroundings are designed to optimally support the modeled rotamers [60, 74].

For Phe binding, this method is adapted, treating the geometric constraints of the Phe backbone as enzyme matching constraints, supporting an interaction instead of catalysis. Modeling of the matching interaction was based on a *T. thermophilus* phenylalanine tRNA synthetase (PDB ID 1b70), allowing Asp/Glu and Ser/Thr for binding of the amine group and Gln and Lys for recognition of the carboxyl group.

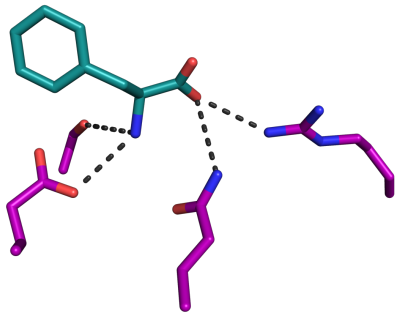

Figure 2.1: Binding mode for phenylalanine in the tRNA synthetase used as a template for phenylalanine binding. Side chains from the protein in purple, ligand in teal.

2.2 Experimental procedures

2.2.1 Cloning and yeast cell-surface display (YSD)

Genes encoding each of the designs were optimized for E. coli expression using DNAWorks [28], and ordered from Gen9Bio (Boston) as double-stranded DNA oligos. Each DNA sample was amplified by PCR to include at least 30 bases overlap with the vector. The pETCON plasmid [25] was linearized using NdeI/BamHI (NEB), and 100 ng of linearized plasmid and 50 μl of PCR product of each gene were transformed into

yeast EBY100 cells by lithium-acetate transformation, essentially as described [9]. Plasmid was then extracted using the Zymoprep Yeast Plasmid Miniprep I kit (Zymo Research) and transformed into Ecloni cells. Single clones were selected for sequence verification, and the verified plasmids then transformed into yeast again for yeast surface display. Yeast-display experiments were conducted essentially as described [9]. Yeast cells were taken from the transformation plate and grown in selective medium (SDCAA: 20 g dextrose, 6.7 g Difco yeast nitrogen base, 5 g bacto casamino acids, 5.4 g Na_2HPO_4, 8.56 g NaH_2PO_4*H_2O, ad 1 l H_2O, sterilized by filtration) over-night at 30°C. From the over-night culture, another 5 ml culture in SDCAA was inoculated and grown at 30°C to OD(600)=0.6. The cells were then harvested and resuspended in 10ml induction medium (SG/RCAA: 20g galactose, 20 g raffinose, 6.7 g Difco yeast nitrogen base, 5 g bacto casamino acids, 5.4 g Na_2HPO_4, 8.56 g NaH_2PO_4*H_2O, ad 1 L H_2O, sterilized by filtration) and incubated at 20°C for 20 h.

For diversification of initial weak binders, libraries were generated according to [5]. Briefly, insert DNA was generated by error-prone PCR (GeneMorph II, Agilent), cleaned by PCR purification (QIAquick, Qiagen), and further amplified by PCR. vector DNA was prepared as described above. Yeast cells were grown to OD(600)=1.6, harvested, washed, and conditioned in 0.1 M lithium acetate/10 mM DTT. After additional washing, cells were electroporated with 3 μgof plasmid and 12 μgof insert DNA at 1.9-2.5 kV, and after recovery in full medium, grown on selective media. Dilutions were plated to determine diversity of the library, resulting in approximately 10^7 variants per library.

100 μl culture was then used for yeast-cell surface display experiments: Cells were harvested, washed twice in ice-cold PBS-F (1 g bovine serum albumin, 100 ml Dulbecco's phosphate buffered saline X10, ad 1 L H_2O, sterilized by filtration), and subjected to primary antibody (mouse monoclonal IgG_1 anti-c-Myc (9E10) sc-40, Santa Cruz Biotechnology) for expression monitoring and biotinylated ligand at the desired concentration in ice-cold PBS-F for 20 min at 4°C. The cells were then washed twice in ice-cold PBS-F and underwent a second staining with fluorescently labeled secondary

antibody (AlexaFluor488 - goat-anti-mouse IgG_1 (Life Technologies) for scFv labeling, Streptavidin-APC (SouthernBiotech) for ligand labeling) for 20 min at 4°C. After another wash in ice-cold PBSF, the cell fluorescence was measured on an Accuri C6 flow cytometer.

2.2.2 Evaluation of YSD results

To determine expression levels, fluorescence of AlexaFluor488 bound to the C-terminal c-Myc tag on the expressed scFvs was measured in stained cells and compared to the fluorescence levels measured in an unstained control sample. The percentage of cells exhibiting higher fluorescence than the control was recorded as expression level.

2.2.3 Gene diversification and selection

Designs that showed initial binding underwent genetic diversification by error prone PCR using GeneMorph II Random Mutagenesis Kit (Agilent) and subsequent library generation [5]. The libraries contained 10^7-10^8 clones with an average of 2-3 mutations per clone. The libraries were then sorted three times on an Aria II fluorescence-activated cell sorter (BD) and 10^5 cells collected each time. After the third sort, single clones were isolated and evaluated separately.

2.2.4 Ligand expression and purification

A gene encoding ACP2 optimized for E. coli expression was ordered from GenScript (Piscataway, NJ) in pET29 and co-transfected with the biotin-ligase BirA in pET28 for in vivo biotinylation in BL21 cells. An overnight culture was grown in LB with 50 ng/mL kanamycin, 30 ng/ml chloramphenicol at 37°C, and a 1L culture with 50 ng/mL kanamycin and 0.5% glucose (to prevent premature expression of the proteins) inoculated from it. The culture was incubated at 200 rpm/37°C until OD_{600}=0.6, when protein expression was induced by addition of 1 mL 1 M IPTG. For in vivo biotinylation, at the same time, D-biotin was added to a final concentration of 50 nM from stock

(5 mM in 10 mM bicine, pH=8.3, sterile filtered). The culture was further incubated at 37°C over-night and harvested in the morning. Cells were resuspended in 20 mL HBS (20 mM Hepes, 150 mM NaCl, pH=7.4) lysed by sonication, the lysate cleared by centrifugation (13,000 g, 20 min, 4°C) and loaded onto Ni-NTA resin (Ni-NTA His-Bind Resin, Merck). After two washing steps with HBS/100 mM imidazole, the protein was eluted with HBS/500 mM imidazole, and the protein dialyzed against HBS. After concentration, the protein was aliquoted and stored at -80°C. For experiments with insulin binders, commercially available biotinylated insulin was used (Immunological and Biochemical Testsystems).

2.2.5 IgG and Fab production

Adherent HEK293T cells were co-transfected with the heavy and light IgG chains or with the heavy and light Fabs (heavy chain fused to C-terminal His tag) using linear PEI as a transfection reagent (12.5 μg : 12.5 μg : 50 μg, respectively per 15 cm plate). 72 h post transfection the medium containing the secreted protein was collected (1.5 L).

2.2.6 Fab purification from IgG (5ins14, 2acp12)

The filtered medium was concentrated to 100 mL using a diafiltration device (QuixStand Benchtop System, GE Healthcare). The medium composition was exchanged to buffer containing 50 mM Tris pH 8.0 and 100 mM NaCl using the same device. The clarified supernatant was subsequently loaded onto a 5 mL HisTrap_Protein A_HP column (GE Healthcare). The IgG was eluted from the column with buffer containing 20 mM citric acid (pH 3.0). 1.5 mL fractions were collected into wells containing 200 μlneutralizing buffer (1 M Tris, pH 8.8). Fractions containing IgG were dialyzed against sample buffer (20 mM sodium phosphate, 10 mM EDTA, pH 7.0) and lyophilized. Immobilized papain (Thermo Scientific) was activated according to the manufacturer's instructions with digestion buffer (sample buffer containing 20mM cys-

teine at pH 7). The lyophilized IgG (10 mg) was dissolved in 1 mL digestion buffer and incubated with 1 mL activated papain beads for 5 h at 37°C. Dilution with 10 mM Tris (pH 7.5) stopped the reaction and prepared the cleavage products for subsequent purification on Protein A column (equilibrated with 10 mM Tris pH 7.5). Fab fragments were collected from the flow-through.

2.2.7 Secreted Fab purification (5ins16)

1.5 L filtered medium was loaded onto a HiTrap_chelating_HP_5ml column (GE Healthcare) equilibrated with 50mM Tris (pH=8), 100mM NaCl. The column was washed with the same buffer containing 100 mM imidazole and the Fab fragments were eluted in one step with 500 mM imidazole and desalted on a HiPrep_26/10_desalting column (GE Healthcare) equilibrated with 50 mM Tris pH 8, 100 mM NaCl. The pure Fab was concentrated to 20 mg/mL for crystallization.

2.2.8 Periplasmatic expression and purification (5acp14)

The design was transformed into RH2.2 plasmid for expression as Fabs where the heavy chain was N-terminally His tagged and the light chain expressed as a separate protein. Both chains contain a secretion sequence for direction to the periplasmatic space, where they fold and dimerize. Cloning was done by RF cloning using Kapa HiFi Hotstart Readymix (Kapa Biosystems) according to the manufacturer's protocol. Cells were induced with 1 mM IPTG at OD_{600}=0.6, transferred to 16°C, and harvested after 20 hrs. The cells were then resuspended in sucrose buffer (50 mM Tris, 1 mM EDTA, 20% w/v sucrose, pH 7.4; 100 mL/L culture) and shaken for 20 min at 4°C, harvested, and the periplasm lysed in low-osmolar solution (500 nM $MgCl_2$; 50 mL/L culture) for 30 min at 4°C. The supernatant was harvested by centrifugation (15,000 g, 15 min), filtered, and 20 x HBS (400 mM Hepes, 3 M NaCl, pH 7.2) added to reach 1x HBS concentration. The osmotic shock fluid was then loaded onto Ni-NTA resin and washed twice with 5 bed volumes 15 mM imidazole in HBS, once with 5 bed volumes

30 mM imidazole in HBS, and eluted with 3 bed volumes of 500 mM imidazole in HBS. Imidazole was removed from the eluate by dialysis against HBS (1 : 500). The sample was then concentrated (Amicon Ultra-15 Centrifugal Filter, Merck) and purified by gel filtration in HBS over a HiLoad 16/600 Superdex 75 pg column.

2.2.9 Cytosolic expression of single-domain antibodies

The designs were cloned into pET29, removing the Avi-tag while maintaining the 6His-tag by RF cloning. Plasmids were transformed into *E. coli* strain BL21 and a 5ml culture grown overnight. 1 L of LB was inoculated from the starter and grown at 37°Cto OD_{600}=0.6, transferred to 16°Cand protein production induced with 1 mM IPTG. Cells were harvested after 20 hrs and pellets frozen until further use.

2.2.10 Purification of single-domain antibodies

Pellets from 1 L of culture were resuspended in 40 mL of lysis buffer (20 mM Citrate, 100 mM NaCl, pH 5.5) and cells lysed by sonication (4-8 pulses, 10 s on/20 s off). The lysate was cleared by centrifugation (10 min at 4,000 g, 4°C), then the supernatant transferred to a new vessel, 1,000 U of benzonase added for digest of residual DNA, and the lysate cleared by another centrifugation step (40 min at 25,000 g, 4°C). The supernatant was filtered and used for affinity chromatography. Ni-NTA beads were transferred to a benchtop column to a final resin volume of 3 mL and washed with 15 mL H_2O. The beads were loaded with 5 mL 500 mM $NiCl_2$, washed with 15 mL H_2O, and equilibrated with 15 mL lysis buffer. The cleared lysate was then loaded onto the column for binding, washed with 15 mL each of 20 mM, 50 mM, and 100 mM imidazole in lysis buffer, and eluted in 300 mM imidazole in lysis buffer. The eluate was concentrated, then purified by gel filtration in lysis buffer over a Superdex 200 Increase 10/300 GL column.

2.2.11 Apparent Tm measurements

The apparent melting temperature of the antibodies was determined by addition of the fluorescent dye SYPRO orange. The principle is that the dye's fluorescence is quenched by the surrounding water. The mixture is then gradually heated, and as the protein unfolds during denaturation, the dye binds to the then-exposed hydrophobic core of the protein, rendering it fluorescent again. All experiments were performed on a ViiA 7 real-time PCR machine (Applied Biosystems). Fabs were diluted to 0.125 mg/mL in a 1:5,000 dilution of SYPRO Orange in HBS and 20 μlof the solution measured. The temperature was ramped from 25°C to 100°C at 0.05 K/s. The maximal unquenching of the fluorophore occurred at 71°C for 2acp12, at 57°C for 5acp14, and at 79°C for 5ins16 (Fig. 3). The slow decrease of fluorescence after the peak in each curve (at high temperatures) is typical of protein aggregation, excluding the dye from formerly exposed hydrophobic core regions. For 5ins16, a 5-bin unweighted sliding-average smooth was applied.

2.2.12 Surface-plasmon resonance

SPR experiments on the ACP binders were carried out on a Biacore T200 (GE Healthcare) at 25°C with HBS-N EP+ (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005% v/v Surfactant P20, pH 7.4). For binding analysis, 1,000-2,800 response units (RU) of Fab were captured on a CM5 sensor chip. Samples of different protein concentrations were injected over the surface at a flow rate of 30 μl/min for 240 s and the chip washed with buffer for 1,200 s. If necessary, surface regeneration was performed with one 30 s injection of 1.5 mM NaOH at a flow rate of 30 μl/min. One flow cell contained no ligand and was used as a reference. The acquired data were analyzed using the device's software, kinetic fits were performed using the 2-state-reaction model to avoid overfitting of the late (asymptotic) dissociation phase. Experiments for the insulin binder were carried out on a ProteOn XPR36 Protein Interaction Array System (BioRad) at 25°C with HBS+ (20 mM Hepes, 100 mM NaCl, 0.005% Tween, pH 7.5). For binding

anaylsis, 8,000 RU of Fab were captured on a GLC Sensor Chip (BioRad). Samples of different concentration were injected over the surface at a flow rate of 30 μl/min for 200 s and dissociation monitored while the chip was flushed with buffer for 2,000 s. One flow cell contained no ligand and was used as a reference. The data were analyzed using the software's kinetic fit for heterogeneous ligands to account for the undirected immobilization of the Fab on the chip surface.

2.2.13 Structure determination and refinement of the 5ins16 and 5ins14 structures

Crystals of 5ins16 and 5ins14 were obtained using the sitting-drop vapor-diffusion method with a Mosquito robot (TTP LabTech). The crystals of 5ins16 were grown from 12% PEG 3,500, 0.1 M $(NH_4)_2SO_4$ and 0.05 M Tris pH 8.5. The crystals formed in the orthorhombic space group $P2_12_12_1$, with 2 monomers per asymmetric unit. A complete dataset to 3.0Åresolution was collected at 100 K on a single crystal on in-house RIGAKU RU-H3R X-ray.

Crystals of 5ins14 were grown from 15% PEG 8,000, 0.1 M $NH_4CH_3CO_2$, 5 mM $Mg(CH_3CO_2)_2$ and 0.05 M Cacodylate pH 6.5. The crystals formed in the trigonal space group $P3_12_1$, with one monomer per asymmetric unit. A complete dataset to 2.1Å resolution was collected the European Synchrotron Radiation Facility (ESRF) beam line, ID23_1.

Diffraction images of the 5ins16 and 5ins14 crystals were indexed and integrated using the Mosflm program [58], and the integrated reflections were scaled using the SCALA program [22]. Structure factor amplitudes were calculated using TRUNCATE from the CCP4 program suite [11]. Both structures were solved by molecular replacement with the program PHASER [48]. The model used to solve the 5ins16 structure was the light chain of the anti human respiratory syncytial virus and the heavy chain of the human CB2 Fab (PDB entries 3qq9 and 2hff, respectively). The model used to solve the 5ins14 structure was the light chain of the anti human respiratory syncytial

virus and the heavy chain of the human phosphoserine (pSAb) scaffold (PDB entries 3qq9 and 4jfz, respectively).

All steps of atomic refinement of both structures were carried out with the CCP4/ REFMAC5 program [52]. The models were built into 2mFobs - DFcalc, and mFobs - DFcalc maps by using the COOT program [21]. Details of the refinement statistics of the 5ins16 and the 5ins14 structures are described in **Table 2.1**. The coordinates of 5ins16 and the 5ins14 were deposited in the RCSB Protein Data Bank with accession codes 5NB5 and 5BNI, respectively. The structures will be released upon publication.

Table 2.1: Data collection and refinement statistics for the designed antibodies 5ins14 and 5ins16. * Values in parentheses refer to the data of the corresponding upper resolution shell.

	5ins16	5ins14
Data collection		
PDB code	5NB5	5NBI
Space group	$P2_12_12_1$	$P3_121$
Cell dimensions		
a,b,c (Å)	63.33, 109.39, 158.90	67.91, 67.91, 182.18
α, β, γ (°)	90, 90, 90	90, 90, 120
No. of copies in a.u.	2	1
Resolution (Å)	90.00-3.00	60.73-2.10
Upper resolution shell (Å)	3.16-3.00	2.21-2.10
Unique reflections	21,280 (738)*	28,196 (4,043)*
Completeness (%)	93.0 (85.9)	98.2 (98.2)
Multiplicity	4.2 (3.8)	10.9 (10.7)
Average I/σ(I)	3.8 (1.1)	6.5 (1.3)
Rsym (I) (%)	19.5 (69.4)	7.4 (59.0)
Refinement		
Resolution range (Å)	90.00-3.00	60.73-2.10
No. of reflections (I/σ(I) $>$ 0)	20,906	27,280
No. of reflections in test set	1,105	1,405
R-working (%)/R-free (%)	22.3 / 25.7	21.2 / 25.7
No. of protein atoms	6558	3303
No. of water molecules	0	8
Overall average B factor (Å^2)	43.2	45.4
Root mean square deviations:		
- bond length (Å)	0.008	0.01
- bond angle (°)	1.376	1.45
Ramachandran Plot		
Most favored (%)	88.6	90.2
Additionally allowed (%)	9.8	9.0
Disallowed (%)	0.3	0.0

Chapter 3

Results

Generation of the design algorithm first presented in [42] was a joint effort between multiple people, and for comprehensiveness, all steps are explained here; all effort is made, however, to give credit to other contributors wherever it is due. I participated in analysis of all results, computationally and experimentally, in this project. I also designed the 78 ACP-targeting conventional antibodies and experimentally characterized them, developed the methods for v_HH design, and designed the 34 v_HHs and experimentally characterized them. As the resulting algorithm constitutes a method, it is described in detail in **section 2.1**, and this chapter is reserved for the experimental results achieved by use of the different versions. A paper describing the main results for conventional antibodies is now under consideration for publication. Some of the data presented here are based on this paper.

3.1 Designed conventional antibodies

Altogether, I designed and tested 78 antibodies against ACP, designed by 3 different design protocols (cycle 2, 3, and 5 as numbered in **chapter 2**). Out of these, two antibodies recognized the ligand ACP, one from the second design cycle, 2acp12, that bore 32 manually introduced mutations, and one from design cycle 5, 5acp14, with only 5 manually introduced mutations.

3.1.1 Low expression levels overall, but one binding design from cycle 2

In this design cycle, 30 designs were selected for experimental testing, with most of them showing poor expression and no binding (see **Table 5.1** for computed scores, **Table 5.2** for sequences). As expression levels correlate with stability, this was a sign that even though the calculated energies of the best designs are comparable to those of natural antibodies, they are far less stable.

The designs obtained in this cycle exhibited several major flaws, mainly in the Ig domain core: The algorithm placed unpaired charges in buried positions, and often failed to pack the core optimally, leaving cavities in both, the core of each Ig domain as well as on the interface between the two chains (see inserts in **Figure 3.4** for examples). The software also placed hydrophobic side chains in exposed positions, creating needless energetically disfavorable patches on the surface and thus impairing stability. During visual inspection, the majority of these flaws were identified and remedied by manual introduction of mutations in FoldIt [13]; as the subject of this study is development of an automated method for computational design, however, a systemic improvement of the algorithm is pursued in the following design cycle.

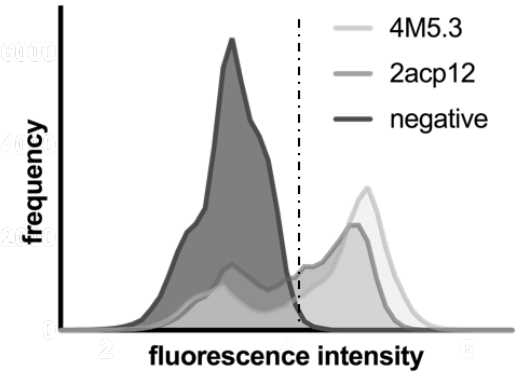

Figure 3.1: Determination of expression levels: Shown are histograms of cells expressing the positive control, 4M5.3 (light gray), the first identified binding design, 2acp12 (medium gray), and unstained cells as negative control (dark gray). The threshold for expression is marked by a dotted line.

In spite of the overall low expression stability in cycle 2, one design bearing 32 manually introduced mutations exhibited strong expression with expression levels comparable to those of 4m5.3. This design, 2acp12, exhibited several favorable features observed in natural antibodies: Tightly packed cores and both, polar and hydropho-

bic interactions within the CDRs stabilize the loops against each other as well as against the framework, and allow the antibody to assume and maintain its modeled conformations. Additionally, the design exhibited several features supportive of ligand recognition: With 2,050 Å^2, its ligand binding interface is exceptionally large, resulting in low binding energy. As a result, this design also recognized its ligand (see **Figure 3.2**, **Table 3.1**) and subsequently underwent affinity maturation and further characterization.

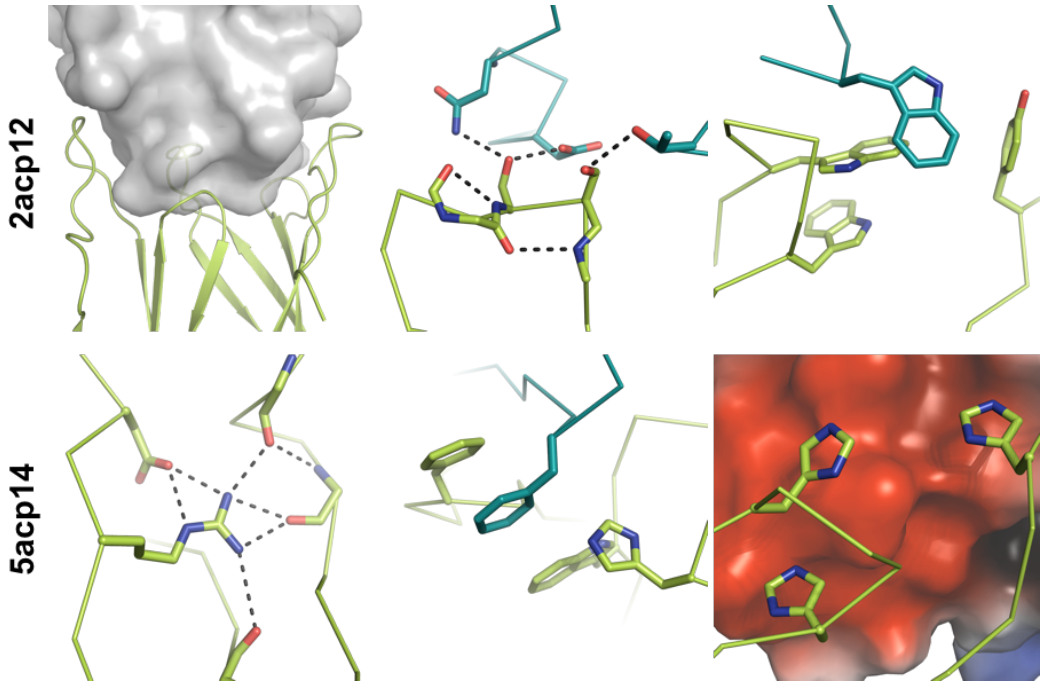

Figure 3.2: Structural features of the modeled ACP binders. (antibody in limon, ligand in teal/gray). (A) shows ACP binding design 2acp12. Long loops in both, LC and HC form a deep cleft, engulfing the ligand and creating a large binding surface (2,050 Å^2). The designed interaction between antibody and ligand incorporates both hydrogen bonds and aromatic stacking. (B) shows 5acp14, where the CDRs are stabilized by buried polar interactions similar to those observed in natural antibodies and ligand binding is mediated by aromatic stacking and charge complementarity.

