Deep Mutational Scans as a Guide to Engineering High Affinity T Cell Receptor Interactions with Peptide-bound Major Histocompatibility Complex*

Received for publication, July 15, 2016, and in revised form, September 15, 2016. Published, JBC Papers in Press, September 28, 2016, DOI 10.1074/jbc.M116.748681

Daniel T. Harris§, Ningyan Wang†, Timothy P. Riley§1, Scott D. Anderson†, Nishant K. Singh§, Erik Procko‡, Brian M. Baker§, and David M. Kranz†,‡

From the †Department of Biochemistry, University of Illinois, Urbana, Illinois 61801 and the ‡Department of Chemistry and Biochemistry and the Harper Cancer Research Institute, University of Notre Dame, South Bend, Indiana 46557

Edited by Peter Cresswell

Proteins are often engineered to have higher affinity for their ligands to achieve therapeutic benefit. For example, many studies have used phage or yeast display libraries of mutants within complementarity-determining regions to affinity mature antibodies and T cell receptors (TCRs). However, these approaches do not allow rapid assessment or evolution across the entire interface. By combining directed evolution with deep sequencing, it is now possible to generate sequence fitness landscapes that survey the impact of every amino acid substitution across the entire protein-protein interface. Here we used the results of deep mutational scans of a TCR-peptide-MHC interaction to guide mutational strategies. The approach yielded stable TCRs with affinity increases of >200-fold. The substitutions with the greatest enrichments based on the deep sequencing were validated to have higher affinity and could be combined to yield additional improvements. We also conducted in silico binding analyses for every substitution to compare them with the fitness landscape. Computational modeling did not effectively predict the impacts of mutations distal to the interface and did not account for yeast display results that depended on combinations of affinity and protein stability. However, computation accurately predicted affinity changes for mutations within or near the interface, highlighting the complementary strengths of computational modeling and yeast surface display coupled with deep mutational scanning for engineering high affinity TCRs.

The process of increasing the affinity of a protein occurs naturally with antibodies, where somatic mutation within the variable region genes is followed by antigen-driven selection of B cells that express membrane-bound antibodies. In contrast, T cell receptors (TCRs) do not undergo somatic mutations and bind to their antigen, a peptide-MHC (pepMHC), with low (micromolar) affinities. However, improvements in TCR affinity to the same levels of antibodies can be achieved by in vitro approaches involving the generation of mutant TCR libraries followed by antigen selection (1–3).

For therapeutic purposes, the affinity of a variety of protein-protein interactions, and especially antibody-antigen interactions, has been enhanced using in vitro directed evolution approaches, including phage, yeast, ribosomal, and mammalian display (e.g. see Refs. 4–7). These methods rely on the generation of large libraries of mutants at residues within the protein-protein interface, followed by several rounds of selection for desired parameters (such as affinity, stability, and expression levels) (8, 9).

Although directed evolution using larger degenerate libraries has been successful, the most recent techniques involving deep sequencing of single-codon libraries have the potential both to provide mechanistic structural information about a binding site and at the same time to provide leads for affinity improvements. Sequence fitness landscapes have successfully been utilized to map protein-DNA interactions (10), protein-peptide interactions (11), and protein-protein interactions (12). Furthermore, using a PDZ protein domain as a model system, McLaughlin and colleagues were able to manipulate ligand-binding specificity through key mutations identified using sequence fitness landscapes (13). Additionally, a sequence fitness landscape of an influenza-binding protein inhibitor was used to enhance affinity and redirect specificity toward a single H1 hemagglutinin subtype (14).

A major goal in protein engineering is to be able to accurately identify mutations that yield improvements in stability or affinity. In addition to directed evolution and sequence fitness landscapes, when the protein structures are known, structure-based computational design has also been used to achieve these goals. Although there have been inspiring successes (reviewed in Ref. 15), advances in computational approaches require a thorough understanding of the rela-

* This work was supported by National Institutes of Health Grants CA178844 and CA187592 (to D. M. K.), GM118166 (to B. M. B.), and CA180723 (to D. T. H.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

† Supported by a fellowship from the Indiana CTSI, funded in part by National Institutes of Health Grant UL1TR001108.

‡ To whom correspondence should be addressed: Dept. of Biochemistry, University of Illinois, 600 S. Matthews Ave., Urbana, IL 61801. Tel.: 217-244-2821; E-mail: d-kranz@illinois.edu.

§ The abbreviations used are: TCR, T cell receptor; pepMHC, peptide-MHC; CDR, complementarity-determining regions; scFv, single-chain fragment variable; MFI, mean fluorescence intensity; SOE, splice overlap extension; RMSF, root mean square fluctuation.

This is an Open Access article under the CC BY license.
The A6 TCR was used previously as a template for directed evolution of a novel peptide specificity, whereby the specificity was switched from Tax-HLA-A2 to the completely distinct cancer peptide, MART1-HLA-A2 (21). The resulting TCR, called RD1-MART1, was further affinity-matured to yield the RD1-MART1\textsuperscript{HIGH} TCR, with a $K_D$ value of about 100 nM. Recently, we generated a sequence fitness landscape of RD1-MART1\textsuperscript{HIGH} by deep sequencing analysis of antigen-selected, single-codon libraries (25). For the sequence fitness landscape and the work conducted in the present study, RD1-MART1\textsuperscript{HIGH} and various mutants were displayed on the surface of yeast as single-chain TCRs containing the V$\beta$ and V$\alpha$ domains linked by a flexible linker (Fig. 1A). HA and c-Myc tags were attached at the N terminus and C terminus, respectively.

To understand the most expeditious strategy to generate higher affinity binding leads from a protein, here we explored three approaches. The first two were based on the substitutions from the single-codon libraries that yielded increases in the number of antigen-selected TCR, characterized as enrichment values (Fig. 1B). In the third, the results of these experimental approaches were compared with an \textit{in silico} computational approach using the structure of the RD1-MART1\textsuperscript{HIGH}, MART1-HLA-A2 complex to predict the structural consequences and the energetic impacts of every individual substitution in the CDRs.

The individual enrichment values for the top substitutions of the RD1-MART1\textsuperscript{HIGH} TCR are plotted in Fig. 1C. To explore the robustness of these results from the sequence fitness landscape, we tested eight individual mutants (arrows in Fig. 1C) within this range of enrichment values, from 10- to 51-fold (T91Q, A99Y, V50D, G28E, S51G, Y50W, N28G, and K97Q). The top binders among this group were then tested in combination to determine whether there was binding additivity (Fig. 1B, left).

