nature chemical biology

Computer-aided design of functional protein interactions

Daniel J Mandell & Tanja Kortemme

Predictive methods for the computational design of proteins search for amino acid sequences adopting desired structures that perform specific functions. Typically, design of 'function' is formulated as engineering new and altered binding activities into proteins. Progress in the design of functional protein-protein interactions is directed toward engineering proteins to precisely control biological processes by specifically recognizing desired interaction partners while avoiding competitors. The field is aiming for strategies to harness recent advances in high-resolution computational modeling—particularly those exploiting protein conformational variability—to engineer new functions and incorporate many functional requirements simultaneously.

Computational protein design—the automated search for amino acid sequences with defined three-dimensional structures—opens exciting possibilities for the engineering of biomolecules with desired shapes, precisely selective functional interactions and new catalytic activities. More than two decades ago, Ponder and Richards proposed a practical solution for designing proteins[1]. First, three-dimensional conformations of each amino acid side chain residue in a protein are picked from a predefined set of conformers. Combinations of these residues are then fit onto a fixed main-chain polypeptide backbone taken from an existing protein structure. The suitability of an amino acid sequence for a given structure is evaluated by the tight fit, or 'packing', of the atoms within the buried core of the protein structure. The procedure is an elegant way of reducing the enormous problem posed by protein design—searching both sequence and structure space simultaneously to find solutions with low free energy—to something much more manageable: matching a sequence to a defined backbone 'template' by only allowing discrete choices for the amino acid side chains. This formulation of 'fixed backbone design' is still the basis for much work in the field.

Since the development of initial protein design concepts[1–3], several studies, each including experimental validation of computationally designed sequences, have produced a series of remarkable achievements. These milestones include predicting sequence changes that alter atomic packing arrangements in the buried core regions of proteins[4–7], increasing protein stability (summarized in ref. 8), and designing a completely new sequence for an existing small protein[9].

Even protein structures not found in nature have been successfully designed, either by creating new arrangements of α-helices in dimers, trimers and tetramers[10], or by composing α-helices and β-sheets into a new fold 'topology'[11]. By definition, creating these new protein structures requires moving beyond existing backbone conformations. The task is not only to find a suitable amino acid sequence, but also to search through backbone conformations and find those that are 'designable'—that is, those that accommodate sequences that would adopt the conformation in nature. Although these studies arguably start with the simplest approach (arranging well-defined regular secondary structure elements with little or no loop regions), they demonstrate accurate modeling of backbone flexibility—a problem of considerable importance for designing functional interactions that we will return to later.

Increasingly, applications of computational design methods aim to create proteins with not only new structures but also modified and novel functions. At the core of 'function' lies, most generally, a physical interaction between a protein and its partner: a small molecule, a substrate, a nucleic acid or another protein. To design functional protein interactions, computational strategies have been applied to create metal binding sites[12,13], protein-protein interactions with new and altered binding preferences[14–20], highly specific small-molecule ligand receptors[21], altered protein-DNA interactions[22] and catalytic activities not observed in naturally occurring enzymes[23,24] by optimizing proteins to be complementary to a modeled transition state for the desired reaction.

Building on this impressive set of proof-of-concept studies, the ability to design protein interactions with defined affinity and selectivity harbors great promise for creating specific and sensitive molecular elements in many practical areas of biological engineering. **Figure 1** highlights some of the concepts for the design of functional protein interactions that we will discuss. Applications include engineered proteins functioning as biological sensors or as actuators that interface with existing cellular machinery to control biological processes. Designer molecules can function as probes to dissect complex cellular protein networks and thus aid in the discovery and validation of disease targets. Protein properties may be engineered to improve the characteristics of biotherapeutics and reduce side effects, and protein mimics of antigenic epitopes can make potent vaccines. Finally, in concert with powerful methods for directed evolution of protein functions[25], protein design can generate useful molecular 'parts' to make biology easier to engineer, both by building

Graduate Program in Bioinformatics and Computational Biology, California Institute for Quantitative Biosciences, and Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, California, USA. Correspondence should be addressed to T.K. (kortemme@cgl.ucsf.edu).

Published online 19 October 2009; doi:10.1038/nchembio.251

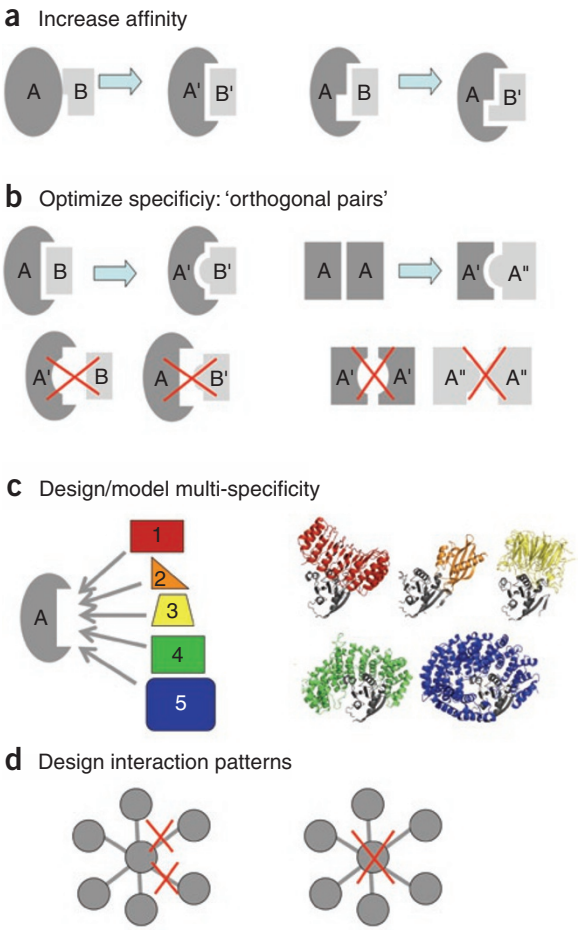

Figure 1 Concepts for protein interface design. (**a**) Designed substitutions to increase interface affinity can either be made in both partners in a complex (left) or just to one partner designed to match a fixed target (right). Such strategies are useful to develop protein interaction inhibitors. (**b**) One of the simplest formulations of the selectivity design problem is to change an existing interface A-B so that the two new partners A' and B' specifically recognize each other in such a way that cross-talk interactions with their original wild-type partners (A'-B and A-B') are avoided (left). Such 'orthogonal pairs' are useful for re-wiring of signaling pathways mediated by protein interactions. The right schematic illustrates the design of a homodimer into a heterodimer. (**c**) Functional design of 'hub' proteins accounts for multi-specificity across shared interfaces (left). Many key signaling proteins, such as GTPases (right), exhibit multi-specificity through overlapping interfaces that bind several partners (GTPase Ran in grey, multiple partners colored). (**d**) Protein design can be useful to dissect individual interactions (left), an approach that is complementary to common knockout approaches of genes or proteins (right).

new functions and by creating part families based on existing functions, where each family member has defined functional characteristics.

Keeping these ambitious long-term goals in mind, we will highlight both opportunities and difficulties in the field, focusing specifically on computational design of functional protein-protein interactions (see also refs. 26–28), although many of the key concepts and unsolved problems apply to protein structure prediction and design more generally. It will be increasingly important to develop innovative approaches to design proteins that are 'fit' to function in complex cellular environments—a considerable challenge given that the criteria for protein 'fitness' are generally not well understood. Moreover, many fitness criteria, such as catalytic activity, are currently difficult to define as objective functions, which are needed for design optimization algorithms. Finally, because both sequence and structure space are vast, a finer mapping of these spaces (computationally and experimentally) is needed to develop more accurate predictive models. In light of these challenges, recent studies have highlighted exciting developments in computational methodologies that are ready for wider-reaching applications.

Key concepts in design methodology

Protein design, and the related problem of protein structure prediction, both require an understanding of the relationship between sequence and structure[29]. Thus, they share common challenges. Structure prediction starts with a known sequence, but involves probing, or 'sampling', many possible protein conformations. Given the vast space of possible structures, sampling conformations that resemble folded and functional proteins ('near-native' conformations) is difficult. A second challenge is positively discriminating near-native protein conformations from the great number of other possibilities that significantly differ from experimentally observed three-dimensional structures of the protein in question. This issue is called the 'scoring problem'. To further complicate matters, sampling and scoring are generally not independent: a 'scoring function' that evaluates the energetics of sampled conformations typically guides the progress of protein structure prediction and design simulations.

