ELSEVIER

Full text provided by www.sciencedirect.com
ScienceDirect

Progress in computational protein design

Shaun M Lippow[1,4] and Bruce Tidor[2,3]

Current progress in computational structure-based protein design is reviewed in the areas of methodology and applications. Foundational advances include new potential functions, more efficient ways of computing energetics, flexible treatments of solvent, and useful energy function approximations, as well as ensemble-based approaches to scoring designs for inclusion of entropic effects, improvements to guaranteed and to stochastic search techniques, and methods to design combinatorial libraries for screening and selection. Applications include new approaches and successes in the design of specificity for protein folding, binding, and catalysis, in the redesign of proteins for enhanced binding affinity, and in the application of design technology to study and alter enzyme catalysis. Computational protein design continues to mature and advance.

Addresses
[1] Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[2] Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[3] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[4] Present address: Codon Devices, Inc., One Kendall Square, Building 300, Cambridge, MA 02139, USA

Corresponding authors: Lippow, Shaun M. (lippow@alum.mit.edu) and Tidor, Bruce (tidor@mit.edu)

Current Opinion in Biotechnology 2007, **18**:1–7

This review comes from a themed issue on Protein technologies
Edited by K. Dane Wittrup and Colin Campbell

0958-1669/$ – see front matter
© 2007 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.copbio.2007.04.009

Introduction

Computational protein design is continually developing as a practical option for solving problems in protein engineering. Much progress has been made from early proof-of-concept redesigns of protein cores to full proteins, with current research addressing a wide diversification of problems. Investigations pursue both scientific and engineering goals in tandem, using design to test and advance our understanding of underlying biophysical interactions.

Developments have been made both toward grand challenges and toward more immediate practical applications. Grand challenges tend to be *de novo* design problems, such as the creation of novel protein folds, binding interfaces, or enzymatic activities. More immediate practical applications involve the redesign of existing proteins, for increased thermostability, altered binding specificity, improved binding affinity, enhanced enzymatic activity, or altered substrate specificity.

An increasingly common limitation in design is the choice of objective function. Few problems can be adequately addressed by the straightforward energy minimization of a single protein state. Instead, multi-objective searches are ideal for designing specificity (to stabilize one or more states relative to others), improving binding affinity (to increase interaction while maintaining folding stability), and designing *de novo* proteins (to avoid alternate structures and aggregation). Furthermore, enzyme design may benefit from more detailed objectives than simply binding the transition state and coordination of key active-site functional groups.

In this review, we address progress in structure-based computational protein design in the past 2 years (since 2005). Other recent reviews provide additional background and viewpoints [1–8]. Here we highlight progress in design methodology and in applications, and discuss some emerging themes.

Progress in methodology

Energy functions

Protein design technology relies on pairing energy functions to evaluate candidates with search algorithms to examine large combinatorial collections of candidates. These interdependent foundational methodologies continue to be improved in ways that promise to increase the accuracy, efficiency, and scope of computational protein design applications. Work on energy functions includes understanding and validating their applicability to design studies, developing new potentials and target functions where appropriate, and improving efficiency through both better algorithms and approximations.

An energy function for nucleic acids and their interactions with proteins has been developed by Siggia and colleagues and validated for protein–DNA binding specificity [9]. Effects are attributed to direct readout through intermolecular electrostatics and packing terms and to indirect readout through intramolecular terms describing the DNA conformation. In another work, a classical mechanical description of interactions for dimanganese centers has been developed by Spiegel *et al.* (including implicit treatment of charge transfer and polarization effects) that accounts for the reduced and oxidized form

Please cite this article in press as: Lippow SM, Tidor B, Progress in computational protein design, Curr Opin Biotechnol (2007), doi:10.1016/j.copbio.2007.04.009

[10]; successful enzyme design requires the surrounding protein to stably bind multiple states of the metal center, as a prerequisite for carrying out catalysis. This important step, as well as additional work for other metal centers, will be beneficial to metalloenzyme design.

Progress has been made in the inclusion of solvent and solvent-mediated effects into protein-engineering computations. For compatibility with discrete search approaches, Mayo and colleagues have developed a pairwise approximation to continuum electrostatics and implemented it with the finite-difference Poisson–Boltzmann model [11]. The true problem is not pairwise because the desolvation of one side chain or the interaction of a pair of side chains depends on the shape of the protein and solvent regions, which is defined by the placement of all other side chains. The approximation involves a reduced representation of the protein structure built from the backbone and a single or a pair of side chains. There is a sense that continuum models appear to be an efficient approach for treating the important effects of the solvent environment, both directly in accounting for important desolvation effects and indirectly through screening of charge, polar, and hydrogen-bonding interactions. Wodak and colleagues have raised questions about the applicability of some implicit solvent models to design, where they found insufficient penalty for the burial of unsatisfied polar groups [12]. Clearly, these questions need to be addressed in future studies.

