ELSEVIER

Computational design of protein–protein interactions

Tanja Kortemme[1,2] and David Baker[1,*]

Computational protein design strategies have been developed to reengineer protein–protein interfaces in an automated, generalizable fashion. In the past two years, these methods have been successfully applied to generate chimeric proteins and protein pairs with specificities different from naturally occurring protein–protein interactions. Although there are shortcomings in current approaches, both in the way conformational space is sampled and in the energy functions used to evaluate designed conformations, the successes suggest we are now entering an era in which computational methods can be used to modulate, reengineer and design protein–protein interaction networks in living cells.

Addresses
[1]Howard Hughes Medical Institute and Department of Biochemistry, University of Washington, Box 357350, Seattle, WA 98195, USA
[2]Department of Biopharmaceutical Sciences and California Institute for Quantitative Biomedical Research, University of California San Francisco, CA 94107, USA
*e-mail: dabaker@u.washington.edu

Current Opinion in Chemical Biology 2004, **8**:91–97

This review comes from a themed issue on
Proteomics and genomics
Edited by Michael Snyder and John Yates III

1367-5931/$ – see front matter

DOI 10.1016/j.cbpa.2003.12.008

Introduction

Protein–protein interactions are central to many processes within cells and organisms, ranging from the assembly of the structural scaffold of cells to immune defense and cellular communication. How do biological molecules recognize their targets, and how do these interactions build up macromolecular complexes and networks responsible for biological regulation and complexity? Tools to alter and interfere with protein interactions offer great promise to help understand and delineate these networks; for example, small-molecule inhibitors of cellular processes have yielded exciting insights into biological regulatory mechanisms [1].

A complementary approach to 'chemical genetics' techniques is to use computational design methods to modulate protein–protein interactions. To develop the required understanding of the physical basis of affinity and specificity in protein interfaces is a major challenge, but such a description would allow the design of new protein–protein interactions and would ultimately open the way to engineer new functions and modulate cellular behavior in a predictive manner. The striking successes with monomeric proteins such as the computational design of protein cores [2–4] (reviewed in [5]), metal binding sites [6], enzyme-like biocatalysts [7], complete proteins [8,9], folding mechanisms [10], and new topologies [11,12] as well as recent impressive work on the design of small-molecule protein receptors [13••] suggests that these techniques have now reached the point that they can be applied and extended to modulate and engineer function in a biological context by altering molecular recognition processes.

We first briefly outline general principles of computational design, with an emphasis on challenges encountered particularly in protein interfaces. We then describe recent successes in the generation of new protein–protein interactions. These results highlight the features of molecular interactions that can and cannot be modeled using current computational approaches and illustrate the potential of the methodology for the redesign of protein interactions in the context of living cells.

Computational protein design

Computational protein design methods seek to identify low-energy amino acid sequences for a specified target protein structure. Two general problems are encountered: first, conformational and sequence space have to be sampled adequately; and secondly, the energy function has to be accurate enough to identify protein sequences for which the desired three-dimensional conformation is at the global free energy minimum. For the design of monomeric proteins, the first problem has been simplified by the use of a fixed polypeptide backbone that was initially taken from an experimentally determined high-resolution structure. More recent variations on this paradigm range from the inclusion of alternative backbones obtained by small perturbations of the backbone torsion angles [14] and the alteration of loop conformations [15] to the design of completely new folds, either by parameterization of the coiled-coil topology [11] or by generating a novel backbone applying techniques borrowed from *ab initio* structure prediction methods [12]. Sequence space is sampled for a given fixed backbone and a library of different rotamers (conformations generated by discrete rotations around the side chain torsion angles) for each amino acid, employing a variety of optimization techniques ranging from deterministic procedures such as dead-end elimination to stochastic methods like Monte-Carlo simulated annealing [16,17].

The need for computationally efficient algorithms to adequately sample sequence space (a small 50 residue

protein comprising the 20 naturally occurring amino acids has $20^{50} = 1.1 \times 10^{65}$ theoretically possible sequences) has consequences for the energy functions applied to protein design. For reasons of speed, the vast majority of protein design energy functions are pairwise-additive. Standard approaches include a combination of van der Waals potentials to describe atomic packing, implicit solvation models, a Coulomb model with dampening of long-range electrostatic effects, explicit hydrogen-bonding terms, statistical terms to describe propensities of amino acids for backbone and side-chain torsion angles and approximations to the conformational entropy of protein side chains and the energy of the unfolded state ensemble [18,19,20•,21].