Current protein design methods make key simplifications in sampling and scoring to enable searching through both sequence and conformational space. Sampling strategies limit accuracy by restricting the number of backbone and side chain conformational degrees of freedom that are varied during a simulation (**Box 1** and **Fig. 2**). Several assumptions in design energy functions limit their accuracy as well (**Box 2** and **Fig. 3**). The contribution of conformational entropy is generally ignored, and for computational efficiency, waters are typically not explicitly considered. Electrostatics and solvation models are often simplified to be pairwise additive, which precludes modeling environment-dependent energies[30–34]. Several recent reviews describe in more depth conformational sampling and sequence optimization methods[27,35] as well as design energy functions[27,36,37].

Achievements and challenges in protein interface design

The objective of designing proteins with not only new structures[35] but also novel functions has been realized by engineering new and modified binding activities into a number of different protein systems. With a focus on the design of protein-protein interfaces, we will highlight achievements using the frameworks illustrated in **Figure 1** and specific experimentally characterized examples (**Fig. 4**).

Making existing interactions stronger. A frequent application of computational interface design is to increase the affinity of interactions[38–43] (**Fig. 1a**). Protein variants binding more tightly to their partners are useful as specific inhibitors for detailed mechanistic studies of existing interactions. They also have potential therapeutic uses, either for target validation or as biotherapeutics themselves. A striking example of a proof-of-concept design success in this context is increasing the affinity of antibody-antigen interactions beyond *in vivo* levels[39].

A basic philosophy underlying several approaches is to target residues for design whose computed per-residue contribution to the binding energy is less favorable than what would be expected, on average, for the same amino acid in a similar protein environment. A straightforward strategy is to increase the amount of hydrophobic surface area buried in the interface[40]. Because design energy functions make substantial approximations, however, correct ranking of designed sequences is a significant challenge, and many predictions

overestimate the stabilizing effect of small-to-large mutations. Both the entropic cost of introducing large side chains and the penalty of burying neighboring polar groups resulting from an increase in hydrophobic surface area are frequently underestimated[39]. Instead, two strategies for electrostatic stabilization have been proposed[39]. In the first, polar groups with poorly satisfied hydrogen bonding donors and acceptors are replaced by hydrophobic residues to decrease desolvation costs. An alternative is to engineer intermolecular charge-charge interactions in the interface periphery where desolvation is minimal. Both approaches may improve designs because estimating their effects primarily involves calculating solvation and medium-range electrostatics that are less sensitive to precise atom locations. A related strategy has focused on increasing on-rates of protein interactions through introduction of electrostatic steering interactions[44].

Most successful strategies to increase interaction affinity perform individual mutations—often selected to be separated from each other in the three-dimensional structure—and then combine those mutations that were experimentally found to stabilize the interface[38,39,41]. This strategy of course ignores potential coupling effects between the individual mutations. It may, however, avoid the problem that one destabilizing mutation out of many designed changes in the interface can 'mask' many correctly predicted stabilizing substitutions[27].

An example for designs requiring multiple, likely coupled mutations is the engineering of an intricate hydrogen bonding network, since all hydrogen bonding donors and acceptors buried in the interface should be satisfied. Such a network has been designed and structurally verified[45], but it remains difficult to estimate the contribution of the designed polar interactions to the binding free energy. Consequently, the designed hydrogen bonding network resulted in an overall weaker interaction affinity of the designed complex, although the predicted hydrogen bonding geometry closely resembled that observed in the crystal structure.

Changing interaction preferences. A fascinating but challenging application of protein interface design is changing the specificity of interactions (**Fig. 1b–d**). Specificity is by definition a property of multiple 'states', such as when a protein adopts different conformations or binds multiple partners. A selectivity design strategy, then, must not only consider optimization for the target, termed 'positive design', but may also involve 'negative design', which selects against 'competitors': alternative conformations (for example an aggregated state) or unwanted binding partners. Specificity then arises from the preference for a given conformation or set of binding partners over others. A general computational strategy to formulate the design of specificity as a multistate optimization problem has been developed and successfully applied to engineer selective coiled-coil pairs[46].

As may be guessed, the necessity for negative design will depend on the potential for cross-reactivity in the design system. In an elegant study converting a homodimer into a heterodimer, both positive and negative design strategies were compared on the same system[47]. In cases where desired and unwanted binding partners have similar structures and sequences, negative design appears essential[15,46,47]. Conversely, positive design alone[18,19] may be sufficient if there are fewer features in common between targets and competitors[47].

A simple proof-of-concept objective for specificity redesign is to convert an existing binary protein interaction into a new selective protein pair by substantially remodeling the interface on both partners (**Fig. 1b**, left). Such 'orthogonal' interacting pairs have been constructed[16,45] and are useful for engineering new signaling pathways, as well as for characterizing the role of cross-talk in existing networks. To construct

Box 1 Sampling concepts in computational protein design

The sampling problem in protein design is typically simplified by the use of a fixed polypeptide backbone that is commonly taken from an experimentally determined protein structure. Increasingly, backbone flexibility is taken into account in design as well, either by designing on an ensemble of backbone conformations, by iterating backbone and sequence optimization or by simultaneously exploring backbones and sequences (see also **Table 3** for details and references on flexible backbone methods). Sequence and side chain conformational space are sampled employing a library of different rotamers[93] (conformations generated by discrete rotations around the side chain torsion angles) for each amino acid. Generally, flexible backbone methods yield a larger diversity of designed sequences and side chain conformations, as can be seen by comparing **Figure 2a** with **b**.

Because complete enumeration of all combinations of sequences and side chain conformations is generally intractable—even when using a fixed backbone—design methods employ a variety of computational optimization techniques for finding low energy sequences[94]. Optimization methods are subdivided into 'deterministic' approaches such as dead-end elimination[68,69,95] that are guaranteed to reach the lowest energy sequence for a given backbone and scoring function (unless the design problem is too large), and 'stochastic' methods that are not. Frequently applied stochastic approaches use Monte Carlo simulated annealing[96], genetic algorithms[6] or the FASTER method[97,98]. As can be expected, there is a tradeoff between accuracy and speed of sequence optimization methods[99].

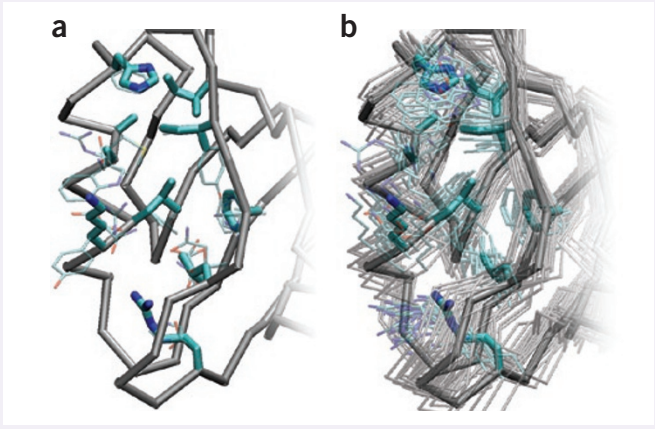

Figure 2 Sampling in fixed and flexible backbone protein design. Shown are low energy sequences and side chain conformations from 40 independent design simulations using Monte Carlo simulated annealing as described in ref. 96 in the core and interface region of a small protein recognition domain (PDZ). The crystal structure backbone (Protein Data Bank (PDB) code 1BE9) is shown as silver Cα trace, the wild-type side chains in thick stick representation, and alternative designed side chains in transparent thin stick representation. (**a**) Side chain sampling on the fixed backbone taken from the crystal structure yields only a small number of different low energy sequences and side chain conformations. (**b**) More diverse low energy sequences and side chain conformations are found using a flexible backbone design protocol, where backbone movements and sequence design steps are iterated. Sampled backbone conformations are shown in thin transparent Cα trace.

Box 2 Scoring concepts in computational protein design

Energy or 'scoring' functions in protein design (**Fig. 3**) generally focus on modeling three fundamental components: (i) detailed atomic packing interactions within and between proteins (**Fig. 3a**), (ii) polar interactions involving ion pairs and hydrogen bonds, often including an explicit hydrogen bonding orientation dependence[100,101] (**Fig. 3b**) and (iii) the interactions of protein atoms with the solvent, including a penalty for the desolvation of buried polar groups (**Fig. 3c**). Many scoring functions used in design frequently also contain statistical terms to describe propensities of amino acids for backbone and side chain torsion angles, approximations to the conformational entropy of protein side chains, and the energy of the unfolded state ensemble[96,102].

While these approximate energy functions have been successfully used in protein structure prediction and refinement as well as design, it should be noted that accurate structure prediction may be easier than correctly ranking designed sequences with similar energies, since near-native structures generally have much lower predicted energies than other structures. For this reason, experimental tests of large numbers of designed sequences to evaluate and improve scoring functions are of great interest.