It is possible that the individual mutation approach, followed by combination of improved mutations, will not identify those collections of mutations that act additively or cooperatively to yield the highest affinity TCRs. Furthermore, positions where there were multiple enriched substitutions (e.g. positions 91$\alpha$ and 99$\beta$) would not be fully sampled by analyzing only eight of the individual mutations. Accordingly, we took an alternative approach in which a combinatorial library at four residues with the highest enrichment values was generated. This combinatorial library was sorted with MART1-HLA-A2 by FACS (Fig. 1B, right) to yield clones that bound with higher affinity.

\textit{Site-specific Mutants Act Individually and in Combination to Enhance Affinity and Expression}—Based on the sequence fitness landscape of RD1-MART1\textsuperscript{HIGH}, 8.3% of the mutations showed enrichment above the wild-type TCR (25). As indicated, eight mutations with enrichment values ranging from 10- to 51-fold (Fig. 1C, arrows) were expressed individually on the surface of yeast and tested for binding to the MART1-HLA-A2 ligand. Interestingly, the four residues with the highest enrichment values were each in different CDR loops (Fig. 2A), raising the possibility that the combined mutations
could yield additive effects. Mutants on yeast were stained with various concentrations of MART1-HLA-A2 monomer to determine affinities (Fig. 3A and Table 1). Previous studies with TCR domains or scFv fragments have shown that $K_D$ values determined by such flow cytometry-based titrations are similar to $K_D$ values determined for the soluble proteins by surface plasmon resonance, at least in the nanomolar $K_D$ range (e.g. see Refs. 18, 28, and 29).

In some cases, we also stained yeast with an excess of MART1-HLA-A2 tetramer to compare relative surface levels of the TCR mutants (Table 1) because surface levels of properly folded TCRs in the yeast display system are known to be influenced by stability of the proteins (18, 30–36), and the multimeric form would be able to assess total surface levels even with lower affinity mutants, due to avidity (37). We also used high levels of the multimeric ligand for this purpose, rather than antibodies to the C-terminal c-Myc tag, because the yeast quality control machinery may allow expression of full-length fusions even if misfolded or partially folded domains are present (38). This would allow us to address the possibility that the enrichment of TCRs with mutations in the single-codon libraries was due to enhanced affinity, higher surface levels of the folded protein, or both.

Of the eight mutations, five showed significant improvements in affinity compared with the RD1-MART1HIGH TCR (Fig. 3A and Table 1). The five mutants exhibited 3-fold (Y50W/H9251) to 11-fold (A99Y/H9252) increases in affinity above the wild-type protein. Although the other three mutants did not exhibit affinity increases, they did appear to be expressed at higher surface levels, which probably accounts for the enhancement values observed in the deep mutational scan.

To test whether the mutations could be combined to yield further improvements, four double mutants and a triple mutant were generated (Figs. 3 and 4 and Table 1). The double mutants yielded increases of 10–20-fold, whereas the triple mutant yielded an increase of almost 100-fold, to an estimated $K_D$ value of 2 nM. Furthermore, combination of these mutations also led to increased TCR surface levels (Table 1), indicating that the...
mutations also may operate by stabilizing the TCR domains. A plot of affinities \((K_D)\) values \textit{versus} enrichment value (Fig. 4C) supports the view that the highest enhancement values in a deep mutational scan can be used to reliably generate higher affinity variants of a protein. Indeed, when the eight mutants were viewed as changes in binding free energy relative to “wild-type” RD1-MART1\(^{HIGH}\), there was a linear relationship where multiple substitutions yield high enrichment values). Indeed, when the eight mutants were viewed as changes in binding free energy relative to “wild-type” RD1-MART1\(^{HIGH}\), there was a linear relationship where multiple substitutions yield high enrichment values). Indeed, when the eight mutants were viewed as changes in binding free energy relative to “wild-type” RD1-MART1\(^{HIGH}\), there was a linear relationship where multiple substitutions yield high enrichment values). Indeed, when the eight mutants were viewed as changes in binding free energy relative to “wild-type” RD1-MART1\(^{HIGH}\), there was a linear relationship where multiple substitutions yield high enrichment values).

**Selection of a Combinatorial Library Yields Further Improvements and Correlates with the Sequence Fitness Landscape**—To compare with individual site-specific mutagenesis, we generated a combinatorial library at the four residues \((28\alpha, 91\alpha, 50\beta,\) and 99\(\beta)\) with the most highly enriched mutations in the deep mutational scan (Figs. 1C and 2A). Selection of this library could yield improved combinations that were not identified in the one-by-one mutational approach. The combinatorial library was sorted three times with MART1-HLA-A2 monomer at concentrations ranging from 1 \(\mu\)M to 1 nM, and the top 1% fluorescent population was collected (Fig. 5A). After the third sort, eight clones were sequenced, and there were three unique sequences identified (represented by clones S3-2, S3-7, and S3-8) (Table 2). Titrations of these three clones were performed and compared with the triple mutant described above (Fig. 5, B and C). Mutants S3-7 and S3-8 were similar in affinity to the triple mutant, but clone S3-2 consistently appeared to have a slightly higher affinity and a higher surface level (Table 1). Based on this result, the combinatorial library/selection approach can yield modest but significant improvements above the single mutation approach (\(e.g.) when there are positions where multiple substitutions yield high enrichment values). The sequences of the mutants were interesting when considered with regard to the sequence fitness landscape and the triple mutant. For example, the T91Q mutation had one of the highest affinities and expression levels among individual mutants (Table 1), and this mutation was present in all three clones from the combinatorial library. Position 50\(\beta\) had either an aspartic or glutamic acid in all three clones identified in the combinatorial library. Position 50\(\beta\) had either an aspartic or glutamic acid in all three clones isolated, consistent with the nearly 4-fold improvement of V50D\(\beta\) (Table 1) and the deep mutational scan results (Fig. 1C). Although all three clones identified in the combinatorial library did not contain the A99Y\(\beta\) mutation that showed the highest affinity individually (Table 1), each had a different highly enriched mutation (A99D\(\beta\), A99E\(\beta\), and A99V\(\beta\)) (Fig. 1C). Interestingly, in clones S3-2 and S3-8, the Gly-28\(\alpha\) position was mutated to arginine despite this residue having a strong preference for glutamic acid in the deep mutational scan (Fig. 1C) and the improvement in affinity of the G28E\(\alpha\) mutant. In fact, the G28R\(\alpha\) mutation was not enriched compared with wild type in the deep mutational scan, suggesting that this mutation may have operated only in the context of one or more of the other three CDR mutations.

