ELSEVIER

Theory and simulation

Editorial overview

Richard Lavery and Kim A Sharp

Current Opinion in Structural Biology 2007, **17**:147–148

Available online 26th March 2007

0959-440X/$ – see front matter
Published by Elsevier Ltd.

DOI 10.1016/j.sbi.2007.03.007

Richard Lavery

Institut de Biologie et Chimie des Protéines, UMR 5086 CNRS, Université de Lyon, 7 passage du Vercors, Lyon 69367, France
e-mail: richard.lavery@ibpc.fr

Kim A Sharp

ER Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104-6059, USA
e-mail: sharpk@mail.med.upenn.edu

''Connect, always connect''. Einstein's well-known quote aptly describes the aim of good science. Connecting, or bridging, as the titles of two reviews explicitly indicate, also serves as a theme for the reviews in this section. One may connect or bridge between descriptions with different resolutions. Thus, Carloni and co-workers review work that links first-principles quantum mechanics with molecular dynamics through the Car–Parrinello method, whereas Shapiro *et al.* discuss recent work bridging the gap between secondary and tertiary (or 3D) levels of RNA structure prediction. At a still higher level of complexity, Stein, Gabdoulline and Wade review work that bridges between molecular structures and biochemical networks.

One may also bridge between different properties of a system at the same level. Examples include structure/function relationships. In this area, Sproviero *et al.* review such a relationship in the specific case of the oxygen-evolving complex of photosystem II. This is an important and well-studied component of bioenergetic pathways, for which structural information must be supplemented with quantum mechanical calculations for a full understanding of the mechanism. Connecting between the Hamiltonian, or energy surface, and thermodynamic or kinetic quantities, and between single-molecule and ensemble properties, is the theme of the two reviews by Meirovitch, and by Lei and Duan. Connecting short and long timescales, and small and large conformational samples is another facet of the work surveyed in these two reviews, and the reviews of Carloni and co-workers, and Ayton, Noid and Voth.

Boas and Harbury review recent progress in developing energy potentials for protein design. They highlight the current unsatisfactory situation in which different potentials must be used for different protein applications. Molecular dynamics, structure prediction and design all use quite different energy functions. The 'connection' between them is the common underlying physics. Differences between potentials reflect what is left out, what is modeled explicitly, and what is modeled implicitly and how it is parameterized. Hopefully, in the not too distant future, there will be a single potential accurate enough to handle all three applications. Boas and Harbury review how molecular mechanics potentials, the most detailed and physics based of the potentials, are being used to advance protein design potentials, a step on the road to this possible unification.

In their review of recent advances in Car–Parrinello molecular dynamics simulations, Carloni and co-workers discuss how advances in methodology and computational power are now enabling, for the first time, the method to be applied to biological systems, although only part of the system (~200 atoms) can be treated quantum mechanically. The emphasis of applications to structural biology has been on treatment of metals. First,

the authors point out that nearly half of known proteins contain metals. Second, quantum treatments are crucial for accurately treating the significant polarization of electron orbitals of and by metals. They also discuss two limitations of the method that are being, or should be, addressed in the near future: inclusion of dispersion forces and generalization beyond plane wave basis sets. The former is required for increased accuracy, whereas the latter will greatly increase the applicability of the method.

Meirovitch, and Lei and Duan both review work that addresses the important problem of sampling, but from different perspectives. Biological systems have a large number of degrees of freedom (DOFs), so adequate sampling is a major challenge. Meirovitch reviews this from the perspective of the calculation of thermodynamic properties, specifically, the calculation of entropy and, by extension, free energy. Lei and Duan focus on sampling techniques for protein folding, kinetics and characterization of protein energy landscapes, including coarse-grained or reduced DOF models. Lei and Duan also point out that many of the most innovative and exciting methods are still in the 'model application' stage, and that next they must be applied to more biological cases. Meirovitch identifies two important issues that need to be addressed in entropy calculations. First, current entropy calculation methods often lack internal estimates of their precision. This is an issue when trying to judge convergence or to compare methods. Second, most methods treat the entropy of systems localized in phase space, for example, a folded, stable protein fluctuating around its native state. Extension of these methods, or the development of new ones, to treat diffusive systems remains a challenge.

Shapiro *et al.* review recent progress in RNA structure prediction, focusing on a particular RNA motif — the pseudoknot — as this is currently the most mature area of structure prediction. The two general strategies are to use multiple sequence alignments or some kind of free energy minimization. Some approaches use a combination of the two. Interestingly, these two strategies recapitulate approaches developed for protein structure prediction, namely homology modeling and *ab initio* protein folding. In a sense, these strategies bracket the spectrum of potential methods, in that sequence alignment is purely knowledge based, whereas *ab initio* folding is purely physics based. It will be interesting to see if RNA equivalents of protein structure prediction methods that combine physics and knowledge-based methods, such as threading, or fragment-based modeling, will prove as powerful as they have for proteins. However, there is no protein equivalent of the key structural and energetic role that Mg^{2+} plays in RNA structure. This remains a major and unique challenge for RNA structure prediction methods. Applying the methods beyond the pseudoknot motif is another clear challenge outlined by Shapiro *et al.*

Multiscale modeling is a rapidly expanding area of simulation technology, driven by the complexity and computational demands of representing biological macromolecules and systems. Multiscale modeling by its nature must cogently connect two or more levels of representation, and therein lie the challenges. Scaling can occur in the time domain, the length domain or both. Many different approaches are being developed and applied, and Ayton, Noid and Voth provide a welcome classification and analysis of these. The principal distinction they make is between serial and parallel multiscale strategies. In the former, different scale representations are used in sequence, with no direct interaction. Interaction occurs typically through the introduction of high-level experimental data, for example, via parameterization. In parallel multiscale simulations, direct information transfer between the different scales occurs during the simulation. This is more challenging to implement rigorously, but potentially more accurate.

Bridging from molecular simulations to biochemical networks, as Stein, Gabdoulline and Wade nicely illustrate in their review, brings in many exciting and diverse theory and simulation tools, ranging from the use of cryo-electron microscopy density maps to refine models of multimolecular complexes, through the simulation of enzyme kinetic parameters using quantum, molecular or Brownian dynamics simulations, to incorporation of such kinetic parameters in simulations of entire biochemical networks. Simulation of entire networks has been tried on the ensemble or large-number level, using, for example, differential equations, or on the small-number, discrete molecule level, using stochastic probabilistic methods. At this point, few metabolic pathways have been characterized sufficiently to either prime or constrain this type of modeling, but this should improve over time. Clearly, multiple approaches are the order of the day in this area, which, perhaps more vividly than any covered in this section, demonstrates the important role that theory and simulation have in connecting different spatial, temporal and complexity scales, and in bridging from experimental data to deeper understanding.