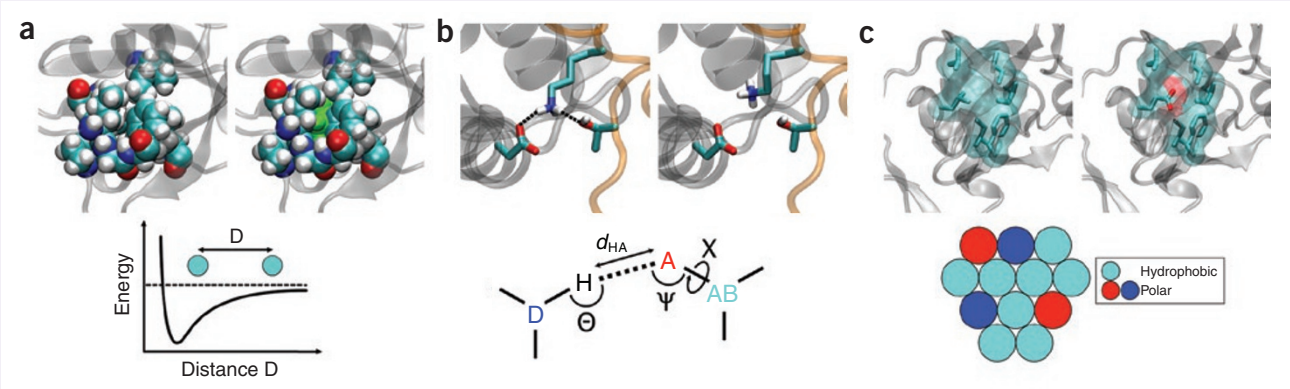

Figure 3 Dominant contributions to protein design scoring functions. (**a**) Atomic packing interactions. The well-packed core of a PDZ domain (top left) would be destabilized by changes in side chain size, such as an alanine-to-phenylalanine substitution (top right, phenylalanine shown in green). Packing interactions are typically described using van der Waals potentials (bottom). The backbone is shown in transparent cartoon representation, and side chains are shown as spheres. (**b**) Hydrogen bonding. The intricate geometry of a hydrogen bond network in the interface between the DNAse E7 (orange backbone) and its inhibitor Im7 (grey backbone) (PDB code 7CEI) (left) would be disrupted by changes in side chain donor/acceptor groups and their orientations (right). The bottom panel shows parameters used in a geometry-dependent hydrogen bonding potential[101]: the distance between the hydrogen (H) and the acceptor (A) (d_{HA}), the angle at the hydrogen (Θ), the angle at the acceptor (Ψ), and the donor (D)–hydrogen (H)–acceptor (A)–acceptor base (AB) torsion angle (X). Side chains are shown in stick representation. (**c**) Solvation. The hydrophobic core of a PDZ domain (left) would be destabilized by an isoleucine-to-glutamate substitution (right). Solvation potentials favor the exposure of polar groups to the solvent and associate a penalty with burial of polar groups (bottom). Core side chains are shown as sticks with a transparent surface representation.

these pairs, one strategy is a 'computational second-site suppressor' approach[16]. First, a mutation is introduced into one of the partners (A′) to destabilize its interaction with its original partner, B. Then mutations in the interface on the second partner are sought (B′) that both compensate for the A′ mutation to form a stable A′-B′ complex and destabilize the A-B′ interaction.

Two patterns for specificity design have emerged through this procedure[16]. In one approach, called a 'steric switch', destabilization and compensation are achieved through 'bump-hole' designs, similar to a strategy for designing specific kinase variant–ATP analog pairs[48]. A second pattern is a 'polarity switch', where an intermolecular polar interaction network is replaced with a hydrophobic one or vice versa[16]. Changes in specificity suggested by computational design have been compared to those observed in naturally occurring homologs of the redesigned pairs, and both existing and new motifs were discovered[16,46].

Designing promiscuity and interaction patterns. In a variation on the theme of selectivity design, it may be desirable to optimize a protein to bind not just one, but multiple target proteins (**Fig. 1c**). Such 'multi-specificity' is a property of many 'hub' proteins in naturally occurring protein-protein interaction networks. Frequently, the hub protein uses a shared interface to bind several of its partners, as could be the case for proposed 'date hub' proteins that interact with their partners at different times or locations[49]. In such cases, in order to dissect the functions of individual interactions, the objective of protein design may be to reengineer a 'multi-specific' interface to selectively recognize some, but not all of its binding partners (**Fig. 1d**). A computational design strategy applied to such multi-specific proteins with shared interfaces suggested that it may be possible to alter their interaction patterns[50] (**Fig. 5**). The sequences of shared interfaces were designed in two different computational experiments. First, the sequence of the shared interface on a hub protein was computationally optimized for each binding partner separately. In the second computational experiment, the design strategy was switched to a 'multi-constraint' approach that aimed to predict hub protein interface sequences that would be consistent with all interaction partners considered. As expected, the multi-constraint approach yielded sequences that were much more similar to the naturally occurring interface sequence, which may be constrained by the requirement to bind its multiple partners. Moreover, a comparison of the resulting 'single' and 'multi-constraint' sequences provides information on, first, how the naturally occurring interface may be optimized for its promiscuity and, second, how the interface could be changed to alter interaction patterns. If it is possible to predict mutations in a shared interface that selectively affect certain interactions and not others (**Fig. 1d**), an upcoming challenge is then to develop experimental methods to characterize the biochemical and *in vivo* consequences of these perturbations. Such a combined

computational and experimental approach may enable mapping of the cellular functions of individual interactions involving hub proteins.

Creating new interfaces. The computational design and structural validation of an entirely new high-affinity protein-protein interface is still an outstanding challenge, but progress has been made toward this goal. One strategy for creating *de novo* interfaces is to derive template backbone arrangements from large-scale protein-protein docking experiments, and then to optimize interface sequences for each of the complex templates. Using this approach, a small domain was converted from a monomer into a heterodimer, although with only modest high-micromolar affinity[51].

A simpler problem than creating a *de novo* interface is to design a new interface between two proteins that do not interact with measurable affinity, but have structurally similar partners that do. This goal was achieved by fusing domains from distantly related homing endonucleases and optimizing their interface for binding (**Fig. 4b**). Introduction of eight computationally designed mutations resulted in a new functional chimeric protein whose crystal structure matched the design predictions[14]. In this study, both partners were redesigned to optimize their interaction. A more common case may be that a specific binding partner is desired for a native target whose sequence cannot be altered simultaneously. For example, the binding site of a PDZ domain, a common mediator of protein-linear motif interactions, was redesigned to match a number of defined peptide targets[17].

Both design studies above were aided by the target interfaces being structurally well defined. In the chimeric protein case, the interface was formed around two well-packed α-helices. In addition, most of the redesigned side chains in the interface area were buried, and the backbone, which was kept fixed in the design simulations, was found to be essentially unchanged when the crystal structure of the designed complex was solved[14]. In the case of PDZ domain–peptide interactions, the backbone is also largely constrained, as the binding peptides complement a β-sheet in the PDZ domain by inserting an extra strand between the protein sheet and an α-helix.

Directed evolution experiments have been very successful in generating many examples of new interfaces, often using antibodies or mimicking their selection strategy by randomizing loops displayed from a protein 'scaffold' and selecting for binding[52]. Computational strategies may provide advantages over directed evolution experiments in cases where the binding interface needs to be defined a priori (for example, because a certain mechanism of activation or inhibition is desired). Second, computational approaches may be preferred if the design strategy has multiple objectives, such as binding to multiple targets and avoiding others. These requirements could also be included in the screening procedure, together with other potentially unknown selection pressures that are implicit in a screen. Nevertheless, if multiple desired objectives can

