Perspectives on: Molecular dynamics and computational methods

Benoit Roux, Guest Editor

Department of Biochemistry and Molecular Biology, Gordon Center for Integrative Sciences, University of Chicago, Chicago, IL 60637

The "Perspectives on: Molecular dynamics and computational methods" in this issue offers a measure of what can now be achieved using advanced computational methods to better understand biological and physiological systems at the atomic level. Protein complexes are molecular nano-machines, able to accomplish many different specific and complex tasks. Ultimately, one needs to be able to visualize how multiple proteins move and change their shape, atom-by-atom as a function of time, while they perform their functions. Atomic motions and conformational transitions governing function involve processes spanning several orders of magnitude in time and space, from femtoseconds to seconds and minutes, from sub-angstroms to millimeters. Identifying the relevant degrees of freedom to understand a physiological process is difficult because there can be strong coupling at all levels of description, starting from the electrons all the way up to the organism. In a naive sense, one would need a tunable virtual microscope, a little bit like Google Earth, linking the different levels of description in a seamless manner and allowing us to focus attention at the relevant lengthscale(s) and timescale(s) for the physiological process of interest.

Obviously, no single computational method is currently able to tackle this task by itself. The four contributions to these Perspectives were, in part, chosen to illustrate how computational methods operate at different levels of representation. Ab initio (first-principles) simulations, such as those presented by Bucher and Rothlisberger, explicitly treat the degrees of freedom from the nuclei and the electrons. Conceptually, it is assumed that the electrons are in their ground state. Thus, the nuclei evolve dynamically on the Born-Oppenheimer (BO) energy surface according to Newton's classical equation of motion, $F = mA$.

One approximation to reduce computer time consists of replacing the BO energy surface from quantum mechanics by an empirical molecular mechanical force field. The latter is a mathematical object constructed from simple analytical functions, which are parameterized and carefully adjusted to reproduce the BO potential energy surface and known experimental results. This model enables one to reduce computer time sufficiently to allow for the calculation of molecular dynamics (MD) trajectories of moderately large biomolecular systems for fairly long times. The simulations presented by Dror et al. show that this all-atom MD method can now be pushed to the point where one can generate multi-microsecond trajectories while treating explicitly all the atomic degrees of freedom of the protein, membrane, ions, and solvent.

In spite of its simplifications, however, the burden of all-atom MD is that the trajectory of a large number of atoms must be calculated, although one might be interested in only a small number among them (e.g., the permeating ions and the selectivity filter of the channel). An attractive approach to focus on the most relevant degrees of freedom is to adopt a "coarse-grained" (CG) representation, which is able to retain a meaningful description of structure and dynamics. The dream of representing the dynamics of a reduced set of degrees of freedom, which is part of a large complex system, has been extensively explored in statistical mechanical theories over the last 50 years. The general idea consists of developing effective dynamical schemes for representing the time evolution of the most "relevant" variables realistically by "projecting out" uninteresting variables. The subset of coordinates that are treated explicitly is expected to evolve according to some effective, nondeterministic (stochastic) dynamics, which incorporates the influence of the rest of the system implicitly. The simulations presented by Ivet Bahar show how important insight can be obtained from such CG models of very large macromolecular assemblies, while achieving a great reduction in the complexity.

The representation of a complex system can be further simplified by describing its dynamical evolution in terms of a small number of "states." The semi-macroscopic approach of Silva and Rudy provides one example of such a strategy. One introduces a discrete-state Markov chain in which the conformational transitions of protein elements are reported as sudden stochastic jumps among a small set of discrete states, making it possible to couple a large number of independent elements to simulate physiology processes.

Correspondence to Benoit Roux: roux@uchicago.edu

These four beautiful contributions summarize the impressive advances made in computational biology. Nonetheless, a few words of caution are in order. Computational models, at any level, are inherently approximate. At one extreme, it is possible to achieve extensive exploration in timescales and lengthscales by simplifying the models, albeit at the risk of representing microscopic interactions by approximations that are too crude. At the other extreme, if the important interactions need to be explored using quantum mechanical approaches, it may be computationally prohibitive to achieve a meaningful sampling of the relevant configurations arising from thermal fluctuations.

Striking a balance between accuracy and relevance in computational studies is challenging, but necessary to draw meaningful conclusions. Even ab initio simulations comprise several approximations. Electrons are negatively charged fermions, stubborn elementary particles that insist on an antisymmetric wavefunction. The density functional theories underlying the first-principle ab initio MD simulations are functional forms that account only approximately for exchange and correlation of the electrons. Likewise, extremely efficient MD such as performed by Dror et al. relies on a potential function that treats electrostatic energy as a sum of pairwise Coulomb interactions between fixed effective atomic charges. Such an approximation, which accounts for many-body–induced polarization effects only in an average way, is expected to be satisfactory for a wide range of systems, but could be less accurate for processes involving the transfer and ex-change of polar and charged moieties between heterogeneous environments. Much effort has been devoted to develop potential functions that will account explicitly for induced polarization, although most current computer simulations of biomolecular systems continue to be based on effective fixed-charge force fields. Lastly, constructing meaningful simplified models, whether they are CG representations of proteins or discrete state Markov models, relies on a corresponding range of approximations.

Clearly, our ability to treat complex biological and physiological problems will continue to be strengthened by a closer integration of different levels of representation. Although much work remains to be done to achieve a complete and seamless multi-scale view of biological systems, the four contributions of these Perspectives illustrate vividly the fascinating results that can be obtained by approaching biological problems at different computational levels.

Letters to the editor related to these Perspectives will be published in the September 2010 issue of the Journal. Letters to the editor should be received no later than Monday, July 19, 2010, to allow for editorial review. The letters may be no longer than two printed pages (approximately six double-spaced pages) and will be subject to editorial review. They may contain no more than one figure, no more than 15 references, and no significant references to unpublished work. Letters should be prepared according to the Journal's instructions and can be submitted electronically, or as an e-mail attachment to jgp@rockefeller.edu.