Table 3.1: Computed scores for the designed ACP binders.

	binding energy (R.e.u.)	shape complementarity	sasa (Å^2)	packstat
2acp12	-38.3	0.74	2,050	0.69
5acp14	-32.3	0.77	1,650	0.74

3.1.2 Minor improvement in expression levels in cycle 3

Our initial approach was to remedy the design flaws by addressing them individually to keep the approach as general as possible. Within the Rosetta framework, the steps of a design protocol can be restricted to specific residues or identities by elements called 'task operations'; as an example for such a restriction, I wrote a task operation to address the improper placement of charged residues in the core of the antibody. The function considered each single residue within the design and determined whether it was in the core by determining the number of neighboring Cα atoms within an 8 Å sphere of this residue's Cα. In order to keep the task operation general and allow it to function even within a secondary structural element, I excluded 10 residues C- and N-terminally of the residue in question for analysis. Each residue with more than 5 other residues in its proximity was then considered core, and design restricted to non-charged identities. This alteration helped to lower the number of manually introduced mutations, but did not significantly increase experimental expression levels or yield any binding designs in cycle 3.

3.1.3 High expression levels and another binding design from cycle 5

As success with the general approach in cycle 3 was very moderate, further alterations aimed at utilizing sequence and conformation context from natural antibodies. This was done by implementing two separate elements: First, segmentation of the antibody for scaffold generation was changed from the original CDR-grafting approach to one that mirrored the V-(D)-J recombination in natural antibodies (**Figure 3.3**). This new segmentation maintained important stabilizing interactions between the CDRs and between each CDR and its supporting framework as well as conserved interactions in the framework that are formed between residues that are in spatially close proximity but distant in sequence (see **Figure 3.9**, Arg74 in 5ins14.). Due to its conservation, the disulfide bond was used as a segmentation point, resulting in one part of the antibody

containing CDRs 1 and 2 as well as most of the framework, and one containing CDR 3 and its stems. This also prevented cavities in the antibody core, as most of the Ig domain core of the scaffold was originally derived from a single, well-packed natural antibody. The second important addition in this cycle was the introduction of PSSMs; the source antibodies were segmented as described above, and the segments aligned and clustered by rmsd. For each cluster, a PSSM was generated, containing a sequence profile for the segment in this specific conformation; the PSSMs for the segments in each scaffold were then combined and used to guide sequence design. Residue choices were more constrained in the framework to ensure stability while the CDRs were given more freedom to allow ligand binding. This addition rendered the general constraints, such as restricting core positions to uncharged identities, obsolete, and allowed a much more functional guidance of sequence design.

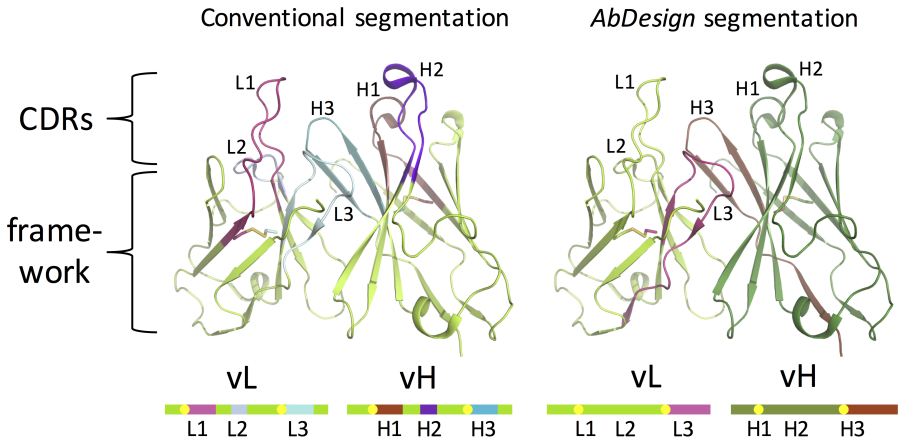

Figure 3.3: Originally, CDR grafting was used ("Conventional segmentation"), maintaining a fixed framework and replacing each CDR separately with one from another source antibodies. In design cycle 5 ("AbDesign segmentation"), antibodies are generated from only two segments in each Ig domain, the first N-terminally to the cystein residue in the stem of CDR3, and the second C-terminally from it, with the disulfide bonds kept constant. Bottom shows the segmentation along the sequence with cysteins marked in yellow. This segmentation closely resembles the V-(D)-J recombination in natural antibodies and yields significantly higher expression levels.

These two alterations yielded a significant increase in expression to levels comparable to 4m5.3 for all designs generated in cycle 5. The improved stability of the designs increased our chances of identifying binders, and thus this design cycle yielded

a second binder for ACP and one for insulin. For overall development of the expression levels over the course of algorithm development, see **Figure 3.4**.

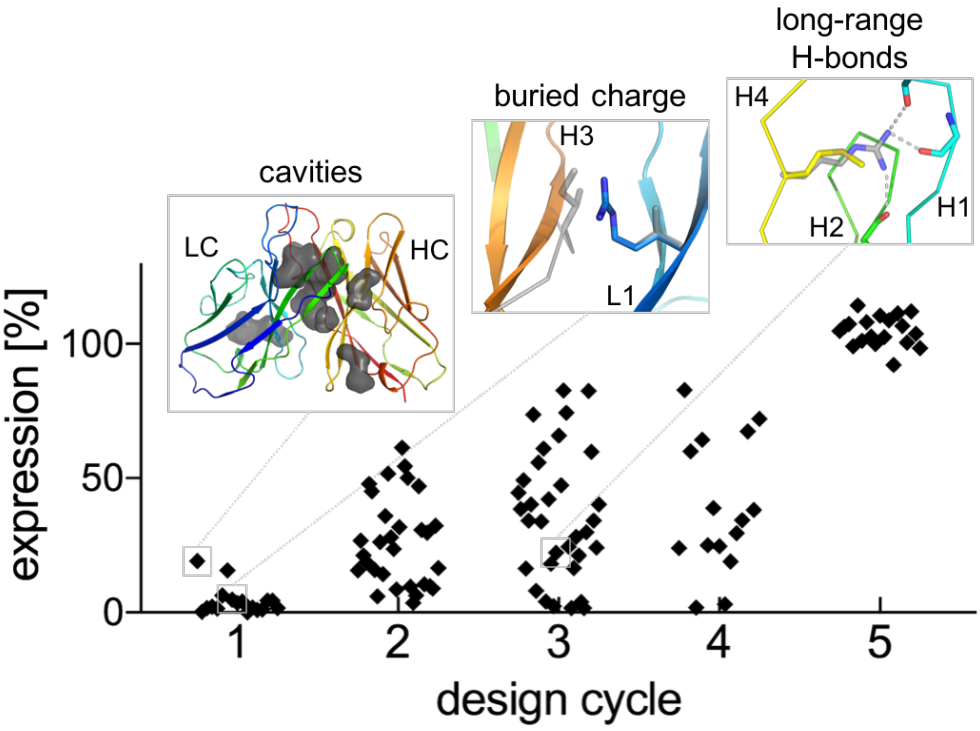

Figure 3.4: Improved expressibility through five design/experiment cycles. Designs were formatted as scFv and their yeast surface expression levels were evaluated in five successive cycles of algorithm development. Molecular representations show flaws observed in early design cycles (from left-to-right): cavities (gray) in the protein core (1ins01), a buried but unpaired arginine (1ins10), and failure to maintain a buried hydrogen-bonding network between segments distant in sequence (3ins17). Sidechains in gray show identities typical of antibodies in the relevant positions. Backbone is rainbow-colored from amino- (blue) to carboxy-terminus (red); LC and HC indicate light and heavy chains, respectively. Y-axis shows expression relative to the positive control 4M5.3. *Shown are expression levels for Insulin, because fewer mutations were introduced manually, thus making them more representative of the quality of the actual algorithm. The overall trend is discernible, yet less clear for ACP designs. All design flaws shown in the panels are similarly present in ACP designs. All experimental work on insulin designs was carried out by Dror Baran.*

3.1.4 Affinity maturation and mutation analysis support the modelled binding mode

In order to improve binding and to assess binding mode of the two ACP binders, *in vitro* evolution was performed as described in subsection 2.2.1 and [5]. None of the

mutations that increase binding in either of the designs are located on the interface, suggesting that the CDRs assume the modelled conformations and, similar to natural affinity maturation, the introduced mutations help stabilize these conformations rather than alter the interface. Affinity before and after maturation was estimated in YSD and a 20-fold improvement from approximately 900 nM to 50 nM observed for 2acp12. Initial affinity for 5acp14 was too low to estimate by YSD, but was estimated to 1 μM afterwards.

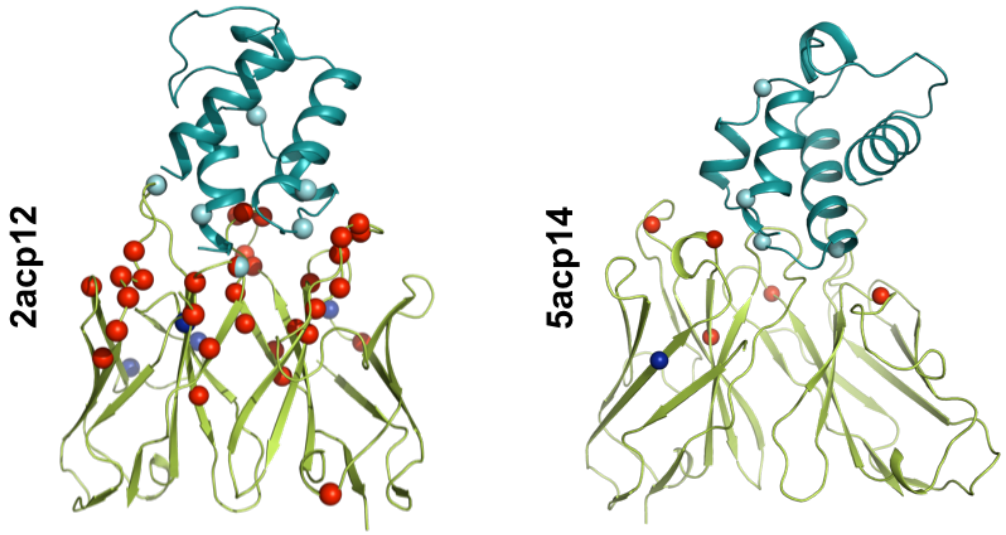

Figure 3.5: Shown are the models of both ACP binders, 2acp12 and 5acp14 (antibody in limon, ligand in teal). For clarity, the sites of mutations are shown as spheres only, for a detailed list of all mutations, see **Table 3.2** Shown are mutations introduced manually in FoldIt (red), mutations that arose during affinity maturation (blue), and mutations that were introduced to assess binding mode (cyan).

To assess binding mode, mutations were also introduced into both interacting partners, scFv and ligand. These mutations introduced either bulky side chains in place of small residues to create clashes or opposite charges to cause electrostatic repulsion. To ensure correct orientation of both partners in the interaction, mutations were introduced on the interface as well as away from it; mutations on the interface are predicted to interfere with binding while mutations further than 8 Å away from it are not expected to impair binding. Due to the low affinity of 5acp14 as scFv, no mutations on

the antibody side were introduced. For both, 2acp12 and 5acp14, mutations on the ligand side were tested against the matured variant of the scFv.

Table 3.2: Listed are all mutations introduced either manually or during affinity maturation. For 2acp12, mutations that arose during in-vitro evolution were tested as single, double, and triple mutants, but only the combination of all four mutations succeeded in the observed 20-fold improvement. For visualization, refer to **Figure 3.5**.

model name	introduced in FoldIt	assessing binding mode	affinity maturation
2acp12	D2N, K25R, K27S, F28Q, K29S, T30L, K31T, Y42F, E73S, A98T, A100S, A103V, Q135S, A142Y, N145L, S160V, T161V, A162G, A34N, S100W, V61A, D66H, H164D, T166S, F167Y, V168T, A208G, H211W, E212S, H215S, H216S, W217S, F220V, S221M, F223Y, V224L	on the scFv: A34N, S100W, D214R on ACP: D45R, A50E, K60D, V71E	V61A, D66H, A96T, F137I
5acp14	H57Y, N96D, N140D, L144S, Q163K	on ACP: D45R, A50E, K60D, V71E	V132A

The mutations introduced on the ligand show clearly that 5acp14 binds ACP on its modeled surface patch. For 2acp12, the picture is similarly clear, with predicted-deleterious mutations on the surface of the antibody impairing binding and one of the mutations on the ligand side of the interface showing the same behavior.

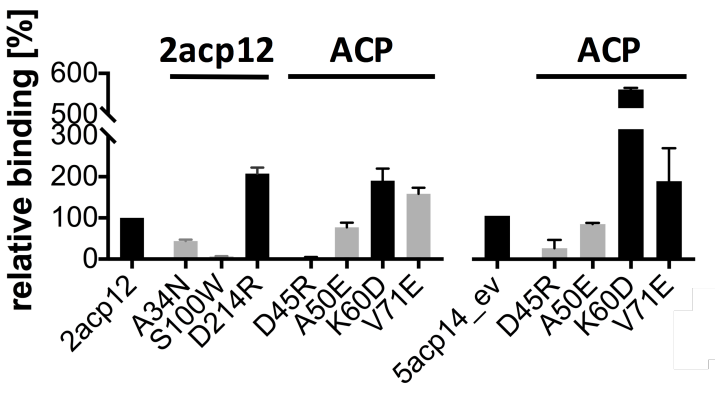

Figure 3.6: Binding mode was assessed by introducing mutations predicted to interfere with binding (light gray) as well as such that were far from the modeled interface (black). For 2acp12, mutations were introduced on both sides of the interface, the antibody as well as the ligand. Independent of interference with the modeled interaction, introduction of negatively charged mutations on the ligand seems to have exerted a positive effect on binding.

Notably, one mutation on the ligand surface, V71E, that should interfere with binding in 2acp12, does not. As we did not succeed in obtaining a bounds structure, we can only speculate on the reasons for this surprising result. One possible explanation is a local error in the model: a backbone shift in CDR H3 could allow the binding site to pack differently, thus accomodating the mutated ligand. Another option is flexibility in the ligand, as a minor backbone shift could already allow a different rotamer that would only cause minor steric clashing. In combination with this, the increased electrostatic attraction between the two proteins; this is already suggested by the increased affinity of the antibody to the K60D variant of the ligand as well as increased affinity upon introduction of an opposite charge on the antibody with the D214K mutation. As our computational minimization algorithm does not perform well enough to detect the small changes in binding energy while given the degrees of freedom to change rigid body orientation, backbone, and packing of the interface, no computational assessment of the mutated interactions was performed.

3.1.5 The designed antibodies bear little similarity to natural antibodies

The manual intervention in the early design rounds was heavily guided by the residue choices in natural antibodies, and as the final design algorithm actually uses natural antibodies to guide sequence design for the models, I wanted to assess how "novel" the designed antibodies were. For this purpose, I compared both antibodies to human and murine germline sequences, and the closest germline sequences were 37 (2acp12) and 46 (5acp14) mutations away. This shows that the designed antibodies are in fact very far from natural antibodies, and the algorithm presented succeeds in generating novel antibodies that not only fold stably into their modeled conformations, but also recognize their ligands. The closest sequences for both models are given in **Table 3.3**

Table 3.3: The closes germline sequences for the designed ACP binders 2acp12 and 5acp14. All sequences are murine in origin.

	light chain	heavy chain
2acp12	IGKV1-110*01 (17 mutations)	IGHV6-3*03 (20 mutations)
5acp14	IGKV3-3*01 (20 mutations)	IGHV6-3*03 (26 mutations)

3.1.6 Characterization of the designed antibodies as Fabs

As stated above, affinity and binding mode were initially tested on scFv formats in YSD. To assess the properties of the antibodies in a more nature-like construct, I proceeded to convert the designed binders into Fabs. The soluble proteins were expressed, purified, and subjected to SPR measurements to obtain more accurate affinity readings as well as thermal shift assays for stability measurements.

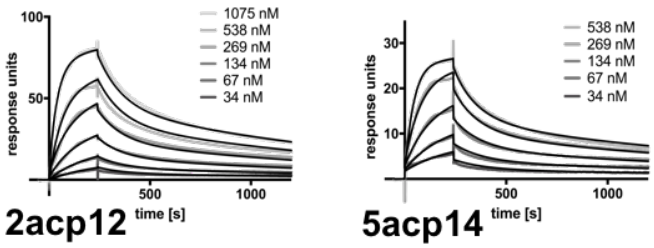

Figure 3.7: Affinity of the binders as Fab was measured by SPR and resulted in a Kd of 100 nM for 2acp12 and 50 nM for 5acp14. The difference between the Kd estimate from YSD and the results from SPR are likely due to the different constructs used (scFv in YSD, Fab in SPR).

For 2acp12, the results were comparable to those obtained by titration in YSD, while for 5acp14, a 20-fold increase in affinity was observed. This was most likely caused by the difference in orientation between the light and heavy chain in the two different constructs: In Fabs, the two domains are held in the correct orientation by the constant domains while the linker in the scFv might have a more detrimental effect; the linker was not modeled, nor was it adapted to the designs. This reinforces the assumption that rigid body orientations is crucial , warranting the implementation of an appropriate minimization scheme into the algorithm.

Apparent Tm measurements

I further assessed the physical stability of the Fabs by conducting thermal shift assays; the slow decrease of fluorescence after the peak in each curve (at high temperatures) is typical of protein aggregation, excluding the dye from formerly exposed hydrophobic core regions.

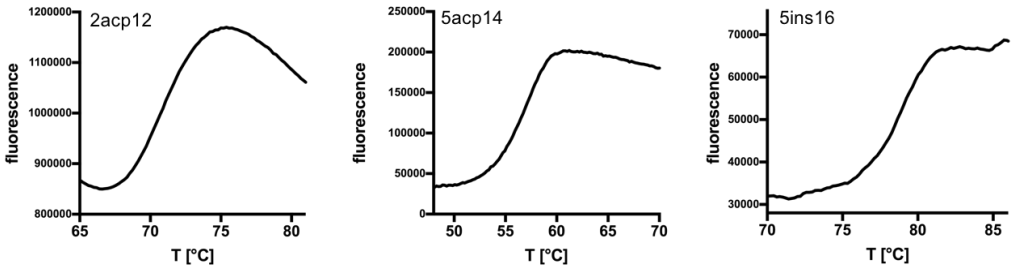

Figure 3.8: Melting curves for the designed binders. Thermal denaturation of the soluble Fabs resulted in apparent Tm of 71°C for 2acp12, 57°C for 5acp14, and 79°C for 5ins16.

Maximal unquenching of the fluorophore occurred at 71°C for 2acp12 and at 57°C for 5acp14. Since unfolding is irreversible, the results do not represent true thermodynamic melting temperatures; we therefore refer to them as apparent melting temperatures.

3.1.7 Structural agreement of models and solved structure

I also attempted to crystallize both ACP binders, but have not succeeded to date. We have in the lab been able to obtain structures that show atomic agreement with the model for two designs against insulin that were generated with the same algorithm.

The overall agreement of model and solved structure is very accurate with an overall deviation of 0.4 and 0.7 Å, respectively. The designed proteins fold correctly and both, stabilizing polar interactions and non-polar packing throughout the cores of both antibodies are formed as modeled (see **Figure 3.9**). The designed binder 5ins16 shows small conformational deviations in H1 and H3, caused by burial of a Trp residue on the LC/HC interface and desolvation of an Arg residue in favor of satisfaction of

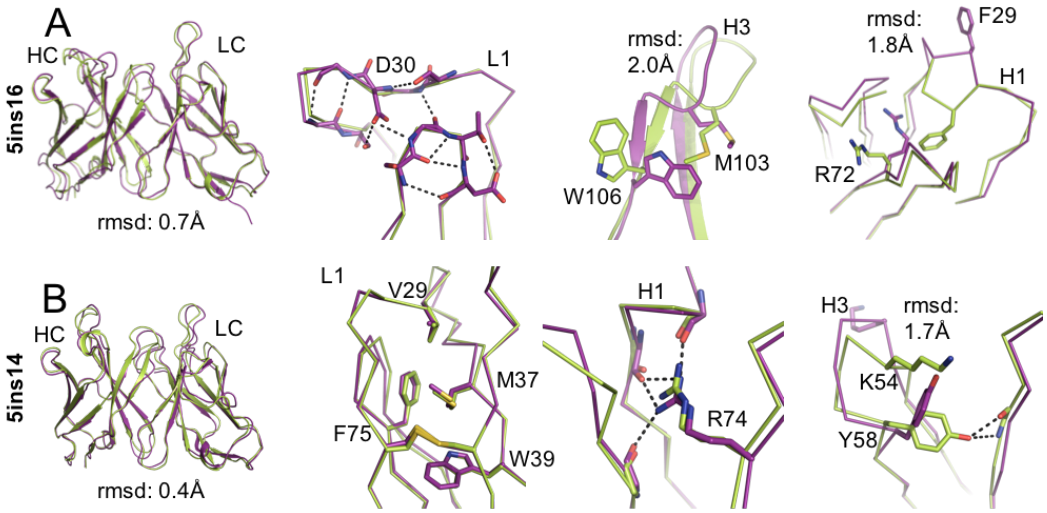

Figure 3.9: Comparison of design models and experimental structures for 5ins16 and 5ins14. In both antibodies, the light chain, the backbone conformation of the framework, and the light-heavy heterodimer interfaces are atomically accurate. (A) 5ins16: Backbone and sidechain packing deviations occur in H1 and H3, but other regions, including buried hydrogen-bonding networks (dashed lines) involving L1 are atomically accurate. (B) 5ins14: Core packing of hydrophobic residues on the framework specifying the L1 conformation as well as a buried polar network specifying H1 are atomically accurate. Model and experimental structures are colored in limon and purple, respectively. Antibodies were crystallized as Fabs, and only the Fvs are shown. *Design against insulin were computed and experimentally tested in the lab by Dror Baran, crystallization and structure solution were performed at the Israeli Structural Proteomics Center.*

buried polar interactions. The conformational changes cause local rmsds of 2.0 and 1.8 Å, respectively, but the changes are confined in both cases restricted to the CDR and its stems and do not propagate through the remainder of the antibody. In 5ins16, the deviant CDRs are also heavily involved in crystal contacts, which likely contributes to the change in conformation **Figure 3.10**.

The non-binding design 5ins14 shows the direct effect of the new segmentation as it maintains a long-range interaction of Arg74 on the non-hypervariable loop of the heavy chain with the stems of H1 and H2. Change of a rotamer in H3 causes a shift in loop conformation, but again, the change does not propagate beyond the stems of the loop, maintaining atomic accuracy throughout the rest of the antibody. For both antibodies, not only each domain in itself is folded correctly, but also the rigid body orientation between LC and HC is highly accurate.

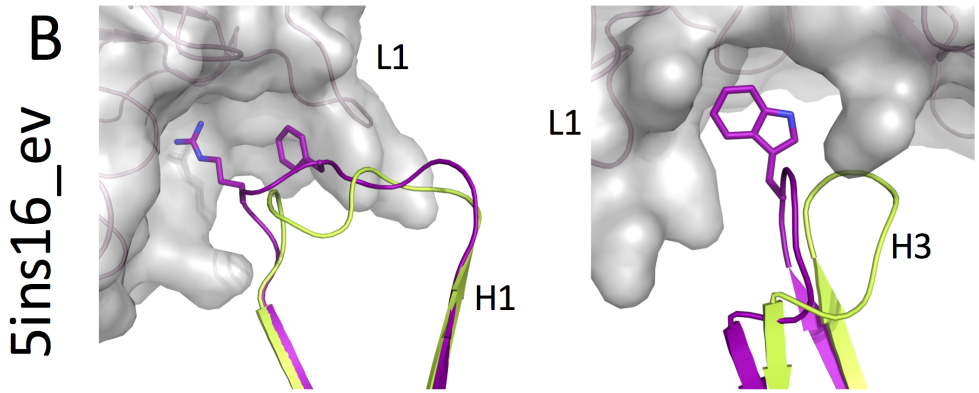

Figure 3.10: Crystal structures of 5ins16 show that CDRs H1 and H3 form packing interactions with crystallographic-symmetry neighbors (shown as surface). This might explain the structural deviations from the model.*Design against insulin were computed and experimentally tested in the lab by Dror Baran, crystallization and structure solution were performed at the Israeli Structural Proteomics Center.*

Sequences and computed scores of all designed antibodies against ACP are listed in **chapter 5**.

