BIOCHEMISTRY
including biophysical chemistry & molecular biology

CURRENT TOPIC

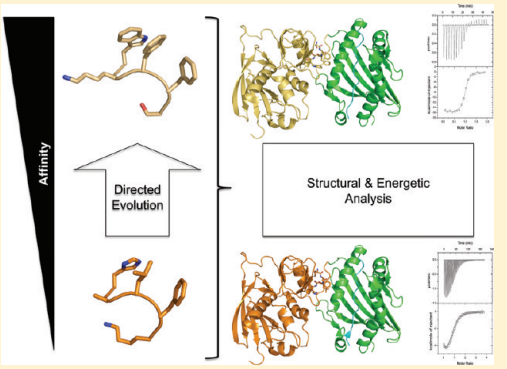
pubs.acs.org/biochemistry

Dissecting Protein−Protein Interactions Using Directed Evolution

Daniel A. Bonsor and Eric J. Sundberg*

Boston Biomedical Research Institute, 64 Grove Street, Watertown, Massachusetts 02472, United States

ABSTRACT: Protein−protein interactions are essential for life. They are responsible for most cellular functions and when they go awry often lead to disease. Proteins are inherently complex. They are flexible macromolecules whose constituent amino acid components act in combinatorial and networked ways when they engage one another in binding interactions. It is just this complexity that allows them to conduct such a broad array of biological functions. Despite decades of intense study of the molecular basis of protein−protein interactions, key gaps in our understanding remain, hindering our ability to accurately predict the specificities and affinities of their interactions. Until recently, most protein−protein investigations have been probed experimentally at the single-amino acid level, making them, by definition, incapable of capturing the combinatorial nature of, and networked communications between, the numerous residues within and outside of the protein−protein interface. This aspect of protein−protein interactions, however, is emerging as a major driving force for protein affinity and specificity. Understanding a combinatorial process necessarily requires a combinatorial experimental tool. Much like the organisms in which they reside, proteins naturally evolve over time, through a combinatorial process of mutagenesis and selection, to functionally associate. Elucidating the process by which proteins have evolved may be one of the keys to deciphering the molecular rules that govern their interactions with one another. Directed evolution is a technique performed in the laboratory that mimics natural evolution on a tractable time scale that has been utilized widely to engineer proteins with novel capabilities, including altered binding properties. In this review, we discuss directed evolution as an emerging tool for dissecting protein−protein interactions.



Protein−protein interactions govern life. Mutations within genes can affect protein complexes as a result of alteration of the affinity of a protein for its binding partner.[1] This can have dire consequences on an organism by disrupting signaling pathways, metabolic and gene regulation, and other cellular processes. Protein−protein complexes span a wide affinity spectrum: millimolar to micromolar values for transient and weak complexes such as electron transfer proteins, micromolar to nanomolar for intermediate complexes involved in antibody−antigen interactions and cell signaling pathways, and nanomolar to femtomolar for strong interactions typically found in nuclease and protease inhibitor interactions. These widely varying affinities are governed by the forces encoded not only in the amino acid sequence but also in the large library of structural elements formed by proteins.

Proteins are not static, in space (e.g., conformational change) or time (e.g., evolution). While we most commonly view proteins as "snapshots" derived from X-ray crystallographic studies, all proteins, to varying degrees, are structurally fluid. This flexibility is essential for energetic communication between distinct sites in a protein, a poorly understood phenomenon that has recently been appreciated as a significant driver of protein binding events. Evolutionarily, spontaneous mutations occur within a gene that produces protein variants within a given population. Certain mutations are beneficial and aid in both the survival of the host organism and the propagation of that particular mutation. The subsequent surviving progeny also undergo mutation, giving rise to a new class of variants that will be selected. While this process occurs slowly within a single generation, over the course of eons accumulated selected mutations lead to profound and fundamental changes in specific proteins that extend to the organismal level.

Directed evolution is a technique that mimics natural evolution in the laboratory. It uses similar principles (i.e., selection of a property of interest within a mutagenized protein population) except on a much shorter time scale. Because the typical selection criterion in the directed evolution of proteins is for increased affinity, it has been employed commonly in the development of high-affinity protein-based therapeutics, technologies, and other applications. Directed evolution, however, need not be constrained to the applied realm of science but can be used by basic scientists investigating fundamental processes involving proteins. Because it mimics the natural process by which proteins alter their function and is conducted on a tractable time scale and the resulting changes can be deciphered using current biochemical and biophysical tools, directed evolution has recently emerged as a powerful instrument in improving our understanding of protein molecular recognition. In this work, we provide an overview of directed evolution platforms and techniques, review some recent progress in using directed evolution to dissect protein−protein

Received: December 20, 2010
Revised: February 16, 2011
Published: February 18, 2011