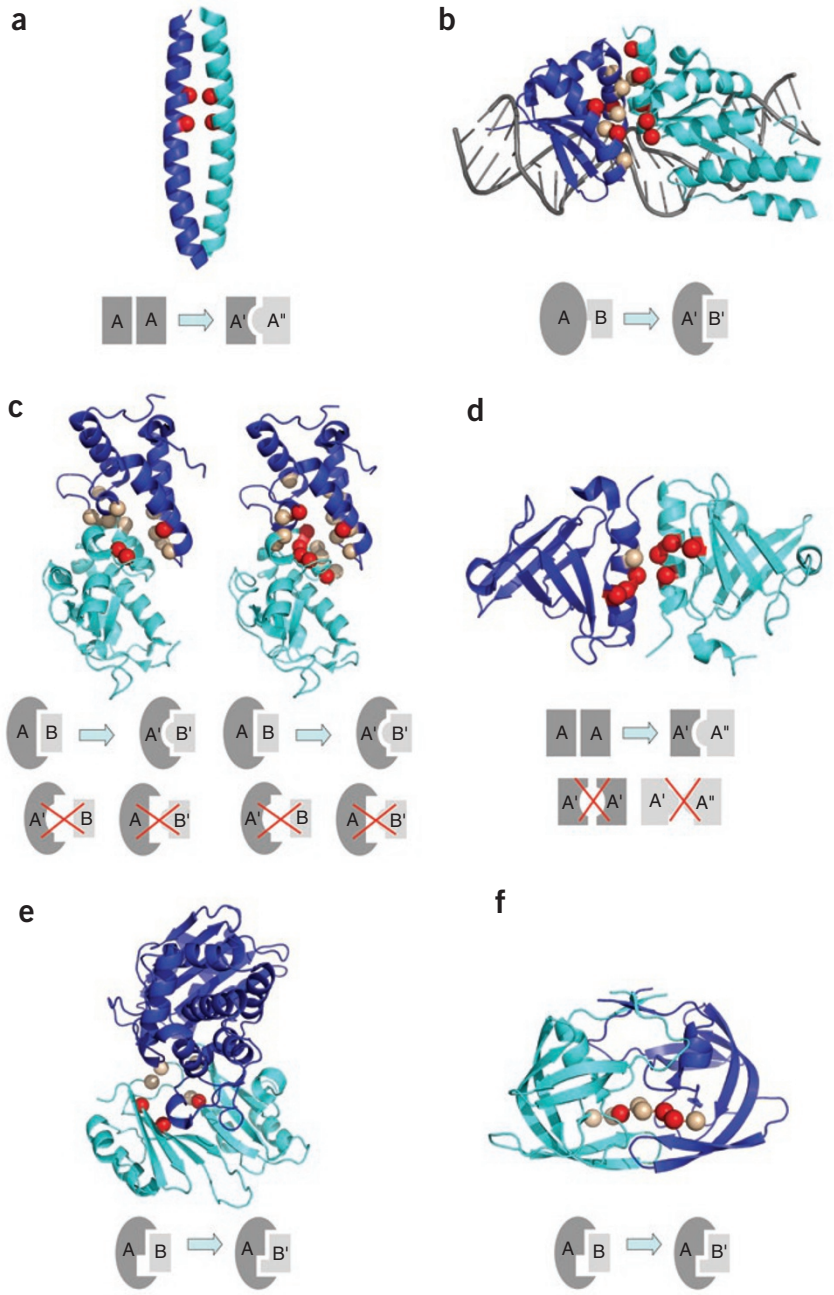

Figure 4 Structurally characterized designed protein-protein interfaces. In each case, different protein chains are shown in cyan and blue in cartoon representation. The residue positions considered in design are shown as color-coded spheres at the Cα position: red positions were mutated from the wild-type residue; wheat positions retained the wild-type amino acid residue type after design. Schematics illustrating the design objective are as in **Figure 1**. (**a**) A designed heterodimeric coiled-coil, PDB code 1KD8 (ref. 84). (**b**) A designed chimeric homing endonuclease, PDB code 1MOW (ref. 14). The protein is bound to DNA shown in grey. (**c**) Two designed pairs of a colicin DNase-inhibitor complex, PDB codes 1UJZ (left)[16] and 2ERH (right)[45]. (**d**) A designed heterodimer derived from the SspB homodimer, PDB code 1ZSZ (ref. 47). (**e**) A designed interface between SHV-1 β-lactamase and β-lactamase inhibitor protein, PDB code 3C4O (ref. 42). (**f**) A peptide redesigned to bind the HIV-1 protease homodimer with increased affinity, PDB code 2NXD (ref. 85).

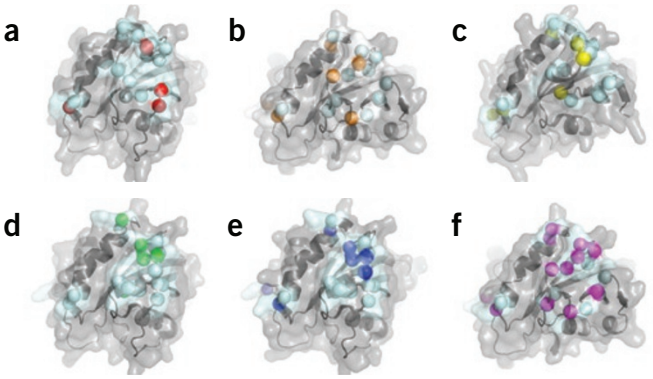

Figure 5 Design of a multi-specific protein-protein interface. (**a**–**f**) Using the GTPase Ran as an example, the figure illustrates the predictions resulting from two computational experiments[50]: a 'single-constraint' simulation, where the shared interface on Ran is optimized using computational design to match binding to several of its partners separately (**a**–**e**, corresponding to Ran in complex with partners 1–5 in **Fig. 1d**), and a 'multi-constraint' simulation, where the shared interface is optimized to be consistent with all 5 binding partners (**f**). Shared interface residues on Ran are defined as positions that are contacted by at least three partners, and are indicated by Cα spheres and light cyan surface shading (Ran binding partners are omitted here to allow viewing the Ran interface). In each case in **a**–**e**, if the design simulation identified the native amino acid residue type as the most favorable, the Cα sphere is color-coded according to the respective partner in **Figure 1c**. The pattern of positions where the design methodology identifies the native amino acid type are varied between partners, suggesting strategies to optimize interaction selectivity. Slight conformational changes of the Ran interface in the different complexes can also be appreciated. The multi-constraint simulation yields a substantially more native-like interface (**f**, magenta), suggesting that multi-constraint design reflects at least to some extent the multiple constraints acting on the Ran interface to bind its partners.

be encoded correctly and efficiently in a design procedure, it may also be possible to characterize how the different objectives shape the resulting sequence 'fitness' across the landscape of input constraints.

Experimental characterization is key. At least two types of information are critical for further development of more accurate computational design methods.

First, experimentally determined three-dimensional structures of the designed molecules are required to assess and improve the quality of computational methods, and often more can be learned from predictions not born out as expected. Structural information is also necessary to judge whether a computational design was successful for the right reasons. **Tables 1** and **2** highlight cases where computational designs were characterized structurally in two different areas relevant to this review: when protein interfaces were redesigned (see also **Fig. 4**), and when flexibility of the protein backbone was explicitly taken into account in the design process (which we will discuss further below). Solved protein structures illuminate some of the key difficulties in modeling protein interfaces. A particular characteristic of interfaces, in contrast to protein core design, is their substantially polar surface that needs to be satisfied by designed interactions. Accordingly, several results from experimental interface design studies highlight the difficulty in modeling the balance between electrostatic and solvation effects, in particular with interface-bound water molecules or complex polar networks[45], where polarization effects cannot be ignored. In some cases, defined water molecules in protein-protein interfaces cannot be displaced upon design, even when specific side chain–side chain interactions were intended to replace the water-mediated interactions[45].

A second source of needed experimental information relates to the design process itself. Fundamentally, computational design seeks to rank sequences with respect to the structural and functional requirements of the design objective. While the major demonstration of success in design has been through experimental proof-of-concept analysis of one or a few sequences, deemed most favorable by a scoring function and manual inspection, a more informative test may be to experimentally characterize as many designed sequences as possible. A comparison of the computational and experimental rankings of designed sequences for different design criteria, such as protein interaction affinity and monomer protein stability, would contribute to a better understanding of successes and failures of design scoring functions.

Translating modeling progress to applications

There are several key areas in high-resolution protein modeling where recent advances may contribute to protein interface design. First, the high-resolution modeling of protein conformational ensembles can broaden and enrich the predicted sequence space for functional proteins. Additionally, more extensive local structural 'remodeling' of protein regions can enable engineering new functional conformations or reshaping protein surfaces to create entirely new interactions. Further, multi-state methodologies for optimizing protein sequences for many structural and functional constraints may aid the design of proteins that perform specific functions in the context of complex biological environments. We next highlight progress in each of these areas.

Backbone flexibility in protein design. The vast majority of the successful applications of computational protein design of structure and function to date have relied on the assumption of a fixed polypeptide backbone onto which amino acid side chains are modeled. Taking into account rearrangements of the protein backbone is a considerable challenge, both because of the enormous increase in computational complexity and owing to the fundamental need for increased accuracy in sampling and scoring methods (**Boxes 1** and **2**). Thus, it has been a long-standing problem to develop and, importantly, adequately test flexible backbone protein design approaches[53]. Assessing these methods is not trivial, as engineering successes generally do not provide enough test cases. Encouragingly, large informative datasets are becoming increasingly available. Examples include collections of structural changes observed upon point mutations[54], as well as sequences compatible with a given protein interaction[55,56], where sequence variation is likely concomitant with slight structural adjustments. Approaches to backbone flexibility are summarized in recent reviews[35,53,57] and include making small random moves of the ψ and ϕ backbone torsion angles[7], sampling backbones from a parametric family of structures[10], performing normal mode analysis[58], gathering backbone ensembles from available crystal structures[59] or computational methods[60–62], and iterating between sequence and structure optimization borrowing techniques from protein structure prediction[11,63–65]. More technical details on different approaches and their assessment are given in ref. 53 and highlighted in **Table 3**.