Energetics at protein interfaces

There are several challenges for the correct description of the energetics at protein interfaces. Simple general rules to identify protein recognition sites and predict energetic hot spots in protein complexes often fail [22], largely because of the extreme diversity in shape and chemical character of protein–protein interfaces [23]. However, physical models have recently had some success rationalizing energetically important interactions in protein–protein interfaces [20•,24–26,27•] and in identifying correctly docked structures of protein–protein complexes from large sets of incorrect arrangements [28]. Common to most of these approaches is the realization that readily computable descriptions of electrostatic and hydrogen-bonding interactions are important for an adequate modeling of energetics and specificity at protein–protein interfaces. Sophisticated approaches such as solving the Poisson-Boltzmann equation have yielded significant insights [29] as well as suggested that protein interfaces are highly optimized electrostatically [30], but are generally unsuitable for protein design algorithms because of the requirement for pairwise additivity. Instead, explicit geometry-dependent hydrogen bonding potentials have been used successfully for protein–protein docking, prediction of energetic effects at protein–protein interfaces and protein design [20•,28,31••]. Interestingly, a comparison of the orientation dependence of hydrogen bonds observed in *ab initio* electronic structure calculations with an orientation dependent hydrogen potential derived from experimental protein structures [32] shows a remarkably close agreement (AV Morozov *et al.*, unpublished data), providing a physical understanding of the empirical hydrogen-bonding potential and suggesting a route to more accurate and computable energy functions by combining the two approaches.

Despite much recent progress, clearly more-accurate models of electrostatics effects in proteins are required to correctly capture the balance between energetically favorable electrostatic and hydrogen-bonding interactions and the unfavorable desolvation of polar groups at protein–protein interfaces [33]. A related challenge is modeling defined water molecules that often form extensive networks of water-mediated hydrogen bonds in protein interfaces which cannot be captured by standard implicit solvation models.

Conformational variability at protein–protein interfaces

In addition to solvation effects, alterations in the conformation and dynamical behavior of a protein upon binding can complicate modeling of molecular recognition processes and lead to considerable entropy–enthalpy compensation. Most proteins exhibit significant conformational changes on binding only at the level of side-chains [34]. However, proteins functioning as sensors or switches in signaling networks can show dramatic rearrangements, such as large hinge motions upon ligand binding (sugar binding sensors) or the reorganization of regions in small GTPases in response to an arriving signal. Even in the presence of multiple crystal structures, the causes and effects of conformational changes are extremely challenging to model. This is even more true in the case of dynamical alterations that are not obvious from static structures; not many high-resolution experimental studies of the dynamical behavior of proteins have been available to assess the accuracy of computational models of these processes. However, with significant recent technical advances in particular in the area of measuring dynamical changes in proteins by NMR spectroscopy [35], this situation is changing. NMR relaxation studies of backbone and side-chain dynamics have reported both decreases, as intuitively expected, as well as increases in protein mobility upon binding that might contribute significantly to the free energy of binding [36].

Despite the many challenges in the accurate modeling of energetic and conformational subtleties (and their interplay) in molecular recognition, there are a growing number of successful computational designs of protein–protein interfaces as described in the next sections.

Design of interfaces in coiled-coil and helical bundle systems

Studies of coiled-coils and helical bundle proteins have contributed significantly to our understanding of conformational specificity [37]. These insights have led to the design of coiled-coil systems as receptors for a peptide helix derived from calcineurin [38], and as an interleukin-4 antagonist based on the GCN4 structure [39]. A dramatic example of successful computational design by Harbury *et al.* was the engineering of new helical bundle topologies [11]. The novelty of this study lies in the use of a family of different alternative structures generated by an algebraic parameterization of the helical bundle structure, which allowed the exploration of structures not seen in naturally occurring proteins. A crystal structure of a designed helical tetramer and biophysical data on different designs assembling into dimeric and trimeric

conformations give insights into how relatively simple hydrophobic/polar amino acid side chain patterns give rise to formation of specific interfaces between the helical peptides.

