REVIEW

Progress in Modeling of Protein Structures and Interactions

Ora Schueler-Furman,[1,2] Chu Wang,[1] Phil Bradley,[1] Kira Misura,[1] David Baker[1,3]*

The prediction of the structures and interactions of biological macromolecules at the atomic level and the design of new structures and interactions are critical tests of our understanding of the interatomic interactions that underlie molecular biology. Equally important, the capability to accurately predict and design macromolecular structures and interactions would streamline the interpretation of genome sequence information and allow the creation of macromolecules with new and useful functions. This review summarizes recent progress in modeling that suggests that we are entering an era in which high-resolution prediction and design will make increasingly important contributions to biology and medicine.

In 1973, Anfinsen demonstrated that the amino acid sequence of a protein completely specifies its three-dimensional structure and hence that the native structures of proteins are likely to correspond to global free-energy minima (*1*). Since then, the de novo structure prediction problem has been well posed—find the lowest free-energy conformation for an amino acid sequence. Yet at the start of the structural genomics efforts in the late 1990s, computational methods remained far from achieving the high-resolution structures available from x-ray crystallography and nuclear magnetic resonance (NMR), and hence embarking on large-scale experimental structure determination with its associated high cost was well warranted. We suggest that recent progress in high-resolution modeling of biomolecules is such that, although the de novo folding problem is far from solved, computational structural biology is reaching a stage where it can contribute both to determining structures of naturally occurring biomolecules and the creation of new ones.

There have been two distinct areas of development in molecular modeling methodology and software—the first aimed at simulating macromolecular dynamics, the second at prediction and design. We focus on the second area, and for the first we refer to recent reviews (*2*, *3*). Two blind tests provide a gauge of progress in prediction of the structures of proteins and protein-protein complexes, namely CASP [http://predictioncenter.org (*4*)] and CAPRI [Critical Assessment of Protein Interactions; http://capri.ebi.ac.uk/ (*5*, *6*)]. Results obtained in the recent CASP and CAPRI experiments, together with recent design results, highlight progress in modeling. This review describes these recent results and outlines the physical basis for the new generation of computational models, the origins of improvement in modeling, and current challenges and bottlenecks.

Prediction Versus Design

Prediction and design are inverse problems: The prediction problem is to find the lowest energy structure for a specified sequence, and the design problem, to find the lowest energy sequence for a specified structure (Fig. 1A). Success in both efforts requires development of an accurate potential function—a quantitative model of the energetics of macromolecular interactions. Evaluation and improvement of the model can be spurred by using the same potential for both problems because the scope of the applicable tests is thereby increased considerably. The commonalities extend to the optimization methods (Fig. 1B). We have taken advantage of this in the development of the ROSETTA software package (used in the prediction and design examples from our laboratory described here), which uses essentially the same protein representation, potential function, and optimization methodology for prediction and design (*7*).

Energy Function

A large collection of experimental data on the effects of point mutations on protein stability (*8*–*11*) has highlighted the critical contribution of tight complementary packing in the core to protein stability. This is likely to derive both from attractive van der Waals interactions among protein atoms and from the dependence of the solvation free energy on the size of the cavity occupied by the protein. Experimental data have also highlighted the contribution of hydrogen bonding and hydrophobic/polar partitioning to protein stability. Alterations in the charge of surface side chains generally have little effect on stability, suggesting that long-range electrostatic interactions are substantially screened by both dynamic and static induced polarization effects.

In light of these data, successful approaches have focused on packing interactions, hydrogen bonding, and solvation effects represented with implicit solvation models (*12*, *13*) that favor burial of nonpolar atoms and exposure of polar atoms. These approaches have borrowed much from the molecular mechanics force fields used to simulate dynamics (*14*–*16*), notably the classical description in which energies are computed as sums over interactions between a relatively small set of different atom types, the use of a Lennard Jones potential to describe van der Waals interactions between atoms, and parameters for ideal bond lengths and angles. There are also important differences. Whereas the parameterization of molecular mechanics force fields relies primarily on experimental data on small molecules, the new force fields also derive parameters from experimental structural and thermodynamic data on proteins. In contrast to most molecular mechanics force fields, which represent hydrogen bonding as a dipole-dipole interaction, the orientation dependence that arises from the partially covalent character of the hydrogen bond is treated explicitly (*17*), both on the basis of geometrical distributions observed in proteins and quantum chemistry calculations on simple model systems (*18*). Longer range electrostatic interactions are generally dampened considerably. Torsional potentials, which are notoriously difficult to determine in molecular mechanics force fields, are obtained by directly inverting probability distributions from protein structures, and the representation of the protein chain is much stiffer—bond lengths and angles are generally kept rigid, and side-chain conformations are restricted to the vicinity of the rotameric states observed in protein structures. The stiffer representation reduces the frequency of false attractors by considerably reducing the size of configurational space. These differences have been driven by the con-

[1]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. [2]Department of Molecular Genetics and Biotechnology, Hebrew University, Hadassah Medical School, Jerusalem 91120, Israel. [3]Howard Hughes Medical Institute, Seattle, WA 98195, USA.