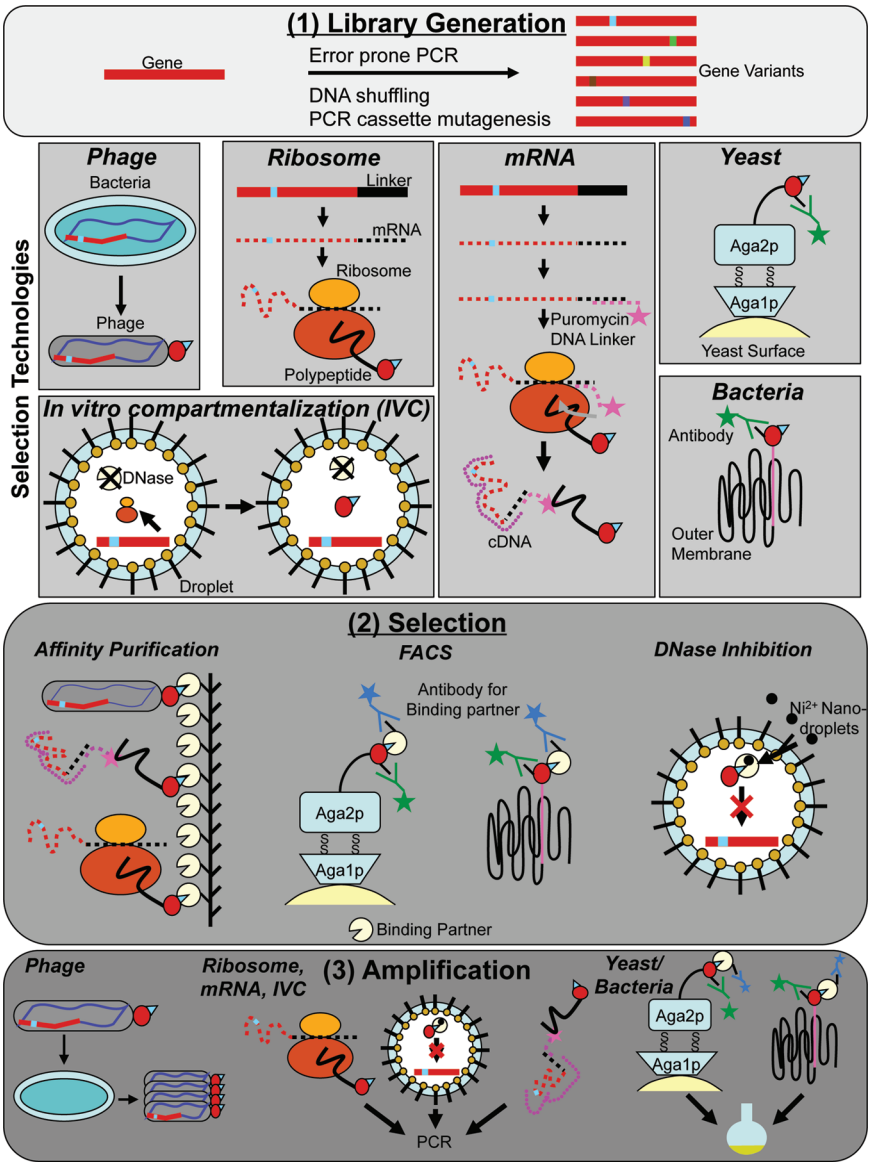


Figure 1. Directed evolution strategy of protein−protein interactions. (1) Library generation. The gene of interest (red bar) is subjected to evolution through the use of error prone PCR, DNA shuffling, and/or PCR cassette mutagenesis to create a library of gene variants (colored squares denote mutations). The following selection technologies were used. For phage display, the gene library is inserted into phage genome and transformed into *Escherichia coli*. Hosts generate the phage library leading to the protein of interest (red circle) with desired mutations (cyan triangle). For ribosome display, the library is fused to a linker (black bar) that lacks a stop codon. The library is transcribed to mRNA (dashed line) that is transcribed by ribosomes (yellow/brown circles), leading to the protein of interest bound to the ribosome due to the lack of a stop codon. For mRNA display, the library is fused to a linker and transcribed to mRNA where a cDNA strand (magenta dashed link) is annealed to the linker region. The DNA strand is covalently attached to puromycin (magenta star). Ribosomes transcribe the library but remain bound to the ribosome because of the DNA linker. The puromycin is transferred (gray arrow) to the protein, and the ribosomes are washed away. A cDNA strand (purple circles) is then synthesized for the library. For yeast display, the library is fused to Aga2p, which is bound to Aga1p displayed on the yeast surface. The protein of interest has a C-terminal tag, which is recognized by an antibody (green Y with a star). For in vitro compartmentalization (IVC), the library, ribosomes, and apo-DNase (yellow pie chart) are mixed with a water emulsion system. The library is transcribed by the ribosomes to produce the protein of interest. For bacterial display, the library is fused to the AIDA-1 gene, which forms a β-barrel in the outer membrane. The protein of interest passes through the lumen of the barrel and is presented extracellularly where an antibody recognizes a C-terminal tag. (2) Selection. For affinity purification, phage, mRNA, and ribosome display technologies are incubated with an immobilized binding partner. Higher-affinity binders are retained, wherease weaker binders are washed away. For FACS, the fluorescence-labeled binding partner (blue Y with a star) is incubated with yeast or bacteria display technologies. Fluorescent activated cell sorting (FACS) is used to screen for high-affinity binders with an appropriate sort gate. For DNase inhibition, nickel nanodrops (black circles) are added to the emulsion, which bind to the DNase. Library variants that bind and inhibit the DNase survive, whereas weak binders lead to the destruction of the DNA within that compartment. (3) Amplification. Phage are reintroduced into fresh *E. coli* hosts that are amplified and then sequenced. Ribosome, mRNA, and in vitro compartmentalization technologies are amplified by PCR and then sequenced. Yeast and bacteria display systems are introduced to selective media to amplify cells, which are then sequenced.

dx.doi.org/10.1021/bi102019c |*Biochemistry* 2011, 50, 2394–2402

interactions, and discuss emerging trends and key problems in the field remaining to be solved.

OVERVIEW OF DIRECTED EVOLUTION

Directed evolution experiments generally consist of three stages: (1) library generation, (2) selection, and (3) amplification (Figure 1). The first stage involves the creation of a genetic library consisting of a randomized set of mutations in the gene of interest. These libraries represent only a small fraction of all possibilities at each position in a protein sequence (20^N, where N is the number of residues). Accordingly, typical libraries of $\sim 10^{10}$ can cover only every amino acid possibility at no more than seven residues.[2] To create the library, the gene of interest is commonly mutated by error prone PCR or DNA shuffling.[3] The second stage involves screening the library to isolate the few variants that show an improvement in the desirable attribute. Many different technologies exist for this stage and are discussed in more detail below. The third and final stage in directed evolution is an amplification process by which selected variants are enriched. For phage, this is simply the infection of bacteria. Ribosomal and mRNA display, as well as in vitro compartmentalization, utilizes PCR and reverse transcription to generate the original template, whereas yeast and bacterial display requires growth of cells in their selective media. Iterations of these three stages of directed evolution are performed typically to the point where the selective property (e.g., affinity) is no longer appreciably improved.

Phage display was created by Smith in the early 1980s and is the oldest and most widespread method employed for directed evolution.[4] Phage are viruses that infect bacteria to achieve replication. To hijack the replication machinery of a bacterium, the phage needs to penetrate the outer membrane using their F-pilus. Attachment begins with the N-terminal domain of the minor coat protein, pIII. This triggers the coat proteins to merge with the cell membrane, uncoating the single-stranded DNA that enters the cytoplasm. A cDNA strand is synthesized and replicated, and the genes encoded are transcribed to generate new progeny that bud from the hijacked cell (reviewed in ref 5). For the purposes of directed evolution, the library ($\sim 10^6 - 10^{11}$) is expressed as a fusion to either the major or minor coat proteins by insertion of the library in the phage genome in bacterial hosts. The hosts generate the phage library, which is isolated from the bacteria and selected for improved binding by affinity purification. Unbound phage are washed away, while the phage that display peptides or proteins that recognize the target are retained. This panning process can be performed under increasingly stringent conditions in later rounds to isolate only the tightest binders among the library.

Since the advent of phage display, several other technologies have been developed for directed evolution. Yeast surface display arose from the work of Wittrup and co-workers that utilizes the adhesion subunit of the yeast agglutinin protein Aga2p, which normally mediates cell−cell contact during mating.[6] In a typical experiment, the library ($\sim 10^7 - 10^9$) of interest is fused to Aga2p. The fusion protein is expressed and attached to the yeast surface through the formation of disulfide bonds to Aga1p projecting it away from the cell surface. Expression levels are monitored by immunofluorescence labeling of epitopes flanking the protein of interest. Screening for improvement in affinity is achieved by incubating the cells with a fluorophore-labeled binding partner and screening using fluorescence-activated cell sorting.[7] A similar system can be employed using bacteria (e.g., bacterial display or autodisplay).[8] Here the library is fused in-frame 3′ to a signal peptide and 5′ to a linker and the β-barrel of AIDA-1. When expressed, the fusion protein is secreted to the periplasm, where the signal peptide is removed and the β-barrel folds in the outer membrane. The linker allows the unfolded protein of interest to pass through the lumen of the barrel and be presented extracellularly in a folded state.