3.2 Designed v_HH scaffolds from camelid templates

With the adapted design algorithm described in **subsection 2.1.12**, a total of 34 antibodies were generated. Initially, I assessed whether the different segmentation yielded stable antibodies and generated 24 designs solely for stability; only after yielding high expression levels for 22 of them, I proceeded to designing v_HHs for function.

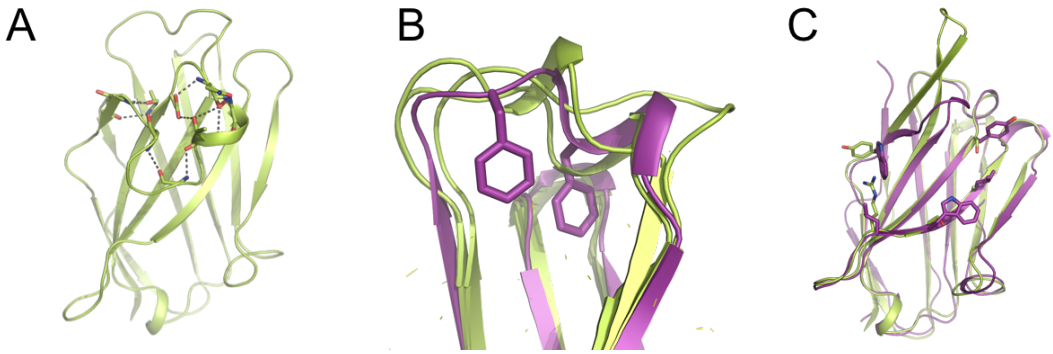

Figure 3.11: Structural features of the designed v_HHs (limon). Like in natural v_HH, the long H3 loops cover most of what would be the interface with the LC. In our designs, these loops are largely stabilized by interactions within themselves rather than by interactions with the surroundings, as would be the case in conventional antibodies (A). Similarly, our designs recapture the natural diversity in H1 conformations (B; for comparison in purple: H1 from 4m5.3 with the conserved stabilizing Phenylalanine side chains). In designs where H3 does not cover the would-be interface sufficiently or at all, solubilizing mutations are introduced (C; for comparison: the corresponding residues in a conventional antibody (tissue factor-binding Fab 5G9, PDB ID 1ahw).

In the next step, v_HHs were designed for function, in this case Phe binding. The metabolic disorder phenylketonurea presents with low levels of the hepatic enzyme phenylalanine hydroxylase, resulting in build-up of the substance to toxic levels. To rate the success of a special diet, accurate determination of Phe levels in the blood is essential, and for this purpose, selective high-affinity binders are required.

From natural Phe-binding proteins, binding modes are known, and the orientations of side chains capable of recognizing this small molecule ligand in such natural examples are copied (see **Figure 3.12**. For each such side chain, rotamers are modeled that allow the interaction with the ligand, and sets of such rotamers placed on the v_HH scaffolds. During this active-site 'matching' (see **subsection 2.1.12**), the algorithm succeeded in placing these residues in several different positions on the antibodies.

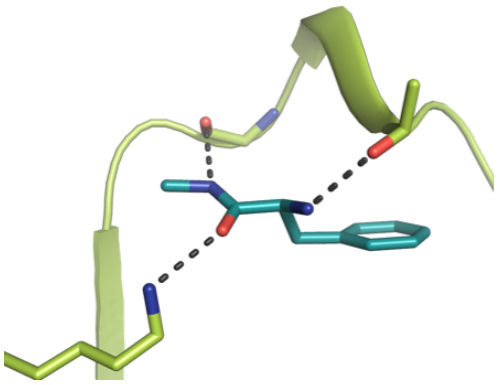

Figure 3.12: An example of the interacting residues Rosetta matches to the scaffolds: backbone carboxyl groups are forming H-bonds to the main chain amides of the ligand and a Lys side chain bonds with the Phe carboxy group. As many polar interactions as possible are formed to ensure specificity of the interaction in addition to the non-polar interaction of aromatic ring desolvation. For design purposes, the C-terminus was extended to prevent burial since this is the attachment site of the biotinylation linker.

Most common is partial burial of the aromatic ring either in a cavity formed by a non-conventional conformation of H1 or in a cleft formed between a long H3 and what would be the LC/HC interface in a conventional Fv. Examples of designs are shown in **Figure 3.13**.

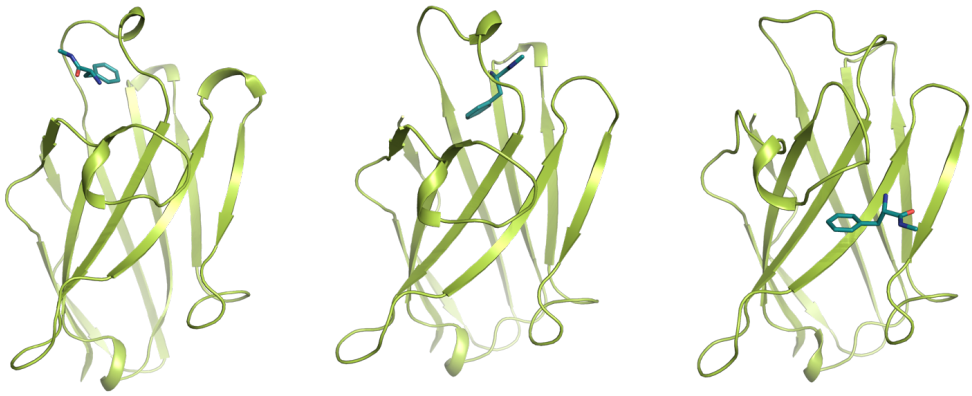

Figure 3.13: Typical Phe binding modes identified by Rosetta. Most natural single-domain antibodies that recognize small molecules bind in the site where, in a conventional antibody Fv, a cleft would be formed between the two chains. Some nanobodies have been identified that bury their ligand in the core of the antibody. The models generated show similar binding modes for Phe.

All designs generated against Phe expressed at high levels, but none of them recognized the ligand. These results suggest that the algorithm was either not fully adapted for design of v_HHs or that recognition of such a small interface area is hard to design computationally at all. Since the experimental methods used in this work require the ligand to remain bound even after attachment of labeled streptavidin to its biotin

moiety, it is also likely that such small interfaces do not provide binding energies low enough to counter the loss in entropy from immobilizing such a large construct.

I succeeded in obtaining stable, soluble protein for four of the original 24 designs (1vHH09, 1vHH11, 1vHH18, and 1vHH25; see **chapter 5** for sequences) and proceeded to perform thermal shift assays as well as attempted to obtain crystal structures.

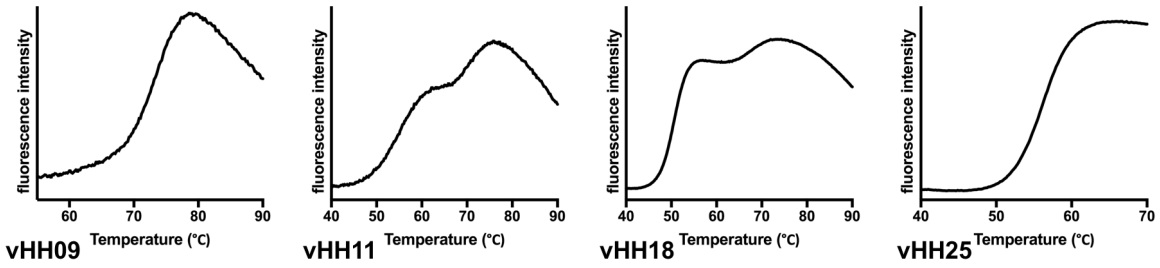

Figure 3.14: Thermal shift assays for the v_HHs expressed solubly reveal apparent Tm of 74°C for vHH09, 55°C for vHH11, 51°C for vHH18, and 56°C for vHH 25. For vHH11 and vHH18, the results reveal a two-step unfolding process.

Apparent Tms for the antibodies were determined to range from 51°C to 74°C; while these values show that I succeeded in designing stable single-domain antibodies, they are mostly below those of natural single-domain antibodies (70-80°C, [19]. Furthermore, the melting curves show a two-step unfolding process for vHH11 and vHH18, which probably indicates noncooperative folding for these two proteins.

Sufficient amounts of soluble protein were obtained and submitted for crystallization attempts at the Israeli Structural Proteomics Center, but so far, no crystals have formed.

Chapter 4

Discussion

4.1 Discussion

There are several different ways in which this work impacts the current protein design landscape. Not only are we presenting a method for generating antibodies to any protein target of solved 3D structure, but much more generally, a strategy that is not restricted to any specific fold, nor to ideal folds at all, but can be applied to any protein family for which sufficient structural and sequence data is available. In addition to those tangible goals, we have also been the first to apply an iterative approach within the field of computational protein design: By using the failures of each experimental testing cycle to improve the design algorithm, we made use of the information gained instead of letting it sink into oblivion.

4.1.1 Design inspired by nature

Over the last decades, protein design has made great advances, not only in design of novel protein folds, but also in re-purposing existing folds to bind new ligands [25, 40, 41]. In both cases, however, the design strategy relies heavily on secondary structural elements for stabilization or formation of a binding surface. In contrast to this, the algorithm presented here demonstrates a way to utilize loops for ligand recognition. This makes it much more flexible not only in terms of the overall architecture

of the binding surface but also in terms of adaptation to new targets. This property makes such 'non-ideal' binders particularly feasible for application with rapidly changing targets, such as viral proteins. The adverse facet, however, is that to date, no reliable way exists to predict or even design loop conformations, and even within the same protein family, insertion of identical sequences at identical positions within the overall fold has not generally been sufficient for yielding comparable backbone conformations. As function in antibodies is mediated entirely through loops, several attempts have been made at relating amino acid sequence to conformation, and as early as two decades ago, Al-Lazikani et al. [1] have succeeded in clustering antibody loops, and expanded over the following years [10, 12, 50], and with exception of the heavy chain CDR3, these clusters give a reasonable correlation between amino acid sequence of the loop and conformation. In our first design rounds, we used these canonical conformations to generate the design scaffolds, but over the course of algorithm development it became apparent, that this strategy did not take into account the crucial interactions between each loop and its surroundings.

The presented design strategy succeeds in recombining loops largely by two key factors:

- Segmentation of the scaffold in a way to maintain essential stabilizing interactions within each segment
- Preservation of amino acid identities crucial for configuration of the backbone

Unlike previous loop grafting strategies, these rules take into account the role that the environment plays in stabilization, for example in the shape of long-range polar networks in the core of the Ig domain or residue identities in some, often non-obvious positions. We are thus, in analogy to the "rulebook" for design of novel folds using secondary structural elements [40], guided by nature: The segmentation used resembles that of the natural V-(D)-J recombination in antibodies, and sequence design is restricted to sequences similar to the ones found in natural antibodies of the same conformation, mimicking the process of affinity maturation introducing few mutations

into a germline sequences. This sequence guidance regime also ensures that antibodies generated with the final algorithm are very similar to natural antibodies in all respects, such as surface charge, distribution of hydrophobic residues on the surface, or hydrogen bonding throughout the core and CDRs.

We have succeeded in solving two designed antibody structures in the lab, both are in almost perfect overall agreement with the model (0.4 and 0.7 Å, respectively). For both, not only the fold, but also the rigid-body orientation between the two chains are very accurate, and for 5ins14 (see **Figure 3.9**), all loops assume the designed conformations. The overall agreement for 5ins16 is similarly high, but in this design, two out of the six CDRs are slightly shifted; the changes are locally confined and do not propagate through the entire structure, showing that the algorithm has succeeded in stabilizing the other CDRs enough to not yield to these perturbations. Since the two loops in question are involved in crystal contacts, it is unclear if they assume the modeled conformation in solution or not; in any case, the inaccuracy reinforces how difficult the task of stabilizing non-ideal structures is, and how imperfect our measures for assessing stability of a structure and affinity of an interaction still are. Hopefully, ongoing research will help us further understand the connection between sequence, physicochemical context, and conformation and allow us to judge the quality of a design more adequately already at the *in silico* stage.

These rules are universally applicable

The basic strategy is not limited to antibodies, but within the guidelines of the two basic rules noted above, it is generally applicable to any protein family of modular fold; it also allows recombination of segments within other binder classes, or even classes of proteins with other functions. In many enzymes, the active site is formed by non-ideal segments that determine substrate specificity. Modular assembly followed by computational design allows much broader coverage of the structural and, subsequently, functional space accessible to these classes of proteins, while sequence design guided by nature ensures optimal stabilization of each segment.

Already, on-going work in our lab has shown that some of the design principles uncovered during antibody design work can be used to improve the physical properties of other proteins: Using PSSMs for sequence design, a method has been developed for stabilization of proteins without infringing on functionality (PROSS, see [26]) and successfully applied to several proteins, including a vaccine immunogen [6]. This achievement has a strong impact on unrelated fields of research as well, where the instability of investigated proteins leads to experimental problems such as poor yields or low signals. Furthermore, the same strategy can not only be applied to improve the physicochemical properties of prospective protein therapeutics, but in theory also to decrease their immunogenicity by guiding sequence design according to non-immunogenic examples. The combination of backbone grafting and guided sequence design has also been applied successfully in allostery design [38], and has shown first successes in enzyme design (unpublished). The results presented here have much wider implications on protein research than just providing a means of designing antibodies; they allow improvement of the physical properties of proteins *in silico*, saving resources for investigation of proteins' function rather than spending them on tedious attempts to make them experimentally accessible at all.

Non-compliance comes at a cost

Coincidentally, the alterations I made for designing v_HHs pose a first attempt at bending the rules: Recombination of loops on a fixed scaffold does not respect the segmentation points of natural antibody recombination. This strategy has been used for conventional antibodies in the past, but without much success; due to their simpler overall structure and higher stability, I expected v_HHs to be more amenable to this less refined strategy, and in combination with the much smaller number of solved structures, decided to use it to increase combinatorial diversity, particularly with regard to HCDR1. And in this case, even this unnatural segmentation proved to be successful. Similarly, the allostery-design mentioned earlier involved grafting and sequence-design of a single loop onto an existing framework, in that case of a conventional antibody [37].

Similar to the allosteric antibody, also the v_HHs I designed are less stable than their natural counterparts, with apparent Tms as low as 51°C compared to over 70°C [2,19]. It is well possible that this decrease in stability is linked to an increase in loop flexibility. If this is the case, it might not only explain why design of function has remained unsuccessful, but also why, in spite of expressing and purifying large quantities of well-behaved protein and testing hundreds of different conditions, I have been unable to obtain crystals.

4.1.2 Computational antibody design turns a new page in protein therapeutics

Antibodies are widely used in both, biomedical research and clinical practice, as diagnostic as well as therapeutic tools. The initial objective of the work presented here was to provide a means of generating stable antibodies, that, unlike currently used methods, does not rely on natural or natural-like systems. While these can be successful, and some antibodies thus generated are currently on the market, they have severe drawbacks: The investment of time and resources is high compared to the success rate - these antibodies are often not specific enough or exhibit too low affinities and rarely fulfil the high standards set for use in humans - and antibodies from animal origin are biased by their respective source organism's immune system and antibody repertoire. Many of the potential targets for therapeutic antibodies are evolutionarily conserved, so self-tolerance hinders generation of high-affinity antibodies.

These restrictions generally do not apply to computationally designed antibodies: Unlike rearrangement in B-cells, the recombination of segments in the presented algorithm is not biased by position within the DNA and even combines segments originating from different species. This diversity allows a much more even spanning of the structural space, largely removing the biases of natural systems and increasing the range of ligands that can be accommodated. Furthermore, computational design requires by far fewer resources, and with the continually decreasing cost of computa-

tional power and simultaneously increasing algorithm quality, is becoming even more cost efficient. As custom DNA synthesis has also become much more affordable, a success rate of few percents as in cycle 5 is already feasible.

Where function is concerned, there is still work to be done: The scores used here to assess the modeled interaction are not sufficient to distinguish high affinity interactions from no binding at all; as more antibodies are designed with this protocol, the increasing amount of data will allow identification of new scoring parameters and subsequently, more accurate prediction of functionality. As most computational methods, this algorithm relies on initial generation of large numbers of designs followed by several layers of filtering; the decrease in manual interference from 32 introduced mutations in cycle 2 to 5 mutations in cycle 5 already demonstrates a big step towards full automation, and introduction of new filters might be able to further reduce or even abolish the need for manual intervention.

The advantages of computational design over conventional methods are clear, and as our capability to assess interactions computationally increases, we will eventually be able to extend the range of ligands from simple proteins to small molecules, other macromolecules, or conformational protein targets. Beyond mere target recognition, this would allow maintaining active or inactive conformations for receptors or otherwise modifying protein function and greatly broaden the scope of applications.

4.1.3 Single domain antibodies: Size does matter

v_HHs present themselves as yet another powerful addition to the toolbox of protein drugs. At just over 120 amino acids in the Fv, their smaller size facilitates diffusion even in dense or otherwise inaccessible tissues [14, 45] and allows binding to small groves and clefts that are accessible to conventional antibodies only after extensive maturation or not at all [67]. It also allows them to bind to active sites more easily, giving them the potential to function as inhibitors and allowing them more easily to bind conformational targets [43, 59]. Another issue that can be addressed by use

of v_HHs is the immunogenicity of antibodies derived from other species. With their similarity to human VH genes, they elicit almost no immune response in humans, and even further humanization appears to be easily attainable [15, 68]. This would allow generation of antidotes against snakebites or other poisonous agents that do not pose the risk of serum sickness.

But v_HHs are promising not only as targets for design as therapeutics; their more favorable physical qualities make them particularly promising targets for further algorithm development. Their inherent stability makes them more forgiving to design flaws and allows soluble expression and purification even of less stable designs; this allows rating of a wider range of designs, and facilitate subsequent identification of further scoring parameters for *in silico* assessment of designs. More such scores would allow another level of filtering on the computational level, eventually increasing the success rate of design for other protein families as well.

With all these advantages, v_HHs take an exceptional position among antibodies; while their affinities are comparable to those of conventional antibodies, there appear to be inherent differences in the range of antigens they recognize, and it is not yet clear what causes these differences [16]. One possible reason is that the lack of a second chain in the Fv causes less conformational constraint in the CDRs, making them more flexible and thus more adaptable to the ligand surface; some single-domain antibodies display large conformational changes upon ligand binding [55], and the variability of the otherwise conformationally conserved HCDR1 support this assumption. If increased flexibility is in fact a reason for their different ligand range, this would make v_HHs less amenable for computational design in the way the presented algorithm suggests; it would, however, make them more apt for generation of an artificial library; such a library would span the conformationally available space and allow isolation of binders from it without the restrictions or biases of an animal's immune system. Generation of such a library *in silico* could easily be performed using the computational methods presented here, and with the help of Type IIS restrictions enzymes, its cloning and application *in vitro* can similarly be done without undue expenditure.

4.1.4 Learning by trial and error

The final objective of our joint effort was introducing an iterative-learning approach that is so common in experimental areas into the ratio-driven field of computational protein design: We deduced rules by the toilsome process of exclusion, using the small successes of each design cycle as stepping stones for the next cycle, and the failures as examples of what not to do again. We were able to demonstrate success by means of repeated failure.

To date, many promising approaches to design tasks have been abandoned because algorithms that did not succeed immediately were regarded as dead ends rather than bumps in the road, and failed designs as disincentives rather than lessons; we want to challenge these views and usher in this more laborious, but eventually rewarding strategy.

Chapter 5

Additional Data

Table 5.1: Computed scored for all designs tested experimentally in the course of the entire study. Designs for insulin were computed, selected, and tested experimentally by Dror Baran.

design name	total score	binding energy	Packstat	sasa	Sc	Expression (%)
4m5.3						72.1
1ins01	-496	-41.1	0.61	1709	0.68	13.8
1ins02	-456	-29.8	0.62	1345	0.63	0.2
1ins03	-451	-39.4	0.60	1768	0.70	1.2
1ins04	-491	-44.2	0.67	1817	0.73	1.5
1ins05	-486	-31.1	0.64	1485	0.50	1
1ins06	-480	-32.0	0.60	1635	0.66	4.6
1ins07	-505	-40.6	0.58	1927	0.62	11.3
1ins08	-483	-31.2	0.64	1563	0.77	3.4
1ins09	-438	-28.7	0.57	1631	0.64	2.6
1ins10	-498	-22.6	0.66	1378	0.59	2.9
1ins11	-505	-33.0	0.57	1660	0.65	0.1
1ins12	-493	-31.0	0.66	1326	0.70	1.3
1ins13	-483	-28.6	0.62	1387	0.62	0.7
1ins14	-495	-40.6	0.58	1685	0.73	0.8
1ins15	-487	-36.7	0.60	1816	0.70	3.2
1ins16	-483	-30.8	0.68	1712	0.65	3.2
1ins17	-501	-36.7	0.61	1721	0.62	1.2
2acp01	-489	-32.3	0.65	1588	0.73	10.2
2acp02	-536	-29.4	0.67	1938	0.65	76.1
2acp03	-535	-22.3	0.63	1761	0.63	40.1
2acp04	-524	-18.8	0.63	1532	0.59	71.9
2acp05	-529	-32.4	0.64	1837	0.77	73.2
2acp06	-492	-32.5	0.67	1700	0.60	62.2
2acp07	-502	-24.5	0.68	1495	0.70	41.9
2acp08	-552	-28.3	0.62	1593	0.71	0.7
2acp09	-514	-34.5	0.67	2033	0.75	15.8
2acp10	-449	-21.4	0.67	1466	0.75	8.5
2acp11	-511	-25.5	0.68	1497	0.69	46.4
2acp12	-529	-30.5	0.64	1522	0.74	71.4
2acp13	-558	-36.3	0.70	1664	0.71	71.8

2acp14	-533	-31.7	0.62	2011	0.70	43.6
2acp15	-544	-29.2	0.65	1596	0.63	59.6
2acp16	-523	-22.2	0.65	1703	0.56	73.8
2acp17	-542	-20.1	0.63	1480	0.61	65
2acp18	-526	-24.3	0.63	1748	0.56	67
2acp19	-535	-19.2	0.67	1389	0.60	46.3
2acp20	-542	-23.8	0.65	1723	0.60	19.7
2acp21	-549	-26.1	0.68	1506	0.64	45.4
2acp22	-536	-25.8	0.64	1652	0.64	70.9
2acp23	-524	-23.6	0.68	1664	0.73	51.9
2acp24	-517	-20.0	0.65	1459	0.67	77.3
2acp25	-560	-22.8	0.64	1588	0.56	77.6
2acp26	-560	-30.4	0.69	1786	0.62	66.2
2acp27	-537	-21.9	0.65	1441	0.53	54.6
2acp28	-547	-25.7	0.62	1709	0.63	62.7
2acp29	-550	-26.8	0.67	1773	0.67	57.7
2acp30	-550	-24.4	0.69	1420	0.73	75.7
2ins01	-454	-20.3	0.59	1727	0.62	11.32
2ins02	-488	-29.3	0.66	1719	0.62	19.3
2ins03	-473	-33.1	0.64	1704	0.68	15.3
2ins04	-448	-32.9	0.63	1834	0.60	13.5
2ins05	-505	-39.7	0.66	1907	0.71	34.5
2ins06	-462	-30.2	0.65	1431	0.80	32.5
2ins07	-465	-25.1	0.62	1705	0.67	11.4
2ins08	-487	-20.7	0.66	1512	0.59	4.3
2ins09	-452	-34.1	0.60	1772	0.65	18.9
2ins10	-448	-31.0	0.61	1657	0.58	10.3
2ins11	-507	-41.2	0.69	1879	0.67	25.9
2ins12	-467	-20.5	0.64	1419	0.56	37.2
2ins13	-508	-28.5	0.70	1734	0.69	20.1
2ins14	-496	-33.6	0.63	1709	0.68	17
2ins15	-480	-28.3	0.63	1635	0.66	6.14
2ins16	-505	-36.4	0.61	1927	0.62	22.9
2ins17	-405	-35.0	0.67	1968	0.68	44.2
2ins18	-488	-29.4	0.64	1558	0.73	39.3
2ins19	-482	-22.1	0.66	1469	0.69	36.1
2ins20	-464	-23.2	0.65	1367	0.68	6.9
2ins21	-479	-25.2	0.65	1529	0.69	2.6
2ins22	-457	-31.6	0.66	1683	0.67	4.6
2ins23	-471	-27.2	0.68	1678	0.69	33.9
2ins24	-482	-32.8	0.67	1435	0.70	22.2
2ins25	-493	-37.0	0.64	1736	0.68	7.5
2ins26	-457	-27.3	0.67	1453	0.67	21.4
2ins27	-464	-27.6	0.67	1483	0.67	7.1
2ins28	-485	-34.5	0.61	1961	0.65	6.4
2ins29	-476	-28.4	0.64	1593	0.59	23.3
2ins30	-501	-37.0	0.60	1689	0.66	11.9
3acp01	-558	-32.8	0.69	1702	0.69	14.8
3acp02	-522	-29.2	0.65	1528	0.64	25.6
3acp03	-550	-29.7	0.67	1824	0.64	4.7
3acp04	-523	-25.4	0.67	1861	0.65	5.7
3acp05	-496	-26.2	0.68	1550	0.68	61.5
3acp06	-502	-27.0	0.63	1558	0.71	36.2
3acp07	-579	-34.3	0.69	1547	0.65	44.1
3acp08	-564	-39.2	0.72	1955	0.76	39.8
3acp09	-527	-26.3	0.68	1579	0.71	20.6
3acp10	-558	-28.9	0.67	1731	0.74	19.3