The placement of individual water molecules, particularly bridging protein complexes, is important in natural and designed proteins. Baker and colleagues have introduced a new energetic description of water-mediated hydrogen bonds and combined it with a 'solvated rotamer' approach to place interfacial water molecules using conventional rotamer search techniques [13••]. In our own protein redesign work, we treat buried, crystallographic water molecules with rotameric conformational freedom and the option to be replaced by new side-chain growth (SM Lippow *et al.*, unpublished).

Significant computational expense is required to assemble individual and pair energy contributions for combinatorial search; efficiencies achieved in this area are valuable. A trie data structure was used by Leaver-Fay, Kuhlman, and Snoeyink to eliminate redundant atom–atom calculations in the assembly of pair energies, which led to a fourfold speedup in this portion of the calculation [14].

The approximation of physical potentials with cluster expansion techniques was undertaken by Keating and colleagues [15,16••]. Energies for a subset of the search space were computed and used to train a set of expansion coefficients, which expressed the design energy in terms of sequence as opposed to structure, effectively integrating over rotamers for each residue. The resulting reduced search was reasonably accurate and extremely fast.

Potential energy functions describe the underlying interactions in protein systems, but folding and binding free energies, as well as kinetic binding and catalytic rates, result from an analysis of ensembles and include energetic and entropic contributions. New studies have expanded design considerations from single structures to explicit consideration of ensembles and their associated entropy. Donald and colleagues have formulated the protein design problem using a target function constructed from conformational ensembles rather than a single conformation and validated its use in redesigning substrate specificity [17••]. Kuhlman and colleagues used a different procedure based on Monte Carlo (MC) search to include side-chain conformational entropy in the design of 110 native protein backbones [18••]. They found very little difference in the resulting sequence designs whether entropy was included or not, with the largest differences involving long, flexible side chains. Even if conformational entropy contributions are not dominant in protein design calculations, the use of ensembles is likely to have other benefits in protein design engineering.

Schreiber and colleagues have developed and validated an approach to design mutations leading to faster and tighter binding complexes through enhancement of the electrostatic contribution to the association rate ([19] and references therein), using a computational treatment of the electrostatic interaction energy. Our group has developed an approach for identifying opportunities to introduce noncontacting residues near a binding interface that enhance affinity by virtue of paying very little desolvation penalty yet making larger 'action-at-a-distance' intermolecular interactions [20]. On the basis of different mechanistic principles, there is some but not very much overlap between designs made by this and the Schreiber approach. What is interesting about both approaches is that they are computationally very rapid because they do not require full repacking calculations.

An important question remains concerning which properties need to be explicitly accounted for in design and which others 'come along for the ride'. For example, focusing on protein stability appears to lead to designed proteins with appropriate kinetic pathways to the folded state with perhaps no need for explicit consideration of folding kinetics in the design. However, the same is not true for aggregation properties. Varani and colleagues redesigned, synthesized, and studied a variant of human U1A protein with 65 substitutions in 95 residues [21]. NMR showed that not only the backbone structure but also its dynamics were reproduced in the redesign. As the authors point out, the computation aims to reproduce the existence of a minimum in the free energy surface

Please cite this article in press as: Lippow SM, Tidor B, Progress in computational protein design, Curr Opin Biotechnol (2007), doi:10.1016/j.copbio.2007.04.009

corresponding to the native backbone structure; in doing so, the shape of the surface also appears to have been reproduced. Although more work needs to be done to assess the generality of this result, it suggests a certain insensitivity of backbone dynamics to at least some details of side-chain packing.

Search and optimization procedures

The tremendous advances in protein design studies over the past 10 years result from the maturation of a number of component technologies, including combinatorial discrete search and optimization methodology. This was led by the adoption of guaranteed discrete approaches such as dead-end elimination, A*, and integer programming, as well as faster, non-guaranteed methods including MC and self-consistent mean field theory. New improvements to foundational search methodology continue to drive innovation.

Donald and colleagues have made progress bridging guarantees in discrete search space with the complexities introduced when energy minimization in continuous space is considered for all members of the discrete space [22••]. In another work, Xie and Sahinidis recast the hierarchy used in guaranteed discrete search by explicitly considering residue elimination in addition to rotamer elimination, which speeds calculations by one to two orders of magnitude [23]. Allen and Mayo report two enhancements to the stochastic optimizer FASTER that result in up to two orders of magnitude speedup [24••]. The work points out benefits of selection of appropriate initial configurations and positions for relaxation following a perturbation. Hom and Mayo present a new MC and FASTER implementation for carrying out fixed composition sequence design, which may have benefits due to uncertainties in modeling the unfolded state [25].