*To whom correspondence should be addressed. E-mail: dabaker@u.washington.edu

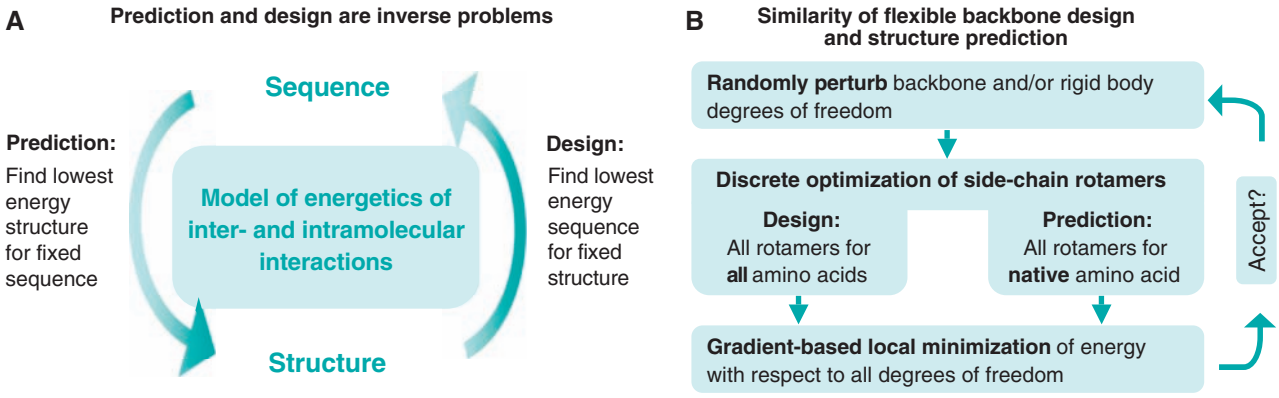

Fig. 1. Prediction and design. (**A**) Structure prediction and fixed backbone design are inverse problems. Completing the cycle corresponds to flexible backbone design, which requires optimization of both sequence and structure. (**B**) Algorithmic similarity of structure prediction, protein-protein docking, and flexible backbone design illustrated by the Monte Carlo minimization (MCM) high-resolution refinement protocol.

stant rigorous testing of the force field and representation by prediction and design calculations, which have the advantage that they can fail quite dramatically and highlight shortcomings in the approach.

Conformational Searching

A good energy function is not enough; a formidable challenge to prediction and design of protein structures and interactions is the very large size of the spaces that must be searched. For protein structure prediction, for example, with as few as three possible states per residue, the number of states of a 100-residue chain is astronomical [this is the often referred to "Levinthal's paradox" (*19*)]. Protein-protein docking requires a search over possible rigid body orientations of the partners, and most current design problems involve a search over protein backbone conformations, as well as amino acid sequences.

An effective approach to conformational sampling is to start with low resolution and go to high resolution. The low-resolution step searches for minima in an energy landscape dominated by hydrophobicity (the burial of nonpolar groups away from solvent), with the sharply varying van der Waals interactions smoothed by spatial averaging. Because sterics (tight complementary packing) is a critical contributor to the specificity of native protein structures and interactions, the native minima cannot be identified reliably with this averaged out representation, and the goal in this step is not to uniquely identify the native state, but to identify a set of energy minima that is almost certain to include the native conformation. Low-resolution approaches to docking, prediction, and design are further described in the next sections.

The second, more computationally intensive step in both the prediction and design of structure and interactions is the search for well-packed low-energy structures in the vicinity of each of the minima of the averaged-out landscape identified in the initial low-resolution step. In this step, all atoms are represented explicitly and steric interactions are not damped or averaged out. In this very rugged landscape, 1 Å deviations of atoms from the native structure can produce overlaps and huge spikes in the energy. This landscape is difficult to explore, but the native minimum is generally substantially deeper than non-native minima (which is not generally the case at the low-resolution stage).

