SOFT MATERIALS
Vol. 2, No. 1, pp. 1–10, 2004

Protein Sequence Optimization—Theory, Practice, and Fundamental Impossibility

Andrew E. Torda*

Zentrum für Bioinformatik, University of Hamburg, Hamburg, Germany

ABSTRACT

In this article, a fundamental aspect of protein engineering called into question: Is it really possible to computationally redesign the sequence of a protein so as to improve its stability? Specifically, the case where almost all protein's sequence is allowed to change to any other residue type is considered. The aim is to highlight two conflicting schools of thought. First, there are compelling fundamental reasons why protein design should not be possible by optimizing a conventional energy-like function. Second, there is evidence that it is possible and has been done in practice.

Key Words: Protein sequence design; Sequence optimization; Protein engineering; Negative design.

INTRODUCTION

In the early 1980s,[1] it was suggested that one should be able to design proteins, almost at will, with techniques from molecular biology at hand to synthesize them. Two decades later, there are some clearly conflicting propositions, results, and interpretations. There are good theoretical grounds to believe the problem is very difficult, but also experimental evidence to believe it is almost solved. It is not clear who is correct, lucky,

*Correspondence: Andrew E. Torda, Zentrum für Bioinformatik, University of Hamburg, Bundesstr. 43, D-20146 Hamburg, Germany; Fax: +49-40-42838-7332; E-mail: torda@zbh.uni-hamburg.de.

DOI: 10.1081/SMTS-120030139 1539-445X (Print); 1539-4468 (Online)

or wrong. It is certainly time to highlight opinions in the literature that are completely conflicting or results that disagree with reputable theory.

Protein design means different things to different researchers. Sometimes, a molecular biologist can change a few residues in a protein sequence, measure a change in activity or stability, and call the process "protein design." This is not the definition used here. Instead, we consider a specific question. Given a protein of known structure, can one permit all the residues to change and apply a computational method to automatically produce a better protein sequence, not necessarily related to the original (native) sequence? The word "better" will mean different things to different researchers. Here, we only consider thermostability or the resistance to unfolding under native conditions. We do not consider making an enzyme active at a different pH, tolerant of organic solvents, or able to bind to a specific ligand. We do not consider modifying a protein so as to change kinetic properties such as folding rate. The reason for this narrow perspective is that some kind of simple stability is the property most often addressed by experiment or calculation and, most important for this discussion, it is the property most related to an energy-like function. It is also the property that leads to a disagreement between what has been experimentally determined and what would be expected from simple reasoning.

There are two very separable aspects to the problem of finding a better protein sequence. First, one needs some kind of score, cost, or merit function. This is the heart of the problem and is discussed below. Second, one needs some way to search through the rather large search space. The problem has been formally shown to be NP-complete,[2] but intuitive arguments make the point. If one is dealing with proteins, there are 20 amino acids and 20^N possible sequences for a protein of length N. If one is dealing with some other heteropolymer or a reduced protein alphabet, the problem still has exponential growth. Even in the classic toy case of a hydrophobic/polar model with just two monomer types, a molecule of 100 residues has 10^{30} possible sequences. Here, problem of searching is not considered because it is the score (energy) aspect that leads to contradictions. Furthermore, one could argue that the search problem has been solved. On a small protein, Dahiyat and Mayo[3] used a pruning algorithm[4] to reduce the search space to one that could be tackled exhaustively. Skorobogatiy et al.[5] took an analytical approach to the problem and if one wants a distribution of protein sequences, one could use configurational-biased Monte Carlo[6] or even Monte Carlo biased with a self-consistent mean field approach.[7]

Searching for a protein or polymer sequence does have some practical and principled differences to other problems in computational chemistry. In protein design, one is given a good answer (the native sequence) before the start of the calculation. One does not know if this native/natural sequence is in some sense optimal, but it is probably a reasonable solution. This is important because many of the search methods can be readily persuaded to include a bias toward this answer, and some reports have included such a bias.[8,9] The problem of protein design also spans a continuum of generality or difficulty. At one extreme, permitting every residue at every site to change to every other residue type is most general and most challenging. A small restriction on the problem would be to fix some residues before the calculation so as to maintain an enzyme's active site. Next, one could argue that only hydrophobic residues should be permitted in the core of a protein, so the problem is computationally simpler. Last, in the least general case, one may pick a small number of residues and vary only these computationally. The consequence of this range of problems is that it is not possible to fairly compare different reports in the literature.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