Saven and Yang present improved sampling techniques based on MC and biased MC with replica exchange for use in extracting residue-specific probability distributions for protein design [26]. The probability distribution formulation is particularly relevant for the design of protein libraries for analysis by selection or high-throughput screening approaches. Saven and colleagues have fully connected probabilistic design and library construction through the development of altered machine protocols for automated DNA sythesizers to produce a pool of DNA corresponding to the desired protein sequence distribution [27]. Maranas and colleagues present an iterative procedure to design combinatorial libraries using mixed-integer linear programming and demonstrate it to explore multiple mutations of a starting sequence to improve the properties of the resulting protein [28]. Taken together, new algorithmic approaches provide greater efficiencies for exploring larger spaces and phrasing different optimization problems.

Progress in applications

Specificity

The term specificity describes several protein phenomena: selective binding to certain ligands, enzymatic activity for particular substrates, and the overall protein fold that a particular sequence adopts. In some cases, the design challenge is in creating new recognition, but at other times, the difficulty is in avoiding undesired recognition. Designing specificity often requires a combination of positive and negative design, but positive alone may be sufficient in some cases. Negative design poses a more difficult search problem, as one needs to directly address the structure prediction problem.

The design of specificity would be simplified if one only needed to consider positive design of a desired state, with specificity for that state a convenient by-product of optimization. Sauer and colleagues made a head-to-head comparison of a pure positive-design protocol and an explicit specificity approach for the redesign of a homodimer into a heterodimer [29••]. Their specificity design protocol yielded heterodimer specificity, but at the cost of protein stability. Baker and colleagues also explored two strategies for the design of a specificity switch at a protein–protein interface [30••]. Their direct affinity design protocol led to the creation after two design rounds of a 300-fold specificity switch over one of the non-cognate interactions. Sampling known variation in rigid-body binding orientation probably contributed to the specificity, and despite the creation of a novel hydrogen-bond network, the majority of the specificity switch was due to a single hydrophobic mutation. In addition, explicit negative design was used by Baker and colleagues to successfully redesign the specificity of an endonuclease [31••], by Keating and colleagues to convert a homotetramer into a heterotetramer [32], and by Jasanoff and colleagues to alter calmodulin specificity [33]. These data together support the need for elements of negative design for creating protein–protein interaction specificity.

De novo protein design often aims for an amino acid sequence that will fold to a specific, single structure, yet it is protein conformational change that often mediates function. Ambroggio and Kuhlman designed a single sequence to adopt two distinct folds, using direct stability design simultaneously for both states [34••]. The new protein switches from a zinc finger-like fold to a trimeric coiled-coil fold depending on pH or transition metals. The protein is aggregation prone from hydrophobic residues on the surface, reflecting the greater need for negative design to avoid undesired interactions.

Affinity

For the redesign of improved protein-binding affinity, energy function accuracy is critical. Binding affinity is usually modified within a few orders of magnitude, making calculations of single kcal/mol changes important. Redesign from nanomolar to picomolar affinities

Please cite this article in press as: Lippow SM, Tidor B, Progress in computational protein design, Curr Opin Biotechnol (2007), doi:10.1016/j.copbio.2007.04.009

is a particular challenge for a variety of maturation technologies.

Several groups have made progress toward protein–protein or protein–peptide binding affinity redesign. Springer and colleagues redesigned the low-affinity ICAM-I/LFA-I interaction using a variety of structure-based techniques, achieving 20-fold improvement to 12 nM by combining designed single mutations [35]. However, predictions from visualization-based expert design included the majority of the affinity-enhancing mutations, and exhibited a higher success rate than the computational methods. Sood and Baker designed N-terminal or C-terminal extensions to increase protein–peptide interaction using a novel technique that combines backbone and side-chain sequence and conformational search [36]. The results were modest though, with 1.3- and 2.3-fold improvement in their two test cases. Van Vlijmen and colleagues achieved eightfold improvement to 850 pM by combining four single mutations designed from a variety of available computational techniques [37]. Their success rate was 12% across 83 constructed mutants but would have been 26% with a retroactive analysis. Dahiyat and colleagues redesigned the Fc/FcγR interaction, yielding an Fc variant with over 100-fold improvement to 2 nM, after greater than 200 Fc variants were tested [38]. These results illustrate a need for reliable redesign methods and indicate an absence of redesign to high-affinity picomolar interactions.

