Molecular Basis of a Million-Fold Affinity Maturation Process in a Protein–Protein Interaction

Daniel A. Bonsor¹, Sandra Postel¹, Brian G. Pierce², Ningyan Wang³, Penny Zhu¹, Rebecca A. Buonpane³, Zhiping Weng², David M. Kranz³ and Eric J. Sundberg¹*

¹Boston Biomedical Research Institute, 64 Grove Street, Watertown, MA 02472, USA
²Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA
³Department of Biochemistry, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA

Protein engineering is becoming increasingly important for pharmaceutical applications where controlling the specificity and affinity of engineered proteins is required to create targeted protein therapeutics. Affinity increases of several thousand-fold are now routine for a variety of protein engineering approaches, and the structural and energetic bases of affinity maturation have been investigated in a number of such cases. Previously, a 3-million-fold affinity maturation process was achieved in a protein–protein interaction composed of a variant T-cell receptor fragment and a bacterial superantigen. Here, we present the molecular basis of this affinity increase. Using X-ray crystallography, shotgun reversion/replacement scanning mutagenesis, and computational analysis, we describe, in molecular detail, a process by which extrainterfacial regions of a protein complex can be rationally manipulated to significantly improve protein engineering outcomes.

© 2011 Elsevier Ltd. All rights reserved.
engineer recombinant proteins to bind a nearly limitless repertoire of potential targets with relatively high specificity and affinity.

Mammalian immune systems are encoded with natural protein engineering tools. Antibodies, as products of the adaptive immune system, are exceptionally well suited for combating infectious diseases: hypervariability and a natural affinity maturation process allow for the recognition of diverse antigens;\(^5\) a constant region triggers potent immune mechanisms. Antibodies have therefore been commonly used as therapeutic molecules, either in their natural state or after further engineering.

Alternatives to antibodies—largely in the form of diverse scaffolds for the design and engineering of recombinant proteins that often serve as high-affinity steric inhibitors of deleterious protein interactions—have been developed (reviewed by Binz et al.\(^6\) and Binz and Pluckthun\(^7\)). Like antibodies, these scaffold proteins are often able to bind a wide range of target proteins. The natural protein binding partners of drug targets, although more restricted in their specificity than generic scaffold proteins, provide additional alternatives for protein engineering that can lead to highly effective therapeutic proteins.\(^8\)–\(^10\)

Protein engineering methods have been restricted traditionally to amino acid sequence variations of the initial protein architecture, and not expansion or modification of the architecture itself. Recent studies, however, have shown that engineering strategies that dispense of natural protein architectures, through the recombination and rearrangement of protein domains and modules, can result in engineered proteins that exhibit unique molecular recognition properties and novel functions (reviewed by Koide\(^11\)). Hybrid methods that exist somewhere between maintaining and dispensing of the initial protein architecture, by diversifying not only the sequence but also the length of protein loops within or near the protein–protein interface, have now been utilized successfully in numerous molecular systems. This is yet another engineering strategy that mimics nature, especially the length diversity of complementarity-determining region (CDR) 3 loops in antibodies and T-cell receptors (TCRs).\(^12\) Several groups, using as scaffold the 10th type III domain of human fibronectin, which has an immunoglobulin-like β-sandwich fold and CDR-like loops, have exploited loop length diversity to achieve significant affinity gains.\(^13\)–\(^16\) Here we present the detailed molecular basis of an amino acid residue and a loop length diversity protein engineering process that, when combined, resulted in a greater-than-million-fold affinity increase in an engineered TCR/superantigen (SAG) complex.

