Theory and simulation

Editorial overview

Julia M Goodfellow* and Ronald M Levy†

Addresses
*Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1 7HX, UK; e-mail: j.goodfellow@bbk.ac.uk
†Department of Chemistry, Rutgers University, PO Box 939, Piscataway, NJ 08855-0939, USA; e-mail: ronlevy@lutece.rutgers.edu

Current Opinion in Structural Biology 2000, **10**:137–138

0959-440X/00/$ – see front matter

This year's section on Theory and simulation begins with two reviews of *ab initio* protein structure prediction. The results of the third Critical assessment of methods of protein structure prediction (CASP3) meeting provide the background for the reviews by Lazaridis and Karplus (pp 139–145) and by Osguthorpe (pp 146–152). As commented on in both reviews, a remarkable and promising outcome of CASP3 is that *ab initio* methods of protein structure prediction proved to be competitive with fold recognition methods for targets of intermediate difficulty. Most of the current *ab initio* methods rely on statistical potentials; in contrast, the use of physics-based potentials is less well developed. Both reviews compare the relative strengths and weaknesses of these two approaches. Although reduced-atom models have traditionally relied on statistical potentials, Osguthorpe describes the development of a simplified off-lattice model based on a physical potential. Lazaridis and Karplus summarize recent developments in scoring models based on physical effective energy functions. These models treat the protein in nearly full atomic detail, while the solvent is treated implicitly. A variety of fast implicit solvent models that approximate both the electrostatic reaction field and the nonpolar/hydrophobic effects of the solvent are being tested on the 'decoy' problem — discriminating correctly folded proteins from cleverly misfolded 'decoys'. As Osguthorpe points out, the enormous increase in computer power, culminating in IBM's announcement of their intention to build a petaflop (10^{15}) computer (the "Blue Gene" project), helps to insure that continued advances in *ab initio* protein folding will depend less on computer cycles and more on the design of appropriate models. These models must contain sufficient detail to capture the essential physical features that are responsible for protein stability, but sufficient simplicity so that meaningful parameters can be found through a combination of statistical and physical approaches.

The paper by Sheinerman, Norel and Honig (pp 153–159) reviews recent structural and theoretical studies of electrostatic aspects of protein–protein interactions, from both thermodynamic and kinetic perspectives. The electrostatic models employed are based on either the Poisson–Boltzmann framework or even simpler continuum solvent approximations. Uncertainties concerning the modeling of conformational changes and the related problem of choosing dielectric constants continue to limit the accuracy of these calculations; qualitative trends are of greater significance. Several studies highlight the overall destabilizing effect of electrostatic interactions as a result of the desolvation penalty paid by polar and charged groups at the buried protein–protein interface; however, individual ion pairs have been found that enhance association. The recent demonstration that the extent of desolvation is less in the encounter complex between associating protein pairs than in the final structure accounts for the puzzling observation that electrostatics can enhance association rates while destabilizing the final complex.

Two of the reviews in this section concern the use of molecular dynamics simulations of proteins to analyze long timescale motions (Daggett [pp 160–164]) and collective modes of protein motion (Berendsen and Hayward [165–169]). Increased computer power is allowing us to follow trajectories on longer and longer timescales. Although we saw the publication of a microsecond simulation of a 36-residue protein in 1998, there have only been a limited number of simulations longer than 5 ns reported in 1999. These have led to insights into folding intermediates of α-helical bundles and into the role of correlated motions in the catalytic activity of dihydrofolate reductase. The *tour de force* remains the 3.5 μs simulation of the villin headpiece, which indicated that hydrophobic collapse took place on a 60 ns timescale. The problem of extracting statistically meaningful information from rare events observed in single trajectories is formidable. Furthermore, microsecond protein simulations are not likely to become routine for some years and we are going to be dependent on multiple shorter simulations for developing biological insight. Berendsen and Hayward consider the problem of developing more insightful ways to analyze the complex motions of atoms extracted from molecular dynamics trajectories. In order to relate function to structural flexibility, the key is to separate the large collective motions from the multitude of small and (presumably) less important ones. This can be accomplished by normal model or principal component analysis of the motions, or by the analysis of the 'essential dynamics', which focuses on the backbone atoms and assumes that the key motions of functional importance are to be found in the essential subspace for the first few eigenvectors. Such analyses have been applied to a number of proteins in order to illuminate dynamical behavior during protein folding and ligand binding, and its role in substrate specificity. Also, the review by Phelps, Speelman and Post (pp 170–173) reports