In contrast to the display methods mentioned above, both ribosome and mRNA display are wholly in vitro methods. They allow for the screening of a larger library ($\sim 10^{12} - 10^{14}$) as both are limited only by the amount of ribosome present. Ribosomal display[9] takes advantage of fusing the library to the 5′-end of a spacer that lacks a stop codon. When the library is expressed as mRNA, the protein is synthesized by the ribosome and folds when it emerges from the ribosomal tunnel. The spacer ensures that the entire protein has escaped the tunnel and that the protein and mRNA are still attached to the ribosome when purified by affinity purification. mRNA display[9,10] is very similar to ribosomal display. Here the library is transcribed to mRNA and covalently attached to a short cDNA linker that carries a puromycin moiety at its 3′-end. When incubated with ribosomes, the polypeptide is synthesized but does not escape the ribosome because of the covalently attached DNA. At this point, the puromycin is close enough to be transferred to the ribosomal A site and then the protein. The mRNA−protein fusions are then isolated, and the first cDNA is synthesized and then isolated by affinity purification.

A more recently developed method, in vitro compartmentalization,[11] takes advantage of a water in oil emulsion system. Here, an in vitro transcription/translation mixture is added to the emulsion to create $\sim 10^{10}$ aqueous drops per milliliter with each drop containing, on average, a single gene of the library. An advantage of the in vitro compartmentalization system is that the selection process can be tailored to a protein system for a specific attribute.[14] In the example depicted in Figure 1, the oil in water emulsion system is created in the presence of an apo-metal-dependent DNase. The library consists of the DNase inhibitor. Successful mutants that result in an increase in binding affinity will protect the DNA inside each compartment from degradation when nickel nanodroplets are introduced into the system.

Each of these technologies can be used to screen large libraries of variants for increased affinity, the most common selective handle. Other attributes, though, can also be screened, such as stability (e.g., improvement in melting temperature) or catalytic activity and/or selectivity. Below, we describe some of the advances in understanding protein−protein interactions using the above platforms for directed evolution.

USING DIRECTED EVOLUTION TO DISSECT PROTEIN−PROTEIN INTERACTIONS

Defining the Minimal Requirements of a Protein Interface. Proteins, as strings of 20 chemically distinct building blocks with varying lengths that fold into unique three-dimensional shapes, are inherently complex. When two or more proteins come together to form an association, that complexity only increases, because both structural heterogeneity and energetic heterogeneity exist within the interface. This leads to a scenario in which many different amino acids can be energetically important for binding in a given complex, but no one amino acid type always is critical in all complexes. Some residues that appear to make few contacts within an interface can contribute significantly to binding energetics.

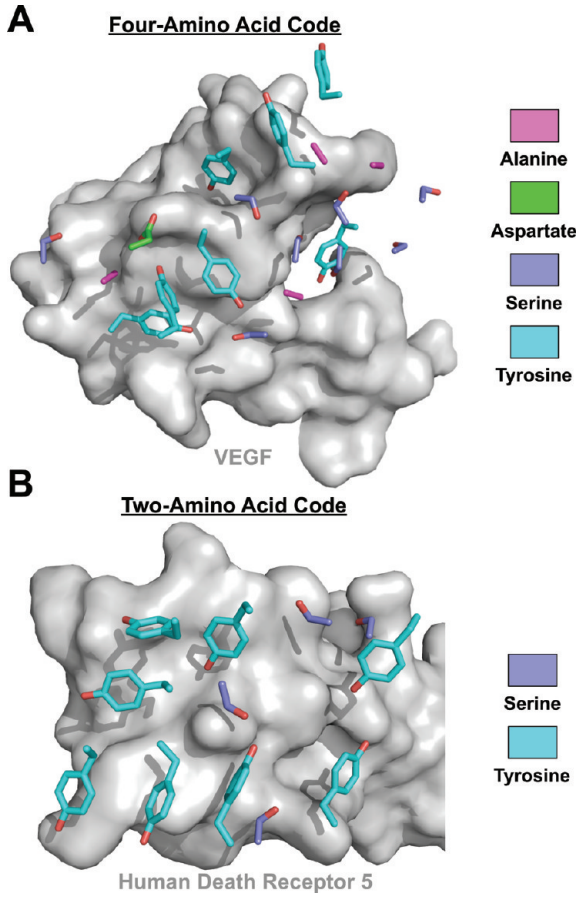


Figure 2. Directed evolution of antibody complementarity-determining regions (CDRs) using four- and two-amino acid codes. (A) Phage display was used to isolate antigen binding fragments (Fabs) that bind specifically to the extracellular domain of human receptor tyrosine kinase ErbB2 with a restriction on the number of amino acids that could be used: alanine (magenta), aspartate (green), serine (blue), and tyrosine (cyan). The crystal structure of the high-affinity antibody(Fab)−VEGF complex (Protein Data Bank entry 1tzh) is shown. VEGF is shown as a gray surface with mutated residues of Fab that form contacts shown as sticks. The residues that were mutated account for 98% of the buried surface area. (B) An antibody directly evolved via restriction of mutations at the CDRs using just two amino acids [serine (blue) and tyrosine (cyan)] to bind the human death receptor (DR5). The high-affinity antibody(Fab)−DR5 crystal structure (Protein Data Bank entry 1za3) is displayed with the human death receptor shown as a gray surface and mutated residues of the antibody that form contacts as sticks. These residues of serine and tyrosine form 79% of the buried surface area, with the remainder originating from the CDR1 loop of the Fab light chain, which was not mutated.

These effects are sometimes due to destabilization of the unbound proteins[12] and not to stabilization of the bound complex itself.

This begs the question of just what are the minimal requirements of a protein interface. This has been addressed using directed evolution with libraries comprised of reduced genetic codes, based on findings that energetically important residues in protein interfaces are relatively enriched in certain amino acids[13] and that antibody paratopes are biased toward the presence of serine and tyrosine residues. Sidhu and colleagues showed that a synthetic antibody with diversity restricted to the four-amino acid code of alanine, aspartate, serine, and tyrosine could be manipulated by phage display to bind vascular endotheilial growth factor (VEGF) with nanomolar affinity[14] (Figure 2A). Restricting diversity even further to a two-amino acid code of serine and tyrosine, these investigators showed that nanomolar affinity antibodies could be engineered to a variety of antigens[15] (Figure 2B) and that tyrosine residues appear to play a privileged role in protein molecular recognition.[16] Driving home the universality of productive molecular interactions deriving from even the most limited chemical diversity, Koide et al. showed that this same two-amino acid code could be used within a distinct scaffold, that of "monobodies" (i.e., fibronectin type III domains with loop sequence and length diversity), to generate specific and high-affinity binders for diverse targets, including maltose binding protein and human small ubiquitin-like modifier 4.[17] All of these studies suggest that low chemical diversity can be offset by high conformational diversity to drive high affinity in protein−protein interactions, as well as to create entirely novel protein complexes.