We previously used a semirational protein engineering strategy that incorporated structure-based knowledge concerning protein complexes that are homologous to the targeted complex in order to increase the extent and degree of affinity maturation.\(^17\) We generated an engineered TCR variant named G5-8, derived from the mouse TCR Vβ8.2 (mVβ8.2) chain, that binds to the bacterial SAG staphylococcal enterotoxin B (SEB) with a 3-million-fold increase in affinity relative to the wild-type mVβ8.2, with measured binding affinities of 48 pM\(^17\) and 150 μM,\(^18\) respectively. Additionally, we showed that G5-8 acts as an inhibitor of SEB-mediated T-cell activation and is completely protective in vivo when administered to animals challenged with a lethal dose of SEB.\(^17\) A brief overview of our semirational structure-based protein engineering strategy is presented here as a guide for the structural and energetic bases of the engineered affinity maturation described below.

Initially, following a standard directed evolution strategy, we created libraries with genetic variability in the mVβ8.2 region but without sequence length changes within the targeted mVβ8.2/SEB complex molecular interface and selected affinity-matured variants, using yeast display (Fig. 1a). Genetic diversity was focused entirely on the CDR2 loop of mVβ8.2, since it forms the majority of the protein–protein interface with SEB and contains several hot-spot contacts.\(^19\) From this process, we generated G2-5, a variant of mVβ8.2 that binds SEB with an affinity of 650 pM,\(^17\) an approximate 200,000-fold increase relative to the wild-type complex.

Subsequently extending from the G2-5 platform, we followed a semirational directed evolution engineering strategy (Fig. 1b) that takes advantage of the structure-based knowledge of a homologous TCR/SAG complex—the human TCR Vβ2.1 (hVβ2.1) domain in complex with streptococcal pyrogenic exotoxin C (SpeC). SpeC interacts with hVβ2.1, forming intermolecular contacts with each TCRβ hypervariable loop\(^20\) (Fig. 1b, left), while SEB contacts only the mVβ8.2 CDR2 and HV4 loops\(^21\) (Fig. 1a, left and middle). The CDR1 loop of hVβ2.1 includes a noncanonical single amino acid residue insertion, which acts to push several residues C-terminal to it closer to the SpeC molecular surface to make numerous intermolecular interactions that have been shown to augment the affinity of the hVβ2.1/SpeC complex.\(^22\) Conversely, residues from the shorter CDR1 loop of mVβ8.2 are located at too great a distance from SEB to make specific interactions (Fig. 1b, middle).

With this more comprehensive structural understanding of TCR/SAG interactions and seeking to functionalize the CDR1 loop of mVβ8.2 as a meaningful contributor to increased SEB binding affinity, we extended the standard directed evolution approach (Fig. 1a, right) by generating additional mVβ8.2 libraries that included randomized CDR1 loops with either one or two additional amino acid residues relative to the wild-type sequence and
by selecting for affinity-matured variants (Fig. 1b, right). After exhaustive iterative rounds of mutagenesis and selection for mVβ8.2 variants with a modified CDR1 loop length, most of the isolated variants contained a single additional residue. One of these, variant G5-8, incorporated the additional CDR1 loop residue (Ser27a\(^{G5-8}\)) as well as two CDR1 loop variant residues (Tyr28\(^{G5-8}\) and Phe29\(^{G5-8}\)). G5-8 binds to SEB with an affinity of 48 pM, \(^{17}\) 3-million-fold higher than the wild-type mVβ8.2/SEB complex and more than 10-fold higher than G2-5, the highest-affinity variant with a conserved sequence length relative to the template for directed evolution, wild-type mVβ8.2.

The structural basis of this rationalized protein engineering method is now revealed by our 2.95-Å X-ray crystal structure of the G5-8/SEB complex, combined with a mutational analysis (see Supplementary Information, Materials and Methods). Crystallographic and refinement statistics for this structure are listed in Table 1. There were eight G5-8/SEB complexes per asymmetric unit in this crystal. The interface G5-8 variant residues and the SEB residues that they contact from all of these complexes superimpose essentially perfectly (Supplementary Fig. 1), even though the noncrystallographic constraints on all interface residues were relaxed in the final stages of refinement. Electron density maps also clearly delinate the side-chain positions of these residues (Supplementary Fig. 2). When the G5-8/SEB structure is superimposed onto the wild-type mVβ8.2/SEB structure, \(^{21}\) the two main chains of the complexes are nearly indistinguishable, except for the CDR1 loops of G5-8 and mVβ8.2 (Fig. 2a). A schematic interaction map of the wild-type mVβ8.2/SEB and affinity-matured G5-8/SEB protein–protein interfaces is shown in Supplementary Fig. 3. Three G5-8 residues (Ser27a\(^{G5-8}\), Tyr28\(^{G5-8}\), and Phe29\(^{G5-8}\)) replace two