There are several classes of design problems that require a treatment of backbone flexibility to varying degrees (**Table 2**). First, the ability to predict backbone conformational changes at high resolution is needed to model how proteins react structurally to the mutations introduced in protein design. This is critical, as the inability to accurately model conformational changes is a major reason that design predictions fail. Even when structural adjustments are relatively small, the energetic effects of these rearrangements can be substantial. In this respect, it is encouraging that methods to model small structural changes in response to mutations have shown improvements in predicting the effects of point mutations on protein structure[54,66] and stability[67].

Additionally, backbone flexibility is needed to explore conformations around a desired structural template, particularly in cases where the target protein topology is not observed in nature. In this case, it is not clear whether the intended structure is designable, and iterating between optimizing a sequence for a fixed structure and optimizing the structure for a fixed sequence—which may be necessary to sample a designable conformation—has been successful[11,64].

In related applications, it has been shown that designing on ensembles of near-native sampled conformations can change the characteristics of the predicted sequences in comparison to designing on the native structure alone. Sequences derived from flexible backbone approaches are typically lower in predicted energy[68,69] and in many cases more closely resemble sequences in the natural family of the target protein[59,60,65,70]. Thus, even though differences in selection pressures complicate comparisons between natural and designed sequences, flexible backbone methods may nevertheless open exciting avenues to fundamentally improve understanding of the relationships between sequence, structure, dynamics and function in natural protein families[53].

Most practically, flexible backbone methods broaden the design sequence space and thus may increase the probability of selecting functional sequences. A recent study has demonstrated that designing sequences for a protein-protein interface on an ensemble of backbone conformations improved the correspondence between amino acid type distributions selected computationally[61] and from comprehensive phage display selection experiments[55]. Thus, to maximize success with challenging protein interface engineering projects, structure-based computational methods may become valuable in designing sequence libraries enriched in functional members that can then be experimentally screened[71]. Additionally, improvements to sampling methods may in turn help to detect and overcome weaknesses of scoring functions.

Remodeling proteins for function. In general, the flexible backbone approaches described above produce deviations from starting structures that are quite small. However, important protein interface engineering objectives require development and assessment of more aggressive approaches to modeling protein conformational variability. These tasks

Table 1 Structurally validated protein interface designs

Goal/system	Design strategy	Biochemical and structural validation	Ref.
Create a new, chimeric DNA cleaving enzyme (**Fig. 4b**)	Fusion of two domains from natural endonucleases and redesign of the chimeric interface	Designed 14 interface positions to produce an enzyme with native-like cleavage rates for a chimeric DNA target site by introducing 8 mutations A crystal structure matched the design to 0.8 Å Cα r.m.s. (root mean square) deviation	14
Alter specificity of colicin DNAse–immunity protein pairs: design complex mutants that recognize each other but not their wild-type partners (**Fig. 4c**)	'Computational second-site suppressor' strategy: design of interface mutations intended to destabilize interactions with the wild-type partners while stabilizing the mutant complex	3 designs were demonstrated to have the desired switched interaction selectivity (although weaker affinity than the original complex) through *in vitro* binding assays and functional assays *in vitro* and in *E. coli* The crystal structure of one mutant complex agreed with the design to 0.5 Å Cα r.m.s. deviation Some differences between the predicted and experimental structures were attributed to water molecules with specific interactions not included in the design calculations	16
Increase the specificity switch between mutant and wild-type colicin DNAse–immunity protein complexes (**Fig. 4c**)	Design of interface residues on a series of systematically sampled rigid-body orientations of the two protein partners to optimize specificity for a binding mode; optimization of the design based on a crystal structure	Design of a new hydrogen bonding network in a protein interface, and confirmation by X-ray crystallographic analysis The crystal structure of the design revealed a water molecule maintained in the mutant interface, resulting in conformational strain and packing defects Further design based on insights from the mutant crystal structure conferred at least a 300-fold specificity switch against the wild-type partners	45
Convert the homodimeric SspB adaptor protein into a stable, specific heterodimer (**Fig. 4d**)	Asymmetric design of 8 positions at the dimer interface using one strategy that accounted only for heterodimeric stability (positive design), and another that explicitly penalized competing homodimeric states (negative design)	The stability-only calculations produced a complex that was more stable but also existed in homodimeric states, while taking negative design into account was necessary to obtain specificity for heterodimers, at the expense of stability A crystal structure of the heterodimer showed that designed side chains were accurately predicted, although two side chains differed at the χ_2 angle, and one side chain with a high B-factor differed at the χ_1 angle	47
Increase affinity and specificity of β-lactamase inhibitor protein (BLIP) for SHV-1 β-lactamase (**Fig. 4e**)	Redesign of clusters of residues in the BLIP–SHV-1 interface to increase the stability of that complex	Several designs increased affinity for SHV-1 while decreasing affinity for TEM-1 β-lactamase Crystal structures of the two highest affinity complexes showed that several designed side chain conformations were correctly predicted, but a critical interfacial salt bridge was missed in both models; the authors suggested that flexible backbone methods with a larger rotamer library might have predicted the salt bridge	42
Increase the affinity of RT-RH peptide for inactivated HIV protease (**Fig. 4f**)	Combination of predictions from charge optimization and sequence design to produce single, double and triple mutants at 8 residue positions	Isothermal titration calorimetry showed that one of three designs improved affinity by tenfold; observed affinities arose from reducing the entropic cost of binding at the expense of lowering enthalpic gains compared to the wild-type sequence Crystal structures of two of the designed peptides showed good agreement with the models, while a third showed rotameric deviations at designed positions	85
Predict the structures and stabilities of a family of heterodimeric coiled coils (**Fig. 4a**)	Sequence design of hydrophobic residues at opposing positions in the oligomeric interface, and enforced heterodimerization though oppositely charged interface residues; designs used an ensemble of backbone conformations derived from parameterized models of coiled coils	Predicted stabilities of minimized designs showed good agreement with values from chemically induced unfolding experiments Crystal structures were within 0.7 Å r.m.s. deviation of the models over all heavy atoms on average	84

Table 2 Structurally validated designs employing flexible backbone approaches

Goal/system	Design strategy	Biochemical and structural validation	Ref.
Design dimeric, trimeric and tetrameric helical bundles with right-handed topology	Hydrophobic-polar residue patterning and side chain packing calculations applied to algebraic models of helical bundle backbones with a right-handed superhelical twist	Sedimentation equilibrium analysis showed the oligomerization states matched the designed topologies Circular dichroism (CD) confirmed helical structure A crystal structure of the tetrameter matched the core atoms of the design to 0.20 Å r.m.s. deviation	10
Design a globular protein fold not observed in nature	Iteration between sequence design for a fixed backbone and structure refinement for a fixed sequence, using techniques from *de novo* structure prediction	NMR and CD experiments showed the protein was well folded and thermostable A crystal structure matched the design to 1.17 Å r.m.s. deviation over all backbone atoms	11
Introduce new loop sequences and structures into the protein tenascin	Grafting of 10-residue loops from the PDB and iteration of sequence design and structure refinement	NMR and CD showed that three designs were well folded and two had near-wild-type thermostabilities A crystal structure of one loop matched the design to 0.46 Å backbone r.m.s. deviation	64
Switch enzyme substrate specificity in guanine deaminase by redesigning an active site loop	Design and remodeling of a shortened 4-residue loop with constraints to position functional groups on side chains	The redesigned loop produced a 2.5×10^6-fold substrate specificity switch from guanine to ammelide with moderate catalytic activity A crystal structure of the loop was within 1 Å Cα r.m.s. deviation of the model	77

include creating proteins that are selective in the context of complex cellular networks, engineering new high-affinity binders for specified targets, and designing 'switches' that change their conformation in response to a signal. The overarching goal is to produce reliable predictive methods to engineer new and functional conformations into proteins.

A key idea used in several strategies for engineering new protein functions has been to 'transplant' elements from naturally occurring proteins with a given function into a new context. Many of these concepts were developed to place sets of residues into proteins forming the precise geometry of metal binding sites[13]. Related ideas were used to graft the putative binding epitope of interleukin-4 for its receptor onto a helical face of the leucine-zipper domain of the yeast transcription factor GCN4 as a scaffold[72], and to transplant 'hot spot' residues for the interaction of erythropoietin with its receptor onto an unrelated small protein, a PH domain[73]. A similar geometric matching strategy was applied to replace a set of residues in the TEM1 β-lactamase–inhibitor interface with a five-residue motif from a different protein[74]. Interestingly, the chosen motif was not derived from another protein-protein interface, but from a protein core region[74].