Chimeric proteins by interface redesign

An interesting application for computational protein design is the creation of complex chimeric proteins that encompass the functions of several protein modules in a single protein assembly. Evolution has used this principle of 'mixing and matching' functional protein entities (domains) many times to generate complexity in multifunctional macromolecular assemblies and machines. This has been accomplished both by combining isolated proteins into larger polypeptide chains that link the different functional entities with intervening, often flexible sequences, or through the intimate fusion of protein domains via highly specific protein interfaces.

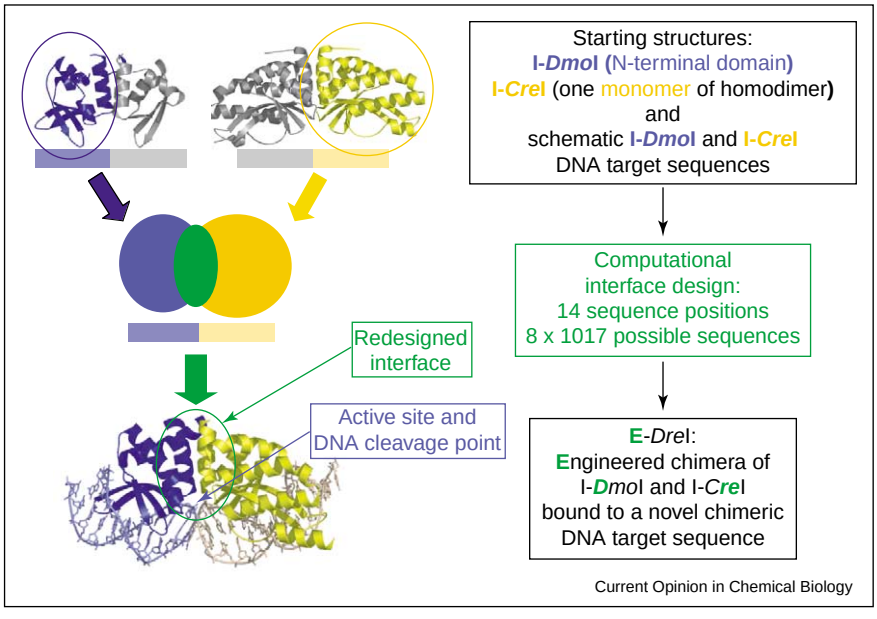
These ideas have been explored recently in two fascinating studies rewiring cellular interaction networks by recombining modules from different pathways [40,41]. Both studies borrowed naturally occurring protein domains and interfaces but linked them in novel combinations to achieve an altered cellular input/output connectivity. Extending these approaches, computational design offers the possibility to create novel interfaces that would go beyond the interaction capabilities of independent modules.

Linking protein functionalities by a well-defined interface might be particularly valuable in a case where two different binding capabilities are to be arranged in a spatially defined manner. In this case, a flexible linker would not be sufficient. A recent study explores this approach for the design of a new chimeric DNA-binding protein [31••]. Computational interface design was used to intimately fuse two domains of distantly related homing endonucleases that each carry a recognition site for a specific DNA target half-site (Figure 1). 14 positions in the protein–protein interface between the two domains were computationally allowed to vary to all naturally occurring 20 amino acids except for cysteine. The resulting eightfold variant specifically recognizes a chimeric DNA target sequence consistent of the two half-site recognition sequences contributed from each of the protein domains. The design procedure not only succeeded in the engineering of a stable 1400 $Å^2$ interface between the domains, but also retained catalytic activity in the chimeric protein that precisely cleaves its new DNA target sequence. These results indicate that computational protein design may be applied to recombine protein domains by reengineering the interfaces between them, resulting in assemblies with new complex functionalities.