---

**Fig. 1.** Directed evolution strategies. (a) Steps in the standard directed evolution strategy. Shown are the structure of the wild-type mVβ8.2/SEB complex (top left), interacting surfaces on mVβ8.2 when bound to SEB (bottom left), close-up view of mVβ8.2 CDR2 loop interactions with SEB (middle), and the iterative process of mutagenesis and selection (right). (b) Steps in the rationalized directed evolution strategy. Shown are the structure of wild-type hVβ2.1/SpεC (top left); interacting surfaces on hVβ2.1 when bound to SpεC (bottom left); comparison of hVβ2.1 and mVβ8.2 CDR1 loop structures and intermolecular contacts with SpεC and SEB, respectively (middle); and the iterative process of mutagenesis and selection beginning with the structure-based modification of extrainterfacial residues (right).
mVβ8.2 residues (Asn28<sup>mV</sup>β8.2 and His29<sup>mV</sup>β8.2), resulting in a longer CDR1 loop with a distinct conformation (Fig. 2b). The structural effect of these CDR1 loop sequence and length changes in G5-8 is that the side chain of Tyr<sup>G5</sup>β8 is pointed directly toward SEB (Fig. 2c), confirming the rational basis of our engineering strategy.

Mutations in the CDR2 loop of G5-8, which are similar to those in G2-5, include two relatively large amino acid side chains that replace minimal side chains. These include substitutions of Ala<sup>52</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 with Ile<sup>52</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 and Arg<sup>53</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2, respectively (Fig. 2d). Together, the variant CDR1 and CDR2 loop residues in G5-8 that make intermolecular contacts with SEB (Tyr<sup>G5</sup>β8, Ile<sup>52</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2, and Arg<sup>53</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2) encompass the β-sandwich domain of SEB, extending from the interdomain cleft to its periphery (Fig. 2c), a well-documented region of energetic importance for TCR/SAG complexes.19,23

The variant residues in G5-8 from both the CDR1 loop and the CDR2 loop form numerous intermolecular contacts with SEB that are absent in the wild-type mVβ8.2/SEB complex. This results in relative increases in buried TCR surface area (805 Å<sup>2</sup> versus 561 Å<sup>2</sup> for the mVβ8.2/SEB complex. The CDR2 loop mutations Ile<sup>52</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 and Arg<sup>53</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 form van der Waals contacts and a hydrogen bond with a trio of SEB asparagine residues (Asn<sup>31</sup>SEB, Asn<sup>66</sup>SEB, and Asn<sup>88</sup>SEB). These three variant residues in the G5-8 CDR1 (Tyr<sup>28</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 and CDR2 (Ile<sup>52</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 and Arg<sup>53</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2)) loops comprise the majority of the increased buried surface area and intermolecular contacts in the G5-8/SEB complex relative to the wild-type complex and form a contiguous interface with SEB centered around Asn<sup>60</sup>SEB (Fig. 2c).

In protein–protein interactions, not all noncovalent contacts in the interface are energetically equivalent.24,25 To determine which mutations in G5-8 resulted in significant energetic changes in complex formation with SEB, relative to the mVβ8.2/SEB complex, we assessed relative binding affinity changes for reversion mutations of each variant residue, as well as alanine and/or phenylalanine mutations of Tyr<sup>28</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 and Arg<sup>53</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 (see Supplementary Information, Materials and Methods). As others have combined phage display with alanine scanning mutagenesis to create “shotgun” alanine scanning mutagenesis,26-28 we combined yeast display and reversion/replacement mutagenesis, as we had performed previously with another TCR/SAG interaction (hV<sub>β</sub>2.1 in complex with toxic shock syndrome toxin-1<sup>β</sup>), for a facile and efficient method for the energetic evaluation of individual amino acid residues, or individual atoms thereof, in an evolved protein (Fig. 3a).