**FIGURE 2. Structure of RD1-MART1\(^{HIGH}\) TCR-MART1-HLA-A2 complex and models for affinity-enhancing mutations.** A, view of the TCR-MART1-HLA-A2 complex with the four CDR residues with the highest enrichment values highlighted in yellow. B, close-up view of Val-50\(\beta\) (left) and the modeled Asp-50\(\beta\) structure (right). Asp-50\(\beta\) is predicted to introduce a long range electrostatic interaction with Arg-65 of the HLA-A2 \(\alpha\) helix. C, close-up view of Thr-91\(\alpha\) (left) and the modeled Gln-91\(\alpha\) structure (right). Gln-91\(\alpha\) is predicted to form an electrostatic interaction with Arg-94\(\beta\) of the neighboring CDR3\(\beta\) loop as well as to form an intraloop hydrogen bond with the backbone of Tyr-93\(\alpha\). Both could stabilize the loop in a binding-competent conformation. D, close-up view of Ala-99\(\beta\) (left) and the modeled Tyr-99\(\beta\) structure (right). Tyr-99\(\beta\) is predicted to pack against Thr-9 of the peptide and Val-76 of the HLA-A2 \(\alpha\) helix. E, close-up view of Gly-28\(\alpha\) (left) and the modeled Glu-28\(\alpha\) structure (right). Glu-28\(\alpha\) is predicted to interact with Trp-167 of the HLA-A2 \(\alpha\)2 helix.
TABLE 1

Affinity and MFIMAX ratios of RD1-MART1HIGH mutations identified from sequence fitness landscape

| TCR          | \( K_D \) (nm) | \( K_D \) (S.D.) | MFIMAX ratio (tetramer) | MFIMAX ratio (monomer) |
|--------------|----------------|-----------------|-------------------------|------------------------|
| RD1-MART1HIGH| 5              | 137             | 126                     | 1                      |
| G28Eα        | 2              | 62              | 27                      | 1.2                    |
| T91Qα        | 3              | 52              | 23                      | 1.6                    |
| V50Dβ        | 2              | 66              | 32                      | 0.8                    |
| A99Yβ        | 3              | 30              | 8                       | 0.7                    |
| Y50Wα        | 3              | 76              | 21                      | 0.6                    |
| N28Gβ        | 2              | 218             | 56                      | 1.3                    |
| S51Gα        | 2              | 129             | 55                      | 0.7                    |
| K97Qα        | 2              | 188             | 54                      | 2.3                    |
| A99Wβ        | 2              | ND              | ND                      | ND                     |
| S94Ma        | 2              | ND              | ND                      | ND                     |
| G28Wα        | 2              | ND              | ND                      | ND                     |
| G28Eα        | 2              | ND              | ND                      | ND                     |
| G28Eα/T91Qα  | 2              | 10              | 3                       | 5.3                    |
| G28Eα/V50Dβ  | 2              | 18              | 8                       | 4.1                    |
| G28Eα/A99Yβ  | 2              | 15              | 3                       | 3.7                    |
| T91Qα/A99Yβ  | 2              | 16              | 14                      | 3.9                    |
| G28Eα/T91Qα/A99Yβ | 4 | 2.1          | 2.5                      | 3.2                    |
| S3-2         | 4              | 3.7             | 0.3                     | 5.5                    |
| S3-7         | 3              | 3.3             | 1.6                     | 1.7                    |
| S3-8         | 3              | 1.3             | 1.2                     | 3.2                    |

In conclusion, the combinatorial approach yielded the greatest improvement, as shown by the S3-2 mutant, which exhibited a \( K_D \) value of 0.7 nm, an increase in affinity of about 250-fold compared with RD1-MART1HIGH. The improvement in affinity of the S3-2 mutant was also validated by a separate approach using soluble single-chain TCRs refolded from Escherichia coli inclusion bodies. Previous surface plasmon resonance measurements of soluble RD1-MART1HIGH TCR yielded a \( K_D \) value of 250 nm (21). Here we examined soluble S3-2 TCR by surface plasmon resonance, which yielded a \( K_D \) value of 2.2 nm (Fig. 5D).

Evaluation of the Sequence Fitness Landscape and Individual Mutations via an in Silico Approach—We next evaluated the same residues from the sequence fitness landscape computationally, examining all 20 amino acids at the 39 CDR loop positions explored in the RD1-MART1HIGH TCR. We used our previously described in silico structure-guided design approach and scoring function developed and trained using TCR-peptide-MHC binding data (24, 39). (Note that values of <1 are considered candidates for improved affinity). There was a general, qualitative agreement between prediction and experiment, because the majority (92%) of the mutations predicted to weaken affinity were also selected against (Fig. 6A). However, 25% of the negatively enriched mutations were also predicted to improve binding.

The majority (78%) of the negatively enriched mutations predicted to improve binding were mutations that reduced polarity, with 62% involving introduction of a hydrophobic amino acid. This suggests that some of the discrepancy between computational prediction and sequence enrichment can be attributed to changes in protein stability. As noted above, this is a well characterized selection parameter in yeast display (18, 30–36) but is not an output of a computational design approach focused solely on improving affinity. A good example is found in the mutation of Ala-99β. Mutation of Ala-99β to Tyr improved affinity, as shown in Table 1. Mutation of Ala-99β to Trp was predicted to have an even more substantial impact on affinity because the modeling indicated the Trp side chain would pack more tightly against the peptide. However, substitution of Ala-99β with Trp is predicted to result in the added exposure of ~50 Å² of hydrophobic surface area when the TCR is displayed on the yeast surface (compared with substitution with Tyr). Using well recognized estimates of 20–30 kcal/mol in free energy per Å² of exposed hydrophobic surface area (40, 41), the resultant destabilization could be as high as ~1.5 kcal/mol.

A wider comparison of enrichment versus predicted changes in hydrophobic solvent-accessible surface further suggested that sequence enrichment in the fitness landscapes was at least in part associated with reduced exposed hydrophobic surface area, and vice versa (Fig. 6B). Of the residues with log₂ enrichment ratios ≥1, 68% are predicted to have reduced solvent-
exposed hydrophobic surface. In contrast, for those residues with enrichment ratios ≤1, 66% are predicted to have increases in exposed hydrophobic surface.