In our own lab, in collaboration with K. Dane Wittrup, we redesigned multiple antibody–antigen interactions (SM Lippow *et al.*, unpublished). Using novel selection criteria based on calculations of improved binding electrostatics, we achieved a success rate for single mutations of over 60%. We combined single or double mutations to improve the lysozyme-binding antibody D44.1 by 140-fold to 30 pM, and the anti-EGFR therapeutic antibody cetuximab (Erbitux) by 10-fold to 52 pM. Our methods also identified known affinity-enhancing mutations in the anti-fluorescein antibody 4-4-20 and the anti-VEGF therapeutic antibody bevacizumab (Avastin).

The design of novel protein–protein interactions or small-molecule binding sites presents additional challenges for conformational search. DeGrado and colleagues developed a procedure to create a new protein framework for binding a specified cofactor and designed a four-helix bundle that binds a metalloporphyrin cofactor [39••]. Yang and colleagues designed a calcium-binding site into the cell adhesion protein CD2 using an approach that evaluates potential binding sites for compatibility with ideal geometry [40••].

Enzymes

It is not well understood how natural enzymes function, and thus the optimization objectives for computational enzyme design are unclear. Factors that may be important include binding to transition state, accommodation of substrate, release of product, protein flexibility and dynamics, and active-site catalytic residues. Chakrabarti *et al.* found that they could recover the majority of wild-type enzyme sequences by optimizing enzyme–substrate binding affinity while imposing geometric constraints on catalytic side-chain conformations [41••]; however, it is unclear if this is sufficient for design. Current work has enforced known key active site contacts [41••,42] or a known description for transition-state and functional-group geometry [43,44].

Mayo and colleagues explored enzyme redesign by stabilizing transition-state binding and only mutating second-shell positions. For the redesign of *Escherichia coli* chorismate mutase, one of five single mutations predicted to maintain activity increased efficiency by 60% [42]. Separately, redesign of the imipenemase IMP-1 predicted a mutation that removes a hydroxyl group and a double mutation that transfers the hydroxyl group [45]. Hydroxyl transfer altered substrate specificity, whereas the presence of both hydroxyl groups turned out to increase catalytic efficiency. Methods remain to be developed for the computational improvement of catalytic activity.

For the *de novo* design of enzymes, conformational search of both active-site residue placement and small-molecule rigid body placement complicates calculations. Baker and colleagues developed methods for the placement of a predefined active site and developed an *in silico* benchmark for 10 chemical reaction types [43]. Their procedure searches a protein for candidate active-site locations and then designs the surrounding protein for binding to the transition state. Mayo and colleagues developed methods for small-molecule placement in enzyme design [44]. They incorporated small-molecule rotational and translational search into protein design, and added energy biasing to favor side-chain–ligand contacts deemed necessary for catalysis or binding. Work by Chakrabarti *et al.* may also be useful for guiding the search for protein scaffolds suitable for introduction of *de novo* activity [46]. Computational enzyme design remains a significant challenge, with a rare success reported by Hellinga and colleagues [47••].

Though computational design has been used to stabilize proteins, enzyme stabilization is complicated by the need to maintain catalytic activity. Stoddard and colleagues thermostabilized the enzyme yeast cytosine deaminase by 10 °C through combination of three synergistic mutations [48••]. They optimized for protein stability while fixing the active site and contacting side chains. It remains unclear to what degree distal residues play a role in catalysis; this work demonstrates that the enzyme core can be modified independently of the active site functionality.

Please cite this article in press as: Lippow SM, Tidor B, Progress in computational protein design, Curr Opin Biotechnol (2007), doi:10.1016/j.copbio.2007.04.009

Discussion

Electrostatics

A common difficulty reported in computational design efforts is the accurate evaluation of electrostatic solvation and interaction terms. Electrostatics in protein design has been previously reviewed [7], and here we highlight continued challenges as exemplified by recent design work.

Electrostatics has affected design methods in various ways. On the one hand, designed structures have been subsequently discarded due to unsatisfied or suboptimal hydrogen bonding for altered protein–protein specificity [30••] or protein–peptide binding [36]. On the other hand, protein–DNA designs have been selected based on the hydrogen-bonding energy contribution, which was 'more highly predictive of the specificity of the native enzyme than the total energy of the complex', though the relative binding affinity prediction for the redesigned endonuclease was inaccurate [31••]. In addition, we found in our redesign of interactions that the electrostatic term of the binding energy was a better predictor than the total energy for affinity improvements (SM Lippow *et al.*, unpublished).