Because the landscape is so rugged, optimization is challenging. A number of methods have been described to address this problem (*20*–*22*). The method used in ROSETTA couples Monte Carlo minimization (MCM) with discrete side-chain optimization (Fig. 1B). For each attempted move, an initial random perturbation of the backbone torsion angles (protein structure prediction) or rigid body degrees of freedom (protein-protein docking) is followed by discrete optimization of side-chain rotamer conformations and then by gradient-based local minimization on all degrees of freedom. MCM, which effectively flattens all barriers to the height of the nearest local minimum, has been found to be a powerful global optimization method for a broad range of problems (*23*, *24*).

Protein-Protein Docking and CAPRI

Rigid backbone protein docking is less challenging in terms of conformational searching than structure prediction or design because there are fewer degrees of freedom to be sampled. The low-resolution search can be performed using an elegant fast Fourier transformation (FFT)–based approach (*25*) or by real space MC. For high-resolution refinement, it is advantageous to simultaneously optimize both side-chain and rigid body degrees of freedom; MCM-based methods such as that outlined above (Fig. 1B) and that described in (*24*) have proven particularly effective.

Accurate predictions by several groups who entered CAPRI (*5*, *26*) indicate that high-resolution modeling methods are beginning to work for protein docking that does not involve pronounced backbone conformational changes. As an example we show results for two CAPRI targets, for which we carried out the MCM with side-chain flexibility protocol many independent times using different random-number seeds. The resulting energy landscape for target 12 (cohesin-dockerin) is shown in Fig. 2A. Three complexes have considerably lower energies than the others (Fig. 2A, inset), and trajectories starting from these low-energy complexes reveal that they lie at the bottom of a fairly narrow energy funnel (Fig. 2A, main panel).

The lowest energy predicted structures for the cohesin-dockerin complex, as well as for CAPRI target 15: the colicin-immunity protein complex, are shown in Fig. 3, A and B, superimposed on the experimentally determined crystal structures, which were released after the predictions were submitted to CAPRI. Not only the rigid body orientation, but also the conformations of almost all of the side chains, are predicted correctly. Other groups achieved similar successes (*5*, *26*). Current research is directed at incorporating backbone flexibility into protein-protein docking [e.g., (*27*)], which is closely related to the high-resolution protein structure prediction problem.

Structure Prediction and CASP

Many creative approaches to the de novo prediction of protein structure at low resolution have been described (*28*–*32*). The approach in ROSETTA to de novo structure prediction seeks to recapitulate the trade-off between local and nonlocal interactions during protein folding, by allowing short segments of the chain to flicker between alternative low-energy local conformations while searching for the lowest energy overall conformation of the chain (*33*). The search space is confined to that defined by the local conformational preferences of the protein sequence, and energy minima are identified using MC sampling with the low-resolution representation of the chain described earlier. Plausible candidate structures with primarily

hydrophobic cores and paired β strands can be rapidly generated (~1 min on 1 CPU for a 100-residue protein), but because atomic detail is neglected, the accuracy is generally low and any individual model is likely to be globally incorrect. Conformational sampling at this stage can be improved by generating structures not only for the protein of interest, but also for sequence homologs, which each have somewhat different low-resolution energy landscapes (*34*). High-resolution refinement of the low-resolution models is again carried out using the MCM with side-chain packing protocol described in Fig. 1B.

Progress in de novo structure prediction was highlighted at CASP6, where the first moderately high-resolution de novo structure prediction was made using the two-stage procedure protocol described above, with the initial low-resolution search followed by the flexible side-chain MCM optimization protocol. The root mean square deviation (RMSD) to the native structure after the low-resolution search was 2.2 Å and decreased to 1.6 Å during the flexible side-chain MCM refinement step (Fig. 3C). Although clearly at a lower level of accuracy than the docking predictions shown in Fig. 3, A and B, it is encouraging that some features of the native side-chain packing arrangement are correctly recapitulated. More recently, the same protocol was found to produce accurate predictions for a subset of small protein domains (*34*). The goal of current work on protein structure prediction is to consistently achieve the accuracy of the docking predictions in Fig. 3, A and B.