Specificity in Protein−Protein Interactions. It seems intuitive that there would be a correlation between affinity and specificity in macromolecular interactions (i.e., between how tightly a protein binds a particular binding partner and its ability to discriminate between that binding partner and all possible others). However, there may not be a simple, correlative relationship between affinity and specificity in protein−protein interactions, perhaps because within an interface the forces responsible for driving affinity increases can also drive binding promiscuity.[18] Thus, the relationship between affinity and specificity appears to be case-specific for any given protein binding its cognate and noncognate binding partners (reviewed in ref 19). Understanding this relationship may be critically important when designing therapeutic and diagnostic proteins as their engineered functions could often depend on both affinity and specificity. An effective tool for understanding this affinity−specificity relationship in protein−protein interactions is directed evolution, and numerous investigators have utilized its selective capacities to improve their understanding of specificity in distinct protein systems.

Many proteins interact with numerous binding partners as is evident from protein interactome studies, which have identified numerous "hub" proteins that have an exceptional number of interacting proteins. Proteins involved in cell signaling and regulation often exhibit this characteristic of cross-reactivity, promiscuity, and multispecificity. Because of the complex and unclear relationship between affinity and specificity, it is challenging to determine the molecular basis of specificity among numerous cognate interactions, let alone to manipulate that specificity toward a more or less restricted subset of interactions. Recent progress, however, has been made in engineering proteins that bind multiple similar proteins either cross-reactively (and with concomitant significantly increased affinities) or selectively, providing the tools to begin to understand the complex relationship between affinity and specificity.

At times, it is desirable to develop protein therapeutics that exhibit a broad range of specificity for an entire family of targets. Superantigens are bacterial toxins that simultaneously bind T cell receptor (TCR) and major histocompatibility complex molecules to activate T cells, hyperstimulate the immune system, and cause toxic shock syndrome.[20] For more than a decade, Kranz and coworkers have been using yeast display to engineer TCR variants that bind with increased affinity to superantigens to be used as antagonists of these interactions in vivo.[21] Even though numerous superantigens bind the same TCR, evolution of the common TCR to bind a particular superantigen most often results in weakened binding to nontargeted superantigens (i.e., more specificity).

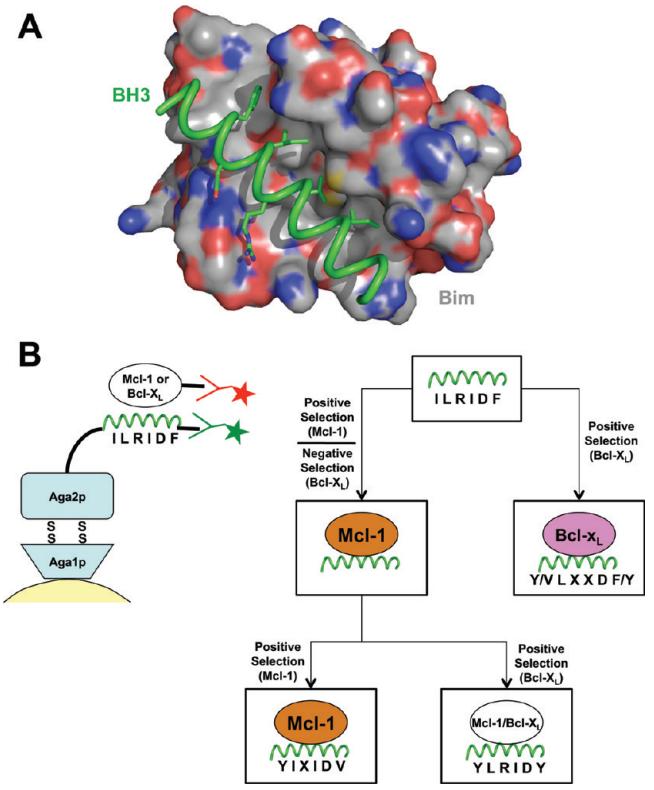


Figure 3. Yeast display of BH3 peptide to determine the binding specificity for Mcl-1 and Bcl-x_L. (A) Crystal structure of the Bim−BH3 complex (Protein Data Bank entry 2pqk) showing BH3 as a helix (green) bound on the surface of Bim. Residues of BH3 that underwent directed evolution are shown as green sticks. (B) Library of BH3 displayed on the surface of yeast cells with a C-terminal FLAG tag. The sequence of wild-type BH3 that underwent directed evolution is shown. The library was sorted for peptides that high affinity and specificity for Mcl-1, Bcl-x_L, or both. Sequencing of these variants highlighted important motifs that allow BH3 to selectively bind to one of the proteins (X is denoted as any amino acid).

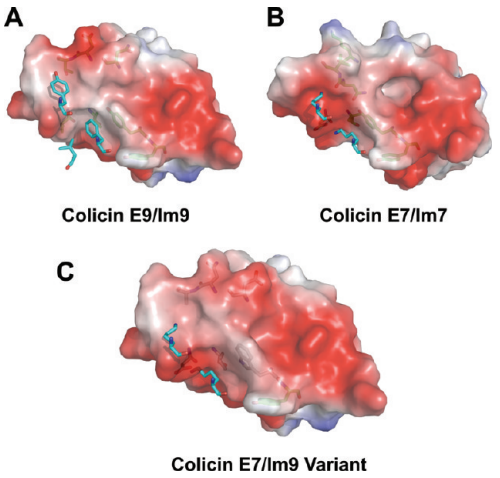


Figure 4. In vitro compartmentalization-directed evolution of a DNase inhibitor. Electrostatic surfaces of immunity proteins (A) Im9 (Protein Data Bank entry 1emv) and (B) Im7 (Protein Data Bank entry 2jb0) bound to their cognate DNase molecules colicin E9 and colicin E7, respectively. Negative electrostatic potential is colored red and positive electrostatic potential blue. Colicin and immunity protein residues that form intermolecular contacts are shown as cyan and green sticks, respectively. (C) Electrostatic surface of a model of the Im9 variant that binds colicin E7, which was found to protect 97% of *E. coli*-producing colicin E7.

In a recent study, however, two variants, termed G5-8 and KKR, that derived from the same TCR origin and evolved separately to the individual superantigens staphylococcal enterotoxin B (SEB) and streptococcal pyrogenic exotoxin A (SpeA), respectively, were shown to bind both superantigens with significantly increased affinity.[22] The G5-8 variant, whose target was SEB, bound SEB with 3.1×10^6-fold and SpeA with 1.2×10^3-fold increased affinity relative to that of the wild-type TCR. Likewise, the KKR variant, whose target was SpeA, bound SpeA with 2.2×10^4-fold and SEB with 2.2×10^3-fold increased affinity relative to that of the wild-type TCR. Furthermore, both the G5-8 and KKR variants were shown to be highly protective in animals challenged with lethal doses of SpeA.

At other times, it is preferable to engineer proteins that could discriminate exquisitely between highly homologous members of a family of cognate binding partners. Prosurvival members of the Bcl-2 protein family can regulate apoptosis by interacting with proapoptotic Bcl-2-homology-3 (BH3)-only family members via binding of the BH3-only α-helical region to a conserved hydrophobic groove on the prosurvival proteins. Diverse specificities for BH3-only/Bcl-2 prosurvival proteins exist, with some family members binding highly promiscuously and others selectively. To improve their understanding of the molecular determinants of these interactions, Keating and co-workers used yeast display and a novel selection scheme to engineer BH3 α-helical peptides that bound to either Bcl-2 prosurvival protein Bcl-XL or Mcl-1 selectively or to both with increased affinity[23] (Figure 3). In combination with X-ray crystallography and SPOT peptide array analysis, they were able to propose a simple model that predicted much of the specificity in their selective Bcl-XL versus Mcl-1 binding variants according to the sequence of variable residues in the BH3-only peptides.