Results from computational design indicate additional progress needed in the treatment of electrostatics. In affinity redesign, successful mutations were 'almost exclusively nonpolar or aromatic, suggesting that packing interactions are predicted more accurately than electrostatic or polar interactions' [35]. For wild-type side-chain placement, 'structural accuracy is somewhat lower for polar and charged side chains compared with nonpolar side chains', though the authors attribute overall success in part to the use of a continuum electrostatic model [46]. In enzyme thermostabilization, 'redesigns involving incorporation or alteration of polar or charged residues in the core ... were less successful than mutations involving substitution of one hydrophobic side chain for another. ... Furthermore, modeling of interactions involving buried polar and charged side chains in the enzyme core is an area for future development' [48••]. Additionally, authors have pointed out the challenge of predicting polar residues engaged in hydrogen bonds with ligand [41••], the implied need for an improved treatment of electrostatics [42], the need for an energy-biasing step to make up for many shortcomings including electrostatics and solvation modeling [44], and the challenge that protein design presents implicit solvation models [12].

Human intervention

Most design methods are not free of human intervention. The use of hand curation is common for selecting or refining designs, as opposed to a fully automated methodology. For instance, designs have been removed that had unsatisfied hydrogen bonding [30••], 15% of structures with best binding energy were discarded because of visual inspection of suboptimal packing or hydrogen bonding [36], predictions were 'inspected visually, and if we observed an improvement in packing, additional intermolecular hydrogen bonds, or an increase in intermolecular hydrophobic contacts, we decided to make and express the mutants' [37], 'the designed Ca^{2+}-binding site in CD2 (Ca-CD2) was finally selected after careful evaluation' [40••], and N-terminal and C-terminal helix-capping residues were added to a four-helix bundle design [39••]. Hand curation of designs can be critical for success but limits the transferability of methods for use in new systems or by other researchers. Furthermore, these methods limit the ability to investigate and improve our understanding of the underlying biophysical interactions.

Bound water molecules

Structure determination of computationally designed proteins has revealed deficiencies in the modeling of explicit, bound water molecules found at interfaces and binding sites. A water molecule was not predicted yet found crystallographically at a protein–protein interface [30••] and a protein–DNA interface [31••]. Redesign has failed to displace a bound water molecule [30••], and wild-type redesign has failed to recapitulate a water molecule [36]. In our own work, we designed high-affinity improvements including a double mutation predicted to displace a bound water molecule; a structure of our mutant complex has yet to be determined (SM Lippow *et al.*, unpublished). We expect that improved handling of the conformational freedom of explicit, bound water molecules will continue to play an important role in design.

Independent designs and subsequent combination

Several successful design efforts have used an iterative procedure. In a first step, many small, independent designs are carried out, and predictions are experimentally validated. In a second step, successful mutations are combined for greater improvement. In redesigning binding affinity, Springer and colleagues were unsuccessful in their simultaneously designed mutations yet successful in combining single mutations [35]. The same was true for enzyme thermostabilization [48••]. The combination of separate, smaller designs was also used by our own group (SM Lippow *et al.*, unpublished) and others [37,38,42]. This divide, conquer, and recombine technique is powerful in that it reduces combinatorial complexity and isolates potentially destabilizing or unbeneficial mutations; however, the capability of computational design to search vast sequence spaces is not taken advantage of fully.

Outlook

Computational protein design is thriving, with more ambitious challenges being achieved together with the development of improved methodology that is on its way to becoming robust. Design using physics-based energy functions provides a more direct test of our understanding of biophysical interactions and might be applicable to a

Please cite this article in press as: Lippow SM, Tidor B, Progress in computational protein design, Curr Opin Biotechnol (2007), doi:10.1016/j.copbio.2007.04.009

broader class of problems, yet knowledge-based functions are more commonly used. Future development of the field will be advanced more by an understanding of failures than successes, and the widespread adoption of fully automated design (removing human intervention) will lead to better estimation of the inherent robustness and transferability of design technology. Current approaches are efficient and enable many practical protein-engineering applications already; future advances will expand the realm of possibilities and increase reliability.

Acknowledgements

This work was supported by a National Science Foundation Graduate Fellowship to S.M.L. and grants from the National Institutes of Health (GM65418 and CA96504).

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Rosenberg M, Goldblum A: **Computational protein design: A novel path to future protein drugs**. *Curr Pharm Design* 2006, **12**:3973-3997.

2. Poole AM, Ranganathan R: **Knowledge-based potentials in protein design**. *Curr Opin Struct Biol* 2006, **16**:508-513.

3. Ambroggio XI, Kuhlman B: **Design of protein conformational switches**. *Curr Opin Struct Biol* 2006, **16**:525-530.

4. Koder RL, Dutton PL: **Intelligent design: The *de novo* engineering of proteins with specified functions**. *Dalton Trans* 2006, **25**:3045-3051.