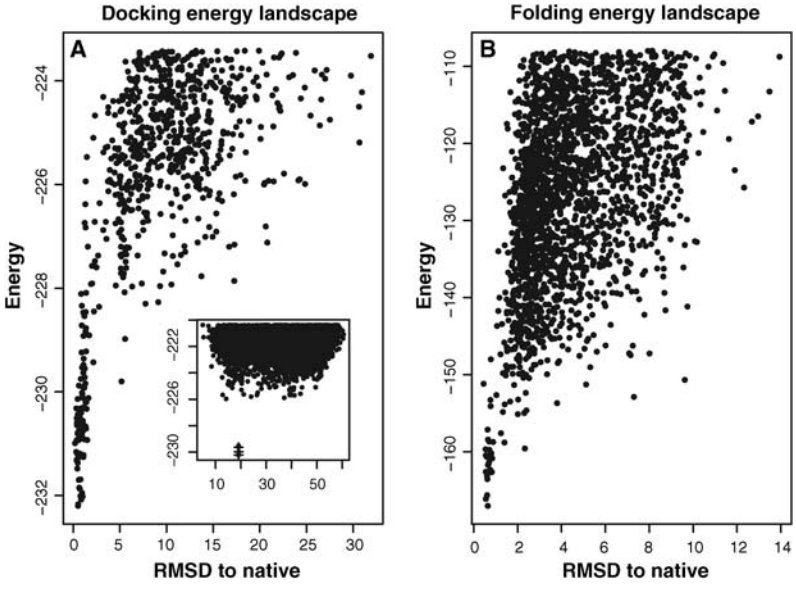

Fig. 2. Energy landscapes. Each point represents the lowest energy structure sampled in a single MCM trajectory. (**A**) Docking energy landscape for Capri Target 12 [cohesin-dockerin complex; PDB (Protein Data Bank) ID 1ohz (*66*)]. (**Inset**) In a large collection of trajectories starting from different random orientations that were carried out for the CAPRI experiment, a small number of structures (+) are distinguished from the background population by a significant energy gap. The *x* axis is the RMSD to an arbitrary reference orientation. (Main panel) Trajectories starting from these low-energy structures map out a narrow energy funnel. The *x* axis is the RMSD to the native structure. A deep energy funnel, as in this example, is a strong indicator that a prediction is correct. (**B**) Folding landscape for double-stranded RNA binding protein [PDB ID 1di2 (*67*)]. The backbone RMSD is to the native structure. The energy function (units are in kcal/mol) includes entropic contributions from solvation effects, but not the configurational entropy associated with protein vibrational and side-chain degrees of freedom, and hence is not the true free energy.

Examples of Protein Design

Protein design has a long history, starting from the realization that side-chain conformations in proteins could to a first approximation be treated as a set of discrete rotameric states and that new sequences and conformations could be derived by combinatorial optimization of this set (*35*). Mayo and co-workers showed that the lowest energy sequence computed for a small naturally occurring structure adopted a structure very close to that of the target starting structure (*36*). Harbury and co-workers (*37*) showed that new helical bundle structures could be created by designing sequences for a set of parametrically generated bundle arrangements. These studies correspond to the right arrow in Fig. 1A. Conceptually similar rotamer search-based design methods have been used to design new protein-protein interfaces, which have been confirmed by x-ray crystallography (*38*–*41*).

More recently, a globular protein fold was designed by alternating between sequence and structure optimization (the complete cycle in Fig. 1A) (*42*). The low-resolution and high-resolution optimization is similar to that used for protein structure prediction; the only notable differences are that in the low-resolution search constraints are added to favor the desired topology, and in the high-resolution search the discrete side-chain optimization is over the conformations of all 20 amino acids (Fig. 1B). The designed protein is exceptionally stable, with a free energy of folding roughly twice that of most naturally occurring proteins in its size range. The high-resolution crystal structure of the designed protein showed that its structure is similar (1.2 Å RMSD) to that of the computer-generated design model (Fig. 3D). The accuracy of the design and the high stability are both likely to stem from the fact that the sequence of the designed protein was optimized entirely for stability; in contrast to most naturally occurring proteins, it contains no regions that are locally suboptimal owing to functional constraints. Consistent with this, prediction of the structure of the designed protein from sequence results in a more accurate prediction than for almost all naturally occurring proteins (*43*).

Recent years have seen important milestones in the design of existing proteins with new functions. A series of small-molecule receptors have been designed that respond to specific ligands—a particularly spectacular achievement is a receptor that causes bacteria to turn green when exposed to TNT (*44*). As in the flexible side-chain docking protocol, these calculations coupled side-chain repacking with rigid body sampling. There has also been exciting progress in the design of new enzymes (*45*–*47*).