SCORE/ENERGY FUNCTIONS

If one is going to computationally optimize the sequence of a protein or polymer, the critical decision will be the type of score or energy function to use. The earliest, truly designed, peptides were based on sets of rules interpreted by hand,[10] but that approach does not scale well to a 20^N space, for even modest values of N. A more applicable approach would be to work with the estimates of potential energy from a classical atomistic force field. Unfortunately, this leads to a serious complication. Use of an atomistic force field assumes one knows where the atoms are to atomic accuracy. Most amino acids, however, have rotatable degrees of freedom in their side chains. This implies that one must add the problem of predicting side-chain geometry to the original problem of optimizing the amino acids. In the context of protein design, Hellinga and Richards[11] described what has become a standard approach to this problem. They assumed that the side chains would only occupy one of several preferred rotamer positions and allowed Monte Carlo moves to select the side-chain conformations. One could also use a self-consistent mean field approach to find preferred, consistent rotamers[12] or treat each rotamer of a residue as a separate type of amino acid.[3]

Given the size of the search space and the approximate nature of the problem, it is very common to use a more coarse-grained approach with a protein model that averages over all the side-chain atoms. This would often have the side chain represented by a single interaction site located at the C^β atom or side-chain center of mass. This may not be a tragic loss of detail. It may even be particularly sensible given that protein design methods rarely admit that the protein backbone can move[13–15] and atomic detail may be largely irrelevant. Ultimately, for the purposes of this argument, the interest lies only in the problems that remain, even if one has a perfect model for the energy of a system.

FUNDAMENTAL PROBLEMS

In 1993, Shakhnovich and Gutin optimized protein sequences using a simple score function and search method, but they did not allow the system to move freely through the possible sequences.[8] Instead, they completely constrained the system to its original composition, merely reordering or shuffling its amino acids. In 1994, Jones used a genetic algorithm and simple scoring scheme to optimize protein sequences, but with the restraint that the sequence be penalized as it changes from the native.[9] This was rationalized by noting that there is some correspondence between overall protein composition and structural class. Neither of these calculations would seem really general. If a score function acted as intended, it should not be necessary to confine the search space. In practice, this kind of restraint is potentially damaging. It has been said that protein sequence space is much larger than structure space.[16] Put more usefully, there are often very large numbers of apparently unrelated protein sequences that fold to similar structures. Any sequence searching method that is restrained to its starting sequence will never find unrelated sequences.

The real reason for having to restrict the search of sequences was given by Godzik.[17] One can change a protein sequence, producing a new sequence/structure pair of lower "energy." There is, however, absolutely no reason to expect this sequence/structure pair not to fold into some different structure. Considering this as a picture, Fig. 1 shows

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

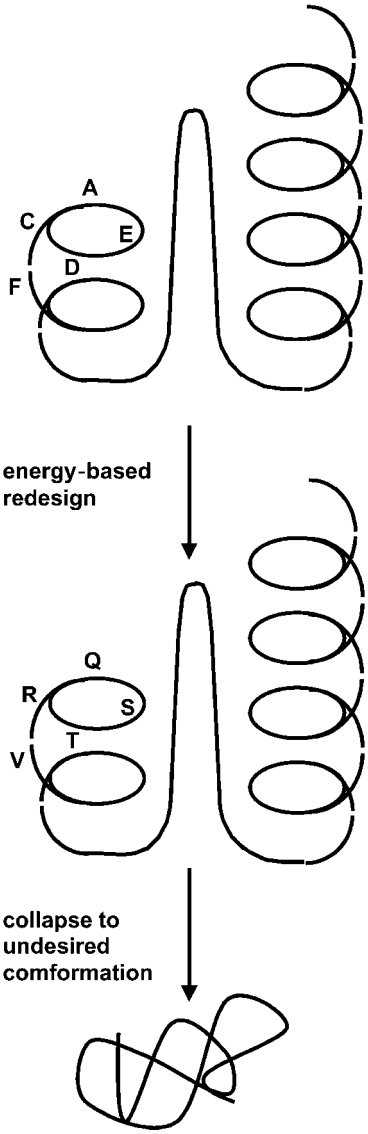

Figure 1. The fundamental problem in energy-based protein sequence design.