recent progress that has been made in examining the macromolecular assembly and structural stability of viral capsids containing 60 to 180 constituent protein copies using molecular dynamics simulations and kinetic models. During the past year, the prediction of an entropic basis for the antiviral activity of hydrophobic compounds, first recognized from compressibility values calculated from trajectories of human rhinovirus, has been corroborated by experimental measurements on poliovirus.

An area of considerable interest is the behavior of membranes and their associated proteins and peptides. Forrest and Sansom's review (pp 174–181) covers the increase in the accuracy and size of simulations of these systems, for which there are still limited experimental structural data compared with that available for aqueous proteins. Improved potentials, long-range electrostatic (particle mesh Ewald; PME) models, different ensembles and longer timescales have all been applied to membrane systems alone and there has been some success in modeling mixed lipid systems. Transmembrane helices continue to be a major focus of computational studies and these have provided some evidence for the importance of conformational flexibility, which may be associated with conformational switches involved in signaling across membranes. Larger systems have also been studied, including model helical bundles in octane slabs or phospholipid bilayers, as well as a trimer of the β-barrel membrane protein comprising the bacterial porin OmpF. However, simpler systems, such as peptides associated with the surface of membranes, continue to offer methodological challenges to current computational approaches.

Beveridge and McConnell (pp 182–196) describe the results of all-atom simulations of DNA and RNA, with an emphasis on studies that serve to document simulation protocols and validate their current level of accuracy. Their accuracy has been greatly advanced by the widespread incorporation into biomolecular simulation packages of methods for modeling long-range coulomb interactions that avoid truncation in these highly charged systems — including PME and periodic fast multipole methods (PFMM). The AMBER force-field has been the most extensively tested on nucleic acid simulations using the PME method. The use of PME leads to very good structural stability; for example, good agreement with 2D NOESY experiments is observed in long (14 ns) simulations of the *Eco*RI dodecamer. Molecular dynamics studies of DNA can now be described as 'second generation', with much improved results obtained over the past three years. The authors suggest that the time is ripe for nucleic acid modelers to develop a community wide approach to selecting prototype systems and protocols for testing force-fields, in order to most efficiently build upon the advances in the predictive power of these tools that have occurred in recent years.

The final review in the Theory and simulation section, that by Case (pp 197–203), reviews progress in constructing a theoretical framework for the structural interpretation of NMR chemical shifts and coupling constants in biological macromolecules. During the past decade, databases of chemical shifts and coupling constants in proteins and nucleic acids of known structure have been assembled with the goal of developing empirical rules for their interpretation. At the same time, advances in quantum chemical theory have progressed to the point at which calculations of shifts and coupling constants now produce results of sufficient accuracy to facilitate structure determination. There have been many highlights in this field during the past two years: the analysis of long-range (up to 25 Å) protein structural information using chemical shift anisotropy/dipolar cross-correlated relaxation data; the extraction of structural information from partially ordered protein solutions on the basis of the measured residual dipolar and indirect spin–spin couplings; and the use of density functional calculations to map out the expected distance- and angle-dependence of spin–spin couplings across hydrogen bonds in peptides, proteins and nucleic acids.

In summary, increases in computer power have enabled modelers to increase the length and sophistication of their simulations. Considerable efforts have been focused on the critical assessment of current techniques in order to elucidate their strengths and weaknesses. However, the success of future computational approaches will depend as much on the design of the original model and the appropriateness of the analysis as on the raw computer power brought to bear on the problem.