5. Baker D: **Prediction and design of macromolecular structures and interactions**. *Philos T Roy Soc B* 2006, **361**:459-463.

6. Butterfoss GL, Kuhlman B: **Computer-based design of novel protein structures**. *Annu Rev Biophys Biomol Struct* 2006, **35**:49-65.

7. Vizcarra CL, Mayo SL: **Electrostatics in computational protein design**. *Curr Opin Chem Biol* 2005, **9**:622-626.

8. Chica RA, Doucet N, Pelletier JN: **Semi-rational approaches to engineering enzyme activity: Combining the benefits of directed evolution and rational design**. *Curr Opin Biotechnol* 2005, **16**:378-384.

9. Morozov AV, Havranek JJ, Baker D, Siggia ED: **Protein–DNA binding specificity predictions with structural models**. *Nucleic Acids Res* 2005, **33**:5781-5798.

10. Spiegel K, DeGrado WF, Klein ML: **Structural and dynamical properties of manganese catalase and the synthetic protein DF1 and their implication for reactivity from classical molecular dynamics calculations**. *Proteins* 2006, **65**:317-330.

11. Marshall SA, Vizcarra CL, Mayo SL: **One- and two-body decomposable Poisson–Boltzmann methods for protein design calculations**. *Protein Sci* 2005, **14**:1293-1304.

12. Jaramillo A, Wodak SJ: **Computational protein design is a challenge for implicit solvation models**. *Biophys J* 2005, **88**:156-171.

13. Jiang L, Kuhlman B, Kortemme TA, Baker D: **A 'solvated rotamer' approach to modeling water-mediated hydrogen bonds at protein–protein interfaces**. *Proteins* 2005, **58**:893-904.
•• An orientation dependent hydrogen bonding potential describing water mediated hydrogen bonds was developed from crystallographic observations and validated through estimation of mutational affinity changes and recovery of native sequences in design. Expanded rotamers with pendent water molecules were developed and used in design studies.

14. Leaver-Fay A, Kuhlman B, Snoeyink J: **Rotamer-pair energy calculations using a trie data structure**. *Lecture Notes in Computer Science* 2005, **3692**:389-400.

15. Zhou F, Grigoryan G, Lustig SR, Keating AE, Ceder G, Morgan D: **Coarse-graining protein energetics in sequence variables**. *Phys Rev Lett* 2005, **95**:148103.

16. Grigoryan G, Zhou F, Lustig SR, Ceder G, Morgan D, Keating AE: **Ultra-fast evaluation of protein energies directly from sequence**. *PLoS Comput Biol* 2006, **2**:551-563.
•• Protein design approaches generally evaluate energies of candidate sequences through their structures. Cluster expansion methods were used to directly map sequence to energy for fixed backbones after training. Applications to a coiled coil, a zinc finger, and a WW domain demonstrate the power of the approach. Interestingly, an examination of dominant expansion terms reveals important structural and energetic interactions.

17. Lilien RH, Stevens BW, Anderson AC, Donald BR: **A novel ensemble-based scoring and search algorithm for protein redesign and its application to modify the substrate specificity of the gramicidin synthetase A phenylalanine adenylation enzyme**. *J Comput Biol* 2005, **12**:740-761.
•• An approach to protein design was developed that scores on ensemble-based estimates of free energy, and the method was applied to studying the enzyme GrsA-PheA. Wild type and a known mutant with selectivity for leucine over phenylalanine were analyzed, and design and experimental validation of new leucine specific mutants were performed.

18. Hu XZ, Kuhlman B: **Protein design simulations suggest that side-chain conformational entropy is not a strong determinant of amino acid environmental preferences**. *Proteins* 2006, **62**:739-748.
•• Protein design calculations were carried out on 110 native backbones using two protocols — one using a single optimal structure for each design and the other using a structural ensemble to evaluate energetic and entropic contributions. There was relatively little difference between the sequences designed by the two methods, which led to the suggestion that conformational entropy does not play a large role in determining sequence preferences in protein design.

19. Shaul Y, Schreiber G: **Exploring the charge space of protein–protein association: A proteomic study**. *Proteins* 2005, **60**:341-352.

20. Joughin BA, Green DF, Tidor B: **Action-at-a-distance interactions enhance protein binding affinity**. *Protein Sci* 2005, **14**:1363-1369.

21. Dobson N, Dantas G, Baker D, Varani G: **High-resolution structural validation of the computational redesign of human U1A protein**. *Structure* 2006, **14**:847-856.