Role of High-Performance Computing

There has been a steady increase in CPU power and decrease in the cost of computing resources over the past 20 years, and this is likely to continue in the near future [Moore's law (*48*)]. It is of considerable sociological interest to identify the stage in this growth of computing power at which different scientific problems become tractable. The coupling of side-chain combinatorial optimization with backbone and/or rigid body optimization as in the protein-protein docking MCM procedure described above, the flexible backbone protein design protocol used to design TOP7, and binding site design for very large numbers of different ligand orientations by Hellinga and co-workers (*47*) were enabled by the increase in computer power; the simultaneous optimization of side-chain and backbone/rigid body degrees of freedom would not have been possible with the computing power available 15 years ago. The example in Fig. 2A shows that in 2005, only three solutions close to the native structure were found in 15 processor (3.2 GHz) days. Carrying out the calculation using a 1995 vintage processor (133 MHz) would have required roughly a year of processor time. Even today, the primary bottleneck to high-resolution structure prediction appears to be conformational sampling, because

the native structure almost always has lower energy than the predicted structures (*34*). The anticipated increase in processor power over the next several years will bring closer in reach the refinement of large protein structures and flexible backbone protein-protein docking, which requires simultaneous optimization of side-chain, backbone, and rigid body degrees of freedom.

Energy Landscapes

The successful predictions and designs described here suggest that the energy function underlying the calculations may be accurate enough to provide insights into the general properties of the folding and docking free-energy landscapes. Figure 2 shows the energy landscapes for both the docking and structure prediction problems. Notable features include the pronounced minimum in the vicinity of the native structure, and the sharp increase in energy associated with ~2 Å deviations from the native structure. It is important to note that configurational entropy is not accounted for in these landscapes, so the actual free energy landscapes will have somewhat broader minima. The folding and binding "funnels" (*49*) that have been the subject of much discussion are evident, but only in the immediate vicinity of the global minimum. Physically, the short range reflects the critical contribution to the energy from close complementary side-chain packing: Once the backbone coordinates have diverged by more than ~2 Å, the native side-chain packing arrangement is completely disrupted (*50*) and the energy increases substantially. The narrow aperture to the binding and folding funnels may be a characteristic feature of biomolecular recognition and folding.

Is the short-range nature of landscapes compatible with the observation that complexes form and proteins fold in finite time? In the case of complexes, it has been possible to rigorously answer this question in the affirmative by solving the steady-state diffusion equation for interacting partners with "reactive zones" corresponding to the aperture of the binding funnels—the computed rates are on the order of 10^6 per second (*51*)—consistent with experimentally observed rates for many proteins (faster association rates are likely to reflect long-range electrostatic steering). In the case of monomeric folding, folding can still proceed rapidly if native interactions are on average lower in energy than non-native interactions—the principle of minimal frustration (*49*, *52*)—and indeed for most compact subdomains the native conformation has a lower computed energy than that of essentially all sampled non-native conformations. This is expected given the dominant contribution of short-range packing and hydrogen bonding interactions—a conformation that is very low in energy on a global scale must also be low in energy on a more local scale, just as the optimal solution for a jigsaw puzzle is also optimal (perfectly packed) on a local scale. Indeed, broader energy funnels are observed when the reaction coordinate (*x* axis in Fig. 2) is the fraction of native contacts, rather than the RMSD, consistent with protein folding landscape theory.