Using this approach, we found that several mutations in both the CDR1 loop and the CDR2 loop of G5-8 were energetically important for complex formation (Fig. 3b). Specifically, reversion mutations at CDR1 position 28 (Tyr<sup>28</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 → Asn; Fig. 3b, red) and CDR2 positions 52 through 54 (Ile<sup>52</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 → Ala, Arg<sup>53</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 → Gly, and Asn<sup>54</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 → Ser; Fig. 3b, blue) resulted in significant reductions in binding affinity when displayed on the yeast surface in the context of the G5-8 background.

To further dissect the molecular basis of affinity maturation in the CDR1 loop, we performed a similar mutagenesis analysis with Tyr<sup>28</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 → Phe and Tyr<sup>28</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 → Ala mutations. These two replacement mutations abrogate the hydrogen-bond and pi-stacking interactions, respectively, observed in the crystal structure (Fig. 2e). These assays indicated that the binding energy ascribed to Tyr<sup>28</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 is derived primarily from the pi-stacking interaction between its phenyl ring and Arg<sup>110</sup><sup>SEB</sup> and not from the hydrogen bond formed between its hydroxyl group and Asn<sup>60</sup><sup>SEB</sup> (Fig. 3b, red). Additionally, we observed no relative change in binding for the Ser<sup>27a</sup>G5<sup>G5</sup>8<sup>mV</sup>β8.2 → Ala mutation, confirming that this inserted residue does not itself make energetically significant interactions with SEB. Instead, the single-residue insertion at this position...
probably acts as a spacer to lengthen the CDR1 loop such that Tyr28 can form energetically productive contacts with SEB (see the text below).

The contiguous stretch of CDR2 loop residues 52 through 54 is critically important for affinity maturation (Fig. 3b, blue). The reversion mutation at position 53 contributes most significantly to the affinity maturation process, as might be expected from the ~210-Å increase in buried surface area that results from Arg53 relative to that from Gly53. Likewise, the ~75-Å increase in buried surface area that results from mutating Ala52 to Ile52 makes a significant contribution to binding in the G5-8/SEB complex. Although Asn54 makes no intermolecular contacts with SEB, it may contribute to the affinity maturation process, perhaps through intramolecular interactions that act to stabilize the conformation of the G5-8 CDR2 loop. According to our mutational analysis, several other residues in both the CDR1 loop and the CDR2 loop (including Lys24 and Val55) may contribute to the affinity maturation process in a similar, although less significant, manner (Fig. 3b, gray). The molecular basis of affinity maturation by these residues is uncertain, but it may be due to effects on the conformational flexibility of the CDR loops, as suggested by an Arg53-to-Ala53 mutation that we observed to have an intermediate affinity between arginine and glycine residues at position 53. Based on our crystal structure, it is unlikely that the Ala53 Cα atom can contact SEB, indicating that CDR2 loop entropy and/or flexibility may contribute to G5-8/SEB binding.

In addition to the experimental mutational analysis described above, we performed a computational analysis of the same set of individual reversion and replacement mutations using the Rosetta program (see Supplementary Information, Materials and Methods). The results of this computational analysis strongly corroborate the experimental results. A plot of Rosetta ΔΔG scores versus experimentally
Fig. 3 (legend on next page)
measured binding free energies for all 18 CDR1 and CDR2 loop mutants (Fig. 3c) exhibits a correlation coefficient of 0.91. As with the experimental analysis, the computational analysis clearly distinguished those mutations that had profound versus insignificant binding effects and clearly implicated the pi-stacking interaction of Tyr28\(^{\text{CS-8}}\) to be of greater energetic importance than the hydrogen bond formed between its hydroxyl group and Asn66\(^{\text{SEB}}\) (Supplementary Table 1). Only those residues for which we observed a small energetic effect experimentally, including Lys24\(^{\text{CS-8}}\) and Val55\(^{\text{CS-8}}\), were in poor agreement with the computational results. These residues do not make specific intermolecular contacts with SEB (Fig. 2e) and, thus, their energetic effects likely involve backbone conformational changes, as mentioned above, which are more difficult to model using computational algorithms.\(^{20}\)