Redesign of protein–protein interaction specificity

Protein–protein interactions are often part of intricate molecular recognition pathways and circuits. To make

Figure 1



Engineering of a new chimeric protein by computational interface design. Two domains of the homing endonucleases I-*Dmo*I and I-*Cre*I, proteins consisting of two individual domains or subunits that bind to long (7–20 base pairs) individual DNA target half-sites, were fused, and the interface between them was computationally redesigned. The designed chimeric protein E-*Dre*I contains eight mutations in the interface, and is an active enzyme cleaving a new chimeric DNA target sequence at specific position in the center of the DNA target, but is inactive towards the original I-*Dmo*I and I-*Cre*I target sites. (All structure figures were created using PYMOL: www.pymol.org.)

use of computationally designed protein–protein interfaces delineating and modulating these complex networks in living cells, computational methods have to be able to capture the molecular bases of specificity. How do proteins discriminate their correct binding partners form many other possible ligands with similar sequence and structure? Computational methods for the prediction of interaction specificity have recently reported significant advances [42–46]. Complementing and extending these approaches, a stringent test of current understanding of molecular recognition is to apply the theory to the actual engineering of systems with designed specificities. Four recent studies have addressed the computational design of protein–protein interaction specificities in different ways: by making a fairly promiscuous interface specific for one of its ligands [47••], by designing novel protein–peptide pairs that can be useful in analytical and cell biological applications [48••], by developing a new computational approach for the automated design of protein–protein interaction specificity [49••], and by applying a computational 'second-site suppressor' approach to the design of new protein pairs that are functional and specific *in vivo* (T Kortemme *et al.*, unpublished data). We describe these approaches in the following sections, and highlight what can be learned from each study towards achieving the goal of engineering functional protein–protein networks in living cells.

Shifman and Mayo [47••] applied an automated methodology that had been successfully used for the first completely computational sequence redesign of an entire small protein [8] to the design of specificity in a calmodulin–ligand peptide complex. Whereas wild-type calmodulin evolved to recognize a variety of different ligand sequences with reasonable affinity, a designed eightfold calmodulin variant now prefers one cognate ligand with up to 86-fold higher affinity over its other binding partners. The authors demonstrated that computational protein interface design can be used to reach affinities similar to the wild-type interaction in the nanomolar range, and that binding specificity can be achieved by explicitly optimizing the binding interface for one ligand.

Reina *et al.* computationally reengineered a small protein–protein interaction motif, the PDZ domain, to bind novel peptide target sequences [48••]. The first target was a naturally occurring peptide sequence at the C-terminus of a kinesin-like molecule; the design goal was to create a 'sensor' PDZ domain specific to this particular cellular protein. The second set of targets consisted of two sequences with different chemical character (charged or hydrophobic); the aim was to engineer new orthogonal PDZ-peptide pairs. A computational design procedure with significant manual intervention was used to create the three different PDZ constructs containing six or seven mutations that bound the three target peptides in *in vitro* fluorescence polarization assays with affinities up to two orders of magnitude higher than the affinity of the wild-type complex, and recognized their partners specifically in a yeast two-hybrid assay. This study illustrates that computational methods are applicable to reengineer existing interfaces to exhibit altered specificity, and that the designed molecules are specific enough in the cellular context that they can be useful for biotechnological applications such as western blotting, affinity chromatography and pull-down experiments.

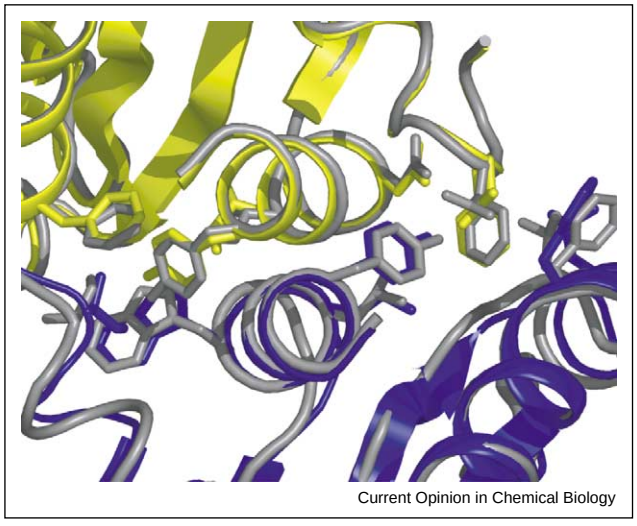
Havranek and Harbury [49••] concentrated on the development of an automated computational design strategy that explicitly takes negative design into account. Using an ensemble of alternative conformations in addition to the target structure, this algorithm maximizes the free energy difference between the desired and other competing conformations. Experimental tests using 13 different coiled-coil interfaces as model systems confirmed the computational results. Comparison between predicted and calculated binding free energies using different energy functions indicated the importance of fine sampling of conformational space at the side-chain level, using a large number of rotamers that were additionally optimized for the target backbone conformation. The computational framework demonstrates the feasibility of an explicit selection of cognate sequence pairs against cross-reactivity with other interfaces a protein might encounter in the biological context.