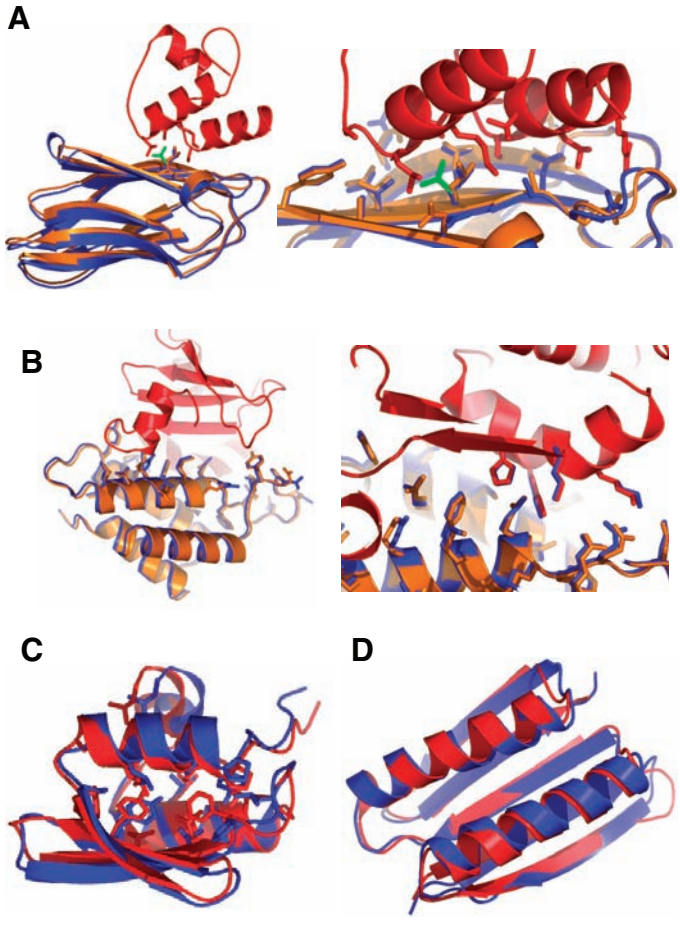

Fig. 3. Examples of high-resolution prediction and design. (**A**) CAPRI Target 12 [dockerin–cohesin (*66*); interface residue backbone RMSD = 0.27 Å]. The lowest energy structure in Fig. 2A, main panel, is shown here. The side chain of Leu-83 (green in the free monomer) changes conformation upon binding. Side-chain conformations in red were provided; those in blue were predicted. (**B**) CAPRI Target 15 [ColicinD–Immunity protein D (*68*); interface residue backbone RMSD = 0.23 Å]. No side-chain information was provided for either partner. (**C**) CASP6 de novo structure prediction Target 0281 [hypothetical protein from *Thermus thermophilus* Hb8, PDB ID 1whz (*69*); backbone RMSD = 1.59 Å]. (**D**) TOP7 (RMSD = 1.2 Å) (*42*). (A) and (B) are adapted from figure 1 of (*70*). Blue: models; red and orange: x-ray structures.

The narrow energy funnels around native structures, and the close correspondence of both prediction and designs with experimentally determined structures, have direct bearing on the fundamental question of the range of structures adopted by proteins in solution. Some simulation studies have suggested that in solution, proteins populate a broad range of conformations up to 4 Å RMSD from the crystal structure. In contrast, the short range of the folding and binding funnels suggests that protein cores are confined to within ~1.5 Å of the experimentally determined crystal structures; the substantial increases in energy accompanying larger deviations suggest that such structures are populated at relatively low levels. The coincidence of the energy minima in the prediction and design calculations with the location of the experimentally determined structure further supports this conclusion: If crystal packing interactions randomly selected out a member of a broad ensemble of structures spanning 3 to 4 Å RMSD, one would not expect RMSDs of much less than this between predicted structures and crystal structures or between design models and crystal structures (*53*). Experimental support for this conclusion comes from comparison of protein structures determined in different crystal forms, which suggests backbone variation of less than 1 Å (*54*, *55*), and from recent NMR studies (*56*) that show that agreement between the dipolar couplings and structure is not limited by the intrinsic dynamic behavior of proteins.

Current Challenges

Considerable challenges remain for high-resolution modeling—this review should not be taken to suggest that the critical problems are by any means solved, but rather that accurate modeling now appears to be an achievable goal. Shortcomings in potential functions include the treatment of buried polar interactions, which are complicated by polarization effects, and the delicate balance between the cost of desolvating a polar or charged group and the favorable hydrogen bonding and electrostatic interactions the group may make upon burial. Consistent prediction and design of polar active sites will require progress in modeling such interactions (*57*). On the sampling side, improvements in methodology and increases in computing capabil-

ities will be necessary for problems with large numbers of degrees of freedom, such as prediction of the structures of large proteins.

To conclude, we list a series of challenges that will spur the next phase of method development. (i) High-resolution refinement of models built by comparative modeling. It is well established that the accuracy of comparative models built by copying a template structure identified using sequence comparison methods decreases steadily with increasing sequence divergence between the sequence being modeled and the protein template owing to structural divergence during protein evolution (*58*, *59*). Recent CASP tests have shown that the best comparative models are currently built by experts such as Ginalski (*60*) who are able to achieve near-perfect alignments of the sequences to proteins of known structure, a critical precondition for further modeling, and go further to improve models using an assortment of protein modeling tools such as MODELLER (*61*), PSIBLAST (*62*), ROSETTA (*7*), PSIPRED (*63*), SQWRL (*64*), and VERIFY3D (*65*) together with visual structural inspection. The challenge is to automatically produce even more accurate structures by high-resolution refinement; the loss of human intuition will have to be compensated by generating starting models for refinement by large-scale sampling of alternative alignments on the basis of alternative homologous structure templates. The methods described in this review are directly applicable to this problem, and it is also a good test of widely used molecular dynamics simulations methods because the starting template can be within 3 Å of the correct structure. (ii) Consistent prediction of the structures of small proteins at atomic-level resolution. (iii) Protein docking with backbone flexibility. This is a formidable challenge because both rigid body and internal backbone degrees of freedom need to be searched. (iv) Prediction and design of the specificity of protein–nucleic acid interactions. (v) Design of new enzymes catalyzing reactions not catalyzed by naturally occurring enzymes. (vi) Prediction of the structures of multidomain and multisubunit protein complexes. We look forward to progress in all of these areas.