two steps. The process begins with the native sequence on the desired structure. Next, a calculation or energy-based redesign yields a "better" sequence for the desired structure, exactly as intended, and the exercise should be finished. Unfortunately, a real protein would now be free to explore conformational space. It may well find some alternative conformation or prefer the entropic attraction of the poorly structured mess shown at the bottom of Fig. 1. This argument is extraordinarily general, plausible, and empirically observed. As Godzik and many others noted, most simple score functions, in the context of protein design, lead to long homopolymeric stretches punctuated by other favored residues

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

around loops and turns. In a very artificial sense, these sequences are of lower energy than the native sequence. They simply will not fold to the desired structure. Godzik's point is even more general. Considering Fig. 1, it probably does not matter if one is dealing with potential energy or free energy approximations. There is no reason to believe that a sequence designed for a structure using some free energy approximations will not be able to find some even better conformation than the intended one. With such a fundamental problem, it is possible that discussions about the relative importance of various terms to protein design[18,19] are premature.

NEGATIVE DESIGN

There is a clear resolution to the problem, even if it seems intractable. One should not merely optimize a sequence for a given structure. One must also choose the sequence so its score on the native structure is better than every other alternative conformation it could ever visit. This idea has taken the name "negative design" and implies some thresholds one could debate. How much better should the desired conformation be, or what should be the energy gap to the nearest competing structure or the average of competing structures? What actually constitutes a different structure? None of this matters to the central problem. What was a difficult problem involving the choice of sequence on one conformation is really a ferocious problem involving a sequence, the desired structure, and every other conformation that contributes to the partition function.

The challenge of negative design has been tackled by methods of ever-increasing cunning. While changing monomer types in the protein, one must somehow incorporate the influence of all other conformations. A true brute force approach will not be practical, so a number of ideas have been tested, usually on lattice and model systems. One may try to do simultaneous Monte Carlo steps of both sequence and conformation,[20] or try to identify and consider the set of most important, competing conformations.[21–23] One could search for an approximation (low-order expansion) that encapsulates the contributions of alternative conformations.[24–26] Working with real proteins instead of model systems, some groups have explicitly incorporated ideas of negative design, but have dealt with a limited number of computational mutations or a small number of alternative configurations.[27,28]

Coming from a different philosophy, a very elegant approach is to try to build an all-encompassing score function that includes the concept of negative design, but encodes it as a function of sequence and the single desired structure. Chiu and Goldstein took a lattice model protein with a defined, pairwise, "true" energy function.[29] The system was small enough that, for a given sequence, all conformations could be visited. They then used numerical optimization to search for a new set of parameters that not only found a sequence of best energy, but also tried to ensure the desired sequence was of better score than competing conformations. In hindsight, it is not surprising that the best function for optimizing a sequence is not the "true" energy function. It is some different function that reflects the influence of all possible configurations. The authors make the point that, for protein sequence design, "the ability of increasingly accurate energy functions... is limited."

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Given the intuitive arguments, simulation results, and quantity of literature devoted to the process of negative design, one may well conclude that the optimization of protein sequences is a pastime doomed to failure on real protein systems.

PRACTICAL RESULTS

Protein sequence optimization is intuitively difficult, formally NP-hard, and the perfect score function does not seem to correspond to energy, potential energy, or any easily defined property. As if to defy common sense and physics, Dahiyat and Mayo[3] took the structure of a small protein and used a pruning algorithm[4] to completely redesign its sequence using a quite conventional, classical, atomistic score function.[30] As a tour de force demonstration, they synthesized the protein, used nuclear magnetic resonance to show that it folded to the desired structure, and by normal sequence searching methods showed that it was not obviously similar to the native sequence. There was apparently no consideration of negative design, and the score function was a conventional, classical, atomistic molecular mechanics force field with slightly reduced van der Waals radii and an extra term to account for solvation. This is not the only work to have used an atomistic force field, but it is remarkable for the synthesis and testing of the results and the method's apparently general nature. It raises an important question. Have other groups been able to achieve similar success and have they been able to tackle this general problem, allowing almost all residues in a sequence to change in an almost unrestricted manner?