We developed a general strategy for the redesign of existing protein–protein complexes to generate new and orthogonal pairs of interacting proteins (T Kortemme *et al.*, unpublished data). In this scheme, both partners in a complex are computationally redesigned in two coupled steps: first, a computational screen identifies mutations in the first partner that significantly destabilize the interface; then, for each mutation in the first partner, compensating sequence changes in the second partner are found by screening for low energy sequences by computational design. Redesigned pairs are then selected that maximize the difference in binding free energy between the designed complex on the one hand and the combinations of one designed protein with the original wild-type partner on the other. Thus the procedure specifically selects for a new interacting pair that is orthogonal to the existing template complex. A test of this procedure on the redesign of DNase-inhibitor protein pairs confirmed the designed specificity switch by *in vitro* assays and crystallographic analysis of a designed interface, and further showed that the designed proteins are functional and specific *in vivo*. This study thus illustrates the potential of computational interface design to create new protein pairs that are both specific and functional in their biological context in living cells.

Structural characterization of designed protein–protein interfaces

Despite these successes, it is important to assess the accuracy of the computationally generated protein–protein

Figure 2



Comparison of the computational model and experimentally determined structure of a computationally designed protein interface. Shown is the 2.4 Å crystal structure of the redesigned chimeric E-*Dre*I homing endonuclease interface (see Figure 1; blue: domain from I-*Dmo*I, yellow: domain from I-*Cre*I) superimposed with the initial computational model generated by the design program (grey). Amino acid side chains forming the interface are shown in stick representation.

interface models. Quite a few high-resolution structures are available for designed monomeric proteins, yielding invaluable feed-back to iteratively improve computational methods in a 'design cycle' [50]. Structural information on redesigned protein–protein interfaces is scarce, but is available for several of the coiled-coil and helical-bundle based designs, and for the redesigned chimeric DNA-binding protein ([31••], Figure 2). In all these cases, there is significant agreement between the computationally modeled and experimentally determined structures, underlining the accuracy of current computational protein design methods.

Several factors can influence the design accuracy, reflecting the approximations made in the design procedure. First, as computational protein design in most cases requires the use of a fixed backbone (or a family of backbones), unanticipated backbone conformation changes can lead to errors. Second, errors in modeling interactions involving polar and charged residues in interfaces will expose inadequacies in current approximate models of electrostatic interactions. Both of these factors are expected to be magnified in protein–protein interfaces relative to monomeric proteins. Conformational and dynamical changes often observed upon binding events can significantly complicate the modeling of the interacting patch. Moreover, protein interfaces can have extensive water-mediated interaction networks that play a significant role in interface energetics [51]. The development of more accurate computational approaches will require the testing of energy functions and sampling methods based on their ability to reproduce the energetic and structural effects of mutations and conformational changes upon binding, and the determination of high resolution structures of designed interfaces.

Preliminary data on two crystal structures of redesigned colicin E7 DNase - Im7 immunity protein interfaces from our laboratory highlight the influence of both factors mentioned above on structural and energetic characteristics of interfaces, and at the same time suggest ways for improvement of energy functions and sampling methodology. Crystallographic analysis of one designed interface revealed slight changes in backbone conformation relative to the template conformation derived from the wild-type complex structure. While the backbone rearrangements were localized around the sites of mutations, they resulted in incorrect prediction of the side-chain rotamers. However, when the correct backbone was used, all rotamers were correctly predicted, and the predicted conformation was much lower in free energy than the incorrect conformation based on the wild-type template. The incorporation of backbone motions localized around the mutated sites in a protein–protein interface, coupled with side-chain repacking techniques and design, might be able to model altered conformations correctly.