22. Georgiev I, Lilien RH, Donald BR: **A novel 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**. *Lecture Notes in Computer Science.* 2006, **3909**:530-545.
•• The dead-end elimination algorithm applies to rigid rotamers in discrete search space. In this work the dead-end criterion was extended to apply to continuous deformation of rotamers through reconceptualizing each rotamer as representing a continuous voxel in a local conformational space. Through computation and appropriate manipulation of minimum and maximum one- and two-body energies for and between romatmeric voxels, a new guaranteed procedure was developed and applied. The new criterion is also applicable to A* searches. The applications demonstrate that configurations removed by rigid discrete search can minimize to energies lower than the minimized discrete global minimum.

23. Xie W, Sahinidis NV: **Residue-rotamer-reduction algorithm for the protein side-chain conformation problem**. *Bioinformatics* 2006, **22**:188-194.

24. Allen BD, Mayo SL: **Dramatic performance enhancements for the FASTER optimization algorithm**. *J Comput Chem* 2006, **27**:1071-1075.
•• FASTER is an efficient stochastic optimizer for protein design that starts by placing the rotamer with the best one-body energy at each position and then refines the design through an iterative series of improvements that involve relaxations and perturbations. The use of a short MC simulation to create a collection of high quality starting structures and focusing relaxation steps on positions that interact strongly with perturbed positions together were found to lead to significant quality and speed enhancements (up to two orders of magnitude faster).

Please cite this article in press as: Lippow SM, Tidor B, Progress in computational protein design, Curr Opin Biotechnol (2007), doi:10.1016/j.copbio.2007.04.009

25. Hom GK, Mayo SL: **A search algorithm for fixed-composition protein design**. *J Comput Chem* 2006, **27**:375-378.

26. Yang X, Saven JG: **Computational methods for protein design and protein sequence variability: biased Monte Carlo and replica exchange**. *Chem Phys Lett* 2005, **401**:205-210.

27. Park S, Kono H, Wang W, Boder ET, Saven JG: **Progress in the development and application of computational methods for probabilistic protein design**. *Comput Chem Eng* 2005, **29**:407-421.

28. Saraf MC, Moore GL, Goodey NM, Cao VY, Benkovic SJ, Maranas CD: **IPRO: An iterative computational protein library redesign and optimization procedure**. *Biophys J* 2006, **90**:4167-4180.

29. •• Bolon DN, Grant RA, Baker TA, Sauer RT: **Specificity versus stability in computational protein design**. *Proc Natl Acad Sci USA* 2005, **102**:12724-12729.

The SspB adaptor protein homodimer was redesigned into a heterodimer using both a stability design protocol and an explicit specificity design. Van der Waals clashes were capped at +5 kcal/mol each to approximate conformational relaxation. The stability design did not achieve heterodimer specificity, but was more stable than wild type. The specificity design did prefer heterodimer formation, but at the cost of protein stability.

30. •• Joachimiak LA, Kortemme T, Stoddard BL, 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* 2006, **361**:195-208.

The authors redesigned the colicin E7 DNase–Im7 immunity protein complex to exhibit a new hydrogen-bond network and at least a 300-fold specificity switch over one of the cognate interactions. They used both a negative and positive design strategy, and sampled rigid-body orientations mimicking the natural specificity determinant. One of 11 initial designs, from the affinity protocol, was crystallized and used as the basis for the second-round design.

31. •• Ashworth J, Havranek JJ, Duarte CM, Sussman D, Monnat RJ, Stoddard BL, Baker D: **Computational redesign of endonuclease DNA binding and cleavage specificity**. *Nature* 2006, **441**:656-659.

The endonuclease I-*Mso*I was redesigned to specifically recognize a target sequence with two base-pair changes. A disrupting base-pair change was found computationally, and then amino acids were redesigned to accommodate the new base pair. This is the first study to computationally switch endonuclease recognition specificity.

32. Ali MH, Taylor CM, Grigoryan G, Allen KN, Imperiali B, Keating AE: **Design of a heterospecific, tetrameric, 21-residue miniprotein with mixed alpha/beta structure**. *Structure* 2005, **13**:225-234.

33. Green DF, Dennis AT, Fam PS, Tidor B, Jasanoff A: **Rational design of new binding specificity by simultaneous mutagenesis of calmodulin and a target peptide**. *Biochemistry* 2006, **45**:12547-12559.

34. •• Ambroggio XI, Kuhlman B: **Computational design of a single amino acid sequence that can switch between two distinct protein folds**. *J Am Chem Soc* 2006, **128**:1154-1161.