The function of some proteins is dependent on their ability to discriminate between cognate and noncognate proteins that are structurally similar. Colicins are bacterial DNase toxins that bind to immunity proteins. Cognate colicin-immunity pairs are some of the highest-affinity protein complexes found in nature ($K_D \leq 10^{-14}$ M), while noncognate colicin-immunity complexes bind on the order of 10^6-fold more weakly, primarily because of faster dissociation rates. The structure of a noncognate complex revealed that the structural basis for weak binding, relative to the related cognate pair, was the disruption of only a few intermolecular bonds[24] (Figure 4A). This negative design principle was probed further by the evolution of an immunity protein from its cognate colicin to a noncognate colicin using in vitro compartmentalization, where the gain in binding the noncognate colicin was not accompanied by weakened binding to the cognate colicin in the absence of negative selection[25] (Figure 4B).

Defining and Assessing the Energetic Roles of Individual Residues. It was not until the mid-1990s that the energetic mosaicity of protein binding sites began to be appreciated. When the structure of a complex of proteins is determined, all of the atoms that make intermolecular contacts are readily identified. That all, or at least most, of these contact residues are energetically favorable for binding had seemed a reasonable assumption. This turns out, in general, not to be the case, as was first determined when Clackson and Wells adopted a strategy of alanine scanning mutagenesis to assess the energetic contributions of individual amino acids in a hormone−receptor complex.[26] In this type of

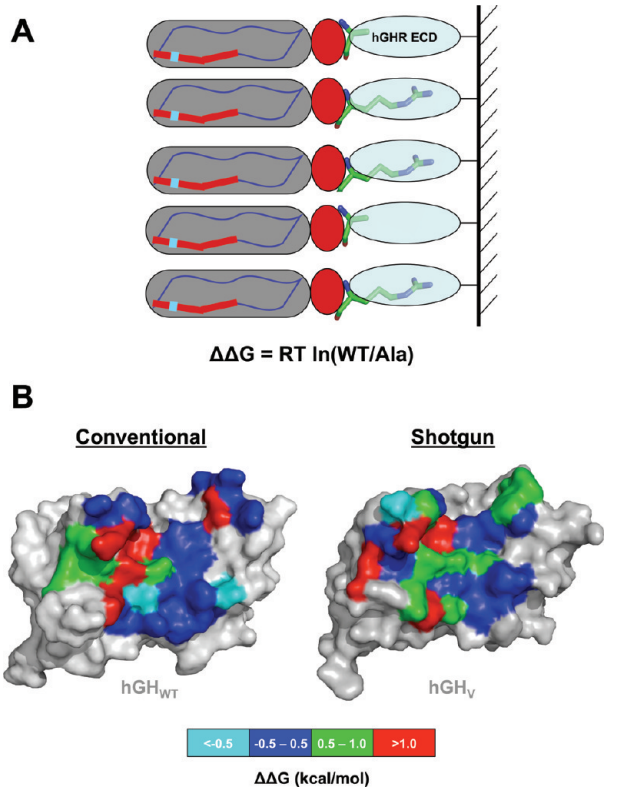


Figure 5. Shotgun alanine mutagenesis of a high-affinity variant of human growth hormone (hGH_V). (A) Schematic of shotgun alanine mutagenisis. For each position of interest, arginine (in this case) is mutated to alanine. Both are incubated with human growth hormone receptor, and the ratio of the wild type (WT) to alanine is calculated from sequencing data. This is related to the change in the Gibbs free energy upon mutation. (B) Surface residues of hGH_{WT} (left) and hHG_V (right) that were scanned by alanine mutagenesis and shotgun alanine mutagenesis, respectively (Protein Data Bank entries 3hhr and 1kf9, respectively). Residues are colored according to $\Delta\Delta G_{Ala\text{-}X}$: red for >1.0 kcal/mol, green for 0.5−1.0 kcal/mol, blue for 0.5 to −0.5 kcal/mol, cyan for less than −0.5 kcal/mol, and gray for untested.

analysis, each interface residue is systematically mutated to alanine (in effect, paring its side chain moiety back to a single methyl group; mutating to glycine to entirely abolish the side chain introduces main chain flexibility that confounds the analysis) and measuring the change in binding energy upon formation of the complex relative to the wild-type complex. Only certain amino acid residues within this hormone−receptor interface contributed significantly to the binding energy and were thus termed "hot spots", while other residues were energetically silent with respect to binding. Within a given protein−protein interface, hot spot residues are more likely to be found in the central portion of the binding site, often surrounded by a ring of less energetically important residues,[27] mirroring the construction of folded globular proteins that serves to exclude water from the sites of energetic importance. However, numerous exceptions to this rule exist, including protein complexes in which no hot spots can be identified[28,29] or interactions in which hot spot residues extend to the periphery of the interface.[30]

While this important energetic assessment of protein interfaces has been performed on many protein complexes to date, in practice it is exceedingly laborious in its traditional approach: one must perform site-directed mutagenesis for each alanine mutant, express and purify each recombinant mutant proteins, and perform some sort of binding assay (e.g., SPR or ITC) using each of them. Directed evolution techniques have been pioneered by Sidhu and co-workers to greatly reduce the experimental effort required for energetic assessment of protein interfaces by alanine scanning mutagenesis. In its original form, shotgun phage display libraries were used to introduce diversity at a single position restricted to either the wild type or an alanine residue, and the energetic consequences of mutation at a given position were determined as a correlation of the ratio of wild-type to alanine sequences at positions among numerous functional clones[31−33] (Figure 5A). A comparison of conventional and shotgun alanine scanning mutagenesis techniques on the similar protein complexes showed a high level of agreement (Figure 5B).

In a permutation of this shotgun method, termed quantitative saturation scanning, library diversity is broadened to include all 20 amino residues but restricted to a limited number of positions.[34] Applying this technique to the human growth hormone−human growth hormone receptor (hGH−hGHR) system revealed, surprisingly, that many mutations that would be considered chemically conservative were not tolerated and vice versa.

Modular Architecture and Networked Energetics within the Interface. Hot spot residues in protein−protein interactions are not necessarily distributed in a random fashion throughout the interface. Instead, they tend to cluster within discrete groups, modules, or "hot regions".[35,36] The resulting decomposition of protein interfaces into such modules, which has been shown both computationally and experimentally, can have significant energetic consequences for protein−protein interactions. Further contributing to the heterogeneity of protein−protein interfaces is the frequent presence of cooperativity, in that the energetic contribution to binding of a protein that has been simultaneously mutated at multiple residues is markedly different than the summation of the changes in binding energy of the single-site mutants.[37−39] Not only can hot spots be of varying energetic significance in and of themselves, but their energetic contributions to binding can vary depending on whether and where other hot spot residues are located in the interface. In many protein−protein interactions, such site-to-site energetic communication is a major contributor to protein binding. Compelling evidence has been mounting of late that the modular architecture that is structurally imprinted on protein binding sites[35,40,41] not only results in a certain roughness to the energetic landscape of the interface but also serves as the primary driver of networked energetic communication in protein−protein interactions.