We attempted to compute changes in protein stability and incorporate them into binding predictions, as performed previously (14). We used multiple approaches to compute stability changes, including quantifying energetic costs of changes in hydrophobic surface exposure and computing stability using both our TCR-specific and more general scoring procedures. These approaches, however, did not provide additional insight beyond that obtained from qualitatively assessing hydrophobicity, as noted above and in Fig. 6C. This result may be related to recognized difficulties in computing protein stability (42), especially in loop structures, as well as an intrinsically low sta-

---

**FIGURE 4. Titrations of mutants of RD1-MART1HIGH identified from the sequence fitness landscape.** A and B, MFI's of the indicated mutants were plotted against the concentration of MART1-HLA-A2 monomer. C, the enrichment values of RD1-MART1HIGH and the indicated mutants were plotted against the calculated binding disassociation constant (affinity, in nM) from titrations with yeast and MART1-HLA-A2 monomer. Error bars are indicative of the range of maximum and minimum values from multiple titrations (RD1-MART1HIGH, n = 5; T91Qα and Y50Wα, n = 3; N28Gβ, S51Gα, K97Qα, G28Eα, V50Dβ, and A99Yβ, n = 2). Binding constants were calculated using non-linear regression of the binding curve (one site-specific binding in GraphPad) based on the highest concentrations of pepMHC as an estimate for maximum MFI. For the N28Gβ and K97Qα mutants, a maximum MFI could not be determined strictly from pepMHC monomer staining; therefore, maximum MFI was estimated based on the ratio of tetramer binding to A99Yβ, for which a maximum MFI was reached with monomer. Inset, a plot of the ΔΔG values calculated for each mutant relative to the wild-type RD1-MART1HIGH interaction, versus enrichment value. The dashed line shows a linear regression to the data. The correlation coefficient (r) is −0.82.
bility of the single-chain RD1-MART1HIGH protein coupled with kinetic, rather than thermodynamic, stability effects that may lead to low yeast surface expression levels (43).

To test whether reduced yeast surface levels of the properly folded mutants were in part responsible for lower enhancement values, we examined four individual mutants. These mutants were predicted to have higher affinities, yet they were observed to have negative enrichment values (Fig. 6A, green). The four mutants (A99Wβ, S94Mα, G28Wα, and G28Fα) were compared with the wild-type RD-MART1HIGH TCR for binding to MART1-HLA-A2 monomers (see Fig. 7A for mutants and Fig. 3A for wild type) and tetramers (Table 1) using flow cytometry. Consistent with the notion that negative enrichment values could be due to lower surface levels of the properly folded proteins, all four mutants, and especially the S94Mα, G28Wα, and G28Fα mutants, were expressed at significantly lower surface levels than wild-type RD-MART1HIGH TCR. Although the A99Wβ mutant appeared to have a similar affinity to the wild-type RD-MART1HIGH TCR, based on the monomer titrations (Fig. 7B), it was not possible to estimate affinities for the other three mutants due to their low surface levels.

**Computational Analysis of Mutations That Yielded Higher Affinities**—If we restrict the focus solely to mutations with binding measurements, after computing binding relative to RD-MART1HIGH using the values in Table 1, moderate agreement between computation and experiment is observed, with a correlation coefficient of 0.68 (Fig. 8A). This analysis, however, includes the N28G and K97Q mutations. Neither of these positions interacts directly with pepMHC either in the wild-type complex or in the models of the mutants through.
either short or long range interactions (the closest predicted approaches for the N28Gβ and K97Qα mutants are 16 and 11 Å, respectively). Excluding these mutants from the comparison substantially improved the correlation between prediction and experiment \( (r = 0.91) \). Prediction scores of the double and triple mutants also corroborated well with experiments, with a correlation coefficient of 0.87 when these were included in the comparison (Fig. 8B).

The accurate recapitulation of the relative binding energies of the RD1-MART1\textsuperscript{HIGH} mutants in Table 1 suggested that the structural models would be of use in rationalizing how the mutations acted to impact binding. The V50Dβ mutation, which improved binding nearly 4-fold, may function by introducing an electrostatic interaction with the side chain of Arg-65 of HLA-A2 (Fig. 2B). The T91Qα mutation occurs in the Vα/Vβ interface and may stabilize the CDR3α loop through an intraloop hydrogen bond and electrostatic interactions within the neighboring CDR3β loop (Fig. 2C). A99Yβ is the most favored mutation and is predicted to improve packing with the peptide and the HLA-A2 α1 helix (Fig. 2D). The new interactions for A99Yβ are largely non-directional contacts, perhaps accounting for why multiple substitutions here are both highly enriched and favorably predicted. G28Eα is predicted to form new van der Waals and electrostatic interactions with Trp-167 of HLA-A2 (Fig. 2E).

It is notable that, with the exception of A99Yβ, the distance of these positions from the interface would tend to place them as a lower priority for investigation in typical structure-guided design studies, with the V50Dβ mutation serving as the best example. This highlights the strength of deep sequencing and yeast surface display in identifying important affinity-enhancing “second shell” mutations, which for TCRs could be important in preserving or even improving antigen specificity (25).

**Discussion**

Proteins often require affinity and stability enhancements for therapeutic purposes. To achieve desired characteristics, various protein-engineering techniques that use in vitro directed evolution have been employed. These include ribosomal, phage, bacterial, yeast, and mammalian display of degenerate mutational libraries, followed by various selection
schemes. More recently, it has been possible to conduct deep mutational scans of every residue in a protein’s binding site to construct a sequence fitness landscape. Here we used information from the sequence fitness landscape of a TCR (RD1-MART1<sup>HIGH</sup>) specific for the cancer antigen MART1/H<sub>18528</sub>HLA-A<sub>2</sub> as a guide to compare various approaches to engineer improvements in the affinity and/or stability. We were able to identify specific mutants that enhanced the affinity 200-fold and increased the yeast surface level expression by 6-fold. Furthermore, we compared these site-specific mutants with structure-based predictions using an in silico approach that modeled the effects of the mutations and predicted the impacts on binding.