3acp11	-564	-27.8	0.66	1617	0.77	19.5
3acp12	-554	-24.1	0.65	1690	0.72	17.9
3acp13	-508	-24.1	0.67	1873	0.69	44.2
3acp14	-553	-26.8	0.66	1634	0.64	17.3
3acp15	-543	-24.9	0.71	1475	0.68	15
3acp16	-522	-26.5	0.64	1973	0.64	46.8
3acp17	-534	-26.5	0.65	2033	0.65	27.1
3acp18	-524	-27.7	0.67	1848	0.64	38.1
3acp19	-563	-32.0	0.71	1936	0.70	23
3acp20	-546	-18.9	0.73	1606	0.71	11
3acp21	-553	-29.0	0.63	1680	0.66	56.3
3acp22	-534	-26.2	0.69	1342	0.76	36.5
3acp23	-543	-26.0	0.70	1363	0.76	17.2
3acp24	-556	-30.9	0.70	1850	0.73	11.5
3acp25	-552	-18.9	0.66	1700	0.55	33.7
3acp26	-542	-26.2	0.69	1663	0.69	59.6
3acp27	-526	-32.7	0.70	1864	0.72	19.8
3acp28	-495	-17.6	0.65	1445	0.69	45.6
3acp29	-543	-28.3	0.69	1651	0.75	7.1
3acp30	-538	-26.6	0.65	1456	0.71	6.8
3acp31	-526	-25.7	0.67	1765	0.66	1.9
3acp32	-538	-25.1	0.69	1567	0.74	48.2
3acp33	-568	-29.4	0.72	1502	0.76	9.2
3acp34	-545	-20.8	0.69	1367	0.75	25.5
3acp35	-534	-22.7	0.68	1463	0.59	3.9
3ins01	-480	-21.0	0.68	1392	0.72	32.08
3ins02	-476	-23.3	0.68	1634	0.71	27.8
3ins03	-498	-29.7	0.61	1658	0.70	35.53
3ins04	-510	-27.3	0.68	1603	0.67	11.82
3ins05	-503	-30.1	0.66	1692	0.68	24.67
3ins06	-490	-22.8	0.65	1381	0.67	28.99
3ins07	-499	-25.8	0.69	1508	0.72	53.1
3ins08	-495	-24.8	0.67	1445	0.66	5.83
3ins09	-501	-24.1	0.65	1327	0.70	40.3
3ins10	-504	-23.3	0.63	1367	0.68	24.4
3ins11	-450	-21.6	0.65	1515	0.73	44.04
3ins12	-456	-31.0	0.64	1667	0.77	2.9
3ins13	-443	-20.6	0.66	1164	0.74	30.38
3ins14	-454	-35.0	0.65	1757	0.70	13.05
3ins15	-458	-14.5	0.68	1344	0.77	1.74
3ins16	-440	-26.0	0.61	1719	0.69	16.07
3ins17	-452	-22.4	0.61	1605	0.68	47.46
3ins18	-458	-18.4	0.65	1224	0.69	34.17
3ins19	-459	-23.4	0.69	1375	0.73	59.55
3ins20	-447	-22.3	0.66	1490	0.66	53.61
3ins21	-462	-23.7	0.66	1425	0.68	17.66
3ins22	-460	-21.7	0.69	1286	0.68	1.13
3ins23	-477	-17.3	0.68	1218	0.71	11.86
3ins24	-469	-23.6	0.66	1590	0.67	20.24
3ins25	-463	-18.0	0.66	1614	0.68	15.33
3ins26	-451	-20.8	0.67	1301	0.76	2.79
3ins27	-435	-18.2	0.63	1643	0.69	1.15
3ins28	-458	-18.5	0.62	1337	0.67	21.57
3ins29	-466	-21.0	0.65	1394	0.65	59.48
3ins30	-438	-21.9	0.69	1467	0.78	43.12
3ins31	-458	-19.9	0.67	1286	0.66	24.59
3ins32	-460	-28.8	0.65	1722	0.68	17.41

3ins33	-443	-17.0	0.60	1312	0.71	29
4ins01	-450	-23.2	0.68	1369	0.72	17.27
4ins02	-482	-25.3	0.69	1471	0.73	59.76
4ins03	-421	-24.9	0.70	1304	0.77	43.25
4ins04	-452	-30.7	0.66	1650	0.72	1.32
4ins05	-452	-24.4	0.67	1506	0.71	46.31
4ins06	-470	-27.4	0.67	1422	0.77	17.98
4ins07	-450	-30.2	0.67	1587	0.73	28.03
4ins08	-484	-29.1	0.66	1732	0.70	17.81
4ins09	-490	-34.4	0.65	1815	0.68	2.21
4ins10	-482	-30.8	0.66	1674	0.73	13.66
4ins11	-466	-25.4	0.64	1458	0.58	21.32
4ins12	-446	-23.9	0.63	1642	0.65	24.79
4ins13	-462	-32.5	0.68	1566	0.81	48.64
4ins14	-444	-21.6	0.65	1532	0.79	27.53
4ins15	-440	-31.1	0.67	1482	0.71	52.04
5acp01	-549	-26.3	0.67	1577	0.69	80.2
5acp02	-554	-39.6	0.71	2336	0.68	78.8
5acp03	-574	-26.0	0.73	1731	0.64	78
5acp04	-566	-26.8	0.67	1593	0.60	78.9
5acp05	-490	-24.3	0.68	1772	0.59	80.9
5acp06	-527	-25.1	0.68	2068	0.63	64.6
5acp07	-494	-29.8	0.69	1783	0.75	71.7
5acp08	-568	-26.3	0.69	1767	0.56	75.9
5acp09	-562	-29.0	0.69	1642	0.66	78.3
5acp10	-554	-33.8	0.68	2016	0.61	71
5acp11	-582	-25.3	0.70	1695	0.54	81.2
5acp12	-513	-20.5	0.69	1470	0.64	75.2
5acp14	-557	-30.9	0.69	1718	0.75	78.2
5ins01	-425	-23.5	0.65	1546	0.67	75.6
5ins02	-425	-29.4	0.63	1588	0.71	76.7
5ins03	-421	-28.2	0.67	1685	0.71	77.5
5ins04	-403	-20.9	0.69	1519	0.72	71.5
5ins05	-416	-24.6	0.70	1553	0.76	82.4
5ins06	-388	-21.0	0.67	1614	0.66	72.9
5ins07	-399	-30.8	0.66	1704	0.69	77.9
5ins08	-416	-17.3	0.66	1889	0.61	74.1
5ins09	-429	-23.4	0.65	1474	0.65	72.1
5ins10	-354	-22.2	0.64	1530	0.57	79.7
5ins11	-421	-25.0	0.69	1743	0.72	74
5ins12	-402	-29.2	0.63	1551	0.73	78.4
5ins13	-365	-21.7	0.63	1624	0.63	66.5
5ins14	-445	-25.3	0.72	1695	0.68	80.3
5ins15	-404	-23.4	0.66	1379	0.67	77
5ins16	-398	-18.2	0.66	1470	0.63	72.5
5ins17	-396	-17.1	0.67	1521	0.64	80.8
5ins18	-411	-24.3	0.65	1456	0.62	74.8
5ins19	-405	-25.8	0.65	1671	0.63	70.9

Table 5.2: Peptide sequences of all designs texted experimentally in the course of the entire study. Designs for insulin were computed, selected, and tested experimentally by Dror Baran.

design name	protein sequence

1ins01	SDVVMTQTPLSLPVSLGDQASISCRSSQSAVLRNGLTFLFWYLQKPGQSPKVLIYKVSNRVSGVP-DRFSGSGFGTDFTLKINRVEAEDLGVYFCAADSWMANENRFGGGTKLEIKSSADDAKKDAAKK-DDAKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCVTLGFNFMEYAMNWVRQSPEKGLEWV-AAFYLNNSAQYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTADARQNQQGNAAVKS-GQGTSVTVS
1ins02	SDVVMTQTPLSLPVSLGDQASISCISADAEAYYSFVMFQLQKPGQSPKVLIYKVSNRVSGVPDRF-YGTGMGILYFLKINRVEAEDLGVYFCGVNWSKSIKFGQGTKLEIKSSADDAKKDAAKKDDAKK-DDAKKDGGVKLDETGGGLVQPGGAMKLSCVTVGFVFAFYLMNWVRQSPEKGLEWVAAFDP-WSDSTEYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCEASESKFNSSSQAAGQGTSVT-VS
1ins03	SDVVMTQTPLSLPVSLGDQASISCIAAKMAISWDGDVFDRWYLQKPGQSPKVLIYKVSNRVSGV-PDRFEGTGRQRMYYLKINRVEAEDLGVYFCGSFGPFGEETSYGFGTKLEIKSSADDAKKDAAK-KDDAKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCTGSGFNVTGYMMAWVRQSPEKGLEW-VGSYAYMEWMVAWADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCRATANGLFTDAQM-GWGTSVTVS
1ins04	SDVVMTQTPLSLPVSLGDQASISCRSSQSLWMLNGVIALAWYLQKPGQSPKVLIYKVSNRVSGVP-DRFSGSGNGNDFTLKINRVEAEDLGVYFCAGAARGYQDFGPGTKLEIKSSADDAKKDAAKKDD-AKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCVTEGFVFGDTIMNWVRQSPEKGLEWVAGF-YFNQDASYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCRSSSMYSKNYAGRGTSVTVS
1ins05	SDVVMTQTPLSLPVSLGDQASISCRSSQYLVMSDGETRLRWYLQKPGQSPKVLIYKVSNRVSGV-PDRFSGSGSGTDFTLKINRVEAEDLGVYFCAAQDSEANRYSFGWGTKLEIKSSADDAKKDAAK-KDDAKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCVTAGFYFGKYLMNWVRQSPEKGLEW-VAFWAYANAAYYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCEAAAYAYDNAYRTA-AAYAGAGTSVTVS
1ins06	SDVVMTQTPLSLPVSLGDQASISCKSKKTSFKDTFWYLQKPGQSPKVLIYKVSNRVSGVPDRFY-GTGEGREYKLKINRVEAEDLGVYFCSQFFSYEVSFGYGTKLEIKSSADDAKKDDAKKDDAKKD-DAKKDGGVKLDETGGGLVQPGGAMKLSCTASGLDSSRAFYMWVRQSPEKGLEWVATYTSGN-LAYYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAIRQSKLNSAMAVMGQGTSVTVS
1ins07	SDVVMTQTPLSLPVSLGDQASISCFAKFMVYDYWKNNSHVAWYLQKPGQSPKVLIYKVSNRVS-GVPDRFYGTGSGRFFRLKINRVEAEDLGVYFCAQRASIPWAATAGGGTKLEIKSSADDAKKDA-AKKDDAKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCRAEGVDDSEMTFEWVRQSPEKGLE-WVAAFTDNNSAAYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCEAASENISNTAFIYIG-DGTSVTVS
1ins08	SDVVMTQTPLSLPVSLGDQASISCRSSLSLKDSSGMTLTAWYLQKPGQSPKVLIYKVSNRVSGVP-DRFKGEGEGKDFVLKINRVEAEDLGVYFCAQFSIAPFTFGFGTKLEIKSSADDAKKDAAKKDDA-KKDDAKKDGGVKLDETGGGLVQPGGAMKLSCVTTGFNFGDTIMNWVRQSPEKGLEWVASFW-SGNLTRYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCAAAKTGLAWWMGDGTSVTVS
1ins09	SDVVMTQTPLSLPVSLGDQASISCFSESSILDNSGSTSSAWYLQKPGQSPKVLIYKVSNRVSGVPD-RFYGSGSGKTFRLKINRVEAEDLGVYFCSDDQSRKKTFGFGTKLEIKSSADDAKKDAAKKDDA-KKDDAKKDGGVKLDETGGGLVQPGGAMKLSCDSQGFWFWDYSFSWVRQSPEKGLEWVATFF-SNNLSAYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCAADTIAKAFSGDGTSVTVS
1ins10	SDVVMTQTPLSLPVSLGDQASISCKSSKYVYQKNFNESKVRWYLQKPGQSPKVLIYKVSNRVSG-VPDRFEGTGSGREYKLKINRVEAEDLGVYFCAAWESDQGRSKFGGGTKLEIKSSADDAKKDAA-KKDDAKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCDTQGFTARKSVFMWVRQSPEKGLE-WVAAFWNDNLAFYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCESASGSTAWVGLYT-GLGTSVTVS
1ins11	SDVVMTQTPLSLPVSLGDQASISCWASTTLIDSSGFTGLAWYLQKPGQSPKVLIYKVSNRVSGVP-DRFSGSGAGTFYTLKINRVEAEDLGVYFCAQNMYAPRTFGGGTKLEIKSSADDAKKDAAKKDD-AKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCVTSGFRFFYYIMNWVRQSPEKGLEWVAQY-EDLTNNYRAMYSDSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCVSDAGAFTGAAWIGKG-TSVTVS
1ins12	SDVVMTQTPLSLPVSLGDQASISCRSSGSLMHSSGWTSTSWYLQKPGQSPKVLIYKVSNRVSGVP-DRFSGSGSQTDFTLKINRVEAEDLGVYFCAQSFQSRWAFGGGTKLEIKSSADDAKKDAAKKDD-AKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCVTEGFTFRNSFMNWVRQSPEKGLEWVAEF-RNLPTNYNTAYSDSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTAASDFSESDWYSGQGT-SVTVS
1ins13	SDVVMTQTPLSLPVSLGDQASISCRSSFSLTSADGFTRAVWYLQKPGQSPKVLIYEVSNRVSGVP-DRFSGSGSGNEFTLKINRVEAEDLGVYFCGAASSDYWSYGRGTKLEIKSSADDAKKDAAKKDD-AKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCVGRGFEFKTYTMNWVRQSPEKGLEWVAG-FKNNNWWRYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTSSTSRQEMYDAMALWM-GQGTSVTVS
1ins14	SDVVMTQTPLSLPVSLGDQASISCKAKFQLYYESNNEARAVWMLQKPGQSPKVLIYKVSNRVSG-VPDRFYGEGSGAEFRLKINRVEAEDLGVYFCAAVSDRSNKSSYSFGGGTKLEIKSSADDAKKDA-AKKDDAKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCYAYGTKARDLIFSWVRQSPEKGLE-WVASFMWNNAEFYADSVEGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCSAASARFSYVIWYSG-DGTSVTVS
1ins15	SDVVMTQTPLSLPVSLGDQASISCFSSDSLINNETNLAITIWYLQKPGQSPKVLIYMVSNRVSGVP-DRFYGEGSGKSYRLKINRVEAEDLGVYFCGGSSNYAARFGGGTKLEIKSSADDAKKDAAKKDD-AKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCYMAGFWFQGVVFMWVRQSPEKGLEWVA-AFYWNNEAAYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCFASIANVTTYVGDGTSVT-VS
1ins16	SDVVMTQTPLSLPVSLGDQASISCRSQENAKSNIAWYLQKPGQSPKVLIYKVSNRVSGVPDRYSG-SGRGKDFTLKINRVEAEDLGVYFCAAYSWRSYKRRFGGGTKLEIKSSADDAKKDAAKKDDAK-KDDAKKDGGVKLDETGGGLVQPGGAMKLSCVTTGFLFYWYQMNWVRQSPEKGLEWVAGFD-PTNGFVMYADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMASWSGDTFYEGQGTSVTV-S
1ins17	SDVVMTQTPLSLPVSLGDQASISCKAAFDLWNFWGEIFLTWFLQKPGQSPKVLIYKVSNRVSGV-PDRFYGTGSGQKFKLKINRVEAEDLGVYFCAQDAWYPSKTQFGPGTKLEIKSSADDAKKDAAK-KDDAKKDDAKKDGGVKLDETGGGLVQPGGAMKLSCYSYGFSSSAYIYTWVRQSPEKGLEWVA-QFTSNPFNYWTYYSDSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAAKFGLSGSLWAG-SGTSVTVS
2acp01	SEVKLDETGGGLVQPGGAMKLSCVVSGFTFGALQMLWIRQSPEKGLEWVAMFESTWHNQRTY-YADSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCLAAFNHVLVYLGEGTSVTVSGGGGSGG-GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRLSARQVGDVYLLWFLQKPGQSPRVLIYYVSRR-VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCLQVTTDPYTLGGGTKLEIK

2acp02	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFGWYEMLWVRQSPEKGLEWVASWSYNNHTTYA- KSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMASSDKNSFNDLRYLGEGTSVTVSGGGGS- GGGGSGGGGSDVVMTQTPLDLPVSLGDQADIDCRSSQSLVSTSGDTQLLWFLQKPGQSPRVLIY- KVSNRVSGVPDRFDGSGSGTDFTLKINRVEAEDLGVYFCVASSQTSETFGRGTKLEIK
2acp03	SEVKLDETGGGLVQPGGAMKLSCVVSGFVFGLYWMLWLRQSPEKGLETVATVYDSRDNFETQ- YSSSVKGRFTISRDDSKDSVYLQMNNLRVEDTGIYYCMGFALENQMVVLGTVSSVTVSGGGGSG- GGGSGGGGSDVVMTQTPLDLPVDLGDQASISCRALYSLVSESGTTELFWFLQKPGQSPKVLIYK- TSERVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCMQWSYWPYTFGGGTKLEIK
2acp04	SEVKLDETGGGTVQPGGAMKLSCVVSGFTFTQFEMNWVRQSPEKGLEWVAQFDPDSDDTNY- ADSVKGRFTISLDTSKSSVYLQMNNLRTEDTGIYYCMAALSELMLYSGQGTDVTVSGGGGSGG- GGSGGGGSDVVMTQTPLDLPVDLGDQADIDCRSSDTLQSDSGQTYVVWWLQKPGQNPKVLIY- RVSNRLSGVPDRFSGSGDGTDFTLKINRVEAEDLGVYFCAQISHTPVTFGGGTKLEIK
2acp05	SEVKLDETGGGLVQPGGAMKLSCVTSGFNFSEYIMQWVRQSPEKGLEWVAAFYEDNTTSYAD- SVKGRFTISRDDSKDSVYLQMNNLRVEDTGIYYCMVLNKGSPWIVLYTGDGTSVTVSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKQDLHSQVAWFLQKPGQSPKTLIYKVSNR- VSGVPDRFSGSGTGRDFTLKINRVEAEDLGVYFCAQFYKDPQTFGGGTKLEIK
2acp06	SEVKLDETGGGLVQPGGAMKLSCTVQGFTFGEYAMDWLRQSPEKGLEWVAAFFSFYWTTYE- RSVKGRFTISMDPSKSSVYLQMNNLRVEDTGIYYCAASSQDSDVATYLGDGTSVTVSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSSSLVSSNGRTYLVWYLQKPGQSPKVLIYKVS- NRQSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAASSDNNDPLTFGGGTKLEIK
2acp07	SEVKLDETGGGLVQPGGAMKLSCVTSGFLFGWYIMLWVRQSPEKGLEWVAAWEYNNLTYYA- NSVKGRFTISRDDSKSSVYLQMNNLRVEDTGRYYCIAFAFGSNAFAYLGDGTSVTVSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSAQALSTRLVWWLQKPGQSPKVLIYKTSNRVS- GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCLQLNNWPWTFGGGTKLEIK
2acp08	SEVKLDETGGGLVQPGGAMKLSCVISGFNFRHYQVLWVRQSPEKGLEWVAGFEPSSGKTSYSD- SVKGRFTINVDPSKSSVYLQMNNLRVEDTGIYYCMAAIFWLPIKAGQGTSVTVSGGGGSGGGGS- GGGGSDVVMTQTPLSLPVSLGDQASISCRSSQPLVHPDGRTYLAWFLQKPGQSPKVLIYRISNRL- SGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCLVYMEASDMVWFGGGTKLEIK
2acp09	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGDYYMLWVRQSPEKGLEWVAVVGPDNSYTNY- ADSVKGRFTISRDDSKSSVYLQMNNLRTEDTGIYYCMGSSWSQDSSSESVMKYLGQGTSVTVSG- GGGSGGGGSGGGGSNVVMTQTPLSLPVSLGDQASISCRSSQSLTAEAGLTVLAWFLQKPGQSPK- VLIYKVSNRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAAWTNSKWVFGGGTKLEIK
2acp10	SEVKLDLTGGGLVQPGGAMKLSCVVSGFRFQLSYMMWVRQSPEKGLEWVAQFVNDAQAAYA- DSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCWAATTSTTTFWSALAVIGQGTSVTVSGGG- GSGGGGSGGGGSDTVMTMTPLSLPVSLGDQASISCRSSQSLVSSTYNTTWVLWYLQLPGQSPKV- LIYLVSNRVSGVPDRFYGTGSGTDFTLKINQVEAEDLGVYFCASATNWSWLSWGGGTLLEIK
2acp11	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFSYYEMLWVRQSPEKGLEWVALFMNNSETYYSD- SVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCLVSLVSSQSYLGDGTSVTVSGGGGSGGGGS- GGGGSRVVMTQTPLSLPVSLGDQASISCRSSAQKVGNSWLAWFLQKPGQSPKVLIYKISNRVSG- VPDRFSGSGSGTDFTLKINRVEAEDLGVYFCVTWLNFYMTFGQGTKLEIK
2acp12	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFGAAVMLWVRQSPEKGLEWVAAYRNDNFNRLT- YYSRSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAAANNFVYFLGNGTSLTVSGGGGSG- GGGSGGGGSEYVMTQTPLDLPVSLGDQADISCRLDSLAGNNAWVSWFLQKPGQSPKVLIYKVR- NRRQGVPDRFDGSGSGTDFTLKINRVEAEDLGVYFCVSFKDEPLRFGEGTKLEIK
2acp13	SEVKLDETGGGLVQPGGAMKLSCTVSGFTFGKSEMVWVRQSPEKGLEWVAQFRSSDHNDETY- YNDNVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCVASVHELLFQLGQGTLVTVSGGGGSG- GGGSGGGGSDVEMTQTPLDLPVSLGDQADIDCRSLYSLVKRYGYTFLMWFLQLPGQSPKVLIY- RVSNRVSGVPDRFDGSGTGTDFTLKINRVEAEDLGVYFCAQASTAPYTFGNGTKLEIK
2acp14	SEVKLDETGGGLVQPGGAMKLSCVVSGFTFGAYTMLWVRQSPEKGLEWVAAFSPATNSTYYA- DSVKGRFTISMDTSKSSVYLQMNNLRVEDTGIYYCVASSETGVIYLGDGTLVTVSGGGGSGGGG- SGGGGSDVVMTQTPLSLPVSLGDQASISCRMSQTLYTTWWYTTELTWMLQLPGQSPKVLIYKV- SNRVSGVPDRFYGSGSGTDFTLKINRVEAEDLGVYYCMSWANERMAFGGGTLLEIK
2acp15	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGWWVVLWVRQSPEKGLEWVAAARTWDDKTY- YAASVAGRFTISLDDSKSSVYLQMNNLRVEDTGIYYCMAAWSDIVWYIGDGTVVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCAALTHNDEHYVVWFLQLPGQSPKTLVYSIKF- RVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSRWPFTFSTGTKLEIK
2acp16	SEVKLDLTGGGLVQPGGAMKLSCVVSGFEFGKATMLWVRQSPEKGLEWVASVSTDPWNAETY- YSDSVAGRFTISLDDSKSSVYLQMNNLRVEDTGIYYCMAAEHLFVFATGDGTVVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASASLQDSDGTIIVMWYLQLPGQSPKTLVHD- TYHRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAAFTTLPAWFSTGTKLEIK
2acp17	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKANMNWVRQSPEKGLEWVAQVKNEYTNKET- RYSDSVAGRFTISLDDSKSSVYLQMNNLRVEDTGIYYCAASWANHIYYTGSGTVVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFKTNHTVIWYLQLPGQSPKVLVMHVSD- RVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCASVYTDKFSFSTGTKLEIK
2acp18	SEVKLDETGGGLVQPGGAMKLSCVTSGFNFGSAIMVWVRQSPEKGLEWVAAFTNDLYNQRTY- YSDSVAGRFTISMDDSKSSVYLQMNNLRVEDTGIYYCMAALSRNIYYTGNGTVVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKYTADQNNVTFYLQLPGQSPKVLFMSTDM- RVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCATSSTWPSRFSIGTKLEIK
2acp19	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGTWNMLWVRQSPEKGLEWVASAAWYNATTW- YAASVRGRFTISMDNSKSSVYLQMNNLRVEDTGIYYCMAARASTAEYTGNGTVVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALFVVSFYVVWYLQLPGQSPKVLVFNVRIR- VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSKWPASFSTGTKLEIK
2acp20	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGSYVMLWVRQSPEKGLEWVASIASNDYLTFRSD- SVAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCVAVSAGSYVLSWTGNATLVTVSSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASHSLKNDYTNQIYIMWYLQLPGQSPKELVFL- TNIRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAAAARWPFTFSTGTKLEIK
2acp21	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGNYVMLWVRQSPEKGLEWVAAVAKNNRSTRYS- DSVRGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCAGWSRGALYMGDGTSVTVSSGGGGSGGG- GSGGGGSDVVMTQTPLSLPVSLGDQASISCRALFNVSSHVVWYLQLPGQSPKVLIYNVDDRVSG- VPDRFSGSGSGTDFTLKINRVEAEDLGVYFCASFNAHPFSFSTGTKLEIK