The crystal structure of a second redesigned interface revealed the presence of several new water molecules that mediated interactions across the interface. Moreover, both the wild-type interface and the designed complex contain several conserved interfacial water molecules. Water molecules were not taken into account explicitly in the specificity redesign procedure, but might be modeled as extensions of side-chain rotamers. Such a model might explain the extremely tight binding affinities seen in some particularly polar protein–protein interfaces such as the barnase–barstar complex [30] and our colicin E7 DNases–Im7 inhibitor protein system, and the weaker affinities of engineered complexes based on the E7/Im7 complex that were designed without accounting for interface water networks.

Conclusions and future challenges

There are clear shortcomings in computational design procedures for protein–protein interactions, both in the free energy function, in particular the treatment of electrostatic interactions and solvation, as well as in the modelling of backbone flexibility. The comparison of computational models with experimentally determined structures of designed interfaces highlights the importance of both of these factors. While current methods are remarkably successful for protein interface *design* despite being approximate, addressing these problems will be critical for a detailed structure-based understanding and *prediction* of protein–protein interaction specificity. This

is particularly challenging in cases where the predictions have to be based on structural models inferred from a homologous protein complex structure. Interestingly, comparatively 'low resolution' sequence- and structure-based methods for the computational prediction of specificity have demonstrated considerable success [43]. However, to apply the power of the methods described in this review to these interactions, detailed and precise structural models generated by more accurate modelling techniques are needed. The ultimate goal of such studies would be to develop an understanding of biological processes ranging from the level of atomic interactions all the way to cellular behaviour.

Another major step for the understanding of molecular interactions will be the development of approaches to model the recognition of nucleic acids by proteins. The use of rotamer-based computational design approaches and simple energy functions similar to those applicable to the prediction and design of protein–protein interfaces [20•,31••,32,44,47••–49••] can highlight the features of protein–nucleic-acid interactions that can and cannot be predicted using current methods. Available structures of protein–nucleic-acid complexes constitute a large database for development and testing of the methods [52], with the aim to extend the ability of computational methods to model and engineer biological regulation both at the genetic (protein–nucleic-acid) and the biochemical (protein–protein) level.

Acknowledgements

This work was supported by the Howard Hughes Medical Institute and NIH. TK would like to thank the Human Frontier Science Program for support.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Specht KM, Shokat KM: **The emerging power of chemical genetics**. *Curr Opin Cell Biol* 2002, **14**:155-159.

2. Desjarlais JR, Handel TM: ***De novo* design of the hydrophobic cores of proteins**. *Protein Sci* 1995, **4**:2006-2018.

3. Ventura S, Vega MC, Lacroix E, Angrand I, Spagnolo L, Serrano L: **Conformational strain in the hydrophobic core and its implications for protein folding and design**. *Nat Struct Biol* 2002, **9**:485-493.

4. Bolon DN, Marcus JS, Ross SA, Mayo SL: **Prudent modeling of core polar residues in computational protein design**. *J Mol Biol* 2003, **329**:611-622.

5. Pokala N, Handel TM: **Protein design – where we were, where we are, were we're going**. *J Struct Biol* 2001, **134**:269-281.

6. Hellinga HW, Caradonna JP, Richards FM: **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* 1991, **222**:787-803.

7. Bolon DN, Mayo SL: **Enzyme-like proteins by computational design**. *Proc Natl Acad Sci USA* 2001, **98**:14274-14279.

8. Dahiyat BI, Mayo SL: ***De novo* protein design: fully automated sequence selection**. *Science* 1997, **278**:82-87.

9. Dantas G, Kuhlman B, Callender D, Wong M, Baker D: **A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins**. *J Mol Biol* 2003, **332**:449-460.

10. Nauli S, Kuhlman B, Baker D: **Computer-based redesign of a protein folding pathway**. *Nat Struct Biol* 2001, **8**:602-605.

11. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS: **High-resolution protein design with backbone freedom**. *Science* 1998, **282**:1462-1467.

12. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D: **Design of a novel globular protein fold with atomic level accuracy**. *Science* 2003, **302**:1364-1368.