To determine the robustness of identifying higher affinity mutations strictly from the results of the sequence fitness landscape, we generated single site-specific mutations with enrichment values ranging from 10- to 51-fold. These mutations were tested as individual and combined mutants. Site-specific mutagenesis was able to increase the TCR affinity by 100-fold when three of the highest affinity mutants were combined. A previous study that aimed to improve the affinity of an influenza inhibitor reported a maximum affinity improvement of 28-fold using a combination of eight mutations identified from a sequence fitness landscape (14).

Although the individual mutations with the highest enhancement values in the sequence fitness landscape yielded higher affinity, and their combinations provided even higher affinity (100-fold), we were interested in determining whether a directed evolution approach based on the same data could yield combined mutations with even higher affinity or stability. This combinatorial approach would include variants with multiple mutations that have additivity or positive cooperativity in binding. To test this, we generated a combinatorial library at the four most highly enriched residues (Gly-28<sup>e</sup>, Thr-91<sup>a</sup>, Val-50<sup>b</sup>, and Ala-99<sup>b</sup>). Three successive selections of this library yielded multiple clones that bound with higher affinity than RD1-MART1<sup>HIGH</sup> to the MART1-HLA-A<sub>2</sub> ligand. One of these clones (S3-2) had the highest affinity (700 pM) and greatest surface level (6-fold above RD1-MART1<sup>HIGH</sup>).

The sequences of this and the other two unique clones revealed that three of the four positions in the library had mutations that were highly enriched in the sequence fitness landscape. Unexpectedly, the Gly-28<sup>e</sup>
position had two clones, including S3-2, with an arginine substitution, whereas the G288Ra substitution was not enriched in the sequence fitness landscape. Because G288Ra is predicted to interact strongly with Glu-1 of the peptide as well as Glu-58 of HLA-A2, this mutation alone may destabilize the TCR, a consequence offset by the other mutations. Accordingly, the use of sequence fitness landscapes to generate higher affinity mutants can benefit by subsequent selections of combinatorial libraries in the codons of enriched amino acid positions.

Our recently described structure-guided computational design approach for engineering TCRs (24, 39) provided an opportunity to evaluate the sequence fitness landscapes computationally and compare enrichment with predicted impacts on binding. Although there was qualitative agreement between those mutations predicted to weaken binding and sequence depletion, the in silico approach did not distinguish well between mutations predicted to improve binding and sequence enrichment, because 25% of mutations selected against were predicted to enhance binding. Much of this may reflect the sensitivity of yeast surface display to changes in protein stability, resulting in some substitutions that are either reduced or enhanced in surface levels. This is supported by our analysis of the modeled structures of each mutant for changes in exposed hydrophobic surface area; sequences enriched showed a much stronger trend to reduce hydrophobic solvent-accessible surface area, and vice versa. There can be a tendency for structure-guided design to select for hydrophobic mutations (12, 24), and our findings reiterate that improvements in in silico approaches may be found from continued attention to electrostatic features, such as polar solvation, hydrogen bonding, and interactions with formal charges (16, 46–49).

For those mutants studied directly, our in silico analysis showed very good agreement between predicted and impacted effects on binding, permitting the use of the structural models in helping to assess how the mutations acted to improve binding. A key observation was that multiple “second shell” mutations appeared to be enriched, and these most likely act via indirect or long range effects (exemplified by the V50D mutation). Such sites are not always considered in structure-guided computational design, highlighting a strength of sequence fitness landscapes and yeast surface display as well as further suggesting how in silico methods might be extended.

The single-codon approach used with sequence fitness landscapes also improves upon the many early studies using directed evolution and error-prone PCR techniques. Although the latter can sample the entire protein interface, it is limited to single-site mutations that are generated by a single base substitution. This is because the probability of having two substitutions in the same codon or in two codons with beneficial substitutions is quite low and often beyond the size of the libraries (or detrimental mutations at a higher error rate would obscure these mutations). As an example, the A99Yα mutant was one of the most highly enriched mutations of RD1-MART11HIGH, but it would require two nucleotide substitutions to mutate from alanine to tyrosine.

Finally, it is worth considering whether information about binding affinities of mutants gained in the yeast display system could be extrapolated to these TCRs expressed in their normal context, T cells. The TCRs are in single-chain form (Vβ-linker-Vα) as Aga2 fusions for yeast display, whereas in T cells, the Vα and Vβ each contain transmembrane-spanning constant regions. Despite these differences, we have invariably observed with other TCRs that the mutations yielding higher affinity in the yeast display system also yielded higher affinity when used as full-length TCRs transferred into T cells (18, 37, 50, 51). Thus, although not formally tested, we have reason to believe that the mutations of RD1-MART11HIGH TCRs identified in the present work would exhibit binding with enhanced affinities when transferred to the full-length TCR context.

In conclusion, deep mutational scans of the entire protein-protein interface provide physical insights into binding, and they will allow improvements in structure-guided in silico analyses. Sequence fitness landscapes also serve as a robust guide for conducting site-specific mutagenesis to enhance affinity. In the future, one might be able to simply combine all of the top enriched mutations into a single mutant and thus bypass the need for any selections. This might be especially true where the residues are located at different locations in the interface. However, the combinatorial library/selection approach, based on the results of the sequence fitness landscape, yielded an improved mutant compared with the single-site mutant strategy. Thus, there may still be advantages to using a directed evolution, selection-based strategy for optimizing improvements.

Experimental Procedures

Reagents and Flow Cytometry—The procedure for induction of yeast to express single-chain TCRs has been described previously (21). HLA-A2 heavy chain and β2-microgobulin were expressed as inclusion bodies in E. coli and refolded with a UV-cleavable peptide (KILGFVFJV) (52). At the C terminus of HLA-A2 heavy chain, a biotinylation site was included for the addition of biotin with BirA enzyme (Avidity). To exchange peptide, refolded HLA-A2 with UV-cleavable peptide was incubated with a 100-fold excess of MART1 10-mer peptide (ELAGIGILTV) during 45 min of UV exposure.