Earlier, in 1994, Hellinga and Richards[11] used a conventional molecular mechanics score function, with an extra solvation contribution, for protein sequence optimization. They only permitted a few core, hydrophobic residues to change and only to other hydrophobic residues. More recently, this kind of calculation has continued, well after one might expect more awareness of the need to penalize undesired, nonnative structures. One group has published more than one sequence optimization calculation simply using the popular CHARMM[31] force field supplemented by an estimation of solvation energy.[32,33] They only allowed a small subset of residues to change and only in a limited fashion. Nevertheless, they refer to their method as a general automatic method for sequence design. Another group, using a classical atomistic force field, refers to their work as "a fully automated protein design strategy," but they do not permit the composition of the protein to change.[34] In their work, they refer to their score function as a free energy. This may or may not be the case, but there is no way to see it as incorporating any explicit penalty for a sequence preferring to fold to some unexpected structure. Continuing in this vein, others based their work on the classic ECEPP force field,[35,36] supplemented by solvation and "entropy" terms, but only allowed a small number of residues to change from the native protein.[37–39] Only when some of the designed sequences are synthesized it will be possible to see how successful the methods were.

RECONCILIATION

There are fundamental reasons to believe protein design should not be possible with a simple energy-like function, without explicitly considering negative design.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

There is a body of work claiming it has been done. It is very difficult to reconcile these views, but there are several possibilities. There could be some fundamental misunderstanding and all the work involving negative design might be unnecessary.[17,20–29] Maybe the literature claiming success should be treated more cautiously and some of the fully automatic general design methods are not as general as they appear. Some of these questions could be answered if the relevant design calculations were allowed to run with an unrestricted choice of sites to change and an unrestricted choice of substituents.

There are some alternative and plausible interpretations. If the energy-function methods are successful and really general, then they must be the all-encompassing design functions that Chiu and Goldstein[29] suggested looking for. They must subtly incorporate features that not only pick the sequence of best score, but also the sequence that will not prefer to be in some other structure. The only obvious changes to conventional functions have been a slight adjustment to Lennard–Jones terms[30] or incorporation of a solvation energy/free energy approximation.[3,11,32,33,37–39] It is completely unclear how these terms could be the key to negative design. Solvation approximations would not do anything to prevent the favoring of hydrophobic homopolymers in the core of proteins and the reduction of van der Waals radii is normally justified as a compensation for the inaccuracies introduced by keeping the protein backbone fixed.

There is another possibility that cannot be confirmed or discounted. Although there is a school of thought that says that one should not chase the details of energy functions, those working with atomic detail spend a vast amount of computational time searching for good atomic packing. This is the *raison d'être* for methods such as the dead-end elimination algorithm.[4] Proponents of these methods may well argue that the secret for confining the system to the correct configuration is careful side-chain packing. One should also note that the most spectacular successes approaching complete sequence redesign have used score functions with much atomistic detail, rather than the coarse-grain methods described above.[3,15]

Last, there are functions that may not be potential, free, or any other energies. If one mixed a modified Lennard–Jones term with statistical rotamer preferences, a solvation approximation,[40] a separate hydrogen-bond term,[41] a statistical approximation to electrostatics,[42] and a statistical preference of amino acid types, one cannot say what kind of score function one has.[43] Most intriguingly, such a function does seem to be successful in protein sequence redesign.[44] The most recent application of this scoring function is more remarkable, in that the authors first designed an artificial structure of 93 residues and then optimized a sequence for it.[15] This result has some important implications. First, there is no possibility of bias toward the native sequence because there was none and the final sequence appears unrelated to anything already known. Next, the calculation was quite unrestrained in sequence terms. The only restriction was that polar residues be on the surface. This leaves one with the conclusion that this function is implicitly incorporating enough negative design without stating it explicitly.

CONCLUSION

In principle, computational protein sequence design should be nearly impossible, but if groups have given their programs names like "DESIGNER"[32,33] one might think the

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

area is a finished story. Unfortunately, after studying the literature, one is still left with the following questions:

1. Is protein sequence design largely a solved problem?
2. If not, is it possible with cost/energy functions similar to those currently in the literature?

Some clues could be given by computational work, without the expense of visiting the laboratory. For example, within the work traveling under the title of protein design, there are many papers with very artificial restrictions on the permitted sequence changes. It would be simple to see what happens when these are removed. The outcome may be already well known, but unpublished. For those methods that appear to work without restrictions, one could acquire some statistical data. A protein sequence optimization calculation may once have taken many hours,[3] but with a different score function, search algorithm or faster machines, now apparently takes minutes.[43] Kuhlman and Baker[43] took advantage of this to optimize 108 sequences. Within this collection, they not only found surprising similarity to native sequences, but also found correlations between residues that seem to correspond to those observed in nature. Clearly, this kind of analysis could be extended to other properties and to more sequences, but more important, could be applied to the methods of other groups.