2acp22	SEVKLDETGGGLVQPGGAMKLSCVASGFDFGQYMMLWVRQSPEKGLEWVAAAFWDPHDNYT- AYSDSVAGRFTISLDDSKSSVYLQMNNLRVEDTGIYYCFAAKAEENAVLVVGNATLVTVSSGGG- GSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASHDLSQNMDNKTHIVWYLQLPGQSPK- TLVYNTYFRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQYWSFPFVFSTGTKLEIK
2acp23	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGHAVMLWVRQSPEKGLEWVAAAYAQTNETAY- SDSVAGRFTISMDDSKSSVYLQMNNLRVEDTGIYYCMAAFHWFIFLTGTRTLVTVSSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASYKANRYVVWYLQLPGQSPKVLVFAINYRVS- GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAAWTQTRYWHFSTGTKLEIK
2acp24	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFSQYIMNWIRQSPEKGLEWVAQISWDASLTAYSD- SVAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMADSNDNSLSLTGDGTLVTVSSGGGGSGGG- GSGGGGSDVVMTQTPLSLPVSLGDQASISCRALYLANDHVVWFLQLPGQSPKVLVANIYWRVS- GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQSSIWPSSFSTGTKLEIK
2acp25	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKSIMNWVRQSPEKGLEWVAAYANNNLTWYSD- SVAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCSAMWWWHTAETNLVIKYSGNGTVVTVSSG- GGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASHSLIYDNGSTYLVIALQLPGQSPKV- LIFHTSWRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQLAMFPYTWSTGTKLEIK
2acp26	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGQTSMLIVRQSPEKGLEWFASIHTDVNWTAYSD- SVAGRFTISLDNSKSSVYLQMNNLRVEDTGIYYCMALFYADNDSVNYVGSATVVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASRALWSHDGSINVVWYLQLPGQSPKVLVYN- VRYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSWTPFSFSTGTKLEIK
2acp27	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGNTWMLWVRQSPEKGLEWVAMVWANRQYTY- YSDSVRGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTAMWKDNNIEYDMVIVYSGNGTVVTVS- SGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFWVSTNHVVWYLQLPGQSPKT- LVYSTFLRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSDNSTFSNGTKLEIK
2acp28	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGQFLMDWVRQSPEKGLEWVALATTNKHNKSTS- YSDSVAGRFTISLDDSKSSVYLQMNNLRVEDTGIYYCMAAEKSFVFYTGTSTVVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALNKTSHFVLWYLQLPGQSPKVLVYYTYKR- VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSEEPASFSNGTKLEIK
2acp29	SEVKLDETGGGLVQPGGAMKLSCVASGFDFGLYVMLWVRQSPEKGLEWVAAADSSHNRTFYS- DSVAGRFTISADSSKSSVYLQMNNLRVEDTGIYYCMALSLIEVEYSGNGTLVTVSSGGGGSGGGG- SGGGGSDVVMTQTPLSLPVSLGDQASISCRSAYVSDAYNIFWFLQLPGQSPKVLVYMVRLRVSG- VPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQLSRWPLSFSTGTKLEIK
2acp30	SEVKLDETGGGLVQPGGAMKLSCVVSGFHFGWTIMVWVRQSPEKGLEWVAAIDASNKRTWYS- DSVRGRFTISLDSSKSSVYLQMNNLRVEDTGIYYCMQAVRSRASAIYTGNGTVVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKYSLVNDDGATNIVWFLQLPGQSPKVLVSS- VNWRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYYCAAFTSYPFTFITGTKLEIK
2ins01	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFSQYIMFWVRQSPEKGLEWVAAFSGTNTSTYYA- DSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMGSGSGGDSSTYSGDGTSVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSQQSVSQLQWYLQKPGQSPKVLISEVSNRVS- GVPDRFSGTGSGTDYTLKINRVEAEDLGVYFCAQEATVPWTFGGGTKLEIK
2ins02	SEVKLDETGGGLVQPGGAMKLSCVASGFTFSKYYMGWVRQSPEKGLEWVAIFQFANTTYYSD- SVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCWGSSDNASMAFFYNGSGTSVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSKSLVTSSGNTYLVWFLQKPGQSPKVLIYD- VSKRVSGVPDRFSGTGSGTDFTLKINRVEAEDLGVYFCAQFTNSTTFFTFGGGTKLEIK
2ins03	SEVKLDETGGGLVQPGGAMKLSCVTSGFRFATYMMMWVRQSPEKGLEWVANFDSDNYTNYA- DSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCSGPKTNSTSSTYLGQGTSVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSSSLSKYLEWYLQKPGQSPKVLIYNTSNRVS- GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCGQWSYWPSTFGGGTKLEIK
2ins04	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFSYFVMVWVRQSPEKGLEWVAAFYSGNSTYYAD- SVKGRFTISADASKSSVYLQMNNLRVEDTGIYYCIGTAGTTYLGQGTSVTVSSGGGGSGGGGSG- GGGSDVVMTQTPLSLPVSLGDQASISCRSSQSLYESLYWYLQKPGQSPKVLIYKVSNRVSGVPDR- FYGSGSGTDYTLKINRVEAEDLGVYFCAAYSSESRAITFGGGTKLEIK
2ins05	SEVKLDETGGGLVQPGGAMKLSCVASGFTFGHYEMVWVRQSPEKGLEWVANFSQSNNTNYAD- SVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAAGSDGAWWAVSTMIYTGQGTSVTVSSG- GGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSKSLVWSSGWTLLSWYLQKPGQSP- KVLIYKVSNRVSGVPDRFSGTGSGTDFTLKINRVEAEDLGVYFCSQSSYVPWTFGGGTKLEIK
2ins06	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFGKTVMGWVRQSPEKGLEWVAAYYWSNITFYA- DSVKGRFTISRDESKSSVYLQMNNLRVEDTGIYYCGGSGGFGEFVFRYLGNGTSVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSQSLVDSDGNTRLAWYLQKPGQSPKVLIY- NVSNRVSGVPDRFSGTGSGTDFTLKINRVEAEDLGVYFCAQSTDAPWTSGGGTKLEIK
2ins07	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFGSYKMSWVRQSPEKGLEWVAAFSPTVNVTSYA- DSVKGRFTISSDGSKSSVYLQMNNLRVEDTGIYYCMAPNPVLMSSLGQGTSVTVSSGGGGSGGG- GSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSQVVSNYLGWYLQKPGQSPKVLIYKVSNRVSG- VPDRFYGTGSGTDFTLKINRVEAEDLGVYFCMADFADSYSFGGGTKLEIK
2ins08	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFGKYTMSWVRQSPEKGLEWVAAFYWVNLTVYA- DSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCVYSSSWSSSMTSNWGQGTSVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSLPLVSRDGSPNLVWFLQKPGQSPKMIIYN- VSNRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAAYSSDLDAFWFFGGTKLEIK
2ins09	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFGWYLMSWIRQSPEKGLEWVAIFAYNNQTRYSDS- VKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCGGLFWIYLGQGASVTVSSGGGGSGGGGSGG- GGSDVVMTQTPLSLPVSLGDQASISCRSSMNVTDDLYWMLQKPGQSPKVLIYKVSQRVSGVPDR- FTGSGQGTDFTLKINRVEAEDLGVYFCAQSAWWPSTFGGGTKLEIK
2ins10	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFGDYSMMWIRQSPEKGLEWVALFSDKNTNYNTY- YASSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMGDYSIYTGQGASVTVSSGGGGSGGGG- SGGGGSDVVMTQTPLSLPVSLGDQASISCRSDEITMGSAVVFWFLQKPGQSPKALIYTTSNRVSG- VPDRFSGSGSGTDYTLKINRVEAEDLGVYFCASWSYASNISFGGGTKLEIK

2ins11	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFGHYSMLWVRQSPEKGLEWVAMFAPADYSTYYA- DDVKGRFTISSDPSKSSVYLQMNNLRVEDTGIYYCWASSYTPNFSSSTADLYSGQGTSVTVSSGG- GGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCSMASDIQPYSALTWYLQKPGQSPKVLIY- KVSLRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCASWNETFYYRTFGGGTKLEIK
2ins12	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFGNYIMNWVRQMPEKGLEWVASFTPKPDNNSTR- YSDSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTASSSTSSSLMTYWGQGTSVTVSSGGG- GSGGGGSGGGGSDVVMTQTPLSLPVSLGDQAQISCRSSSPLSNFLAWYLQKPGQSPKVLIYQVS- NRVSGVPDRFSGSGNGTDFTLKINRVEAEDLGVYFCSQFSVNPPTFGGGTKLEIK
2ins13	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFSKYDMNWMRQSPEKGLEWVASFYDDNDTSYAD- SVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTTSSQDETSSSDFSWYLGQGTSVTVSSGGGG- SGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSHSLTSSSGHTTLAWFLQKPGQSPKVLIY- KVSNRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQSWFSPFTFGPGTKLEIK
2ins14	SEVKLDETGGGLVQPGGAMKLSCVTLGFNFMEYAMNWVRQSPEKGLEWVAAFYLNNSAQYA- DSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTADARQNQQGNAAVKSGQGTSVTVSSGG- GGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSQSAVLRNGLTFLFWYLQKPGQSPKV- LIYKVSNRVSGVPDRFSGSGFGTDFTLKINRVEAEDLGVYFCAADSWMANENRFGGGTKLEIK
2ins15	SEVKLDETGGGLVQPGGAMKLSCTASGLDSSRAFYMWVRQSPEKGLEWVATYTSGNLAYYAD- SVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAIRQSKLNSAMAVMGQGTSVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCKSKKTSFKDTFWYLQKPGQSPKVLIYKVSNR- VSGVPDRFYGTGEGREYKLKINRVEAEDLGVYFCSQFFSYEVSFGYGTKLEIK
2ins16	SEVKLDETGGGLVQPGGAMKLSCRAEGVDDSEMTFEWVRQSPEKGLEWVAAFTDNNSAAYA- DSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCEAASENISNTAFIYIGDGTSVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCFAKFMVYDYWKNNSHVAWYLQKPGQSPKVLI- YKVSNRVSGVPDRFYGTGSGRFFRLKINRVEAEDLGVYFCAQRASIPWAATAGGGTKLEIK
2ins17	SEVKLDETGGGLVQPGGAMKLSCVTTGFDSSKYEMNWVRQSPEKGLEWVAATSPLNYYYSYA- DSVKGRFTISKDNSKSSVYLQMNNLRVEDTGIYYCMASSYGSTAWTWTGNGTSVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALTEKAGNRTVTWYLQKPGQSPKVLIYMV- SNRVSGVPDRFSGSGSGSDYTLKINRVEAEDLGVYFCATYSSDNFKQFFGGGTK
2ins18	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFSKAVMNWVRQSPEKGLEWVAAFVPVKNMTFY- ADSVKGRFTISADTSKSSVYLQMNNLRVEDTGIYYCMGAVWTDNDELDYWVMYTGQGTSVTV- SSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCYAAVDTSRYVAWYLQKPGQSPKV- LIYKVSNRVSGVPDRFSGSGQGRVYTLKINRVEAEDLGVYFCAQYAFWPWSWGGGTKLEIK
2ins19	SEVKLDETGGGLVQPGGAMKLSCVTTGFNFGKSYMNWIRQSPEKGLEWVASFWTLNWTYYA- DSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCAASESASTSTMWIGDGTSVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCKSSSSLVNEKGEIYTSWYLQKPGQSPKILIYKVS- NRVSGVPDRFSGSGSGDTFTLKINRVEAEDLGVYFCAQDSEDPSTWGGGTKLEIK
2ins20	SEVKLDETGGGLVQPGGAMKLSCVTTGFTFGVYFMNWVRQSPEKGLEWVASFTSNPNNNLTY- YSDSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAAFSQRSDNWYVGDGTSVTVSSGGG- GSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCKANDLVWTKGFASLSWYLQKPGQSPKVL- IYQVSSRTSGVPDRFSGSGSGLVFTLKINRVEAEDLGVYFCATSYKYPATFGGGTKLEIK
2ins21	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKFFMNWVRQSPEKGLEWVANVWADNRQTYY- SDSVAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMSALAQKAWWTGDATVVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALYKVAWIVWYLQLPGQSPKVLVYFIYIRVS- GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFFYWPATFSTGTKLEIK
2ins22	SEVKLDETGGGLVQPGGAMKLSCVQSGFDFGQATMNWVRQSPEKGLEWVAAMSSDDVSTSYA- ASVRGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCAASDKDAGWKYMGDATVVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAWWRAWNQVIWYLQLPGQSPRALIVNTNDR- VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQYAASATFSTGTKLEIK
2ins23	SEVKLDETGGGLVQPGGAMKLSCVTSGFVFGLTVMNWVRQSPEKGLEWVAAVITWTNDTMY- SDSVAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAFTTSLTSRSQAFWGQSTLVTVSSGGG- GSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCAALRQAAQAYIYWYLQLPGQSPKVLVFFT- RYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCQYFVSAPALWAEGTKLEIK
2ins24	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGRTNMNWVRQSPEKGLEWVASARDSSNNKDTR- YSDSVRGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCWAAARAFASSGQATVVTVSSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALYAVNNRVVWYLQLPGQSPKVLIYNVSNRVS- GVPDRFSGSGTGTDFTLKINRVEAEDLGVYFCAQDSSEPATFSTGTKLEIK
2ins25	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGTTIMNWVRQSPEKGLEWVAAIKMSALNRQTM- YSDSVAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCWAAARANASSGTSTVVTVSSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFVVRDYVVWYLQLPGQSPKALIFRIRYRVS- GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCSQFAIWPASFSTGTKLEIK
2ins26	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGVDAMNWVRQSPEKGLEWVTRVDSWDNRTEY- SDSVAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCYSDSNWWTGSATVVTVSSGGGGSGGGG- SGGGGSDVVMTQTPLSLPVSLGDQASISCRALYSANNSVTWYLQLPGQSPKVLIWKVSNRVSGV- PDRFSGSGSGQDFTLKINRVEAEDLGVYFCAQDSEKSSYSNGTKLEIK
2ins27	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKSNMNWVRQSPEKGLEWVAFASIFNYATYYA- ASVRGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCWASNNVDYAMMLTGNSTVVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFYMNTFAVFWYLQLPGQSPKVLFSKVF- NRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAVMSNSSSWSTGTKLEIK
2ins28	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYAMVWVRQSPEKGLEWVVSAVTWSNQTSY- SDSVAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCWAQYAGKATTVAGSATVVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFVLTDQFNNKTWVTWYLQLPGQSPKV- LIYKVSNRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAATSWYLSTSYSTGTKLEIK
2ins29	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYFMMWVRQSPEKGLEWVAYTDATNTSTRY- SDSVAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCLAFAADSEWKTKKTGNATVVTVSSGGG- GSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRARYATADFVYWYLQLPGQSPKVLIYKVS- NRVSGVPDRFSGSGRGTDFTLKINRVEAEDLGVYFCTTMTANYYSFSTGTKLEIK

2ins30	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGQYTMNWVRQSPEKGLEWVAAASDNSKMTFY- SDSVRGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCAAALYNWIWYMGQATVVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYSLENAGTAYVLWYLQLPGQSPTLLVYN- TNQRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQASTYPITYSTGTKLEIK
3acp01	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGHYIMIWVRQSPEKGLEWVAAVKSHRNLTWYS- DSVAGRFTISIDISKSSVYLQMNNLRVEDTGIYYCMAAFARALWYIGTGTVVTVSSGGGGSGGG- GSGGGGSDVVMTQTPLSLPVSLGDQASISCAALYDALQFVIWFLQLPGQSPKVLVAWNRWRVS- GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCLQTTEFPWTFSTGTKLEIK
3acp02	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFSKFVMLWLRQSPEKGLEWVAAVITNFNWTYYS- DSVRGRFTISYDISKSSVYLQMNNLRVEDTGIYYCVAAQHWTLWYIGDATVVTVSSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYVLINNDGEIHIVWYLQLPGQSPKVLVYNV- DQRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQVTIWPFSYSTGTKLEIK
3acp03	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYAMNWVRQSPEKGLEWVASVLYNEQATYYS- DSVAGRFTISRDVSKSSVYLQMNNLRVEDTGIYYCLAALSWTSWYFGSATVVTVSSGGGGSGGG- GSGGGGSDVVMTQTPLSLPVSLGDQASISCAALYATWAHVVWYLQLPGQSPKVLVYNNDFRVS- GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSTYPFTFSTGTKLEIK
3acp04	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGRYVMVWLRQSPEKGLEWVAAAFTEAFNRETR- YSDSVAGRFTISIDHSKSSVYLQMNNLRVEDTGIYYCAAVYSKVVWYIGTATVVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCAALYSARWFVVWYLQLPGQSPKVLVYKTFYR- VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYYCAQFAIWPAHYSTGTKLEIK
3acp05	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGSYAMNWLRQSPEKGLEWVASIDSTSNQTWYA- ASVAGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCAAAWKAAYSVLYLGDGTLVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFDAWEAVTWYLQLPGQSPKVLVYRTRL- RVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCANLTYMPSTFSNGTKLEIK
3acp06	SEVKLDETGGGLVQPGGAMKLSCVTSGFVFGLTIMAWVRQSPEKGLEWTAAAIPAYNRTWYS- DSVRGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCMAWFLDSAIAIGYGTVVTVSSGGGGSGGG- GSGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYQLNNHVVWYLQLPGQSPKVLVHDTNSRVS- GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFHRWPASYSTGTKLEIK
3acp07	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGSVIMLWVRQSPEKGLEWVAAIKNERNNKETW- YSDSVRGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCFAMHSNSNSLWAIGTGTVVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALISAHRHIVWYLQLPGQSPKVLVFNVRYR- VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCMQTSYWPISYSTGTKLEIK
3acp08	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGSYVMLWLRQSPEKGLEWVAAFYSAANQTRYS- DSVRGRFTISADVSKSSVYLQMNNLRVEDTGIYYCLAVAMSFYSASALVYLGDGTVVTVSSGGG- GSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASKSTEKDGWAWVAWFLQLPGQSPKTL- VYNVYYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFFLWPFSFSTGTKLEIK
3acp09	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYHMFWLRQSPEKGLEWVAAIKSFDNATYYA- ASVAGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCLAAYQAFYSLIYIGNGTVVTVSSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCAAKYVVSAVIWYLQLPGQSPKTLVWNVWIRVS- GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCATFSTYWWFSNGTKLEIK
3acp10	SEVKLDETGGGLVQPGGAMKLSCVASGFDFGNYVMLWLRQSPEKGLEWVAAVSAWSIYTWYS- DSVRGRFTISADISKSSVYLQMNNLRVEDTGIYYCWAVDASTSSFNTWGSGTLVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALYSASQHVLWYLQLPGQSPKVLVYNVNQR- VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQTWYWPFTWSTGTKLEIK
3acp11	SEVKLDETGGGLVQPGGAMKLSCVTSGFSFGRAIMLWVRQSPEKGLEWVAVAFIAKRNYETRY- SDSVRGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCMALWDSHFAMRWLGNGTVVTVSSGGGG- SGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFKLVSSNGQIYVVWYLQLPGQSPKTLV- SHVWVRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQMSFWPFSFSTGTKLEIK
3acp12	SEVKLDETGGGLVQPGGAMKLSCVTSGFHFGFTIMAWLRQSPEKGLEWTAAVIPWLNLTYYSD- SVAGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCVFAAWRAASTILVLGNGTQVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRATNKLQEYVAWFLQLPGQSPKTLVYNVYFR- VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCSQFAQWPWSFSTGTKLEIK
3acp13	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYVMAWVRQSPEKGLEWTAAIDSTNNVTWY- AASVAGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCTAAQFFVVSIFGNGTVVTVSSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAQQSTRNLVVWTLQLPGQSPKALVWNVDYRV- SGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSNFPWSFSTGTKLEIK
3acp14	SEVKLDETGGGLVQPGGAMKLSCVTSGFVFQWTWMVWVRQSPEKGLEWVAAVAPDKKYTY- YSDSVRGRFTISVDASKSSVYLQMNNLRVEDTGIYYCAAAFWYFSRVIVYWGSGTQVTVSSGGG- GSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKQVVNYYVVWTLQLPGQSPKSLVVNS- SQRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCLQLTTFPWSFSTGTKLEIK
3acp15	SEVKLDETGGGLVQPGGAMKLSCVSSGFDFGRTIMAWVRQSPEKGLEWTAATWPWANLTYYYS- DSVRGRFTISWDKSKSSVYLQMNNLRVEDTGIYYCTAAFWMYSNLIVYFGNGTQVTVSSGGGG- SGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFVVWVSVVWLLQLPGQSPKALVVRVS- HRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCLQLWHEPWTWSNGTKLEIK
3acp16	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKTSMLWVRQSPEKGLEWTAAVHTSDNQTYYA- ASVAGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCLWFDYSNSVRIFGNGTVVTVSSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCAATTRVSEHVLWFLQLPGQSPKTLVFNTYYRVS- GVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCVQFSRWPLTFSTGTKLEIK
3acp17	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGTTTMVWVRQSPEKGLEWTASIDTTSNRTWYA- ASVAGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCLWAAWWYTYNVVVWGDGTVVTVSSGGG- GSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRANYSVSRSLDWFLQLPGQSPKTLVFSTYI- RVSGVPDRFSGSGSGTDYTLKINRVEAEDLGVYFCSAWATWSESMSFSTGTKLEIK
3acp18	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGTVWMVWVRQSPEKGLEWTALAIPRLNITLYS- DSVRGRFTISWDESKSSVYLQMNNLRVEDTGIYYCWALLIFAETSLLIGNGTVVTVSSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASRDTSQWVWWFLQLPGQSPKLLVAFVWRRL- SGVPDRFSGSGSGTDYTLKINRVEAEDLGVYFCAAWANSRSPWSWSTGTKLEIK