13. •• Looger LL, Dwyer MA, Smith JJ, Hellinga HW: **Computational design of receptor and sensor proteins with novel functions**. *Nature* 2003, **423**:185-190.
A landmark study demonstrating the successful computational design of several potent and selective protein receptors for an impressive array of small-molecule ligands.

14. Desjarlais JR, Handel TM: **Side-chain and backbone flexibility in protein core design**. *J Mol Biol* 1999, **290**:305-318.

15. Kuhlman B, O'Neill JW, Kim DE, Zhang KY, Baker D: **Accurate computer-based design of a new backbone conformation in the second turn of protein L**. *J Mol Biol* 2002, **315**:471-477.

16. Desjarlais JR, Clarke ND: **Computer search algorithms in protein modification and design**. *Curr Opin Struct Biol* 1998, **8**:471-475.

17. Voigt CA, Gordon DB, Mayo SL: **Trading accuracy for speed: a quantitative comparison of search algorithms in protein sequence design**. *J Mol Biol* 2000, **299**:789-803.

18. Gordon DB, Marshall SA, Mayo SL: **Energy functions for protein design**. *Curr Opin Struct Biol* 1999, **9**:509-513.

19. Kuhlman B, Baker D: **Native protein sequences are close to optimal for their structures**. *Proc Natl Acad Sci USA* 2000, **97**:10383-10388.

20. • Kortemme T, Baker D: **A simple physical model for binding energy hot spots in protein-protein complexes**. *Proc Natl Acad Sci USA* 2002, **99**:14116-14121.
'Computational alanine scanning' employing a fast computational protocol correctly identifies 79% of the energetically important interactions (binding energy hot spots) in protein–protein interfaces. The results highlight successes and failures of simple protein design energy functions for protein interface energetics.

21. Mendes J, Guerois R, Serrano L: **Energy estimation in protein design**. *Curr Opin Struct Biol* 2002, **12**:441-446.

22. Bogan AA, Thorn KS: **Anatomy of hot spots in protein interfaces**. *J Mol Biol* 1998, **280**:1-9.

23. Conte LL, Chothia C, Janin J: **The atomic structure of protein-protein recognition sites**. *J Mol Biol* 1999, **285**:2177-2198.

24. Sharp KA: **Calculation of HyHel10-lysozyme binding free energy changes: effect of ten point mutations**. *Proteins* 1998, **33**:39-48.

25. Massova I, Kollman PA: **Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies**. *J Am Chem Soc* 1999, **121**:8133-8143.

26. Huo S, Massova I, Kollman PA: **Computational alanine scanning of the 1:1 human growth hormone-receptor complex**. *J Comput Chem* 2002, **23**:15-27.

27. • Guerois R, Nielsen JE, Serrano L: **Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations**. *J Mol Biol* 2002, **320**:369-387.
A fast computational method to predict the effect of mutations on the stability of proteins and protein complexes with good accuracy.

28. Gray JJ, Moughon SE, Kortemme T, Schueler-Furman O, Misura KM, Morozov AV, Baker D: **Protein-protein docking predictions for the CAPRI experiment**. *Proteins* 2003, **52**:118-122.

29. Honig B, Nicholls A: **Classical electrostatics in biology and chemistry**. *Science* 1995, **268**:1144-1149.

30. Lee LP, Tidor B: **Barstar is electrostatically optimized for tight binding to barnase**. *Nat Struct Biol* 2001, **8**:73-76.

31. •• Chevalier BS, Kortemme T, Chadsey MS, Baker D, Monnat RJ, Stoddard BL: **Design, activity, and structure of a highly specific artificial endonuclease**. *Mol Cell* 2002, **10**:895-905.
The application of computational interface redesign to the engineering of a functional chimeric protein that combines the two different binding capabilities of the parent proteins. The crystal structure of the designed interface confirms the accuracy of the design algorithm.

32. Kortemme T, Morozov AV, Baker D: **An orientation-dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein-protein complexes**. *J Mol Biol* 2003, **326**:1239-1259.

33. Morozov AV, Kortemme T, Baker D: **Evaluation of models of electrostatic interactions in proteins**. *J Phys Chem B* 2003, **107**:2075-2090.