Several computational strategies have been developed to allow the matching of desired features—hot spot residues, interaction motifs or loops—to a given scaffold[2,73–75]. In general, these methods find locations in the backbone of a potential scaffold that have the correct relative spatial arrangement and directionality of C_α-C_β vectors matching the geometry of the desired motif. Such a strategy was also used to generate proteins with new catalytic activities[23,24] by finding scaffold proteins that could support the arrangement of putative catalytic residues around a modeled transition state[75].

In all cases described above, the polypeptide backbone of the scaffold stayed fixed, and the computational strategies employed varying degrees of design and remodeling restricted to side chain conformations around the introduced motif. In the general case, 'motif-directed design', or grafting of multiple protein segments to reengineer entire binding sites, may require remodeling of the protein backbone, either to accommodate the precise geometry of the motif or to reshape the area around it to support and stabilize the introduced substitutions[76]. Toward these goals, entire loops have recently been grafted and redesigned, while allowing some backbone flexibility, to create new loop sequences with correctly modeled structures at high resolution[64] and to modulate the substrate specificity of an enzyme[77].

A promising approach to more extensively 'reshape' protein surfaces and mediate new interactions could take advantage of recent developments in the high-resolution modeling of protein regions devoid of regular secondary structure. New methods employing 'inverse kinematics' can rapidly move loops or defined protein segments to new conformations with prescribed endpoints, bond lengths and bond angles[78]. The basic principles are borrowed from robotics, where calculating the accessible conformations of objects subject to constraints, such as determining the possible positions of the interior joints of a robot arm given fixed positions for the shoulder and fingertips, is a well-studied problem. For proteins, the task becomes modeling the 'mechanically accessible' conformations of segments in the structure, and evaluating the larger number of sampled conformations using a physically realistic energy function. If these methods of segment reshaping can be successfully integrated with protein design, they open avenues to mimic the natural ability of proteins to use loop regions for achieving tight surface complementarity in interfaces. Such methods for surface-matching could be used in combination with grafting hot spot motifs for engineering new tight binding partners for existing targets, for reshaping peripheral interface regions to increase selectivity, or for creating new functional conformational switches.

Modeling context and function as constraints. To engineer new proteins that function in realistic biological contexts, techniques are needed to optimize protein sequences for 'fitness' within a complex environment, where proteins must form desired interactions while avoiding unwanted competitors. Other fitness constraints may require proteins to adopt multiple conformations to function. A signaling switch protein, such as a small GTPase, has many functional requirements, some of which are unknown. The GTPase needs to adopt different conformations and transition between them in response to interactions with nucleotides and other proteins, while specifically recognizing a variety of downstream binding partners responsible for propagating the switch signal. While designing such complex biological systems computationally is a daunting task, many recent computational developments mark progress toward this goal. These developments include methods to model at high resolution how protein switch loop regions change in response to binding different partners[78], to design specific small-molecule sensors[21] and biocatalysts with activities not observed in nature[23,24], and to integrate multiple positive and negative constraints on protein binding[15,46,50,79,80].

As discussed above, negative design is particularly important in cases of strong structural and sequence similarity between targets and competitors[15,46,47]. Such cases frequently arise because many protein interactions in cells are mediated by large families of domains using recurring protein-protein and protein-peptide interaction motifs.

Table 3 Approaches to modeling backbone flexibility in protein design

Method	Features	Computational benchmark	Experimental validation	Refs.
Parameterized models of coiled coils	Enables modified hydrophobic packing arrangements for alternative topologies		Circular dichroism, sedimentation equilibrium experiments, hydrogen-deuterium exchange and X-ray crystallography confirmed the predicted superhelical twists	10
Random ϕ, ψ and ω perturbations	Generalizes to any topology	Agreement with experimentally measured stabilities of T4 lysozyme mutants		7
Insertion of peptide fragments	Reduces search space by capturing experimentally determined structural preferences	R.m.s. deviation to the native structure for *ab initio* models of 7 small proteins		86,87
Fragment insertion with compensating torsion moves	Reduces nonlocal perturbations due to fragment insertion	R.m.s. deviation to the native structure for a large dataset of 4- to 34-residue regions lacking secondary structure	Circular dichroism, NMR experiments and X-ray crystallography confirmed the design of a new protein fold	11,88
Fragment insertion with small torsion moves and the cyclic coordinate decent (CCD) loop closure method	Enables closure of breaks in peptide chains from fragment insertion and small torsion moves	R.m.s. deviation to the native structure of *ab initio* 8- and 12-residue loop reconstructions, and r.m.s. deviation to the native structure of complexes docked during loop modeling	Design of 3 new loop structures and sequences in tenascin; remodeling and redesign of a loop to switch substrate specificity in guanine deaminase	64,77,89
Backbone ensemble generation by normal mode analysis	Relatively inexpensive and generally applicable if a template structure is available		Peptides designed to bind Bcl-xL showed selective binding compared to other Bcl family members in pull-down assays	58
Integer linear programming on backbone templates	Simultaneously samples sequences and backbone conformations	Comparison of sequences designed for human β-defensin 2 to naturally occurring homologous sequences		90
Dead-end elimination (DEE) with random ϕ, ψ moves	Finds the lowest energy sequence and conformation within a predefined volume	Comparison of predicted energies of sequences designed for the β1 domain of protein G and gramicidin synthetase A to a fixed backbone DEE method		68
DEE with 'backrub' moves inspired by conformational variability in crystal structures[91]	Finds the lowest energy sequence within conformations derived from local nonoverlapping tripeptide backrub moves	Comparison of predicted energies of sequences designed for the β1 domain of protein G and gramicidin synthetase A to a fixed backbone method		92
Monte Carlo with backrub moves	Enables small local moves for peptide chains of arbitrary length	Comparison of the accuracy of predicted side chain conformations of point mutants to a fixed backbone method; agreement of protein side chain and backbone conformational variability with NMR dynamical measurements; comparison of designed sequences to natural protein homologs for ubiquitin; agreement of predicted human growth hormone sequences designed to fold and bind human growth hormone receptor with experimental screening		61,62,66,70
Monte Carlo with kinematic closure	Analytically determines 6 ϕ, ψ torsions of peptide chains while efficiently sampling bond angles, bond lengths and remaining torsions for peptide chains of arbitrary length	R.m.s. deviation to the native structure for *ab initio* reconstruction of the conformations of 12-residue loops in proteins and interfacial loops in protein-protein complexes		78

Hence, distinguishing desired from undesired binding partners may require consideration of many structurally similar states. A remarkable recent study did just that, creating peptide partners for targets from 19 families of human basic-region leucine-zipper transcription factors that were selective in the context of competing factors from the other families[15]. A key to solving this challenging engineering problem was the development of elegant and fast computational methods that could evaluate the different states[80] and model tradeoffs between stability and selectivity, aided by the relatively regular and well-characterized interaction geometries of the leucine-zipper interfaces[15].

Outlook

As methods are being developed and validated to model multiple constraints and achieve selectivity among sets of structurally similar binding partners, computational design is becoming a useful tool for controlling biological processes mediated by functional protein interactions. Toward engineering and analysis of *in vivo* protein function, there are several key areas in need of progress.

Typical protein-protein interfaces involve large, flat interacting surfaces with nonlinear epitopes and conformationally coupled buried side chain interactions. These features endow energetics that require more accurate and tractable models of the physical basis of interaction affinity and selectivity. In parallel, higher accuracy methods to predict the dynamical responses that proteins undergo upon binding or mutation are required. Continued cooperative development between structure prediction and design methodologies[29] will be central to meeting these goals, together with large-scale experimental characterizations of ranked lists of designed sequences.

Key next steps are also to derive and demonstrate methods that can optimize the complex interfaces of protein signaling hubs for

functional interactions with partners of varied fold classes while specifically eliminating unwanted interactions. More generally, one ultimately would like to develop methods to optimize 'network fitness' for any type of quantifiable criterion, such as adopting specific conformations or interacting with certain association rates in addition to exhibiting a specified network topology.

In addition to engineering new molecules, much can be learned from applying computational protein design approaches to disrupt specific interactions, and then studying the phenotypic consequences of these perturbations using large-scale physical and genetic interaction mapping methodologies[81–83]. Such a combined approach would be complementary to powerful, commonly used strategies, such as knockout and knockdown methods, in that it may allow specific modulation of a subset of the interactions made by a given protein, instead of eliminating or affecting all interactions (**Fig. 1d**). In this respect, computational protein design as an engineering approach may contribute to a more fundamental understanding of the balance of interaction selectivity and promiscuity regulating key biological processes governed by the atomic details of functional protein interactions.