The authors designed a single amino-acid sequence that can adopt either a zinc finger-like fold or a trimeric coiled-coil, depending upon pH or the presence of transition metals.

35. Song G, Lazar GA, Kortemme T, Shimaoka M, Desjarlais JR, Baker D, Springer TA: **Rational design of intercellular adhesion molecule-1 (ICAM-1) variants for antagonizing integrin lymphocyte function-associated antigen-1-dependent adhesion**. *J Biol Chem* 2006, **281**:5042-5049.

36. Sood VD, Baker D: **Recapitulation and design of protein binding peptide structures and sequences**. *J Mol Biol* 2006, **357**:917-927.

37. Clark LA, Boriack-Sjodin PA, Eldredge J, Fitch C, Friedman B, Hanf KJ, Jarpe M, Liparoto SF, Li Y, Lugovskoy A *et al.*: **Affinity enhancement of an *in vivo* matured therapeutic antibody using structure-based computational design**. *Protein Sci* 2006, **15**:949-960.

38. Lazar GA, Dang W, Karki S, Vafa O, Peng JS, Hyun L, Chan C, Chung HS, Eivazi A, Yoder SC *et al.*: **Engineered antibody Fc variants with enhanced effector function**. *Proc Natl Acad Sci USA* 2006, **103**:4005-4010.

39. •• Cochran FV, Wu SP, Wang W, Nanda V, Saven JG, Therien MJ, DeGrado WF: **Computational *de novo* design and characterization of a four-helix bundle protein that selectively binds a nonbiological cofactor**. *J Am Chem Soc* 2005, **127**:1346-1347.

The authors developed computational design methodology to engineer a protein that binds a specificied cofactor. They designed a four-helix bundle to bind a DPP–Fe cofactor.

40. •• Yang W, Wilkins AL, Ye YM, Liu ZR, Li SY, Urbauer JL, Hellinga HW, Kearney A, van der Merwe PA, Yang JJ: **Design of a calcium-binding protein with desired structure in a cell adhesion molecule**. *J Am Chem Soc* 2005, **127**:2085-2093.

A Ca^{2+}-binding site was designed into the cell adhesion protein CD2, which retained its ability to associate with the natural target molecule. This is the first study to design a Ca^{2+}-binding protein with specified properties.

41. •• Chakrabarti R, Klibanov AM, Friesner RA: **Computational prediction of native protein ligand-binding and enzyme active site sequences**. *Proc Natl Acad Sci USA* 2005, **102**:10153-10158.

The sequences of protein ligand-binding and enzyme active sites were predicted by optimizing binding affinity. Constraints based on known catalytic mechanism were used to capture the active sites of diverse enzymes.

42. Lassila JK, Keeffe JR, Oelschlaeger P, Mayo SL: **Computationally designed variants of *Escherichia coli* chorismate mutase show altered catalytic activity**. *Protein Eng Des Sel* 2005, **18**:161-163.

43. Zanghellini A, Jiang L, Wollacott AM, Cheng G, Meiler J, Althoff EA, Rothlisberger D, Baker D: **New algorithms and an in silico benchmark for computational enzyme design**. *Protein Sci* 2006, **15**:2785-2794.

44. Lassila JK, Privett HK, Allen BD, Mayo SL: **Combinatorial methods for small-molecule placement in computational enzyme design**. *Proc Natl Acad Sci USA* 2006, **103**:16710-16715.

45. Oelschlaeger P, Mayo SL: **Hydroxyl groups in the beta beta sandwich of metallo-beta-lactamases favor enzyme activity: A computational protein design study**. *J Mol Biol* 2005, **350**:395-401.

46. Chakrabarti R, Klibanov AM, Friesner RA: **Sequence optimization and designability of enzyme active sites**. *Proc Natl Acad Sci USA* 2005, **102**:12035-12040.

47. •• Dwyer MA, Looger LL, Hellinga HW: **Computational design of a biologically active enzyme**. *Science* 2004, **304**:1967-1971.

Triose phosphate isomerase activity was introduced into ribose-binding protein. Substrate recognition was first introduced, and then catalytic residues were incorporated. This is the first report of a computationally designed enzyme.

48. •• Korkegian A, Black ME, Baker D, Stoddard BL: **Computational thermostabilization of an enzyme**. *Science* 2005, **308**:857-860.

The enzyme yeast cytosine deaminase was redesigned to be 10 °C more thermostable while maintaining its wild-type catalytic efficiency. The authors fixed the active site and contacting sites, and combined three separate single mutations for the overall improvement.

Please cite this article in press as: Lippow SM, Tidor B, Progress in computational protein design, Curr Opin Biotechnol (2007), doi:10.1016/j.copbio.2007.04.009