3acp19	SEVKLDETGGGLVQPGGAMKLSCVTSGFNFGNYTMSWVRQSPEKGLEWTASTTTSNQRTWY- AASVRGRFTISHDVSKSSVYLQMNNLRVEDTGIYYCAAAHSWFIWYWGNGTVVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALLNSHSVIWYLQLPGQSPKVLIFSTWWRVS- GVPDRFSGSGSGTDYTLKINRVEAEDLGVYFCAQVTHWPWSFSTGTKLEIK
3acp20	SEVKLDETGGGLVQPGGAMKLSCVASGFDFGQYTMMWVRQSPEKGLEWTAATASYNLQTWY- AASVRGRFTISFDVSKSSVYLQMNNLRVEDTGIYYCMAVSVRWMVWGQGTQVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALYSTDQFVVWYLQLPGQSPKVLVVKNNQR- VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFQAWPASFSTGTKLEIK
3acp21	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGTTVMMWLRQSPEKGLEWVAAAKTEANNYRT- YYSDSVAGRFTISMDRSKSSVYLQMNNLRVEDTGIYYCLAAWYNHVLYIGTGTQVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALQDLWRHVVWFLQLPGQSPKVLVYDNNL- RVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFAWYPASFSTGTKLEIK
3acp22	SEVKLDETGGGLVQPGGAMKLSCVTSGFVFGLTIMVWLRQSPEKGLEWTAAAIPADNRTYYSD- SVRGRFTISVNVSKSSVYLQMNNLRVEDTGIYYCMAWFLDSLIALGTGTVVTVSSGGGGSGGGG- SGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYQLNNQVVWFLQLPGQSPKVLVHDTNSRVSG- VPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFHRWPASFSTGTKLEIK
3acp23	SEVKLDETGGGLVQPGGAMKLSCVTSGFVFGLAVMVWLRQSPEKGLEFTAAVIPVENNTYYSD- SVRGRFTISFDISKSSVYLQMNNLRVEDTGIYYCLALFFKTFLFVGTGTVVTVSSGGGGSGGGGS- GGGGSDVVMTQTPLSLPVSLGDQASISCRALYKTNNFVVWFLQLPGQSPKVLVFDTNSRVSGVP- DRFSGSGSGTDYTLKINRVEAEDLGVYFCATFYLWPVHFSNGTKLEIK
3acp24	SEVKLDETGGGLVQPGGAMKLSCVTSGFVFGFMFWMLWLRQSPEKGLEWVALVIPWLNTTYYS- DSVAGRFTISVDISKSSVYLQMNNLRVEDTGIYYCWAFTLFAFNALYIGNGTVVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCAASRSTYISDIDRAIVWWMLQLPGQSPKILVLY- TNYRLSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCVQFAYFPFSFSTGTKLEIK
3acp25	SEVKLDETGGGLVQPGGAMKLSCVTSGFYFGWFTMSWLRQSPEKGLEWTASIMPTNNNTYYS- DSVAGRFTISIDISKSSVYLQMNNLRVEDTGIYYCAASSVKSKRLSQIGNGTQVTVSSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASADLKTSSGNIFIVWFLQLPGQSPKVLVEQNS- QRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCVQFAYWPWTFATGTKLEIK
3acp26	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGNYIMAWVRQSPEKGLEWTAAIDPSNNRTWYS- DSVAGRFTISVDKSKSSVYLQMNNLRVEDTGIYYCLIRAVTFWGNGTVVTVSSGGGGSGGGGSG- GGGSDVVMTQTPLSLPVSLGDQASISCRANKNVYRYVYWFLQLPGQSPKLLVYWTSARVSGVP- DRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFAYWPWSFSTGTKLEIK
3acp27	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGQTWMSWVRQSPEKGLEWVATATTASHNFTT- YYSDSVAGRFTISMDISKSSVYLQMNNLRVEDTGIYYCAAMNHKNNLVFFWGDGTVVTVSSGG- GGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKQVVYKYSVAWFLQLPGQSPKALVV- TTYIRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCLQVSIWPLTFSNGTKLEIK
3acp28	SEVKLDETGGGLVQPGGAMKLSCVASGFDFGLAIMAWVRQSPEKGLEWTAAIVPAKNLTYYSD- SVAGRFTISADISKSSVYLQMNNLRVEDTGIYYCVVFLFNSVHFWGDGTVVTVSSGGGGSGGG- SGGGGSDVVMTQTPLSLPVSLGDQASISCRALYSVNNHVLWLLQLPGQSPKVLVSDTNSRVSGV- PDRFSGSGSGTDFTLKINRVEAEDLGVYFCVQLYQYPLHFSTGTKLEIK
3acp29	SEVKLDETGGGLVQPGGAMKLSCVASGFHFGWTIMLWVRQSPEKGLEWVASVRPFQQQTYYS- DSVAGRFTISMDISKSSVYLQMNNLRVEDTGIYYCVALFYYYSNVVVAWGSGTQVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSLLDVSNRVVWLLQLPGQSPKVLVVNTRHR- VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFYLFPYSFATGTKLEIK
3acp30	SEVKLDETGGGLVQPGGAMKLSCVASGFNFGYTWMAWVRQSPEKGLEWVATAWDNNKYTY- YSDSVAGRFTISRDVSKSSVYLQMNNLRVEDTGIYYCTALFWYYSRTITAWGNGTQVTVSSGGG- GSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKVTVSNHVIWVLQLPGQSPKVLVVNT- DYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCVQLTSEPWTFSTGTKLEIK
3acp31	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGLMTMMWLRQSPEKGLEWVASVNAETNQTRY- SDSVAGRFTISMDISKSSVYLQMNNLRVEDTGIYYCLAAAWATVIMFIGHGTQVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALQVVSIKVIWMLQLPGQSPKVLVYMVYYR- VSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCLQVAFHPWSFANGTKLEIK
3acp32	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGNFVMVWVRQSPEKGLEWTAAVDPTNNRTFYS- DSVRGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCLAKQVALIGHGTVVTVSSGGGGSGGGGSG- GGGSDVVMTQTPLSLPVSLGDQASISCRALLDVNDVVVWFLQLPGQSPKMLVYWARARLSGVP- DRFSGSGSGTDFTLKINRVEAEDLGVYFCSQVRQFPYHFSTGTKLEIK
3acp33	SEVKLDETGGGLVQPGGAMKLSCVASGFHFGWTIMAWVRQSPEKGLEWTAAYTNDSNNNLTY- YSDSVRGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCLAAFAFYSNVVVYWGQGTQVTVSSGG- GGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKYVTSVDVVWLLQLPGQSPKVLIVRV- RHRVSGVPDRFSGSGSGTDYTLKINRVEAEDLGVYFCAQFYYWPWSFSTGTKLEIK
3acp34	SEVKLDETGGGLVQPGGAMKLSCVTSGFVFGLTIMAWVRQSPEKGLEWVAAILALRNVTWYA- ASVRGRFTISMDISKSSVYLQMNNLRVEDTGIYYCVAFFMNAVVHWGDGTVVTVSSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALYAVNNQVMWLLQLPGQSPKVLVRDANSRV- SGVPDRFSGSGSGTDYTLKINRVEAEDLGVYFCMQSSTFPFSFSTGTKLEIK
3acp35	SEVKLDETGGGLVQPGGAMKLSCVTSGFVFGLSVMVWLRQSPEKGLEWVAVIAAYQASTHYA- ASVRGRFTISMDYSKSSVYLQMNNLRVEDTGIYYCMGSNSNRNVLEVIGTGTQVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALYSVKDEVVWFLQLPGQSPKTLVFRTRIRV- SGVPDRFSGSGSGTDLTLKINRVEAEDLGVYFCLQTSIWPWSWSTGTKLEIK
3ins01	SEVKLDETGGGLVQPGGAMKLSCVTSGFSFGTTIMNWVRQSPEKGLEWVASITSAEYNFATYY- SDSVRGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCMASYNAYTAIGNGTVVTVSSGGGGSGGG- GSGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYVTNQTVYWYLQLPGQSPKVLVSRTSLRVSG- VPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQNSQWPQSYSTGTKLEIKGGGSEQKLISEEDL
3ins02	SEVKLDETGGGLVQPGGAMKLSCVTSGFNFGLYYMNWVRQSPEKGLEWVAAKTPDNNDSYYS- DSVRGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCMAASSKSKTVSLWGSGTVVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASRQLYIQNLNSSYTVWYLQLPGQSPKVLVM- NVNFRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFAWWPWTYSTGTKLEIKGGGSE- QKLISEEDL

3ins03	SEVKLDETGGGLVQPGGAMKLSCVTSGFNFGATIMNWVRQSPEKGLEWVAATDPENNRTWYSDSVAGRFTISYDISKSSVYLQMNNLRVEDTGIYYCMAAAYLSAAFIGNGTLVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYSLKVDQDNTAVISWYLQLPGQSPKTLVFSTDKRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFAIWPATFSTGTKLEIKGGGSEQKLISEEDL
3ins04	SEVKLDETGGGLVQPGGAMKLSCVTSGFNFGKTIMNWVRQSPEKGLEWVAAATNADNNKLTYYSDSVRGRFTISIDESKSSVYLQMNNLRVEDTGIYYCMAAWQKFAAYIGTGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASRRLVSSSGNIYIAWYLQLPGQSPKLLVFNVRYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFAYWPWTWSTGTKLEIKGGGSEQKLISEEDL
3ins05	SEVKLDETGGGLVQPGGAMKLSCVTSGFFFGFYDMNWVRQSPEKGLEWVASVSKTNDYTYYSDSVAGRFTISIDISKSSVYLQMNNLRVEDTGIYYCMASSWAKTTTVIGSGTLVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALLSLKDSNNNEIWIVWYLQLPGQSPKVLVSNNNWRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFAWWPFSYSTGTKLEIKGGGSEQKLISEEDL
3ins06	SEVKLDETGGGLVQPGGAMKLSCVTSGFNFGNAWMNWVRQSPEKGLEWVAAVRPEQNLTWYSDSVRGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCAAAMRKEAATITLGTGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYSLKDDNNNKIYVTWYLQLPGQSPSTLVYNVRYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCSQFWAWPATYAEGTKLEIKGGGSEQKLISEEDL
3ins07	SEVKLDETGGGLVQPGGAMKLSCVTSGFVFGWSIMNWVRQSPEKGLEWVAAADSDNSRTWYSDSVAGRFTISEDRSKSSVYLQMNNLRVEDTGIYYCWSSSNDRTSASWLGSGTLVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASRALVSRYGTSNVSWYLQLPGQSPKTLVTNVSYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFTYYPFTFSTGTKLEIKGGGSEQKLISEEDL
3ins08	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGNFQMNWVRQSPEKGLEWVAAASDNNASTYYSDSVAGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCAAISWGLNSYSYWGNGTLVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSTDLKSSEGKINVLWYLQLPGQSPKTLVFFVSARVTGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQYWREPYSFSNGTKLEIKGGGSEQKLISEEDL
3ins09	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGNYIMNWVRQSPEKGLEWVAAIKANENKTWYSDSVAGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCAAMSWGSSSLSWWGNGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCAALYTAAWWYIMWYLQLPGQSPKELVFATSARTSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQSSQWPSTFSTGTKLEIKGGGSEQKLISEEDL
3ins10	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGQAIMNWVRQSPEKGLEWVASVKTAKYNYETRYSDSVRGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCAASFWNSSWLIGTGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASRDLKSSNGRYWVVWYLQLPGQSPKTLVFWVSARVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFTYKATYSTGTKLEIKGGGSEQKLISEEDL
3ins11	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKTIMNWVRQSPEKGLEWVAAVTSAAYNWLTYYSDSVAGRFTISWDKSKSSVYLQMNNLRVEDTGIYYCWASARSFASWGSGTAVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSRSSSDNRVAWYLQLPGQSPKALVVNTYIRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSRWPISWSTGTKLEIKGGGSEQKLISEEDL
3ins12	SEVKLDETGGGLVQPGGAMKLSCVASGFDFGQTVMNWVRQSPEKGLEWVAAVDTTNNRTYYSDSVAGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCWASASLYRSTAAFGTGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSRSSSNNIVTWYLQLPGQSPKTLVWFTYWRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCSQNSQNPQTFSNGTKLEIKGGGSEQKLISEEDL
3ins13	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGRTIMNWVRQSPEKGLEWVAATSRDNAETRYSDSVAGRFTISQDRSKSSVYLQMNNLRVEDTGIYYCMYLAYEASLEQYRYVKYIGNATVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFVADKTFIVWYLQLPGQSPKTLVVWTYIRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQYSQWPATFSNGTKLEIKGGGSEQKLISEEDL
3ins14	SEVKLDETGGGLVQPGGAMKLSCVTSGFHFGAFIMAWVRQSPEKGLEWVAAAVPWNNLTYYSDSVAGRFTISFDISKSSVYLQMNNLRVEDTGIYYCMMFWEHVSVVLGTGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKYVISRDVMWYLQLPGQSPKTLVYNVSLRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQSYIWPITYSTGTKLEIKGGGSEQKLISEEDL
3ins15	SEVKLDETGGGLVQPGGAMKLSCVTSGFRFGLFWMSWVRQSPEKGLEWVATVWMYNWATYYSDSVRGRFTISRDVSKSSVYLQMNNLRVEDTGIYYCMAYSIYWYWYSVAVWGHATVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALYKTNYTVYWYLQLPGQSPKVLVFLVMFRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCASWATVSNSMSFSTGTKLEIKGGGSEQKLISEEDL
3ins16	SEVKLDETGGGLVQPGGAMKLSCVSSGFSFGTTWMNWVRQSPEKGLEWVASVWAEQFNKLTYYSDSVRGRFTISMDISKSSVYLQMNNLRVEDTGIYYCAASTAKASTYLGNATVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFDSSRFVAWYLQLPGQSPKTLVFSVYYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCATYASFPFSYSTGTKLEIKGGGSEQKLISEEDL
3ins17	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKFIMNWVRQSPEKGLEWVASYVPQYNLTYYSDSVRGRFTISADISKSSVYLSMNNLRVEDTGIYYCAAAAAYERALSYKIGNGTAVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALFDTSNYVAWYLQLPGQSPKALVVSTWIRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQSSSFPSTFSTGTKLEIKGGGSEQKLISEEDL
3ins18	SEVKLDETGGGLVQPGGAMKLSCVTSGFAFGYTWMNWVRQSPEKGLEWVAAVWTWDNATYYSDSVAGRFTISVDFSKSSVYLQMNNLRVEDTGIYYCMAMFATATNLSMAAIKYFGNATVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALYATSNYVIWYLQLPGQSPKTLVFLTNIRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQVSLFPFSYSTGTKLEIKGGGSEQKLISEEDL
3ins19	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKTIMAWVRQSPEKGLEWTAAVDPNNNRTWYSDSVRGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCTASTSSYSSSWGDGTAVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALLDVSSNVSWHLQLPGQSPKTLVSHAKTRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCSQSYFWPWSFSNGTKLEIKGGGSEQKLISEEDL
3ins20	SEVKLDETGGGLVQPGGAMKLSCVTSGFFFGKTIMSWVRQSPEKGLEWTASIKPEENQTWYSDSVRGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCSFFAYDNSSTFWGNATQVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKYATNNHVTWYLQLPGQSPKALVTDTDRRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSHWPASFSTGTKLEIKGGGSEQKLISEEDL

3ins21	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGSHVMTWVRQSPEKGLEWTAAVTSQLWNRLTYYSDSVRGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCAASSWYASAYLGNGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYVVSISVAWYLQLPGQSPKTLVVRVYIRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFTTFPWTFSTGTKLEIKGGGSEQKLISEEDL
3ins22	SEVKLDETGGGLVQPGGAMKLSCVTSGFNFGEYIMMWVRQSPEKGLEWVAAITSTNVTFYSDSVAGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCASSSRSRAVMYFGNGTQVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCAASRKLWIWAYLAAFVVWYLQLPGQSPKALVVLVNVRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFAYWPASYATGTKLEIKGGGSEQKLISEEDL
3ins23	SEVKLDETGGGLVQPGGAMKLSCVTSGFLFGLYVMAWVRQSPEKGLEWVAAIAWDRSSTYYAASVRGRFTISEDRSKSSVYLQMNNLRVEDTGIYYCWAWLLWWNSAMWWGSGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASHSLVDDTRNKAFIVWYLQLPGQSPKVLVWATRYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQYARMPATFSTGTKLEIKGGGSEQKLISEEDL
3ins24	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYIMSWVRQSPEKGLEWTASVTDQLSNWKTYYSDSVRGRFTISWDKSKSSVYLQMNNLRVEDTGIYYCTGARAKYTMYWGDGTQVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCAAAYSTWVWQYMTAFVVWTLQLPGQSPKTLVHTVNDRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQMTTFPWTFSTGTKLEIKGGGSEQKLISEEDL
3ins25	SEVKLDETGGGLVQPGGAMKLSCVTSGFYFGRFVMMWVRQSPEKGLEWVASAKSNNKSTYYAASVAGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCVAVYWALSLWGSGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCAASVDLKWATGEIWVSWYLQLPGQSPKELVYNVRLRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAAVAVISNSVSFATGTKLEIKGGGSEQKLISEEDL
3ins26	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGRYWMSWIRQSPEKGLEWTATVDSTSNVTWYAASVAGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCAAEQSYNSSTLAIGDGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALLSASNSLAWYLQLPGQSPKVLVYDVNRRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCATYNYWPHSFSTGTKLEIK
3ins27	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYWMTWVRQSPEKGLEWTAMVDSTSNSTAYAASVAGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCMAAFAYTSAAFGTGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSYDVDSVAWYLQLPGQSPKALVLDAKNRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFKYFPFSWATGTKLEIK
3ins28	SEVKLDETGGGLVQPGGAMKLSCVTSGFIFGLSVMNWVRQSPEKGLEWVAAIEPVLNLTYYSDSVAGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCAATVEQYFSSYSSAAIIGQGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALFSSSAAVAWYLQLPGQSPKTLVFKVSIRVSGVPDRFSGSGDGTDFTLKINRVEAEDLGVYFCASFLRFPFTFSTGTKLEIK
3ins29	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKTQMAWIRQSPEKGLEATAIIDSTNNRTWYAASVAGRFTISVDISKSSVYLQMNNLRVEDTGIYYCSFFAYDNSSTFWGNGTQVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYATNDNVTWYLQLPGQSPKTLVTNTNQRVSGVPDRFSGSGSGTDYTLKINRVEAEDLGVYFCAAYSDMAAFFSTGTKLEIK
3ins30	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYVMSWVRQSPEKGLEWTASVIPQRNQTYYSDSVRGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCAAALAHFMWYIGNGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCAASANTDYVTWYLQLPGQSPKVLVWWIYYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQTTKYPWTFSTGTKLEIK
3ins31	SEVKLDETGGGLVQPGGAMKLSCVASGFTFGLASMNWVRQSPEKGLEWVASTTPFLNLTWYAASVAGRFTISVDLSKSSVYLQMNNLRVEDTGIYYCAASSSWTSSNASVAAWSTGTQVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALFSASMVVFWYLQLPGQSPKLLVFWVRYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQASRFPVTFSTGTKLEIK
3ins32	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGLMTMAWVRQSPEKGLEFVAAAWPARNLTYYSDSVRGRFTISMDISKSSVYLQMNNLRVEDTGIYYCTYSQFIADSWHDSTIVYWGNGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASHDSKRDVIWTLQLPGQSPKTLVYNVYYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCASTLSHPFTFSTGTKLEIK
3ins33	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGMFIMNWVRQSPEKGLEWVAAYVTVANLTFYSDSVRGRFTISADISKSSVYLQMNNLRVEDTGIYYCAASKSWWSDSWYIGNATVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASKSVNSVAWYLQLPGQSPKVLVYEVYKRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAAWAVDSLSASFSTGTKLEIK
4ins01	SEVKLDETGGGLVQPGGAMKLSCVTSGFNFGKSTMTWVRQSPEKGLEWTAAVSSQDDNWKTYYSDSVAGRFTISWDKSKSSVYLQMNNLRVEDTGIYYCASSTSSYSSSWGDGTAVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFEVKTSVAWTLQLPGQSPKTLVSHVSWRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCSQSYQFPWTFSTGTKLEIK
4ins02	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKFVMMWVRQSPEKGLEWTAAVTNDNNNWATYYSDSVAGRFTISWDKSKSSVYLQMNNLRVEDTGIYYCLASSHNTTIYIGNGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALLATNDNVTWWLQLPGQSPKALVVWTSIRVSGVPDRFSGSGSGTDYTLKINRVEAEDLGVYFCAQFTYWPASFSTGTKLEIK
4ins03	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGSSIMAWVRQSPEKGLEWVAAVMSDLNLTYYSDSVAGRFTISMDISKSSVYLQMNNLRVEDTGIYYCIAWAYMNVSVSWGNGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSTHVLNQHVAWTLQLPGQSPKTLVYMNYYRTSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCSQYEIMPASYSTGTKLEIK
4ins04	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGQYAMAWIRQSPEKGLEWVATITTLQYSTVYAASVKGRFTISMDISKSSVYLQMNNLRVEDTGIYYCMATSDLFNFSLFWWGHGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFVTSLHVVWFLQLPGQSPKTLVYDTYYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQTFIVPWSFSTGTKLEIK
4ins05	SEVKLDETGGGLVQPGGAMKLSCVTSGFWFGSTKMAWTRQSPEKGLEWVASADPNSNSTYYSDSVRGRFTISWDESKSSVYLQMNNLRVEDTGIYYCMAFSDNAALTIGTGTVVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSHDSKRIVMWWLQLPGQSPKTLVYEVRKRTSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFTTYPWTYATGTKLEIK
4ins06	SEVKLDETGGGLVQPGGAMKLSCVSSGFDFGYYWASWVRQSPEKGLEWVASTHSSNMWTLYAASVRGRFTISADKSKSSVYLQMNNLRVEDTGIYYCSFHSKTSMTTTVWGNGTQVTVSSGGGGSGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALYDTWWIVTWYLQLPGQSPKTLVYNTSIRVSGVPDRFSGSGSGTDYTLKINRVEAEDLGVYFCAQYSRWPASFSTGTKLEIK