ACKNOWLEDGMENTS

We thank members of the Kortemme group and the Rosetta developers community for many stimulating discussions and important contributions. This work was supported by a CAREER award from the US National Science Foundation (T.K.) and a PhRMA Foundation Predoctoral Fellowship (D.J.M.).

Published online at http://www.nature.com/naturechemicalbiology/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Ponder, J.W. & Richards, F.M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. *J. Mol. Biol.* **193**, 775–791 (1987).
2. Hellinga, H.W. & Richards, F.M. Construction of new ligand binding sites in proteins of known structure. I. Computer-aided modeling of sites with pre-defined geometry. *J. Mol. Biol.* **222**, 763–785 (1991).
3. Richardson, J.S. & Richardson, D.C. The de novo design of protein structures. *Trends Biochem. Sci.* **14**, 304–309 (1989).
4. Bolon, D.N., Marcus, J.S., Ross, S.A. & Mayo, S.L. Prudent modeling of core polar residues in computational protein design. *J. Mol. Biol.* **329**, 611–622 (2003).
5. Dahiyat, B.I. & Mayo, S.L. Probing the role of packing specificity in protein design. *Proc. Natl. Acad. Sci. USA* **94**, 10172–10177 (1997).
6. Desjarlais, J.R. & Handel, T.M. De novo design of the hydrophobic cores of proteins. *Protein Sci.* **4**, 2006–2018 (1995).
7. Desjarlais, J.R. & Handel, T.M. Side-chain and backbone flexibility in protein core design. *J. Mol. Biol.* **290**, 305–318 (1999).
8. Kuhlman, B. & Baker, D. Exploring folding free energy landscapes using computational protein design. *Curr. Opin. Struct. Biol.* **14**, 89–95 (2004).
9. Dahiyat, B.I. & Mayo, S.L. De novo protein design: fully automated sequence selection. *Science* **278**, 82–87 (1997).
10. Harbury, P.B., Plecs, J.J., Tidor, B., Alber, T. & Kim, P.S. High-resolution protein design with backbone freedom. *Science* **282**, 1462–1467 (1998).
11. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302**, 1364–1368 (2003).
12. Hellinga, H.W. Metalloprotein design. *Curr. Opin. Biotechnol.* **7**, 437–441 (1996).
13. Hellinga, H.W., Caradonna, J.P. & Richards, F.M. Construction of new ligand binding sites in proteins of known structure. II. Grafting of a buried transition metal binding site into *Escherichia coli* thioredoxin. *J. Mol. Biol.* **222**, 787–803 (1991).
14. Chevalier, B.S. *et al.* Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell* **10**, 895–905 (2002).
15. Grigoryan, G., Reinke, A.W. & Keating, A.E. Design of protein-interaction specificity gives selective bZIP-binding peptides. *Nature* **458**, 859–864 (2009).
16. Kortemme, T. *et al.* Computational redesign of protein-protein interaction specificity. *Nat. Struct. Mol. Biol.* **11**, 371–379 (2004).
17. Reina, J. *et al.* Computer-aided design of a PDZ domain to recognize new target sequences. *Nat. Struct. Biol.* **9**, 621–627 (2002).
18. Shifman, J.M. & Mayo, S.L. Modulating calmodulin binding specificity through computational protein design. *J. Mol. Biol.* **323**, 417–423 (2002).
19. Shifman, J.M. & Mayo, S.L. Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc. Natl. Acad. Sci. USA* **100**, 13274–13279 (2003).
20. Palmer, A.E. *et al.* Ca2+ indicators based on computationally redesigned calmodulin-peptide pairs. *Chem. Biol.* **13**, 521–530 (2006).
21. Looger, L.L., Dwyer, M.A., Smith, J.J. & Hellinga, H.W. Computational design of receptor and sensor proteins with novel functions. *Nature* **423**, 185–190 (2003).
22. Ashworth, J. *et al.* Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature* **441**, 656–659 (2006).
23. Jiang, L. *et al.* De novo computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 (2008).
24. Röthlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
25. Bloom, J.D. & Arnold, F.H. In the light of directed evolution: pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci. USA* **106** (suppl. 1), 9995–10000 (2009).
26. Kortemme, T. & Baker, D. Computational design of protein-protein interactions. *Curr. Opin. Chem. Biol.* **8**, 91–97 (2004).
27. Lippow, S.M. & Tidor, B. Progress in computational protein design. *Curr. Opin. Biotechnol.* **18**, 305–311 (2007).
28. Karanicolas, J. & Kuhlman, B. Computational design of affinity and specificity at protein-protein interfaces. *Curr. Opin. Struct. Biol.* **19**, 458–463 (2009).
29. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K. & Baker, D. Progress in modeling of protein structures and interactions. *Science* **310**, 638–642 (2005).
30. Street, A.G. & Mayo, S.L. Pairwise calculation of protein solvent-accessible surface areas. *Fold. Des.* **3**, 253–258 (1998).
31. Marshall, S.A., Vizcarra, C.L. & Mayo, S.L. One- and two-body decomposable Poisson-Boltzmann methods for protein design calculations. *Protein Sci.* **14**, 1293–1304 (2005).
32. Pokala, N. & Handel, T.M. Energy functions for protein design I: efficient and accurate continuum electrostatics and solvation. *Protein Sci.* **13**, 925–936 (2004).
33. Wisz, M.S. & Hellinga, H.W. An empirical model for electrostatic interactions in proteins incorporating multiple geometry-dependent dielectric constants. *Proteins* **51**, 360–377 (2003).
34. Lazaridis, T. & Karplus, M. Effective energy function for proteins in solution. *Proteins* **35**, 133–152 (1999).
35. Butterfoss, G.L. & Kuhlman, B. Computer-based design of novel protein structures. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 49–65 (2006).
36. Boas, F.E. & Harbury, P.B. Potential energy functions for protein design. *Curr. Opin. Struct. Biol.* **17**, 199–204 (2007).
37. Vizcarra, C.L. & Mayo, S.L. Electrostatics in computational protein design. *Curr. Opin. Chem. Biol.* **9**, 622–626 (2005).
38. Haidar, J. *et al.* Structure-based design of a T-cell receptor leads to nearly 100-fold improvement in binding affinity for pepMHC. *Proteins* **74**, 948–960 (2009).
39. Lippow, S.M., Wittrup, K.D. & Tidor, B. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat. Biotechnol.* **25**, 1171–1176 (2007).
40. Sammond, D.W. *et al.* Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. *J. Mol. Biol.* **371**, 1392–1404 (2007).
41. Song, G. *et al.* Rational design of intercellular adhesion molecule-1 (ICAM-1) variants for antagonizing integrin lymphocyte function-associated antigen-1-dependent adhesion. *J. Biol. Chem.* **281**, 5042–5049 (2006).
42. Reynolds, K.A. *et al.* Computational redesign of the SHV-1 beta-lactamase/beta-lactamase inhibitor protein interface. *J. Mol. Biol.* **382**, 1265–1275 (2008).
43. Hao, J. *et al.* Identification and rational redesign of peptide ligands to CRIP1, a novel biomarker for cancers. *PLoS Comput. Biol.* **4**, e1000138 (2008).
44. Schreiber, G., Shaul, Y. & Gottschalk, K.E. Electrostatic design of protein-protein association rates. *Methods Mol. Biol.* **340**, 235–249 (2006).
45. Joachimiak, L.A., Kortemme, T., Stoddard, B.L. & Baker, D. Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J. Mol. Biol.* **361**, 195–208 (2006).
46. Havranek, J.J. & Harbury, P.B. Automated design of specificity in molecular recognition. *Nat. Struct. Biol.* **10**, 45–52 (2003).
47. Bolon, D.N., Grant, R.A., Baker, T.A. & Sauer, R.T. Specificity versus stability in computational protein design. *Proc. Natl. Acad. Sci. USA* **102**, 12724–12729 (2005).
48. Bishop, A.C. *et al.* A chemical switch for inhibitor-sensitive alleles of any protein kinase. *Nature* **407**, 395–401 (2000).
49. Han, J.D. *et al.* Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* **430**, 88–93 (2004).
50. Humphris, E.L. & Kortemme, T. Design of multi-specificity in protein interfaces. *PLoS Comput. Biol.* **3**, e164 (2007).
51. Huang, P.S., Love, J.J. & Mayo, S.L. A de novo designed protein protein interface. *Protein Sci.* **16**, 2770–2774 (2007).
52. Sidhu, S.S. & Koide, S. Phage display for engineering and analyzing protein interaction interfaces. *Curr. Opin. Struct. Biol.* **17**, 481–487 (2007).
53. Mandell, D.J. & Kortemme, T. Backbone flexibility in computational protein design. *Curr. Opin. Biotechnol.* published online, doi:10.1016/j.copbio.2009.07.006 (24 August 2009).
54. Bordner, A.J. & Abagyan, R.A. Large-scale prediction of protein geometry and stability changes for arbitrary single point mutations. *Proteins* **57**, 400–413 (2004).
55. Pal, G., Kouadio, J.L., Artis, D.R., Kossiakoff, A.A. & Sidhu, S.S. Comprehensive and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. *J. Biol. Chem.* **281**, 22378–22385 (2006).
56. Tonikian, R. *et al.* A specificity map for the PDZ domain family. *PLoS Biol.* **6**, e239 (2008).
57. Ambroggio, X.I. & Kuhlman, B. Design of protein conformational switches. *Curr. Opin. Struct. Biol.* **16**, 525–530 (2006).
58. Fu, X., Apgar, J.R. & Keating, A.E. Modeling backbone flexibility to achieve sequence diversity: the design of novel alpha-helical ligands for Bcl-xL. *J. Mol. Biol.* **371**, 1099–1117 (2007).