The validity of computational methods will be measured by laboratory work, but progress is slow. The volume of literature on computational sequence optimization is much bigger than the amount of experimental verification. Presumably, this is because computer time is cheaper than laboratory work.

At the moment, all spectators in this field must wait to see if the automatic and general design methods are as general as hoped. If so, it will be of interest to see if anyone can identify the aspects of successful score functions that implicitly include the properties of negative design.

REFERENCES

1. Drexler, K.E. Molecular engineering: an approach to the development of general capabilities for molecular manipulation. Proc. Natl Acad. Sci. U.S.A. **1981**, *78*, 5275–5278.
2. Pierce, N.A.; Winfree, E. Protein design is *NP*-hard. Protein Eng. **2002**, *15*, 779–782.
3. Dahiyat, B.I.; Mayo, S.L. De novo protein design: fully automated sequence selection. Science **1997**, *278*, 82–87.
4. Desmet, J.; Demaeyer, M.; Hazes, B.; Lasters, I. The dead-end elimination theorem and its use in protein side-chain positioning. Nature **1992**, *356*, 539–542.
5. Skorobogatiy, M.; Guo, H.; Zuckermann, M.J. A deterministic approach to the protein design problem. Macromolecules **1997**, *30*, 3403–3410.
6. Cootes, A.P.; Curmi, P.M.G.; Torda, A.E. Biased Monte Carlo optimization of protein sequences. J. Chem. Phys. **2000**, *113*, 2489–2496.
7. Zou, J.M.; Saven, J.G. Using self-consistent fields to bias Monte Carlo methods with applications to designing and sampling protein sequences. J. Chem. Phys. **2003**, *118*, 3843–3854.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

8. Shakhnovich, E.I.; Gutin, A.M. A new approach to the design of stable proteins. Protein Eng. **1993**, *6*, 793–800.
9. Jones, D.T. De-novo protein design using pairwise potentials and a genetic algorithm. Protein Sci. **1994**, *3*, 567–574.
10. Gutte, B.; Däumigen, M.; Wittschieber, E. Design, synthesis and characterisation of a 34-residue polypeptide that interacts with nucleic acids. Nature **1979**, *281*, 650–655.
11. Hellinga, H.W.; Richards, F.M. Optimal sequence selection in proteins of known structure by simulated evolution. Proc. Natl Acad. Sci. U.S.A. **1994**, *91*, 5803–5807.
12. Koehl, P.; Delarue, M. Application of a self-consistent mean-field theory to predict protein side-chains conformation and estimate their conformational entropy. J. Mol. Biol. **1994**, *239*, 249–275.
13. Kraemer-Pecore, C.M.; Lecomte, J.T.J.; Desjarlais, J.R. A de novo redesign of the WW domain. Protein Sci. **2003**, *12*, 2194–2205.
14. Harbury, P.B.; Plecs, J.J.; Tidor, B.; Alber, T.; Kim, P.S. High-resolution protein design with backbone freedom. Science **1998**, *282*, 1462–1467.
15. Kuhlman, B.; Dantas, G.; Ireton, G.C.; Varani, G.; Stoddard, B.L.; Baker, D. Design of a novel globular protein fold with atomic level accuracy. Science **2003**, *302*, 1364–1368.
16. Holm, L.; Sander, C. Mapping the protein universe. Science **1996**, *273*, 595–602.
17. Godzik, A. In search of the ideal protein-sequence. Protein Eng. **1995**, *8*, 409–416.
18. Gordon, D.B.; Marshall, S.A.; Mayo, S.L. Energy functions for protein design. Curr. Opin. Struct. Biol. **1999**, *9*, 509–513.
19. Pokala, N.; Handel, T.M. Review: protein design—where we were, where we are, where we're going. J. Struct. Biol. **2001**, *134*, 269–281.
20. Seno, F.; Vendruscolo, M.; Maritan, A.; Banavar, J.R. Optimal protein design procedure. Phys. Rev. Lett. **1996**, *77*, 1901–1904.
21. Micheletti, C.; Maritan, A.; Banavar, J.R. A comparitive study of existing and new design techniques for protein models. J. Chem. Phys. **1999**, *110*, 9730–9738.
22. Micheletti, C.; Seno, F.; Banavar, J.R.; Maritan, A. Learning effective amino acid interactions through iterative stochastic techniques. Proteins **2001**, *42*, 422–431.
23. Micheletti, C.; Seno, F.; Maritan, A.; Banavar, J.R. Protein design in a lattice model of hydrophobic and polar amino acids. Phys. Rev. Lett. **1998**, *80*, 2237–2240.
24. Irbäck, A.; Peterson, C.; Potthast, F.; Sandelin, E. Monte Carlo procedure for protein design. Phys. Rev. E **1998**, *58*, R5249–R5252.
25. Irbäck, A.; Peterson, C.; Potthast, F.; Sandelin, E. Design of sequences with good folding properties in coarse-grained protein models. Struct. Fold. Des. **1999**, *7*, 347–360.
26. Deutsch, J.M.; Kurosky, T. New algorithm for protein design. Phys. Rev. Lett. **1996**, *76*, 323–326.
27. Havranek, J.J.; Harbury, P.B. Automated design of specificity in molecular recognition. Nat. Struct. Biol. **2003**, *10*, 45–52.
28. Summa, C.M.; Rosenblatt, M.M.; Hong, J.K.; Lear, J.D.; Degrado, W.F. Computational de novo design, and characterization of an A_2B_2 diiron protein. J. Mol. Biol. **2002**, *321*, 923–938.
29. Chiu, T.L.; Goldstein, R.A. Optimizing potentials for the inverse protein folding problem. Protein Eng. **1998**, *11*, 749–752.
30. Dahiyat, B.I.; Mayo, S.L. Protein design automation. Protein Sci. **1996**, *5*, 895–903.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

31. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. CHARMM—a program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. **1983**, *4*, 187–217.
32. Wernisch, L.; Hery, S.; Wodak, S.J. Automatic protein design with all atom force-fields by exact and heuristic optimization. J. Mol. Biol. **2000**, *301*, 713–736.
33. Ogata, K.; Jaramillo, A.; Cohen, W.; Briand, J.P.; Connan, F.; Choppin, J.; Muller, S.; Wodak, S.J. Automatic sequence design of major histocompatibility complex class i binding peptides impairing cd8(+) t cell recognition. J. Biol. Chem. **2003**, *278*, 1281–1290.
34. Koehl, P.; Levitt, M. De novo protein design. I. In search of stability and specificity. J. Mol. Biol. **1999**, *293*, 1161–1181.
35. Momani, F.A.; Mcguire, R.F.; Burgess, A.W.; Scheraga, H.A. Energy parameters in polypeptides vii. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions and intrinsic torsional potentials for the naturally occurring amino acids. J. Phys. Chem. **1975**, *79*, 2361–2381.
36. Nemethy, G.; Pottle, M.S.; Scheraga, H.A. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen-bond interactions for the naturally-occurring amino-acids. J. Phys. Chem. **1983**, *87*, 1883–1887.
37. Reina, J.; Lacroix, E.; Hobson, S.D.; Fernandez-Ballester, G.; Rybin, V.; Schwab, M.S.; Serrano, L.; Gonzalez, C. Computer-aided design of a pdz domain to recognize new target sequences. Nat. Struct. Biol. **2002**, *9*, 621–627.
38. De La Paz, M.L.; Lacroix, E.; Ramirez-Alvarado, M.; Serrano, L. Computer-aided design of beta-sheet peptides. J. Mol. Biol. **2001**, *312*, 229–246.
39. Fisinger, S.; Serrano, L.; Lacroix, E. Computational estimation of specific side chain interaction energies in alpha helices. Protein Sci. **2001**, *10*, 809–818.
40. Lazaridis, T.; Karplus, M. Effective energy function for proteins in solution. Proteins **1999**, *35*, 133–152.
41. 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. **2003**, *326*, 1239–1259.
42. Simons, K.T.; Ruczinski, I.; Kooperberg, C.; Fox, B.A.; Bystroff, C.; Baker, D. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. Proteins **1999**, *34*, 82–95.
43. Kuhlman, B.; Baker, D. Native protein sequences are close to optimal for their structures. Proc. Natl Acad. Sci. U.S.A. **2000**, *97*, 10383–10388.
44. 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.

Received November 6, 2003
Accepted December 3, 2003