4ins07	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKYFMTWIRQSPEKGLEWVATTASYNMWTHY- AASVRGRFTISADVSKSSVYLQMNNLRVEDTGIYYCAADMSVYNDSLAIGTGTVVTVSSGGGGS- GGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRALYSADRTVSWMLQLPGQSPKVLVYHVRE- RVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCSAVRGDKISFGGGTKLEIK
4ins08	SEVKLDETGGGLVQPGGAMKLSCVASGFIFGLYIMAWVRQSPEKGLEWTAAVSSTSNLTWYAA- SVAGRFTISADVSKSSVYLQMNNLRVEDTGIYYCTAWYYWVATIWGDGTVVTVSSGGGGSGG- GGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYVLRFKTSDLSIIAWTLQLPGQSPKTLVYNT- RARVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQYSIMPATYSTGTKLEIK
4ins09	SEVKLDETGGGLVQPGGAMKLSCVTSGFSFGNSWMAWVRQSPEKGLEWVASARNASHNHKTF- YSDSVRGRFTISMDESKSSVYLQMNNLRVEDTGIYYCAAAVAQAITRIGTGTVVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCAASMDLWITSHNSAYVIWYLQLPGQSPKVLVF- NTYYRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSYYPWTYSTGTKLEIK
4ins10	SEVKLDETGGGLVQPGGAMKLSCVTSGFHFGAFIMSWIRQSPEKGLEWVASASRYSVYTWYSD- SVRGRFTISQDISKSSVYLQMNNLRVEDTGIYYCVASWDNTIMLIGNGTVVTVSSGGGGSGGGG- SGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYVLYLSDGSEHLTWWLQLPGQSPKLLVYNTRL- RVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSYWPWTWSTGTKLEIK
4ins11	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGKHIMLWTRQSPEKGLEWTAAVASALWNKQT- WYSDSVAGRFTISIDISKSSVYLQMNNLRVEDTGIYYCWAFFYTADNLFFIGNGTVVTVSSGGGG- SGGGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRASRMLISSDGFVIVWWYLQLPGQSPKVLV- AYVRWRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQFSYWPATYSTGTKLEIK
4ins12	SEVKLDETGGGLVQPGGAMKLSCVTSGFDFGSTIMLWIRQSPEKGLEWTAAVASMRNQTWYA- ASVAGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCWSSIATYTGTGTLVTVSSGGGGSGGGGSG- GGGSDVVMTQTPLSLPVSLGDQASISCRAKFKTRFVVWYLQLPGQSPKVLVYYIYYRVSGVPDR- FSGSGSGTDYTLKINRVEAEDLGVYFCAQYSIWPATFSTGTKLEIK
4ins13	SEVKLDETGGGLVQPGGAMKLSCVTSGFHFGAFQMSWTRQSPEKGLEWTATVDPTNNRTWY- SDSVRGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCTAQWRSLSFYGSGTVVTVSSGGGGSGGG- GSGGGGSDVVMTQTPLSLPVSLGDQASISCRAAYVTSHDVTWHLQLPGQSPKTLVYSTRLRDSG- VPDRFSGSGSGTDFTLKINRVEAEDLGVYFCSQSYISPTTYSTGTKLEIK
4ins14	SEVKLDETGGGLVQPGGAMKLSCVTSGFNFGAYIMAWVRQSPEKGLEWTAAVDPSTNTTWYS- DSVRGRFTISVDVSKSSVYLQMNNLRVEDTGIYYCTAHLIRYDLYIGTGTVVTVSSGGGGSGGG- GSGGGGSDVVMTQTPLSLPVSLGDQASISCRASHDVLAVAWHLQLPGQSPKTLVYWIYWRVSG- VPDRFSGSGSGTDFTLKINRVEAEDLGVYFCAQYSRWPATYSTGTKLEIK
4ins15	SEVKLDETGGGLVQPGGAMKLSCVTSGFWFGKSIMQWVRQSPEKGLEWVATARSEDWNKLT- FYSDSVRGRFTISIDVSKSSVYLQMNNLRVEDTGIYYCTASSSSYASMWGDGTQVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRAKFDTSSEVSWTLQLPGQSPKTLVMVTLTRV- SGVPDRFSGSGSGTDYTLKINRVEAEDLGVYFCSQNTTDPATFSTGTKLEIK
4M5.3	SEVKLDETGGGLVQPGGAMKLSCVTSGFTFGHYWMNWVRQSPEKGLEWVAQFRNKPYNYET- YYSDSVKGRFTISRDDSKSSVYLQMNNLRVEDTGIYYCTGASYGMEYLGQGTSVTVSSGGGGSG- GGGSGGGGSDVVMTQTPLSLPVSLGDQASISCRSSQSLVHSNGNTYLRWYLQKPGQSPKVLIYK- VSNRVSGVPDRFSGSGSGTDFTLKINRVEAEDLGVYFCSQSTHVPWTFGGGTKLEIK
5acp01	SQVQLQETGGGLVQPGASMKLSCKASGFNFSNYGVSWVRQRPGQGLEWVAWVSPDGGTTYY- NDKVKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCAVGNSFNASSTFWGQGTTVTVSSGGGG- SGGGGSGGGGSDIVMTQTPLTLPVSIGDQASISCRASQSTLHSSGDHNVLWFLQKPGQSPKLLVF- RMSRRFSGVPDRFSGSGSGTDFTLKISRVEAEDFGVYYCSVGWRWPWTFGGGTKLEIK
5acp02	SQVQLQETGGGLVQPGASMKLSCVASGFDFSDWQMLWVRWSPGRGMEWVAAIGHKGNNHAT- YYAPSVKGRFTISRDDSQNMVYLQMNNLRAEDTGIYYCAAFWGQFAHYWGPGTTVTVSSGGG- GSGGGGSGGGGSNIVMTQTPLTLPASPGDRITISCRASQQAYKNVEWFRQKPGQAPKLLVYWA- NQLASGVPSRFSGSGSGTDYTLTVSNVQPEDFGTYYCSQHTRHPHSFGGGTKVEIK
5acp03	SQVQLQETGGGLVQPGASMKLSCVTSGFTFSNFWMSWVRQPPGKGMEWTASIANKPNNHAT- WYAPSVKGRFTISRDDSQNIVYLQMNNLRAEDTGIYYCAGGWGPNTGAVYWGQGTTVTVSSG- GGGSGGGGSGGGGSSIVMTQTPLTLPASPGDRVTISCKASQKVSTHVVWFVQKPGQSPKLLVY- NGSEQAPGVPSRFSGSGSGTDFTLTISSVQPEDFGTYFCMQGNTYPWTFGQGTKVEIK
5acp04	SQVQLQETGGGLVQPGGSMKISCVTSGSTFSDYTASWYRQPPGKGLEWTGHISNKSNNKATWY- APSVKGRFTISKDDSQNRLYLQMNNLRAEDTGIYYCALSGDNGYYWATDYWGQGTSVTVSSG- GGGSGGGGSGGGGSDIVMTQTPLTLPASPGDRITISCRASQDVGQHVVWYRQKPGQSPKLLVY- GGSNLYSGVPSRFSGSGSGTDFTLTVSNVQPEDFATYYCQQGWGYPFTTGGGTKVEIK
5acp05	SQVQLQETGGGLVQPGGSMKISCVASGFNFSNASMSWVRQPPGKGMEWVGSISNKSNGHATYY- APSVKGRFTISRDDSQNIVYLQMNNLRAEDTGIYYCASTSGSPGSAIIWWGQGTTVTVSSGGGG- SGGGGSGGGGSDIVMTQTPLTLPASPGDRITISCRASTDVGKSVAWFRQKPGQSPKLLVLSGSW- LYRGVPSRFSGSGSGTDFTLTVSNVQPEDFATYYCMQGYDTHSFGGGTKVEIK
5acp06	SGVQLQETGGGLVQPSQSMSLTCVTSGFNFSKTHAAWVRQPPGKGLEWTARIRNKSNGHATE- YAPSVKGRFTISRDDSQNRVFLQMNNLRAEDTGIFYCAISGWRHPGSPFHTWGQGTTVTVSSG- GGGSGGGGSGGGGSDVVLTQTPLTLPMSVGEKVTITCKSSQSLTNSNYETTWHAWYKQKPGQ- SPKLLVYWASWRESGVPRRFTGSGSGTDFTFTISSVQAEDIAVYFCAQYKNSTSFGGGTKVEIK
5acp07	SQVQLQETGGGLVQPGGSMKISCATSGFTFQHQRMFYLRQPPGKAMEFLGMIYNQSNGFTTYY- APSVKGRFTMSKDDSQDRLYLQMNNMRAEDTGTFYCATGNGYWWGQGTTVTVSSGGGGSG- GGGSGGGGSDIVMTQTPLTLPASPGERATITCRASHTVRSSQFHFFEQKPGQSPKLIIWGDEHLA- SGVPDRFSGSGSGTDYSMTIGSVEPEDFAVYYCMQGHEFPFTFGQGTKLEIK
5acp08	SQVQLQETGGGLVQPGASMKLSCKTSGFNFSDWGVLWVRQRPGQGMEWVAAISPGGTITFYN- DKVKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCAVRDRDWDSGITLFGQGTTVTVSSGGGG- SGGGGSGGGGSDIVLTQTPLTLPVSPGQRATISCRASQSTDNNGNSHVTWYEQKPGQPPKLKV- YNASNVESGVPARFSGSGSGTDFTLTIHPVEPEDFATYFCLQGNKVPYSWGGGTKLEIK
5acp09	SQVQLQETGGGLVQPGASMKLSCKASGFDFSSTGMEWVRQRPGQGLEWVAWISPGGGSHAYN- NKVKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCAAFDGNNWDYWGQGTTVTVSSGGGGSG- GGGSGGGGSNIVLTQTPLTLPVSPGQRATISCRAAQSTDTNGNSHMDWYEQKPGQPPKLKVYN- ASNVESGVPARFSGSGSGTDFTLTIHPVEPEDFATYFCSQNNTLPRTFGGGTKLEIK

5acp10	SEHQLQETGGGLVQPGASMKLSCVTSGFTFSHYWQSWTKQSPGRGMEYLATIRNKSHNYATY- YAPSVKGRFTISRDDSQNMLYLQMNNLRAEDTGIYYCATGWHAAYMGQGTLVTVSSGGGGSG- GGGSGGGGSDVVLTQTPLTLPVSPGQRATISCRASQSTDNNGTSNTTWYQQKPGQPPKLLVSS- AQFVQSGVPARFSGSGSGTDFTLTIHPVEPEDFATYFCSQHYSVPWSFGGGTKLEIK
5acp11	SEVQLQETGGGLVQPGASMKLSCVVSGTDFSNNAALWVRQSPGRGLEWTAAISSKPNRYATYY- APSVKGRFTISRDDSQNMVYLQMNNLRAEDTGIYYCALGSGASSSDSGPSYWGQGTTVTVSSG- GGGSGGGGSGGGGSDIVLTQTPLTLPVSPGQRATISCRASQSTNSNGTSHTHWYQQKPGQPPK- LLVHSASTVQSGVPARFSGSGSGTDFTLTIHPVEPEDFATYYCMQSWQTPPTFGGGTKLEIK
5acp12	SGVQLHETGGGLVQPSQSMSLTCVTSGFNFTSHHFSWFRQPPGKGLEYVGSIENNSNGWATYY- APSVKGRFTISRDDSQNRVYLQMNNLRAEDTGIYYCAASAQSNPGSTMAHWGQGTLVTVSSGG- GGSGGGGSGGGGSDIVLTQTPLTLPVSPGQRATISCRASQSTDENGISWTAWLQQKPGQPPKLL- VAWARIVYSGVPNRFHGSGSGTDFTLTINPVEPEDFATYFCMQGHESPYSFGGGTKLEIK
5acp14	SQVQLQETGGGLVQPGASMKLSCVTSGFTFSNTWMLWFKQSPGRGMEYVATIRDQSHNYATY- YAPSVKGRFTISRDDSQNMLYLQMNNLRAEDTGIYYCAGTAAFSGGGGWWGQGTLVTVSSGG- GGSGGGGSGGGGSDIVLTQTPLTLPVSPGQRATISCRAAQSVDYNGHSYMHWYQQKPGQPPKL- LVHNASHVQSGVPARFSGSGSGTDFTLTIHPVEPEDFAMYYCQAWNGYAAVWGQGTKLEIK
5ins01	SQVQLQETGGGLVQPGASMKLSCKASGFNFTSSGMAWVRQRPGQGLEWVAWISPGGGTTHYN- DKVKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCAADPWNSYGFNYWGQGTLVTVSGGGGS- GGGGSGGGGSDIVLTQTPLTLPVSPGQRATISCRASQSVDNSGNSYMHWYQQKPGQPPKLLVY- RASNLESGVPARFSGSGSGTDFTLTIHPVEPEDFAMYYCLQGYSYPFTFGQGTKLEIK
5ins02	SQVQLQETGGGLVQPGASMKLSCKASGFTFTDYSMSWVRQRPGQGMEWVANISPNGGSRSYN- DKVKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCARLDSSGDSGFAYWGQGTLVTVSGGGGS- GGGGSGGGGSDIVMTQTPLTLPASPGDRVTISCKASQDVGKYVAWFQQKPGGSPKLLIYYASN- LYSGVPSRFSGSGSGTDFTFTVSNVQPEDFGTYYCVQGSYAPPTFGQGTKVEIK
5ins03	SQVQLQESGGGLVQPGASMKLSCKASGFTFTRYAMHWVRQRPGQGLEWVAGISPDGGETHYN- DKVKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCAREYWHSLDYWGQGTTVTVSGGGGSGG- GGSGGGGSDIVLTQTPLTLPVSVGQRATISCRSSQSLVDSNGNTYLNWYLQKPGQSPKLLIYQVS- NRFSGVPDRFSGSGSGTDFTLKISRVEAEDFGVYFCQQGYRHPLTFGQGTKLEIK
5ins04	SQVQLQESGGGLVQPGASMRLSCKTSGFTFTSYAISWVKQAPGQGFEWIGSINPGSGNTHYNDK- FKGRATITADTSSNTAYLQLNNLTSEDTAVYYCAASDYRSHGALDYWGQGTQVTVSGGGGSG- GGGSGGGGSDIVMTQTPLTLPASPGDRVTISCKASQNIYKHIIWYQQKPGQAPKLLVYYASNLY- SGVPSRFSGSGSGTDYTLTISNVQPEDFGTYYCAQHYSHPLTFGQGTKVEIK
5ins05	SQVQLQETGGGLVQPGASMKLSCKASGFTFTDYGMIWVRQRPGQGLEWVAYISPGGGYTYYN- DKVKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCARDDGYGAMDYWGQGTTVTVSGGGGSG- GGGSGGGGSDIVMTQTPLTLPVSPGQRATISCRASQSVDSNGYSHVNWYQQKPGQPPKLLVYL- ASNLESGVPARFSGSGSGTDFTLTIHPVEPEDFATYFCMQGYKWPYTFGQGTKLEIK
5ins06	SQVQLQETGGGLVQPGASMKLSCKASGFTFTRYGMSWVRQRPGQGLEWVASISPDGGETDYN- DKVKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCASSHFDYWGQGTTVTVSGGGGSGGGGSG- GGGSDIVLTQTPLTLPVSPGQRATISCRASQSVDYYGYSFMHWYQQKPGQPPKILVYRGSNLYS- GVPARFSGSGSGTDFTLTIHPVEPEDFATYYCQQSYAVPPTFGQGTKLEIK
5ins07	SQAQLQETGGGLVQPGASMKLSCKASGYTFTDYGMSWVRQRPGQGLEWVASISPDGGTTYYN- DKVKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCARMDFTSYQDSAVDYWGQGTSVTVSGG- GGSGGGGSGGGGSDIVMTQTPLTLPASPGDRVTISCKASRDVGNYVAWYQQKPGQSPQLLIYS- ASNLHSGVPSRFSGSGSGTDFTLTISSVQPEDFGDYYCMQVSHSPYTFGQGTKVEIK
5ins08	SQVQLQETGGGLVQPGASMKLSCKASGFTFTDYGMSWVRQRPGQGLEWVAWISPNGGRTHY- NDKVKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCAMTGDYGYAFTLDYWGQGTTVTVSGG- GGSGGGGSGGGGSDIVMTQTPLTLPASPGDRVTISCRASQKVSESVAWYQQKPGQSPKLLIYSA- SNLYDGVPSRFSGSGSGTDFTFTISSVQPEDFATYYCMQSSSYPYTFGQGTKVEIK
5ins09	SQVQLQETGGGLVQPGASMKLSCKASGFTFTRYSMNWVRQRPGQGLEWVASISPGGGSTHYN- DKVKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCAASDRGYYGALGYWGQGTLVTVSGGGG- SGGGGSGGGGSDIVLTQTPLTLPVSPGQRATISCRASQSVDDNGYSYMAWYQQKPGQPPKLLV- YAGSNLESGVPARFSGSGSGTDFTLTIHPVEPEDFATYYCMQMYSTPLTFGSGTKLEIK
5ins10	SEVQLQETGGGLVQPSQSMSLTCTVSGDSFSNYAWSWFRQTPGNRLEYVGSISGGGSTYYHDSV- KGRFTISRDTSKNQVFLQMNSLTTEDTAVYYCARSNHAAFDYWGQGTTVTVSGGGGSGGGGS- GGGGSDIVLTQTPLTLPVSPGQRATISCRASQSVDSYGYSFMDWYQQKPGQPPKLLVYLASNLY- SGVPARFSGSGSGTDFTLTIHPVEPEDFATYFCQQHSEFPYTFGQGTKLEIK
5ins11	SEVQLQESGGGLVQPGASMKLSCKTSGYTFTDYWMHWVRQRPGQGLEWVGWIDPGNGSTNY- NNKFKGRATITADTSSNTAYMQMSSLTSEDTAVYYCARAGGGYRVPYGMDYWGQGTTVTVS- GGGGSGGGGSGGGGSDIVLTQTPLTLPVSPGQRATISCRASESVDSNGYSFMMWYQQKPGQPP- KLLVYLASNLYSGVPARFSGSGSGTDFTLTIHPVEPEDFATYFCMQGHSSPWTFGQGTKLEIK
5ins12	SQVQLQETGGGLVQPGASMRLSCKASGFTITDHGMSWVRQRPGQGFEWIGGIYPGGGNTHYN- DKFKGKATITVDTSSNTAYMQLNSMTSEDTAVYYCARNFWAALDDWGQGTTVTVSGGGGSG- GGGSGGGGSDIVMTQTPLTLPASPGDRVTISCKASQKVGKHVAWFQQKPGQSPKLLIYNASNL- HSGVPSRFSGSGSGTDFTLTISSVQPEDFATYFCVQTHGWPYTFGQGTKVEIK
5ins13	SEVQLQESGGGLVQPGASMKLSCKTSGFTFTDHYMHWVKQRPGQGLEWVAGIDPGNGSTQYN- NKFKGKATFTADTSSNTAYMQLSSLTSEDTAVYYCARSYDAAMDDWGQGTTVTVSGGGGSGG- GGSGGGGSDIVLTQTPLTLPMSVGEKVTITCKSSRSLFSSGTQKNYMTWYQQKPGQPPKLLVY- WASTRESGVPDRFTGSGSGTDFTLTVSSVQAEDIAVYFCMQSHSSPFTFGQGTKVEIK
5ins14	SQVQLQETGGGLVQPGGSMKISCVTSGFNFSHAWMSWVRQSPGKGLEWVAEIRNKSDGYTTY- YAPSVKGRFTVSRDDSQNMVYLQMNNLRAEDTGIYYCAQSSPYSRAMDYWGQGTTVTVSGGG- GSGGGGSGGGGSDIVLTQTPLTLPVSPGQRATISCRASQSVDDNGHSFMNWYQQKPGQPPKLL- VHAASYVKSGVPARFSGSGSGTDFTLTIHPVEPEDFATYYCQQGYSHPWTFGGGTKLEIK
5ins15	SQVQLQESGGGLVQPGASMKLSCKASGFTFTRYAMHWVRQRPGQGLEWVAWISPNGSNTRYN- SKVKGRATITRDKSSNTAYLQMSSLTSEDTAVYYCARDWGASMDYWGQGTTVTVSGGGGSGG- GGSGGGGSDIVMTQTPLTLPASPGDRVTISCKASQDIGTSVAWYQQKPGQSPKLLIYSASQLYD- GVPSRFSGSGSGTDFTFTISSVQPEDFATYFCMQSYTTPYTFGQGTKVEIK

5ins16	SQVQLQETGGGLVQPGASMKLSCKASGFTFTRSGMSWVRQRPGQGLEWVAWISPNGGSTDYN-DKVKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCARGWGGMQYWGQGTTVTVSGGGGSGG-GGSGGGGSDIVLTQTPLTLPVSPGQRATISCRASQSVDYRGYSFMHWYQQKPGQPPKLLVYWG-SNLESGVPARFSGSGSGTDFTLTIHPVEPEDFATYFCQQSYSTPWTFGGGTKLEIK
5ins17	SQVQLQETGGGLVQPGASMRLSCVTSGFDITSSAMHWVRQRPGQGLEWVGGISPYGGETHYA-PKFKGKATITVDTSSNTAYMQMSSLTSEDTAVYYCARRENYAEAGFSYWGQGTTVTVSGGGG-SGGGGSGGGGSDIVMTQTPLTLPASPGDRVTISCRASQDVGKYVAWYQQKPGQSPKLLIYYAS-RRAPGVPSRFSGSGSGTDFTLTISNVQPEDFGTYFCQQGYSWPYTFGQGTKVEIK
5ins18	SQVQLQETGGGLVQPGASMKLSCKASGFTFTEYAMSWVRQRPGQGMEWVAWISPGGGNTHY-NDKVKGRATITRDTSSNTAYLQMSSLTSEDTAVYYCAAMGRHSTGAMDYWGQGTQVTVSGG-GGSGGGGSGGGGSDIVLTQTPLTLPMSVGEKITITCKSSQSLFNSRSQKNYLAWYQQKPGQSPK-LLIYWASTRESGVPDRFTGSGSGTDFTFTVSSVQAEDIAVYFCQQGYHHPYTFGQGTKVEIK
5ins19	SQVQLQETGGGLVQPGASMRLSCKASGFTITDYTMHWVRQRPGQGLEWVGYINPGSGNSYFN-NKFKGRATFSVDNSSNTAYMQMSSLTSEDTAVYYCAAKSNGRIGAFGYWGQGTQVTVSGGGG-SGGGGSGGGGSDIVLTQTPLTLPVSPGQRATISCRASQSVDYYGYSFMQWYQQKPGQPPKLLV-YLGSNLESGVPARFSGSGSGTDFTLTIHPVEPEDFAMYYCQQSHAHPPTFGQGTKLEIK

Table 5.3: Expression levels and peptide sequences of the v_HH designs. I tested a total of 34 v_HH designs, 24 to assess stability achieved by the combinatorial strategy I used, and ten designed to bind Phe. The antibodies were expressed in YSD and expression monitored by labeling the C-terminal c-Myc tag (sequence not given).

design name	expression level (%)	peptide sequence
1vHH01	76.7	SQVQLQESGGGLVQAGGSLRLSCAASGASNNIEWMSWFRQAPGKEREYVAAINT-GGGNTIYASSVRGRFTITRDKAKNLMFLQMNSLKPEDTAIYWCAAGGYEVVATSF-GRWGKGTQVTVS
1vHH02	75.4	SQVQLQESGGGLVQAGGSLRLSCTASGYSISTYVIGWFRQAPGKERESVAAINSAN-GTTWYASSVRGRFTITKDDARNSVYLQMNSLKPEDTAIYWCGAGKGGSWNAWG-KGTQVTVS
1vHH04	62.6	SQVQLQESGGGLVQAGGSLRLSCTASGQISSNWNMGWFRQAPGKERESVAAINW-ATGWTWYASSVRGRFTITQDNSQNSVYLQMNSLKPEDTAIYWCAAQDKNKLGT-VGYWGQGTQVTVS
1vHH05	77.8	SQVQLQESGGGLVQAGGSLRLSCTASGQISSNWNMGWFRQAPGKEREYVAAINW-GSGTTWYGSSVRGRFTITQDNSQNSVYLQMNSLKPEDTAIYWCAASREGWHLPM-NSYGFDYWGKGTQVTVS
1vHH06	76.6	SQVQLQESGGGLVQAGGSLRLSCSASGQSGSSYNMGWFRQAPGKEREYVAAINW-GSGTTWYASSVRGRFTITRDKSRNTVYLQMNSLKPEDTAIYWCAAGGYEVKATS-FGRWGQGTQVTVS
1vHH07	74.3	SQVQLQESGGGLVQAGGSLRLSCQATGNTYRLNDMGWFRQAPGKEREYVAAISG-NGKETSYASSVRGRFSITKDKSQNTVYLQMNSLKPEDTAIYWCAAKYRPVKYNEH-PANSNFHWYGQGTQVTVS
1vHH08	78.5	SQVQLQESGGGLVQAGGSLRLSCQASGNTSSINSMGWFRQAPGKEREYVAAINSS-GGTTFYASSVRGRFSITQDNARNTVYLEMNSLKPEDTAIYWCAAKYKPVQYNNSP-ANANFHWYGQGTQVTVS
1vHH09	71.4	SQVQLQESGGGLVQAGGSLRLSCTATGDTVSSHAMGWFRQAPGKEREYVAGINT-SSGTTWYASSVRGRFSITRDNSRNTVYLQMNSLKPEDTAIYWCAAMKGYFLKMPS-ANHFKYWGQGTQVTVS
1vHH10	75.3	SQVQLQESGGGLVQAGGSLRLSCQATGNTSSIKSVGWFRQAPGKEREYVAAIYTD-NGNTWYASSVRGRFSITQNKAQNSVYLEMNSLKPEDTAIYFCAASEMAGYPLDVG-IYNYWGKGTQVTVS
1vHH11	74.8	SQVQLQESGGGLVQAGGSLRLSCSASGFPVGRKSMSWFRQAPGKEREYVAAINTS-SGTTWYASSVRGRFTITRDKSRNTVYLQMNSLKPEDTAIYFCAASEMSGYPLDVG-IFNYWGKGTQVTVS
1vHH12	71.5	SQVQLQESGGGLVQAGGSLRLSCSASGFPVGRFSMSWFRQAPGKERESVAAINSR-DGNTWYASSVRGRFTITRDKSRNTVYLQMNSLKPEDTAIYFCSANMKTWAGLTR-NYYGKGTQVTVS
1vHH13	18.6	SQVQLQESGGGLVQAGGSLRLSCQATGNTYRIKDMGWFRQAPGKERESVAAINT-DSGTTWYASSVRGRFSITRDKSQNTVYLQMNSLKPEDTAIYFCSANVKTWAGLTR-NYYGKGTQVTVS
1vHH14	75.7	SQVQLQESGGGLVQAGGSLRLSCTASGYSISTRAIGWFRQAPGKEREYVAAISQNG-GITYYASSVRGRFTITQDDARNSVYLQMNSLKPEDTAIYWCAAQPRGDYVTANHE-YWGQGTQVTVS
1vHH15	76.6	SQVQLQESGGGLVQAGGSLRLSCTASGRTMSSYAIGWFRQAPGKEREYVAAINSG-SGTTYYASSVRGRFTITRNNAQNTIYLQMNSLKPEDTAIYWCAAKPRGDYVTAYH-EYWGQGTQVTVS
1vHH16	78.7	SQVQLQESGGGLVQAGGSLRLSCQASGNTSSINDMGWFRQAPGKEREYVAAVNT-GSGTTWYASSVRGRFSITQDNARNTVYLEMNSLKPEDTAIYFCAAGGHQMVATSF-GQWGKGSQVTVS
1vHH17	80.4	SQVQLQESGGGLVQAGGSLRLSCSASGQSGTSWNMGWFRQAPGKEREYVAAIFW-RDGNTWYASSVRGRFTITRDKSRNTVYLQMNSLKPEDTAIYFCAADENWITANW-RNGPNYYGQGTQVTVS