59. Larson, S.M., England, J.L., Desjarlais, J.R. & Pande, V.S. Thoroughly sampling sequence space: large-scale protein design of structural ensembles. *Protein Sci.* **11**, 2804–2813 (2002).
60. Ding, F. & Dokholyan, N.V. Emergence of protein fold families through rational design. *PLoS Comput. Biol.* **2**, e85 (2006).
61. Humphris, E.L. & Kortemme, T. Prediction of protein-protein interface sequence diversity using flexible backbone computational protein design. *Structure* **16**, 1777–1788 (2008).
62. Friedland, G.D., Lakomek, N.A., Griesinger, C., Meiler, J. & Kortemme, T. A correspondence between solution-state dynamics of an individual protein and the sequence and conformational diversity of its family. *PLoS Comput. Biol.* **5**, e1000393 (2009).
63. Ambroggio, X.I. & Kuhlman, B. Computational design of a single amino acid sequence that can switch between two distinct protein folds. *J. Am. Chem. Soc.* **128**, 1154–1161 (2006).
64. Hu, X., Wang, H., Ke, H. & Kuhlman, B. High-resolution design of a protein loop. *Proc. Natl. Acad. Sci. USA* **104**, 17668–17673 (2007).
65. Saunders, C.T. & Baker, D. Recapitulation of protein family divergence using flexible backbone protein design. *J. Mol. Biol.* **346**, 631–644 (2005).
66. Smith, C.A. & Kortemme, T. Backrub-like backbone simulation recapitulates natural protein conformational variability and improves mutant side-chain prediction. *J. Mol. Biol.* **380**, 742–756 (2008).
67. Yin, S., Ding, F. & Dokholyan, N.V. Eris: an automated estimator of protein stability. *Nat. Methods* **4**, 466–467 (2007).
68. Georgiev, I. & Donald, B.R. Dead-end elimination with backbone flexibility. *Bioinformatics* **23**, i185–i194 (2007).
69. Georgiev, I., Lilien, R.H. & Donald, B.R. The minimized dead-end elimination criterion and its application to protein redesign in a hybrid scoring and search algorithm for computing partition functions over molecular ensembles. *J. Comput. Chem.* **29**, 1527–1542 (2008).
70. Friedland, G.D., Linares, A.J., Smith, C.A. & Kortemme, T. A simple model of backbone flexibility improves modeling of side-chain conformational variability. *J. Mol. Biol.* **380**, 757–774 (2008).
71. Treynor, T.P., Vizcarra, C.L., Nedelcu, D. & Mayo, S.L. Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. *Proc. Natl. Acad. Sci. USA* **104**, 48–53 (2007).
72. Domingues, H., Cregut, D., Sebald, W., Oschkinat, H. & Serrano, L. Rational design of a GCN4-derived mimetic of interleukin-4. *Nat. Struct. Biol.* **6**, 652–656 (1999).
73. Liu, S. *et al.* Nonnatural protein-protein interaction-pair design by key residues grafting. *Proc. Natl. Acad. Sci. USA* **104**, 5330–5335 (2007).
74. Potapov, V. *et al.* Computational redesign of a protein-protein interface for high affinity and binding specificity using modular architecture and naturally occurring template fragments. *J. Mol. Biol.* **384**, 109–119 (2008).
75. Zanghellini, A. *et al.* New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **15**, 2785–2794 (2006).
76. Havranek, J.J. & Baker, D. Motif-directed flexible backbone design of functional interactions. *Protein Sci.* **18**, 1293–1305 (2009).
77. Murphy, P.M., Bolduc, J.M., Gallaher, J.L., Stoddard, B.L. & Baker, D. Alteration of enzyme specificity by computational loop remodeling and design. *Proc. Natl. Acad. Sci. USA* **106**, 9215–9220 (2009).
78. Mandell, D.J., Coutsias, E.A. & Kortemme, T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods* **6**, 551–552 (2009).
79. Allen, B.D. & Mayo, S.L. An efficient algorithm for multistate protein design based on FASTER. *J. Comput. Chem.* published online, doi:10.1002/jcc.21375 (27 July 2009).
80. Grigoryan, G. *et al.* Ultra-fast evaluation of protein energies directly from sequence. *PLoS Comput. Biol.* **2**, e63 (2006).
81. Gavin, A.C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
82. Krogan, N.J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
83. Schuldiner, M. *et al.* Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–519 (2005).
84. Keating, A.E., Malashkevich, V.N., Tidor, B. & Kim, P.S. Side-chain repacking calculations for predicting structures and stabilities of heterodimeric coiled coils. *Proc. Natl. Acad. Sci. USA* **98**, 14825–14830 (2001).
85. Altman, M.D., Nalivaika, E.A., Prabu-Jeyabalan, M., Schiffer, C.A. & Tidor, B. Computational design and experimental study of tighter binding peptides to an inactivated mutant of HIV-1 protease. *Proteins* **70**, 678–694 (2008).
86. Simons, K.T., Bonneau, R., Ruczinski, I. & Baker, D. Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins* **37** (suppl. 3), 171–176 (1999).
87. Simons, K.T., Kooperberg, C., Huang, E. & Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268**, 209–225 (1997).
88. Rohl, C.A., Strauss, C.E.M., Chivian, D. & Baker, D. Modeling structurally variable regions in homologous proteins with rosetta. *Proteins* **55**, 656–677 (2004).
89. Wang, C., Bradley, P. & Baker, D. Protein-protein docking with backbone flexibility. *J. Mol. Biol.* **373**, 503–519 (2007).
90. Fung, H.K., Floudas, C.A., Taylor, M.S., Zhang, L. & Morikis, D. Toward full-sequence de novo protein design with flexible templates for human beta-defensin-2. *Biophys. J.* **94**, 584–599 (2008).
91. Davis, I.W., Arendall, W.B., Richardson, D.C. & Richardson, J.S. The backrub motion: how protein backbone shrugs when a sidechain dances. *Structure* **14**, 265–274 (2006).
92. Georgiev, I., Keedy, D., Richardson, J.S., Richardson, D.C. & Donald, B.R. Algorithm for backrub motions in protein design. *Bioinformatics* **24**, i196–i204 (2008).
93. Dunbrack, R.L. Jr. Rotamer libraries in the 21st century. *Curr. Opin. Struct. Biol.* **12**, 431–440 (2002).
94. Desjarlais, J.R. & Clarke, N.D. Computer search algorithms in protein modification and design. *Curr. Opin. Struct. Biol.* **8**, 471–475 (1998).
95. Desmet, J., De Maeyer, M., Hazes, B. & Lasters, I. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature* **356**, 539–542 (1992).
96. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proc. Natl. Acad. Sci. USA* **97**, 10383–10388 (2000).
97. Allen, B.D. & Mayo, S.L. Dramatic performance enhancements for the FASTER optimization algorithm. *J. Comput. Chem.* **27**, 1071–1075 (2006).
98. Desmet, J., Spriet, J. & Lasters, I. Fast and accurate side-chain topology and energy refinement (FASTER) as a new method for protein structure optimization. *Proteins* **48**, 31–43 (2002).
99. Voigt, C.A., Gordon, D.B. & Mayo, S.L. Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. *J. Mol. Biol.* **299**, 789–803 (2000).
100. Gordon, D.B., Marshall, S.A. & Mayo, S.L. Energy functions for protein design. *Curr. Opin. Struct. Biol.* **9**, 509–513 (1999).
101. Kortemme, T., Morozov, A.V. & Baker, D. An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes. *J. Mol. Biol.* **326**, 1239–1259 (2003).
102. Pokala, N. & Handel, T.M. Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. *J. Mol. Biol.* **347**, 203–227 (2005).