Yeast stained with HLA-A2 monomer were incubated with various concentrations of MART1-HLA-A2 monomers (at the lowest concentration, a minimum of 2-fold excess monomer relative to the total number of TCRs, estimated to be 50,000 receptors/cell, was used). Cells were washed and stained with streptavidin-PE (1:300) and subsequently analyzed by flow cytometry. To determine binding affinities, the mean fluorescence intensity (MFI) of secondary reagent control was subtracted from the MFI of the yeast cells stained with monomer. Binding constants were calculated using non-linear regression of the binding curve (one site-specific binding in GraphPad). Maximum MFI values were the MFI achieved at the highest concentration of monomer (for titrations that had a plateau), or, in cases where saturation was not reached (e.g. N28Gβ and K97Qα mutants), maximum MFI values were estimated from the ratio of tetramer binding to a TCR run at the same time that did exhibit a plateau. Titrations were performed a minimum of two times, in completely separate experiments.
Yeast were additionally stained with anti-HA (1:50) and anti-c-Myc (1:100), followed by secondary goat anti-mouse Alexa Fluor 647 (1:100) and goat anti-chicken Alexa Fluor 647 (1:100), respectively. For tetramer stains, yeast were stained with 50 nM MART1-HLA-A2 tetramer for 1.5 h, washed, and analyzed by flow cytometry.

Single-site Mutants and Combinatorial Library Design—Single-site mutants of the RD1-MART1<sup>HIGH</sup>-TCR in the pCT302 yeast display vector were created using the QuikChange II site-directed mutagenesis kit (Agilent Technologies). The identity of mutants was confirmed by sequencing. EBY100 yeast were subsequently transformed using a lithium acetate heat shock protocol (53). A four-codon combinatorial library of RD1-MART1<sup>HIGH</sup> spanning four CDR domains was created by splice overlap extension (SOE) PCR. Pre-SOE products were created using the following primers: Splice4/L, 5'-GGC AGC CCC ATA AAC CAC CAG TAT TAT-3'; T7 rev, 5'-TAA TAC GAC TCA CTA TAG GG-3'; V50b-fwd, 5'-TGG GCT TAA GAT TGA TTC ACT ATT CAN NKG GTG TTG GTA CAT CGA TCT TAA GCC CA-3'; A99b-fwd, 5'-TCC CGT CCG GGT TGG ATG NNK GGG GGG GTG GAA CTG T-3'; A99b-rev, 5'-CAT CCA ACC CGG AGC GGA-3'; T91a-fwd, 5'-GCC TCT TTA AAC TGT ACA CAT TCA GAT TGA AAT CAA TCT TAA GCC CAC-3'; G28a-rev, 5'-TCT ATC TGA ATA TGT TGT ACA TAT TCA GAT TGA AAT CAA TCT TAA GCC CAC-3'; G28a-fwd, 5'-GTG CTA TTC ACT ATT CAN NKG GTG TTG GTA CAT CGA TCT TAA GCC CA-3'. Pre-SOE products were combined by SOE PCR using the T7 and Splice 4L primers. Libraries were created using homologous recombination in EBY100 yeast as described previously (54). The theoretical diversity of the four-codon library is 32<sup>4</sup>, or 1 x 10<sup>6</sup>. Based on the number of independent transformants, the estimated size of the combinatorial library was 2 x 10<sup>7</sup>, which should provide coverage of all amino acid combinations of the library. The library was sorted three times with MART1-HLA-A2 monomer at 50, 0.5, and 0.1 nM, respectively, collecting the top 1% of fluorescent cells at each sort. After the third sort, eight high affinity clones were sequenced, of which three unique sequences were identified.

Characterization of Binding by Surface Plasmon Resonance—Recombinant single chain S3-2 TCR and MART1-HLA-A2 were generated from bacterially expressed inclusion bodies, refolded, and purified as described (25). Binding analysis was performed at 25 °C using surface plasmon resonance on a Biacore T200 instrument. 150 resonance units of S3-2 TCR was coupled to a CM5 sensor surface using standard amine coupling. MART1-HLA-A2 analyte was injected in a series of concentrations at a flow rate of 45 μl/min. Data were corrected for injections over a blank surface, and the responses versus time were fit globally to a 1:1 kinetic binding model using Biacore T200 evaluation software version 3.0.

Computational Prediction of Mutations—Structural modeling and energetic scoring followed closely the procedures described by Riley et al. (24). For modeling, Rosetta with the Talaris2013 score function was used (55–57), using the PyRosetta interface (58). The crystallographically absent linker region of the RD1-MART1<sup>HIGH</sup> structure was constructed and modeled using kinematic loop closure (59). Following linker modeling, the entire structure was relaxed through multiple cycles of backbone minimization and rotamer optimization (60). For comparison with the experimental sequence landscapes, 19 amino acids were introduced separately into each position, and the mutants plus the wild-type structures (760 in total) were modeled from three independent Monte Carlo-based simulated annealing trajectories. Modeling utilized Rosetta’s Loop_Mover_Refine_CCD mover with three outer cycles and 10 inner cycles, with an initial metropolis acceptance criterion of 2.2 that decreased linearly to 0.6 (61). For double and triple mutants, mutations were introduced simultaneously, followed by a minimum of six independent minimization trajectories to account for additional structural impacts. For all models, the large number of packing operations introduced some minor variability when scoring. Therefore, the unweighted score terms for the three trajectories were averaged and stored for calculations (62).

Each model was scored in the Rosetta interface with the TR3 score function trained on TCR-specific data. TR3 utilizes dynamic information from the wild-type structure as positional modifiers to account for protein flexibility. Accordingly, molecular dynamics simulations were calculated using the AMBER molecular dynamics suite (63). Starting coordinates for the complex used the relaxed structure described above, and the coordinates for the free TCR were generated by stripping away the pepMHC. After adding solvent and counterions, systems were energy-minimized and heated to 300 K with solute restraints. Solute restraints were gradually relaxed and followed with 2 ns of simulation with no solute restraints for equilibration, after which 100-ns production trajectories were calculated. Trajectory analysis for determining root mean square fluctuation (RMSF) values used ccpptraj from the AMBER suite. The RMSF values were used along with the additional terms of the TR3 function to score each of the 760 modeled structures. The scores of the 114 wild-type control structures determined the baseline score of 0.92 energy units, where lower scores indicated stronger affinity and higher scores indicated weaker affinity. For double and triple mutants, RMSF descriptors were averaged between the positions for scoring purposes (i.e. for double mutant XY, the RMSF of position X is averaged with the position Y RMSF to give an RMSF descriptor for double XY). Solvent-accessible surface areas were calculated using a 1.4-Å probe.

Calculations of protein stability included using structural models to calculate changes in exposed hydrophobic surface area and computing changes in stability using both the Talaris2013 and TR3 score functions. Computing binding probabilities as in Whitehead et al. (14) or by accounting for percentage of bound ligand using both folding and binding partition functions together did not yield more quantitative insight than simply considering changes in exposed hydrophobic surface area.