34. Betts MJ, Sternberg MJ: **An analysis of conformational changes on protein-protein association: implications for predictive docking**. *Protein Eng* 1999, **12**:271-283.

35. Akke M: **NMR methods for characterizing microsecond to millisecond dynamics in recognition and catalysis**. *Curr Opin Struct Biol* 2002, **12**:642-647.

36. Forman-Kay JD: **The 'dynamics' in the thermodynamics of binding**. *Nat Struct Biol* 1999, **6**:1086-1087.

37. Hill RB, Raleigh DP, Lombardi A, DeGrado WF: ***De novo* design of helical bundles as models for understanding protein folding and function**. *Acc Chem Res* 2000, **33**:745-754.

38. Ghirlanda G, Lear JD, Lombardi A, DeGrado WF: **From synthetic coiled coils to functional proteins: automated design of a receptor for the calmodulin-binding domain of calcineurin**. *J Mol Biol* 1998, **281**:379-391.

39. Domingues H, Cregut D, Sebald W, Oschkinat H, Serrano L: **Rational design of a GCN4-derived mimetic of interleukin-4**. *Nat Struct Biol* 1999, **6**:652-656.

40. Park SH, Zarrinpar A, Lim WA: **Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms**. *Science* 2003, **299**:1061-1064.

41. Howard PL, Chia MC, Del Rizzo S, Liu FF, Pawson T: **Redirecting tyrosine kinase signaling to an apoptotic caspase pathway through chimeric adaptor proteins**. *Proc Natl Acad Sci USA* 2003, **100**:11267-11272.

42. Brannetti B, Via A, Cestra G, Cesareni G, Helmer-Citterich M: **SH3-SPOT: an algorithm to predict preferred ligands to different members of the SH3 gene family**. *J Mol Biol* 2000, **298**:313-328.

43. Aloy P, Russell RB: **Interrogating protein interaction networks through structural biology**. *Proc Natl Acad Sci USA* 2002, **99**:5896-5901.

44. Wollacott AM, Desjarlais JR: **Virtual interaction profiles of proteins**. *J Mol Biol* 2001, **313**:317-342.

45. Li L, Shakhnovich EI, Mirny LA: **Amino acids determining enzyme-substrate specificity in prokaryotic and eukaryotic protein kinases**. *Proc Natl Acad Sci USA* 2003, **100**:4463-4468.

46. Brinkworth RI, Breinl RA, Kobe B: **Structural basis and prediction of substrate specificity in protein serine/threonine kinases**. *Proc Natl Acad Sci USA* 2003, **100**:74-79.

47. •• Shifman JM, Mayo SL: **Modulating calmodulin binding specificity through computational protein design**. *J Mol Biol* 2002, **323**:417-423.
Optimizing the fairly promiscuous calmodulin interface for one of its ligands using a successful computational protein design method [7] resulted in a stable interaction in the nanomolar range that is now more specific for the selected ligand. This is the first study showing that computational interface redesign is capable of enhancing the specificity of an interaction.

48. •• Reina J, Lacroix E, Hobson SD, Fernandez-Ballester G, Rybin V, Schwab MS, Serrano L, Gonzalez C: **Computer-aided design of a PDZ domain to recognize new target sequences**. *Nat Struct Biol* 2002, **9**:621-627.
Three PDZ domain constructs were engineered to specifically recognize their new targets peptides in western blotting and pull-down experiments. This important study demonstrates that protein interaction specificity can be altered rationally by computational design.

49. •• Havranek JJ, Harbury PB: **Automated design of specificity in molecular recognition**. *Nat Struct Biol* 2003, **10**:45-52.
This impressive study describes the development and experimental verification of a novel computational protocol that automatically selects for sequences that prefer the desired cognate interaction over alternative partners and conformations. The experimental results on the formation of homodimeric or heterodimeric coiled-coil interfaces confirm the predicted specificities in all cases.

50. Street AG, Mayo SL: **Computational protein design**. *Structure Fold Des* 1999, **7**:R105-R109.

51. Janin J: **Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition**. *Structure Fold Des* 1999, **7**:R277-R279.

52. Nadassy K, Wodak SJ, Janin J: **Structural features of protein-nucleic acid recognition sites**. *Biochemistry* 1999, **38**:1999-2017.