Because the modular architecture of, and networked energetics in, protein interfaces rely on the communication between sets of amino acid residues, it is impossible to assess using techniques that rely on individual, site-directed mutagenesis. The iterative process of wide-ranging mutation and selection in directed evolution describes an affinity maturation pathway of protein variants that, in total, can span many orders of magnitude in affinity. Because numerous mutations are made that together increase the affinity, the dissection of these affinity maturation pathways by interrogating the structural and energetic changes associated with different combinations of mutations makes directed evolution an especially powerful method for investigating biophysical parameters that are combinatorial by definition, such as energetic cooperativity. A number of studies have now been published using directed evolution to dissect energetic cooperativity in protein−protein interactions.

Using a model system consisting of a yeast display affinity-matured TCR protein that exhibited an ~1500-fold affinity increase for the bacterial superantigen staphylococcal enterotoxin C3 (SEC3),[42] Sundberg and co-workers analyzed group and individual TCR maturation and reversion pathway mutations for binding to SEC3 by SPR analysis.[39] Energetic cooperativity was observed within a single hot region, in this case defined by the second complementarity-determining region (CDR2) loop, while combinations of mutations from distinct hot regions were found to be energetically additive. Even though this was one of the most highly affinity-mature complexes characterized at the time, the highest-affinity variant was found to be restricted by negative cooperativity (i.e., the summation of the changes in the binding free energies of the individual mutations exceeds the change in binding free energy of the final, fully evolved variant). Two maturation mutations in particular accounted quantitatively for the entirety of this negative cooperativity. By determining the X-ray crystal structures of several of these variant TCR proteins that define this affinity maturation pathway, they observed that the mutations at these two positions exerted opposing conformational changes on the CDR2 loop, providing a structural basis for this measured negative cooperativity.[43]

In a similar study involving another affinity-mature TCR−superantigen model protein−protein interaction system, the question of whether amino acids separated by long distances and residing at the peripheral extremes of the interface could act in an energetically cooperative manner was addressed.[44] The hVβ2.1 TCR had been previously affinity-matured by yeast display to bind the superatigen toxic shock syndrome toxin-1 (TSST-1) with an increased affinity of greater than 3000-fold relative to that of the wild-type TCR.[30] Analysis of each of the individual residue changes revealed that there were four mutations within the interface that were energetically significant in the affinity maturation process. Three of these positions are located within the CDR2 loop of the TCR and form one hot region, while the fourth is located in the third framework region (FR3) loop and forms a distinct hot region. From the X-ray crystal structure of this TCR−superantigen complex,[45] it is evident that these two hot regions are separated by more than 20 Å and each lies at the periphery of the interface. TCR variants in which every possible combination of these four amino acids as either their wild-type or affinity-matured residue were tested for binding to the superantigen, and the binding free energies of the combinatorial variants were compared to the summation of binding free energies of their corresponding single-site mutants to ascertain the extent of cooperativity. As expected, several of the amino acids within the CDR2 hot region exhibited cooperative energetics. Surprisingly, though, combinations of mutations involving residues from both the CDR2 and FR3 hot regions were also found to be energetically cooperative. Furthermore, the magnitude of this interhot regional cooperativity was significantly greater than the observed intrahot regional cooperativity.[44]

Kossiakoff and co-workers have taken advantage of phage display to dissect cooperative and additive energetics in the hGH−hGHR system. Using alanine scanning mutagenesis, they probed an affinity-mature variant of hGH (hGHv) that had been engineered using multiple phage display libraries targeting different regions of the protein that were subsequently recombined to create the final, highest-affinity variant.[38] Such a model system allowed them to probe energetic communication between distinct secondary structural elements of hGHv. As in other cooperative protein−protein interaction systems, many of the energetic contributions to binding of individual residues were highly context-dependent. Furthermore, the degree of cooperativity of particular residues was correlated to the conformational plasticity of the protein region in which it resided, and cooperativity was observed over large distances.

Focusing on the high-affinity binding site of the hGH−hGHR interaction (site 1), Pal et al. utilized combinatorial shotgun alanine scanning mutagenesis to assess intramolecular additive and cooperative energetics in this interaction.[46] In total, 19 residues were analyzed, yielding 145 statistically significant pairwise data points, of which eight residues were observed to exhibit cooperative energetics. Double mutant cycles were performed to corroborate the additive and cooperative predictions from the shotgun mutagenesis approach. The agreement between the two techniques suggests that such a shotgun, directed evolution-based approach can be used with confidence to assess intramolecular cooperativity in protein−protein interactions.

Assessing Energetic Contributions from Disordered Protein Regions. Another poorly understood property of protein−protein interactions that is dependent on the coordinated behavior of numerous amino acid residues is protein disorder. Many functional proteins are at least partially disordered, and the structural transitions upon binding of these regions can play energetically significant roles in formation of the protein complex. Indeed, single-site alanine scanning mutagenesis studies have clearly established the importance of disordered protein regions in binding.[47,48] However, the contributions of disordered regions to protein−protein interactions cannot be dependent upon single amino acid residues acting in isolation but undoubtedly always involve the concerted structural variations, and consequent energetic effects, of combinations of residues. It is implausible that conformational changes, even when focused primarily on one residue, could involve significant protein backbone movements without mutual movements of neighboring residues.

Once again, directed evolution can provide avenues for assessing such complex factors in protein molecular recognition. Sundberg and co-workers assessed, both structurally and thermodynamically, an affinity-matured superantigen−TCR system as a model for determining the role of a disordered protein region in complex formation.[49] A region of local disorder within the superantigen had been previously evolved by phage display,[50] with the resulting affinity-matured variants differing only in a linear sequence of five residues within a disulfide loop that is disordered in the wild-type protein in the unbound state. Through X-ray structure determination of the protein binding partners before and after complex formation and isothermal titration calorimetric analysis of the interactions, a correlation between protein ordering and binding affinity for complexes along this affinity maturation pathway was observed. Additionally, discrepancies between observed and calculated heat capacities based on buried surface area changes in the protein complexes could be explained largely by heat capacity changes that would result solely from folding the locally disordered region.

EMERGING TRENDS AND FUTURE DIRECTIONS

While the primary goal of many protein engineers using directed evolution may still be the development of proteins with novel properties, a number of researchers have seized upon these methods to provide model systems with which to address fundamental questions in protein molecular recognition. The ability to select for evolved properties that result from combinatorial

changes in a protein complex, rather than targeted single-residue changes, significantly broadens the scope of investigations that can be undertaken to dissect protein−protein interactions.

Directed evolution has already been used to shed new light on the minimal requirements of protein interfaces. Recently, this work has been applied to the development of novel antibodies and monobodies bearing a simple two-amino acid code that bind to a diversity of protein targets, some of which are pharmaceutically important. We may be on the cusp of a new era of synthetic biology in which whole libraries of simplified proteins constructed on some type of universal scaffold are engineered to bind with high affinity to every protein produced by an organism. Such libraries of simplified, synthetic, and high-affinity proteins would have vast potential both as research tools and as therapeutics.