Author Contributions—D. T. H., N. W., T. P. R., S. D. A., and N. K. S. performed experiments and analyzed data. D. M. K., E. P., and B. M. B. conceived and organized the project. D. T. H., N. W., T. P. R., S. D. A., N. K. S., E. P., B. M. B., and D. M. K. wrote the manuscript. All authors reviewed the results and approved the final version of the manuscript.
Acknowledgments—We thank Barbara Pilas and Alvaro Hernandez for assistance with flow cytometry and deep sequencing, respectively, and Cory Ayres for performing molecular dynamics simulations.

References

1. Foote, J., and Eisen, H. N. (2000) Breaking the affinity ceiling for antibodies and T cell receptors. Proc. Natl. Acad. Sci. U.S.A. 97, 10679–10681
2. Holler, P. D., Holman, P. O., Shusta, E. V., O’Herrin, S., Wittrup, K. D., and Kranz, D. M. (2000) In vitro evolution of a T cell receptor with high affinity for peptide/MHC. Proc. Natl. Acad. Sci. U.S.A. 97, 5387–5392
3. Li, Y., Moysey, R., Molloy, P. E., Vuideat, A. L., Mahon, T., Baston, E., Dunn, S., Liddy, N., Jacob, J., Jakobsen, B. K., and Boulter, J. M. (2005) Directed evolution of human T cell receptors with picomolar affinities by phage display. Nat. Biotechnol. 23, 349–354
4. Huse, W. D., Saxtry, L., Iverson, S. A., Kang, A. S., Alting-Mees, M., Burton, D. R., Benkovic, S. J., and Lerner, R. A. (1989) Generation of a large combinatorial library of the immunoglobulin repertoire in phage alpha. Science 246, 1275–1281
5. Hanes, J., and Plückthun, A. (1997) In vitro selection and evolution of functional proteins by using ribosome display. Proc. Natl. Acad. Sci. U.S.A. 94, 4937–4942
6. Boder, E. T., and Wittrup, K. D. (2000) Yeast surface display for directed evolution of protein expression, affinity, and stability. Methods Enzymol. 328, 430–444
7. Chervin, A. S., Aggen, D. H., Raseman, J. M., and Kranz, D. M. (2008) Engineering higher affinity T cell receptors using a T cell display system. J. Immunol. Methods 339, 175–184
8. Hoogenboom, H. R. (2005) Selecting and screening recombinant antibody libraries. Nat. Biotechnol. 23, 1105–1116
9. Van Deventer, J. A., and Wittrup, K. D. (2014) Yeast surface display for antibody isolation: library construction, library screening, and affinity maturation. Methods Mol. Biol. 1131, 151–181
10. Shultzaberger, R. K., Malashock, D. S., Kirsch, J. F., and Eisen, M. B. (2010) The fitness landscapes of cis-acting binding sites in different promoter and environmental contexts. PLoS Genet. 6, e1000142
11. Fowler, D. M., Araya, C. L., Fleishman, S. J., Kellogg, E. H., Stephany, J. J., Malashock, D. S., Kirsch, J. F., and Eisen, M. B. (2010) Yeast polypeptide fusion surface display levels predict thermal stability. Proc. Natl. Acad. Sci. U.S.A. 107, 10651–10656
12. Shusta, E. V., Kieke, M. C., Sundberg, E., Shusta, E. V., Mariuzza, R. A., Wittrup, K. D., and Kranz, D. M. (2001) High affinity T cell receptors from yeast display libraries block T cell activation by superantigens. J. Mol. Biol. 307, 1305–1315
13. Kieke, M. C., Shusta, E. V., Boder, E. T., Tytton, L., Wittrup, K. D., and Kranz, D. M. (1999) Selection of functional T cell receptor mutants from a yeast surface-display library. Proc. Natl. Acad. Sci. U.S.A. 96, 5651–5656
14. Shusta, E. V., Kieke, M. C., Parke, E., Kranz, D. M., and Wittrup, K. D. (1999) Yeast polypeptide fusion surface display levels predict thermal stability and soluble secretion efficiency. J. Mol. Biol. 292, 949–956
15. Orr, B. A., Carr, L. M., Wittrup, K. D., Roy, E. J., and Kranz, D. M. (2003) Rapid method for measuring ScFv thermal stability by yeast surface display. Biotechnol. Prog. 19, 631–638
16. Starwal, S. E., Masteller, E. L., Bluestone, J. A., and Kranz, D. M. (2003) Directed evolution of a single-chain class II MHC product by yeast display. Protein Eng. 16, 147–156
17. Jones, L. L., Brophy, S. E., Bankovich, A. J., Colf, L. A., Hanick, N. A., Garcia, K. C., and Kranz, D. M. (2006) Engineering and characterization of a stabilized α1α2 module of the class I major histocompatibility complex product. J. Biol. Chem. 281, 25734–25744
18. Traxlmayr, M. W., Hasenhindl, C., Hackl, M., Stadlmayr, G., Rybka, J. D., Borth, N., Grillari, J., Rüker, F., and Obing, C. (2012) Construction of a stability landscape of the CH3 domain of human IgG1 by combining directed evolution with high throughput sequencing. J. Mol. Biol. 423, 397–412
19. Traxlmayr, M. W., and Obing, C. (2012) Directed evolution of proteins for increased stability and expression using yeast display. Arch. Biochem. Biophys. 526, 174–180
20. Stone, J. D., Artymov, M. N., Chervin, A. S., Chakraborty, A. K., Eisen, H. N., and Kranz, D. M. (2011) Interaction of streptavidin-based peptide-MHC oligomers (tetramers) with cell-surface TCRs. J. Immunol. 187, 6281–6290
21. Park, S., Wu, Y., Stowell, X. F., Gai, F., Saven, J. G., and Boder, E. T. (2006) Limitations of yeast surface display in engineering proteins of high thermostability. Protein Eng. Des. Sel. 19, 211–217
22. Smith, S. N., Sommermeyer, D., Piepenbrink, K. H., Blevins, S. I., Bernhard, H., Uckert, W., Baker, B. M., and Kranz, D. M. (2013) Plasticity in the contribution of T cell receptor variable region residues to binding of peptide-HLA-A2 complexes. J. Mol. Biol. 425, 4496–4507
23. Smith, S. N., Wang, Y., Baylon, J. L., Singh, N. K., Baker, B. M., Tajkhorshid, E., and Kranz, D. M. (2014) Changing the peptide specificity of a human T-cell receptor by directed evolution. Nat. Commun. 5, 5223
24. Schmitt, T. M., Aggen, D. H., Stromnes, I. M., Dossett, M. L., Richman, S. A., Kranz, D. M., and Greenberg, P. D. (2013) Enhanced-affinity murine T-cell receptors for tumor/self-antigens can be safe in gene therapy despite surpassing the threshold for thymic selection. Blood 122, 348–356
25. Procko, E., Hedman, R., Khan, M. A., Piepenbrink, K. H., Blevins, S. I., Bernhard, H., Uckert, W., Baker, B. M. (2014) Computational design of the affinity and specificity of a therapeutic T cell receptor. PLoS Comput. Biol. 10, e1003478
Engineering Protein-Protein Interactions