Computationally, we also assessed why the CDR1 residues in G5-8 may have given rise to a higher affinity and, therefore, selection in the final round of the directed evolution process. We found that, without the "spacer" residue Ser27\(^{\text{CS-8}}\), the tyrosine residue at position 28 makes very few, and no energetically favorable, contacts with SEB (Supplementary Table 2), supporting the need for CDR1 loop extension to achieve increased affinity. In addition, computational analysis revealed very few amino acids at position 28, other than tyrosine, to be energetically favorable in an extended CDR1 loop, with only tryptophan, phenylalanine (verified experimentally), and glutamic acid predicted to result in affinities commensurate to tyrosine at that position (Supplementary Table 2). The tryptophan is predicted to have favorable packing interactions in an orientation similar to that of Tyr28, while the modeled glutamic acid adopts a conformation for maximal electrostatic interactions with the positively charged side chain of Arg110 on SEB (Supplementary Fig. 4).

The structural and energetic changes that arise from our structure-based semirational directed evolution approach are entirely compatible with our original rationale for modifying the standard engineering strategy. Just as in the wild-type hV\(^{\text{2.1}}\)/SpeC structure,\(^{20}\) we find that lengthening the CDR1 loop of G5-8 by a single additional residue pushes a residue C-terminal to the insertion site closer towards SEB, with which it can form intermolecular contacts that significantly increase binding affinity. The evolution step of this approach is still required, however, as simply increasing the length of the wild-type CDR1 loop would not provide for these energetically productive interactions, since the wild-type residue Asn28\(^{\text{CS-8}}\) would be unable to form similar pi-stacking or hydrogen-bond interactions with SEB. Thus, by rationalizing the directed evolution process in a structure-based manner to augment its evolutionary power, we have achieved an unprecedented level of affinity maturation in a protein–protein interaction that, in turn, resulted in a highly effective protein therapeutic.

**Accession code**

Coordinates and structure factors have been deposited in the Protein Data Bank under accession code 3R8B.

**Acknowledgements**

We thank Dr. Vivian Stojanoff for technical assistance at beamline X6A of the National Synchrotron Light Source. Use of the National Synchrotron Light Source, Brookhaven National Laboratory, was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. DE-AC02-98CH10886. This work was supported, in part, by National Institutes of Health grants AI065690 (to E.J.S.), AI064611 (to D.M.K.), and GM084884 (to Z.W.).

**Supplementary Data**

Supplementary data to this article can be found online at doi:10.1016/j.jmb.2011.06.009

**References**


---

**Fig. 3.** Energetic dissection of the affinity-increasing mutations. (a) Titration binding curves for G5-8 and reversion/replacement mutants thereof, assayed by yeast display and normalized to G5-8 yeast surface expression. (b) Binding affinities of G5-8 and mutants thereof, as measured by yeast display flow titration. Energetically significant mutations are shown in red for CDR1 loop residues and in blue for CDR2 loop residues. Additional mutations that may potentially contribute to the affinity maturation process are shown in gray. (c) Computational analysis versus mutational analysis. Rosetta scores from computational analysis are plotted versus measured binding free energy changes (\(\Delta G\)), which were calculated as \(RT \ln (K_{\text{mut}}/K_{\text{wt}})\), with \(K_{\text{wt}}\) and \(K_{\text{mut}}\) from values in (b). The correlation coefficient (\(R\)) is 0.91. Color coding is the same as in (b).