The use of directed evolution has now greatly streamlined conventional alanine scanning mutagenesis with shotgun alanine scanning using phage display. Because determining the functional epitope, in addition to the structural epitope, of a protein complex is critical for understanding the driving forces governing the interaction, as well as for the design of potential of inhibitors, this significantly higher-throughput method will likely expand in use and could eventually replace the highly laborious practice of the conventional methodology. There is no technical reason that such shotgun analysis should be restricted to phage display but could theoretically be performed using other display techniques. This would remove barriers to applying shotgun alanine scanning to any or all protein−protein interactions in an organism and, in this way, allow for the meaningful detailing of entire interactomes, which, as simple lists of interacting partners, are inherently information-poor.

Directed evolution is now also coming of age in assessing the complex, networked energetic communications that exist within the protein interface, including cooperativity and the impact of disordered protein regions in binding. Evolutionary techniques are exceptionally well-suited to the study of these combinatorial effects in molecular recognition because they are, themselves, combinatorial processes. Biological research and drug development would be markedly accelerated if our computational tools were able to predict, from protein sequences alone, the specificity and affinity of protein−protein interactions with high fidelity. Certainly, state-of-the-art computational algorithms dependably perform this function for protein complexes involving proteins that are relatively conformationally static and energetically additive. However, because structural flexibility and energetic cooperativity are important factors in protein−protein interactions and poorly modeled in current computational algorithms, methods that can provide quantitative information and a more comprehensive understanding of these factors in protein complex formation are critical for developing the next generation of computational tools. The use of directed evolution coupled to structural and energetic analysis of protein−protein interactions may be the key to achieving this.

AUTHOR INFORMATION

Corresponding Author

*E-mail: sundberg@bbri.org. Telephone: (617) 658-7882. Fax: (617) 972-1761.

ABBREVIATIONS

BH3, Bcl-2-homology-3; CDR, complementarity-determining region; FR, framework region; ITC, isothermal titration calorimetry; hGH, human growth hormone; hGHR, human growth hormone receptor; hGHv, affinity-mature variant of human growth hormone; PCR, polymerase chain reaction; SEB, staphylococcal enterotoxin B; SEC3, staphylococcal enterotoxin C3; SpeA, streptococcal pyrogenic exotoxin A; SPR, surface plasmon resonance; TCR, T cell receptor; TSST-1, toxic shock syndrome toxin-1; VEGF, vascular endothelial growth factor.

REFERENCES

(1) Maslon, M. M., and Hupp, T. R. (2010) Drug discovery and mutant p53. *Trends Cell Biol. 20*, 542–555.

(2) Yuen, C. M., and Liu, D. R. (2007) Dissecting protein structure and function using directed evolution. *Nat. Methods 4*, 995–997.

(3) Labrou, N. E. (2010) Random mutagenesis methods for in vitro directed enzyme evolution. *Curr. Protein Pept. Sci. 11*, 91–100.

(4) Smith, G. P. (1985) Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. *Science 228*, 1315–1317.

(5) Smith, G. P., and Petrenko, V. A. (1997) Phage Display. *Chem. Rev. 97*, 391–410.

(6) Boder, E. T., and Wittrup, K. D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol. 15*, 553–557.

(7) Chao, G., Lau, W. L., Hackel, B. J., Sazinsky, S. L., Lippow, S. M., and Wittrup, K. D. (2006) Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc. 1*, 755–768.

(8) Jose, J. (2006) Autodisplay: Efficient bacterial surface display of recombinant proteins. *Appl. Microbiol. Biotechnol. 69*, 607–614.

(9) Lipovsek, D., and Pluckthun, A. (2004) In-vitro protein evolution by ribosome display and mRNA display. *J. Immunol. Methods 290*, 51–67.

(10) Keefe, A. D. (2001) Protein selection using mRNA display. *Current Protocols in Molecular Biology*, Chapter 24, Unit 24, p 25, Wiley, New York.

(11) Griffiths, A. D., and Tawfik, D. S. (2006) Miniaturising the laboratory in emulsion droplets. *Trends Biotechnol. 24*, 395–402.

(12) DeLano, W. L. (2002) Unraveling hot spots in binding interfaces: Progress and challenges. *Curr. Opin. Struct. Biol. 12*, 14–20.

(13) Lo Conte, L., Chothia, C., and Janin, J. (1999) The atomic structure of protein-protein recognition sites. *J. Mol. Biol. 285*, 2177–2198.

(14) Fellouse, F. A., Wiesmann, C., and Sidhu, S. S. (2004) Synthetic antibodies from a four-amino-acid code: A dominant role for tyrosine in antigen recognition. *Proc. Natl. Acad. Sci. U.S.A. 101*, 12467–12472.

(15) Fellouse, F. A., Li, B., Compaan, D. M., Peden, A. A., Hymowitz, S. G., and Sidhu, S. S. (2005) Molecular recognition by a binary code. *J. Mol. Biol. 348*, 1153–1162.

(16) Fellouse, F. A., Barthelemy, P. A., Kelley, R. F., and Sidhu, S. S. (2006) Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. *J. Mol. Biol. 357*, 100–114.

(17) Koide, A., Gilbreth, R. N., Esaki, K., Tereshko, V., and Koide, S. (2007) High-affinity single-domain binding proteins with a binary-code interface. *Proc. Natl. Acad. Sci. U.S.A. 104*, 6632–6637.

(18) Greenspan, N. S. (2010) Cohen's Conjecture, Howard's Hypothesis, and Ptashne's Ptruth: An exploration of the relationship between affinity and specificity. *Trends Immunol. 31*, 138–143.

(19) Schreiber, G., and Keating, A. E. (2011) Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol. 21*, 50–61.

(20) McCormick, J. K., Yarwood, J. M., and Schlievert, P. M. (2001) Toxic shock syndrome and bacterial superantigens: An update. *Annu. Rev. Microbiol. 55*, 77–104.

(21) Buonpane, R. A., Churchill, H. R., Moza, B., Sundberg, E. J., Peterson, M. L., Schlievert, P. M., and Kranz, D. M. (2007) Neutralization of staphylococcal enterotoxin B by soluble, high-affinity receptor antagonists. *Nat. Med. 13*, 725–729.

(22) Wang, N., Mattis, D. M., Sundberg, E. J., Schlievert, P. M., and Kranz, D. M. (2010) A single, engineered protein therapeutic agent

neutralizes exotoxins from both *Staphylococcus aureus* and *Streptococcus pyogenes*. *Clin. Vaccine Immunol. 17*, 1781–1789.

(23) Dutta, S., Gulla, S., Chen, T. S., Fire, E., Grant, R. A., and Keating, A. E. (2010) Determinants of BH3 binding specificity for Mcl-1 versus Bcl-xL. *J. Mol. Biol. 398*, 747–762.