1vHH18	76.1	SQVQLQESGGGLVQAGGSLRLSCSATGFSFTNHAVSWFRQAPGKEREYVAAIYSG-NGTTWYASSVRGRFSISRNKAQNMVYLQMNSLKPEDTAIYFCAADDNWVTANW-RNGPNWYGQGTQVTVS
1vHH19	79.7	SQVQLQESGGGLVQAGGSLRLSCTASGQISSNFDMGWFRQAPGKEREYVAAINW-GSGTTFYASSVRGRFTITQDNSQNSVYLQMNSLKPEDTAIYYCAADENYLTATYR-NGPNYFGQGTQVTVS
1vHH20	78.4	SQVQLQESGGGLVQAGGSLRLSCQASGNTSRINSMGWFRQAPGKEREYVAAVSQS-GDNTSYASSVRGRFSITQDNARNTVYLEMNSLKPEDTAIYWCAAKWQPVQYSEW-PANTNFFEWGQGTQVTVS
1vHH21	78.3	SQVQLQESGGGLVQAGGSLRLSCQASGNTSRIHSMGWFRQAPGKEREYVAAINGD-GKRTSYASSVRGRFSITRDNARNTVYLEMNSLKPEDTAIYWCAAQWQPVRYNEW-PSNTNFFEWGQGTQVTVS
1vHH22	17.2	SQVQLQESGGGLVQAGGSLRLSCQATGNTYRIKDMGWFRQAPGKEREYVAAISG-DGRNTYYASSVRGRFSITKDKSQNTVYLQMNSLKPEDTAIYWCAAKYRPVRYSD-HPANTNFFEWGQGTQVTVS
1vHH23	74.5	SQVQLQESGGGLVQAGGSLRLSCSASGFPVGRWSMSWFRQAPGKEREYVAAISSG-SGTTHYASSVRGRFTITRDKSRNTVYLQMNSLKPEDTAIYWCAANKYSPLVPGPV-PMSSVESWGQGTQVTVS
1vHH24	78.6	SQVQLQESGGGLVQAGGSLRLSCTASGQISSNYNMGWFRQAPGKERESVAAINWG-SGTTWYASSVRGRFTITQDNSQNSVYLQMNSLKPEDTAVYWCAAAPTSTHSLYFN-YYGKGTQVTVS
1vHH25	80.4	SQVQLQESGGGLVQAGGSLRLSCTASGYSISTHAIGWFRQAPGKEREYVAGISSGG-GNTYYASSVRGRFTITRDDARNSVYLQMNSLKPEDTAIYWCAADRYPSLKVVSIIT-DKYNFWGKGTQVTVS
1phe01	65.4	SPQESGGGLVQAGGSLRLSCSTTGGNEWAWSSWTSGWFRQAPGKEREFVAGISG-SGKISKYASSVRGRFSITRDNSQNTDYLEMNSLKPEDTAIYWCAAAKEGWHIPMNS-YGFDQFGKGTQVTVS
1phe02	60.1	SLQESGGGLVQAGGSLRLSCSASGYHFSHKDMGWFRQAPGKEREYVAAIAPKTN-KTHYGSSVRGRFSVSRDEGKNSLYLQMNSLKPEDTAIYWCAASNKQMNYAVVTA-KTFDAWGQGTQVTVS
1phe03	59.1	SLQESGGGLVQAGGSLRLSCQLSGNTSATKSVGWFRQAPGKEREYVASKAQGDG-LSFYASSVRGRFTISRDEAQNSVYLQMNSLKPEDTAIYWCAAGGYEVVATSFGRW-GKGTQVTVS
1phe04	70.2	SLQESGGGLVQAGGSLRLSCTASGRTMSSYAIRWNRQAPGKERENVAAISPNGGIT-WYASSVRGRFTITQNNAQNTIYLQMNSLKPEDTAIYFCDAGTKQSTSTKRNPWD-HWGKGTQVTVS
1phe05	71.9	SLQESGGGLVQAGGSLRLSCASTGNTSSINSVGWFRQAPGKEREYVAAINPSGGNT-WYASSVRGRFTISQDKAQDTIYLQMNSLKPEDTAIYWCAAGGYEVKATTFGRWG-KGTQVTVS
1phe06	68.6	SPQESGGGLVQAGGSLRLSCSLSGKSQDGGAMGSDYRLTAGSTGWFRQAPGKER-ESVAAINTATGWTWYASSVRGRFTITREAAATMVFLQMNSLKPEDTAIYFCSNNM-GYQHFGRGSQVTVS
1phe07	57.4	SLQESGGGLVQAGGSLRLSCTANGYTYSIDWFRQAPGKEREYVATIAPQSNKSNY-ASSVRGRFSITQDKARNTVFLQMNSLKPEDTAIYWCAAKTTSYGGQNPNNWDYW-GKGTQVTVS
1phe08	73.4	SLQESGGGLVQAGGSLRLSCSKSGQSGSSDNVGWFRQAPGKEREYVASIGSSLIGP-NNYYTQYASSVRGRFTITQDNARQTIYLQMNSLKPEDTAIYWCAAGGYAVVETSF-GRWGKGTQVTVS
1phe09	74.0	SLQESGGGLVQAGGSLRLSCSATGFSFTNHEVSWFRQAPGKEREWVATIHSSGNT-NYASSVRGRFSISQNKAQNMVYLQMNSLKPEDTAVYWCAAAPTSTNSLHFNWYG-KGTQVTVS
1phe10	57.4	SLQESGGGLVQAGGSLRLSCQASGNTSSSNSEGWFRQAPGKEREYVAAINPATGT-TYYASSVRGRFTISKDDAQNSIYLQMNSLKPEDTAIYWCASNGYEVVATSYKKWG-KGTQVTVS

Bibliography

[1] B. Al-Lazikani, A. M. Lesk, and C. Chothia. Standard conformations for the canonical structures of immunoglobulins. *Journal of Molecular Biology*, 273(4):927–48, nov 1997.

[2] M. Arbabi Ghahroudi, A. Desmyter, L. Wyns, R. Hamers, and S. Muyldermans. Selection and identification of single domain antibody fragments from camel heavy-chain antibodies. *FEBS Letters*, 414(3):521–526, 1997.

[3] L. Banaszak. *Foundations of structural biology*. Academic, 2000.

[4] R. Baron and J. A. McCammon. Molecular recognition and ligand association. *Annual Review of Physical Chemistry*, 64:151–75, jan 2013.

[5] L. Benatuil, J. J. M. Perez, J. Belk, and C.-M. Hsieh. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Engineering, Design & Selection*, 23(4):9–10, apr 2010.

[6] I. Campeotto, A. Goldenzweig, J. Davey, L. Barfod, J. M. Marshall, S. E. Silk, K. E. Wright, S. J. Draper, M. K. Higgins, and S. J. Fleishman. One-step design of a stable variant of the malaria invasion protein RH5 for use as a vaccine immunogen. *Proceedings of the National Academy of Sciences*, 114(5):998–1002, jan 2017.

[7] P. Carter. Improving the efficacy of antibody-based cancer therapies. *Nature Reviews Cancer*, 1(November):118–129, 2001.

[8] P. Carter and L. Presta. Method for making humanized Antibodies, 2002.

[9] G. Chao, W. L. Lau, B. J. Hackel, S. L. Sazinsky, S. M. Lippow, and K. D. Wittrup. Isolating and engineering human antibodies using yeast surface display. *Nature Protocols*, 1(2):755–68, jan 2006.

[10] C. Chothia, I. Gelfand, and A. Kister. Structural determinants in the sequences of immunoglobulin variable domain. *Journal of molecular biology*, 278(2):457–79, may 1998.

[11] N. . Collaborative Computational Project. The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D Biological Crystallography*, 50(5):760–763, sep 1994.

[12] A. V. J. Collis, A. P. Brouwer, and A. C. R. Martin. Analysis of the antigen combining site: correlations between length and sequence composition of the hypervariable loops and the nature of the antigen. *Journal of molecular biology*, 325(2):337–54, jan 2003.

[13] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović, and Foldit Players. Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756–60, aug 2010.

[14] V. Cortez-Retamozo and N. Backmann. Efficient Cancer Therapy with a Nanobody-Based Conjugate. *Cancer research*, 64:2853–2857, 2004.

[15] V. Cortez-Retamozo, N. Backmann, P. D. Senter, U. Wernery, P. De Baetselier, S. Muyldermans, and H. Revets. Efficient Cancer Therapy with a Nanobody-Based Conjugate. *Cancer Research*, 64(8):2853–2857, 2004.

[16] L. P. Daley, M. A. Kutzler, B. W. Bennett, M. C. Smith, A. L. Glaser, and J. A. Appleton. Effector functions of camelid heavy-chain antibodies in immunity to West Nile virus. *Clinical and Vaccine Immunology*, 17(2):239–246, 2010.

[17] R. Das and D. Baker. Macromolecular modeling with rosetta. *Annual review of biochemistry*, 77:363–82, jan 2008.

[18] W. L. DeLano. Convergent Solutions to Binding at a Protein-Protein Interface. *Science*, 287(5456):1279–1283, feb 2000.

[19] M. Dumoulin and K. Conrath. Single-domain antibody fragments with high conformational stability. *Protein Science*, 11:500–515, 2002.

[20] J. Dunbar, K. Krawczyk, J. Leem, T. Baker, A. Fuchs, G. Georges, J. Shi, and C. M. Deane. SAbDab: the structural antibody database. *Nucleic Acids Research*, 42(D1):D1140–D1146, jan 2014.

[21] P. Emsley and K. Cowtan. Coot : model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography*, 60(12):2126–2132, dec 2004.

[22] P. Evans. Scaling and assessment of data quality. *Acta Crystallographica Section D Biological Crystallography*, 62(1):72–82, jan 2006.

[23] A. Fersht. *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. New York, 2000.

[24] S. J. Fleishman and D. Baker. Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell*, 149(2):262–73, apr 2012.

[25] S. J. Fleishman, T. A. Whitehead, D. C. Ekiert, C. Dreyfus, J. E. Corn, E.-M. Strauch, I. A. Wilson, and D. Baker. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science*, 332(6031):816–21, may 2011.

[26] A. Goldenzweig, M. Goldsmith, S. E. Hill, O. Gertman, P. Laurino, Y. Ashani, O. Dym, T. Unger, S. Albeck, J. Prilusky, R. L. Lieberman, A. Aharoni, I. Silman, J. L. Sussman, D. S. Tawfik, and S. J. Fleishman. Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Molecular Cell*, 63(2):337–346, 2016.

[27] R. B. Hill, D. P. Raleigh, A. Lombardi, and W. F. Degrado. De novo design of helical bundles as models for understanding protein folding and function. *Accounts of Chemical Research*, 33(11):745–754, 2000.

[28] D. M. Hoover and J. Lubkowski. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic acids research*, 30(10):e43, may 2002.

[29] P.-S. Huang, S. E. Boyken, and D. Baker. The coming of age of de novo protein design. *Nature*, 537(7620):320–7, 2016.

[30] P.-S. Huang, K. Feldmeier, F. Parmeggiani, D. A. Fernandez Velasco, B. Höcker, and D. Baker. De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nature Chemical Biology*, 12(November):29–34, 2016.

[31] T. M. Jacobs, B. Williams, T. Williams, X. Xu, A. Eletsky, J. F. Federizon, T. Szyperski, and B. Kuhlman. Design of structurally distinct proteins using strategies inspired by evolution. *Science (New York, N.Y.)*, 352(6286):687–90, 2016.

[32] R. K. Jha, A. Leaver-Fay, S. Yin, Y. Wu, G. L. Butterfoss, T. Szyperski, N. V. Dokholyan, and B. Kuhlman. Computational design of a PAK1 binding protein. *Journal of Molecular Biology*, 400(2):257–70, jul 2010.

[33] P. T. Jones, P. H. Dear, J. Foote, M. S. Neuberger, and G. Winter. Replacing the complementary-determining regions in a human antibody with those from a mouse. *Nature*, 321(29):522–525, 1986.

[34] A. Joshi, R. Grinter, I. Josts, S. Chen, J. A. Wojdyla, E. D. Lowe, R. Kaminska, C. Sharp, L. McCaughey, A. W. Roszak, R. J. Cogdell, O. Byron, D. Walker, and C. Kleanthous. Structures of the Ultra-High-Affinity Protein–Protein Complexes of Pyocins S2 and AP41 and Their Cognate

Immunity Proteins from Pseudomonas aeruginosa. *Journal of Molecular Biology*, 427(17):2852–2866, aug 2015.

[35] J. Karanicolas, J. E. Corn, I. Chen, L. a. Joachimiak, O. Dym, S. H. Peck, S. Albeck, T. Unger, W. Hu, G. Liu, S. Delbecq, G. T. Montelione, C. P. Spiegel, D. R. Liu, and D. Baker. A de novo protein binding pair by computational design and directed evolution. *Molecular Cell*, 42(2):250–60, apr 2011.

[36] O. Keskin, A. Gursoy, B. Ma, and R. Nussinov. Principles of Protein-Protein Interactions: What are the Preferred Ways For Proteins To Interact? *Chemical Reviews*, 108(4):1225–1244, apr 2008.

[37] O. Khersonsky and S. J. Fleishman. Why reinvent the wheel? Building new proteins based on ready-made parts. *Protein Science*, 25:1179–1187, 2016.

[38] O. Khersonsky and S. J. Fleishman. Incorporating an allosteric regulatory site in an antibody through backbone design. *Protein Science*, jan 2017.

[39] P. M. Kirkham, F. Mortari, J. a. Newton, and H. W. Schroeder. Immunoglobulin VH clan and family identity predicts variable domain structure and may influence antigen binding. *The EMBO Journal*, 11(2):603–9, feb 1992.

[40] N. Koga, R. Tatsumi-Koga, G. Liu, R. Xiao, T. B. Acton, G. T. Montelione, and D. Baker. Principles for designing ideal protein structures. *Nature*, 491(7423):222–7, nov 2012.

[41] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, and D. Baker. Design of a novel globular protein fold with atomic-level accuracy. *Science*, 302(5649):1364–8, nov 2003.

[42] G. D. Lapidoth, D. Baran, G. M. Pszolla, C. Norn, A. Alon, M. D. Tyka, and S. J. Fleishman. AbDesign: An algorithm for combinatorial backbone design guided by natural conformations and sequences. *Proteins*, 83(8):1385–406, 2015.

[43] M. Lauwereys, M. A. Ghahroudi, A. Desmyter, J. Kinne, W. Hölzer, E. De Genst, L. Wyns, and S. Muyldermans. Potent enzyme inhibitors derived from dromedary heavy-chain antibodies. *EMBO Journal*, 17(13):3512–3520, 1998.

[44] M. C. Lawrence and P. M. Colman. Shape complementarity at protein/protein interfaces, 1993.

[45] T. Li, J.-P. Bourgeois, S. Celli, F. Glacial, A.-M. Le Sourd, S. Mecheri, B. Weksler, I. Romero, P.-O. Couraud, F. Rougeon, and P. Lafaye. Cell-penetrating anti-GFAP VHH and corresponding fluorescent fusion protein VHH-GFP spontaneously cross the blood-brain barrier and specifically recognize astrocytes: application to brain imaging. *FASEB Journal*, 26(10):3969–79, oct 2012.

[46] X. Liu, K. Fan, and W. Wang. The Number of Protein Folds and Their Distribution over Families in Nature. *Proteins: Structure, Function and Genetics*, 54(3):491–499, 2004.

[47] E. Marcos, B. Basanta, T. M. Chidyausiku, Y. Tang, G. Oberdorfer, G. Liu, G. V. T. Swapna, R. Guan, D.-a. Silva, J. Dou, J. H. Pereira, R. Xiao, B. Sankaran, P. H. Zwart, G. T. Montelione, and D. Baker. Principles for designing proteins with cavities formed by curved β sheets. *Science*, 355(6321):201–206, 2017.

[48] A. J. McCoy. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallographica Section D Biological Crystallography*, 63(1):32–41, jan 2007.

[49] K. S. Midelfort, H. H. Hernandez, S. M. Lippow, B. Tidor, C. L. Drennan, and K. D. Wittrup. Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody. *Journal of Molecular Biology*, 343(3):685–701, oct 2004.

[50] V. Morea, A. Tramontano, M. Rustici, C. Chothia, and A. M. Lesk. Conformations of the third hypervariable region in the VH domain of immunoglobulins. *Journal of Molecular Biology*, 275(2):269–94, jan 1998.

[51] S. L. Morrison, M. J. Johnson, L. a. Herzenberg, and V. T. Oi. Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains. *Proceedings of the National Academy of Sciences of the United States of America*, 81(21):6851–5, nov 1984.

[52] G. N. Murshudov, A. A. Vagin, and E. J. Dodson. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D*, 53(3):240–255, may 1997.

[53] S. Muyldermans. Nanobodies: Natural Single-Domain Antibodies. *Annual Review of Biochemistry*, 82(1):775–797, 2013.

[54] V. K. Nguyen, S. Muyldermans, and R. Hamers. The specific variable domain of camel heavy-chain antibodies is encoded in the germline. *Journal of Molecular Biology*, 275(3):413–8, jan 1998.

[55] V. S. Nguyen, L. Logger, S. Spinelli, P. Legrand, T. Thanh, H. Pham, T. Trang, N. Trinh, Y. Cherrak, A. Zoued, A. Desmyter, E. Durand, A. Roussel, C. Kellenberger, E. Cascales, and C. Cambillau. Type VI secretion TssK baseplate protein exhibits structural similarity with phage receptor-binding proteins and evolved to bind the membrane complex. *Nature Microbiology*, 17103(June):1–9, 2017.

[56] J. Novotný, R. Bruccoleri, and M. Karplus. An analysis of incorrectly folded protein models. Implications for structure predictions. *Journal of Molecular Biology*, 177(4):787–818, 1984.

[57] L. Pauling, R. B. Corey, and H. R. Branson. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, 37(4):205–211, 1951.

[58] R. J. Read and J. L. Sussman, editors. *Evolving Methods for Macromolecular Crystallography*, volume 245 of *NATO Science Series II: Mathematics, Physics and Chemistry*. Springer Netherlands, Dordrecht, 2007.

[59] A. M. Ring, A. Manglik, A. C. Kruse, M. D. Enos, W. I. Weis, K. C. Garcia, and B. K. Kobilka. Adrenaline-activated structure of β2-adrenoceptor stabilized by an engineered nanobody. *Nature*, 502(7472):575–579, sep 2013.

[60] D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. a. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik, and D. Baker. Kemp elimination catalysts by computational enzyme design. *Nature*, 453(7192):190–5, may 2008.

[61] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, and H. J. Wolfson. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic acids research*, 33(Web Server issue):W363–7, jul 2005.

[62] P. Setny, Z. Wang, L. T. Cheng, B. Li, J. A. McCammon, and J. Dzubiella. Dewetting-Controlled binding of ligands to hydrophobic pockets. *Physical Review Letters*, 103(18):1–4, 2009.

[63] A. Stank, D. B. Kokh, J. C. Fuller, and R. C. Wade. Protein Binding Pocket Dynamics. *Accounts of Chemical Research*, 49(5):809–815, 2016.

[64] P. B. Stranges and B. Kuhlman. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Science*, 22(1):74–82, 2013.

[65] Z. Sun, Y. N. Yan, M. Yang, and J. Z. H. Zhang. Interaction entropy for protein-protein binding. *The Journal of Chemical Physics*, 146(12):124124, 2017.

[66] J. L. Sussman, M. Harel, F. Frolow, C. Oefner, A. Goldman, L. Toker, and I. Silman. Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein. *Science (New York, N.Y.)*, 253(5022):872–879, 1991.

[67] J.-L. Teillaud. *From Whole Monoclonal Antibodies to Single Domain Antibodies: Think Small*, pages 3–13. Humana Press, Totowa, NJ, 2012.

[68] C. Vincke, R. Loris, D. Saerens, S. Martinez-Rodriguez, S. Muyldermans, and K. Conrath. General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold. *Journal of Biological Chemistry*, 284(5):3273–3284, 2009.

[69] F. Wang, S. Sen, Y. Zhang, I. Ahmad, X. Zhu, I. A. Wilson, V. V. Smider, T. J. Magliery, and P. G. Schultz. Somatic hypermutation maintains antibody thermodynamic stability during affinity maturation. *Proceedings of the National Academy of Sciences*, 110(11):4261–4266, mar 2013.

[70] S. Warszawski, R. Netzer, D. S. Tawfik, and S. J. Fleishman. A "fuzzy"-logic language for encoding multiple physical traits in biomolecules. *Journal of Molecular Biology*, 426(24):4125–4138, 2014.

[71] J. Wesolowski, V. Alzogaray, J. Reyelt, M. Unger, K. Juarez, M. Urrutia, A. Cauerhff, W. Danquah, B. Rissiek, F. Scheuplein, N. Schwarz, S. Adriouch, O. Boyer, M. Seman, A. Licea, D. V. Serreze, F. a. Goldbaum, F. Haag, and F. Koch-Nolte. Single domain antibodies: promising experimental and therapeutic tools in infection and immunity. *Medical microbiology and immunology*, 198(3):157–74, aug 2009.

[72] T. T. Wu and E. A. Kabat. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *The Journal of experimental medicine*, 132(2):211–50, 1970.

[73] G. I. Yakovlev, V. A. Mitkevich, K. L. Shaw, S. Trevino, S. Newsom, C. N. Pace, and A. A. Makarov. Contribution of active site residues to the activity and thermal stability of ribonuclease Sa. *Protein science : a publication of the Protein Society*, 12(10):2367–73, 2003.

[74] A. Zanghellini, L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. a. Althoff, D. Röthlisberger, and D. Baker. New algorithms and an in silico benchmark for computational enzyme design. *Protein Science*, 15(12):2785–2794, 2006.