References and Notes

1. C. B. Anfinsen, *Science* **181**, 223 (1973).
2. R. Schleif, *Methods Enzymol.* **383**, 28 (2004).
3. M. Karplus, J. A. McCammon, *Nat. Struct. Biol.* **9**, 646 (2002).
4. J. Moult, *Curr. Opin. Struct. Biol.* **15**, 285 (2005).
5. J. Janin, *Protein Sci.* **14**, 278 (2005).
6. J. Janin *et al.*, *Proteins* **52**, 2 (2003).
7. C. A. Rohl, C. E. Strauss, K. M. Misura, D. Baker, *Methods Enzymol.* **383**, 66 (2004).
8. K. A. Bava, M. M. Gromiha, H. Uedaira, K. Kitajima, A. Sarai, *Nucleic Acids Res.* **32**, D120 (2004).
9. B. W. Matthews, *Adv. Protein Chem.* **46**, 249 (1995).
10. J. U. Bowie, J. F. Reidhaar-Olson, W. A. Lim, R. T. Sauer, *Science* **247**, 1306 (1990).
11. J. T. Kellis Jr., K. Nyberg, D. Sali, A. R. Fersht, *Nature* **333**, 784 (1988).
12. T. Lazaridis, M. Karplus, *Proteins* **35**, 133 (1999).
13. M. Feig, C. L. Brooks III, *Curr. Opin. Struct. Biol.* **14**, 217 (2004).
14. A. D. Mackerell Jr., *J. Comput. Chem.* **25**, 1584 (2004).
15. B. Brooks *et al.*, *J. Comput. Chem.* **4**, 187 (1983).
16. W. Cornell *et al.*, *J. Am. Chem. Soc.* **117**, 5179 (1995).
17. D. B. Gordon, S. A. Marshall, S. L. Mayo, *Curr. Opin. Struct. Biol.* **9**, 509 (1999).
18. A. V. Morozov, T. Kortemme, K. Tsemekhman, D. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6946 (2004).
19. C. Levinthal, in *Mossbauer Spectroscopy in Biological Systems: Proceedings of a Meeting Held at Allerton House, Monticello, Illinois*, J. T. P. DeBrunner, E. Munck, Eds. (Univ. of Illinois Press, Urbana, 1969), pp. 22–24.
20. U. H. Hansmann, Y. Okamoto, *Curr. Opin. Struct. Biol.* **9**, 177 (1999).
21. K. Tai, *Biophys. Chem.* **107**, 213 (2004).
22. M. P. Jacobson *et al.*, *Proteins* **55**, 351 (2004).
23. Z. Li, H. A. Scheraga, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6611 (1987).
24. R. Abagyan, M. Totrov, *J. Mol. Biol.* **235**, 983 (1994).
25. E. Katchalski-Katzir *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2195 (1992).
26. R. Mendez, R. Leplae, M. F. Lensink, S. J. Wodak, *Proteins* **60**, 150 (2005).
27. C. Dominguez, R. Boelens, A. M. Bonvin, *J. Am. Chem. Soc.* **125**, 1731 (2003).
28. J. U. Bowie, D. Eisenberg, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4436 (1994).
29. B. Park, M. Levitt, *J. Mol. Biol.* **258**, 367 (1996).
30. D. T. Jones, L. J. McGuffin, *Proteins* **53** (suppl. 6), 480 (2003).
31. J. Skolnick *et al.*, *Proteins* **53** (suppl. 6), 469 (2003).
32. Q. Fang, D. Shortle, *Proteins* **53** (suppl. 6), 486 (2003).
33. K. T. Simons, C. Kooperberg, E. Huang, D. Baker, *J. Mol. Biol.* **268**, 209 (1997).
34. P. Bradley, K. M. Misura, D. Baker, *Science* **309**, 1868 (2005).
35. J. W. Ponder, F. M. Richards, *J. Mol. Biol.* **193**, 775 (1987).
36. B. I. Dahiyat, S. L. Mayo, *Science* **278**, 82 (1997).
37. P. B. Harbury, J. J. Plecs, B. Tidor, T. Alber, P. S. Kim, *Science* **282**, 1462 (1998).
38. T. Kortemme, D. Baker, *Curr. Opin. Chem. Biol.* **8**, 91 (2004).
39. B. S. Chevalier *et al.*, *Mol. Cell* **10**, 895 (2002).
40. T. Kortemme *et al.*, *Nat. Struct. Mol. Biol.* **11**, 371 (2004).
41. J. M. Shifman, S. L. Mayo, *J. Mol. Biol.* **323**, 417 (2002).