(24) Meenan, N. A., Sharma, A., Fleishman, S. J., Macdonald, C. J., Morel, B., Boetzel, R., Moore, G. R., Baker, D., and Kleanthous, C. (2010) The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. *Proc. Natl. Acad. Sci. U.S.A. 107*, 10080–10085.

(25) Levin, K. B., Dym, O., Albeck, S., Magdassi, S., Keeble, A. H., Kleanthous, C., and Tawfik, D. S. (2009) Following evolutionary paths to protein-protein interactions with high affinity and selectivity. *Nat. Struct. Mol. Biol. 16*, 1049–1055.

(26) Clackson, T., and Wells, J. A. (1995) A hot spot of binding energy in a hormone-receptor interface. *Science 267*, 383–386.

(27) Bogan, A. A., and Thorn, K. S. (1998) Anatomy of hot spots in protein interfaces. *J. Mol. Biol. 280*, 1–9.

(28) Roisman, L. C., Jaitin, D. A., Baker, D. P., and Schreiber, G. (2005) Mutational analysis of the IFNAR1 binding site on IFNα2 reveals the architecture of a weak ligand-receptor binding-site. *J. Mol. Biol. 353*, 271–281.

(29) Svensson, H. G., Wedemeyer, W. J., Ekstrom, J. L., Callender, D. R., Kortemme, T., Kim, D. E., Sjobring, U., and Baker, D. (2004) Contributions of amino acid side chains to the kinetics and thermodynamics of the bivalent binding of protein L to Igκ light chain. *Biochemistry 43*, 2445–2457.

(30) Buonpane, R. A., Moza, B., Sundberg, E. J., and Kranz, D. M. (2005) Characterization of T cell receptors engineered for high affinity against toxic shock syndrome toxin-1. *J. Mol. Biol. 353*, 308–321.

(31) Pal, G., Kossiakoff, A. A., and Sidhu, S. S. (2003) The functional binding epitope of a high affinity variant of human growth hormone mapped by shotgun alanine-scanning mutagenesis: Insights into the mechanisms responsible for improved affinity. *J. Mol. Biol. 332*, 195–204.

(32) Vajdos, F. F., Adams, C. W., Breece, T. N., Presta, L. G., de Vos, A. M., and Sidhu, S. S. (2002) Comprehensive functional maps of the antigen-binding site of an anti-ErbB2 antibody obtained with shotgun scanning mutagenesis. *J. Mol. Biol. 320*, 415–428.

(33) Weiss, G. A., Watanabe, C. K., Zhong, A., Goddard, A., and Sidhu, S. S. (2000) Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc. Natl. Acad. Sci. U.S.A. 97*, 8950–8954.

(34) Pal, G., Kouadio, J. L., Artis, D. R., Kossiakoff, A. A., and Sidhu, S. S. (2006) Comprehensive and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. *J. Biol. Chem. 281*, 22378–22385.

(35) Keskin, O., Ma, B., and Nussinov, R. (2005) Hot regions in protein−protein interactions: The organization and contribution of structurally conserved hot spot residues. *J. Mol. Biol. 345*, 1281–1294.

(36) Reichmann, D., Rahat, O., Albeck, S., Meged, R., Dym, O., and Schreiber, G. (2005) The modular architecture of protein-protein binding interfaces. *Proc. Natl. Acad. Sci. U.S.A. 102*, 57–62.

(37) Albeck, S., Unger, R., and Schreiber, G. (2000) Evaluation of direct and cooperative contributions towards the strength of buried hydrogen bonds and salt bridges. *J. Mol. Biol. 298*, 503–520.

(38) Bernat, B., Sun, M., Dwyer, M., Feldkamp, M., and Kossiakoff, A. A. (2004) Dissecting the binding energy epitope of a high-affinity variant of human growth hormone: Cooperative and additive effects from combining mutations from independently selected phage display mutagenesis libraries. *Biochemistry 43*, 6076–6084.

(39) Yang, J., Swaminathan, C. P., Huang, Y., Guan, R., Cho, S., Kieke, M. C., Kranz, D. M., Mariuzza, R. A., and Sundberg, E. J. (2003) Dissecting cooperative and additive binding energetics in the affinity maturation pathway of a protein-protein interface. *J. Biol. Chem. 278*, 50412–50421.

(40) Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H., and Schreiber, G. (2007) The molecular architecture of protein-protein binding sites. *Curr. Opin. Struct. Biol. 17*, 67–76.

(41) Keskin, O., Ma, B., Rogale, K., Gunasekaran, K., and Nussinov, R. (2005) Protein-protein interactions: Organization, cooperativity and mapping in a bottom-up systems biology approach. *Phys. Biol. 2*, S24–S35.

(42) 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.

(43) Cho, S., Swaminathan, C. P., Yang, J., Kerzic, M. C., Guan, R., Kieke, M. C., Kranz, D. M., Mariuzza, R. A., and Sundberg, E. J. (2005) Structural Basis of Affinity Maturation and Intramolecular Cooperativity in a Protein-Protein Interaction. *Structure 13*, 1775–1787.

(44) Moza, B., Buonpane, R. A., Zhu, P., Herfst, C. A., Rahman, A. K., McCormick, J. K., Kranz, D. M., and Sundberg, E. J. (2006) Long-range cooperative binding effects in a T cell receptor variable domain. *Proc. Natl. Acad. Sci. U.S.A. 103*, 9867–9872.

(45) Moza, B., Varma, A. K., Buonpane, R. A., Zhu, P., Herfst, C. A., Nicholson, M. J., Wilbuer, A. K., Seth, N. P., Wucherpfennig, K. W., McCormick, J. K., Kranz, D. M., and Sundberg, E. J. (2007) Structural basis of T-cell specificity and activation by the bacterial superantigen TSST-1. *EMBO J. 26*, 1187–1197.

(46) Pal, G., Ultsch, M. H., Clark, K. P., Currell, B., Kossiakoff, A. A., and Sidhu, S. S. (2005) Intramolecular cooperativity in a protein binding site assessed by combinatorial shotgun scanning mutagenesis. *J. Mol. Biol. 347*, 489–494.

(47) Fasolini, M., Wu, X., Flocco, M., Trosset, J. Y., Oppermann, U., and Knapp, S. (2003) Hot spots in Tcf4 for the interaction with β-catenin. *J. Biol. Chem. 278*, 21092–21098.

(48) Hu, D., Crist, M., Duan, X., and Gimble, F. S. (1999) Mapping of a DNA binding region of the PI-sceI homing endonuclease by affinity cleavage and alanine-scanning mutagenesis. *Biochemistry 38*, 12621–12628.

(49) Cho, S., Swaminathan, C. P., Bonsor, D. A., Kerzic, M. C., Guan, R., Yang, J., Kieke, M. C., Andersen, P. S., Kranz, D. M., Mariuzza, R. A., and Sundberg, E. J. (2010) Assessing energetic contributions to binding from a disordered region in a protein-protein interaction. *Biochemistry 49*, 9256–9268.

(50) Andersen, P. S., Geisler, C., Buus, S., Mariuzza, R. A., and Karjalainen, K. (2001) Role of the T cell receptor ligand affinity in T cell activation by bacterial superantigens. *J. Biol. Chem. 276*, 33452–33457.