39. Riley, T. P., Singh, N. K., Pierce, B. G., Baker, B. M., and Weng, Z. (2016) Computational reprogramming of T cell antigen receptor binding properties. Methods Mol. Biol. 1414, 305–318

40. Chothia, C. (1974) Hydrophobic bonding and accessible surface area in proteins. Nature 248, 338–339

41. Reynolds, J. A., Gilbert, D. B., and Tanford, C. (1974) Empirical correlation between hydrophobic free energy and aqueous cavity surface area. Proc. Natl. Acad. Sci. U.S.A. 71, 2925–2927

42. Thiltgen, G., and Goldstein, R. A. (2012) Assessing predictors of changes in protein stability upon mutation using self-consistency. PLoS One 7, e46084

43. Baker, D., and Agard, D. A. (1994) Kinetics versus thermodynamics in protein folding. Biochemistry 33, 7505–7509

44. Honegger, A., Malebranche, A. D., Röthlisberger, D., and Plückthun, A. (2009) The influence of the framework core residues on the biophysical properties of immunoglobulin heavy chain variable domains. Protein Eng. Des. Sel. 22, 121–134

45. Miller, B. R., Demarest, S. J., Lugovskoy, A., Huang, F., Wu, X., Snyder, W. B., Croner, L. J., Wang, N., Amatucci, A., Michaelson, J. S., and Glaser, S. M. (2010) Stability engineering of scFvs for the development of bispecific and multivalent antibodies. Protein Eng. Des. Sel. 23, 549–557

46. Nielsen, J. E., Gunner, M. R., and García-Moreno, B. E. (2011) The pKₐ Cooperative: a collaborative effort to advance structure-based calculations of pKₐ values and electrostatic effects in proteins. Proteins 79, 3249–3259

47. Li, Z., Yang, Y., Zhan, J., Dai, L., and Zhou, Y. (2013) Energy functions in de novo protein design: current challenges and future prospects. Annu. Rev. Biophys. 42, 315–335

48. Fleishman, S. J., Whitehead, T. A., Strauch, E. M., Corn, J. E., Qin, S., Zhou, H. X., Mitchell, J. C., Demerdash, O. N., Takeda-Shitaka, M., Terashi, G., Moal, I. H., Li, X., Bates, P. A., Zacharias, M., Park, H., et al. (2011) Community-wide assessment of protein-interface modeling suggests improvements to design methodology. J. Mol. Biol. 414, 289–302

49. Stranges, P. B., and Kuhlman, B. (2013) A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. Protein Sci. 22, 74 – 82

50. Holler, P. D., and Kranz, D. M. (2003) Quantitative analysis of the contribution of TCR/pepMHC affinity and CD8 to T cell activation. Immunity 18, 255–264

51. Chervin, A. S., Stone, J. D., Holler, P. D., Bai, A., Chen, J., Eisen, H. N., and Kranz, D. M. (2009) The impact of TCR-binding properties and antigen presentation format on T cell responsiveness. J. Immunol. 183, 1166–1178

52. Rodenko, B., Toebes, M., Hadrup, S. R., van Esch, W. J., Molemaar, A. M., Schumacher, T. N., and Ova, H. (2006) Generation of peptide-MHC class I complexes through UV-mediated ligand exchange. Nat. Protoc. 1, 1120–1132

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

54. Smith, S. N., Harris, D. T., and Kranz, D. M. (2015) T cell receptor engineering and analysis using the yeast display platform. Methods Mol. Biol. 1319, 95–141

55. Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H., and Meiler, J. (2010) Practically useful: what the Rosetta protein modeling suite can do for you. Biochemistry 49, 2987–2998

56. Leaver-Fay, A., O’Meara, M. J., Tyka, M., Jacak, R., Song, Y., Kellogg, E. H., Thompson, J., Davis, I. W., Pache, R. A., Lyskov, S., Gray, J. J., Kortemme, T., Richardson, J. S., Havranek, J. J., Snoeyink, J., et al. (2013) Scientific benchmarks for guiding macromolecular energy function improvement. Methods Enzymol. 523, 109–143

57. Moretti, R., Fleishman, S. J., Agius, R., Torchala, M., Bates, P. A., Kastritis, P. L., Rodrigues, J. P., Trellet, M., Bonvin, A. M., Cui, M., Rooman, M., Gillis, D., Dehouck, Y., Moal, I., Romero-Durana, M., et al. (2013) Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions. Proteins 81, 1980–1987

58. Chaudhury, S., Lyskov, S., and Gray, J. I. (2010) PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. Bioinformatics 26, 689–691

59. Mandell, D. J., Coutsias, E. A., and Kortemme, T. (2009) Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. Nat. Methods 6, 551–552

60. Bradley, P., Misura, K. M., and Baker, D. (2005) Toward high-resolution de novo structure prediction for small proteins. Science 309, 1868–1871

61. Canutescu, A. A., and Dunbrack, R. L., Jr. (2003) Cyclic coordinate descent: a robotics algorithm for protein loop closure. Protein Sci. 12, 963–972

62. Kellogg, E. H., Leaver-Fay, A., and Baker, D. (2011) Role of conformational sampling in computing mutation-induced changes in protein structure and stability. Proteins 79, 830–838

63. Götz, A. W., Williamson, M. J., Xu, D., Poole, D., Le Grand, S., and Walker, R. C. (2012) Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. J. Chem. Theory Comput. 8, 1542–1555