42. B. Kuhlman *et al.*, *Science* **302**, 1364 (2003).
43. M. von Grotthuss, L. S. Wyrwicz, J. Pas, L. Rychlewski, *Science* **304**, 1597 (2004).
44. L. L. Looger, M. A. Dwyer, J. J. Smith, H. W. Hellinga, *Nature* **423**, 185 (2003).
45. D. N. Bolon, S. L. Mayo, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14274 (2001).
46. J. Kaplan, W. F. DeGrado, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11566 (2004).
47. M. A. Dwyer, L. L. Looger, H. W. Hellinga, *Science* **304**, 1967 (2004).
48. G. E. Moore, *Electronics* **38**, 114 (1965).
49. J. N. Onuchic, P. G. Wolynes, *Curr. Opin. Struct. Biol.* **14**, 70 (2004).
50. S. Y. Chung, S. Subbiah, *Pac. Symp. Biocomput.*, 126 (1996).
51. M. Schlosshauer, D. Baker, *Protein Sci.* **13**, 1660 (2004).
52. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, *Proteins* **21**, 167 (1995).
53. The similarity of predictions and designs with crystal structures is not likely to be an artifact of the use of force fields in structure determination: X-ray structure determination at high resolution is almost entirely driven by experimental diffraction data, and the force fields typically used in refinement differ from those used in the prediction and design calculations.
54. X. J. Zhang, J. A. Wozniak, B. W. Matthews, *J. Mol. Biol.* **250**, 527 (1995).
55. E. Eyal, S. Gerzon, V. Potapov, M. Edelman, V. Sobolev, *J. Mol. Biol.* **351**, 431 (2005).
56. T. S. Ulmer, B. E. Ramirez, F. Delaglio, A. Bax, *J. Am. Chem. Soc.* **125**, 9179 (2003).
57. J. W. Ponder, D. A. Case, *Adv. Protein Chem.* **66**, 27 (2003).
58. A. Tramontano, V. Morea, *Proteins* **53** (suppl. 6), 352 (2003).
59. G. Wang, R. L. Dunbrack Jr., *Protein Sci.* **13**, 1612 (2004).
60. K. Ginalski, *CASP6 Abstract Book* (http://predictioncenter.org/casp6/abstracts/abstract.html), 64 (2004).
61. A. Fiser, A. Sali, *Methods Enzymol.* **374**, 461 (2003).
62. S. F. Altschul *et al.*, *Nucleic Acids Res.* **25**, 3389 (1997).
63. D. T. Jones, *J. Mol. Biol.* **292**, 195 (1999).
64. A. A. Canutescu, A. A. Shelenkov, R. L. Dunbrack Jr., *Protein Sci.* **12**, 2001 (2003).
65. D. Eisenberg, R. Luthy, J. U. Bowie, *Methods Enzymol.* **277**, 396 (1997).
66. A. L. Carvalho *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13809 (2003).
67. J. M. Ryter, S. C. Schultz, *EMBO J.* **17**, 7505 (1998).
68. M. Graille, L. Mora, R. H. Buckingham, H. Van Tilbeurgh, M. De Zamaroczy, *EMBO J.* **23**, 1474 (2004).
69. M. Kanagawa, S. Yokoyama, S. Kuramitsu, unpublished data.
70. O. Schueler-Furman, C. Wang, D. Baker, *Proteins* **60**, 187 (2005).
71. The software used in the prediction and design examples described above is available free for academic use in the ROSETTA software package (http://depts.washington.edu/ventures/UW_Technology/Express_Licenses/Rosetta/), and personal computers can be enlisted in the conformational sampling essential to the prediction effort at http://boinc.bakerlab.org/rosetta/.

10.1126/science.1112160

Progress in Modeling of Protein Structures and Interactions

Ora Schueler-Furman *et al.*

Science **310**, 638 (2005);
DOI: 10.1126/science.1112160