



HHS Public Access

Author manuscript

Curr Opin Struct Biol. Author manuscript; available in PMC 2019 April 01.

Published in final edited form as:

Curr Opin Struct Biol. 2018 April ; 49: iv–v. doi:10.1016/j.sbi.2018.04.002.

Editorial overview: Theory and simulation: Interpreting experimental data at the molecular level

Robert B. Best and

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA

Kresten Lindorff-Larsen

Structural Biology and NMR Laboratory & Linderstrøm-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, Denmark

Increases in computational power, as well as the development of new algorithms, have helped to expand the potential applications of theory and simulation to structural biology. From the simulation perspective, it is now possible with all-atom simulations to access longer time scales which are more relevant to biological processes, and these require new methods of analysis in order to obtain quantitative mechanistic information. These longer simulations, however, also reveal the need for more accurate simulation methodology and force fields. At a more coarse-grained scale, it is now possible to treat problems at ever larger length scales, such as ribosomal function, or the structure of chromatin; and at a higher resolution, more accurate QM/MM methods allow the details of enzyme mechanism to be elucidated. From an experimental perspective, lower computational costs permit an analysis of experimental data that embraces the diversity within the experimental ensemble. Thus, the broad conformational distributions of disordered proteins, or the heterogeneous orientations of particles in cryo-electron microscopy can be explicitly accounted for in refinement against experimental data. In this issue, we have tried to compile a set of topics that touches on work in each of these areas.

Boulanger and Harvey discuss QM/MM methods for describing photochemistry, and in particular free energy methods. In QM/MM studies, a small part of the system is treated using quantum chemical methods and the remainder via classical mechanics, in order to capture chemistry or photochemistry which would not be possible with conventional molecular dynamics. However, obtaining free energy surfaces is much more challenging, because of the need to obtain sufficient sampling with a sufficiently accurate energy function at the QM level. They review methods by which the sampling can be done using a simpler, semiempirical, model. Further, the calculation of spectroscopic properties and of excited state dynamics, and the importance of a polarizable treatment of the MM environment for capturing these properties are discussed.

Mlýnský and Bussi address the use of atomistic molecular dynamics simulations to explore RNA structure and function. This is challenging from the outset because RNA folding and

dynamics is known to be slow due to the relatively frustrated energy landscape compared with proteins. This makes it ripe for the application of enhanced sampling methods, and the application of variety of methods ranging from replica-exchange molecular dynamics simulation to coordinate-based methods such as metadynamics is reviewed. In particular, a novel reaction coordinate, e RMSD, is found to be much more discriminating of correct base pairing and orientation than conventional RMSD. Unfortunately, the use of better sampling methods has revealed that some discrepancies between RNA force fields and experiment were due to force field inaccuracies rather than lack of sampling. The authors conclude that improving force fields and sharing negative results in the area should be the focus of the all-atom RNA simulation field.

The topic of RNA force fields is also discussed by **Nerenberg and Head-Gordon** who provides a description of the current state of the art in biomolecular force fields, and highlights current challenges and future directions. The authors highlight the many different strategies taken to optimize force fields and, as exemplified by recent developments in protein force fields, describe how there are many possible solutions to reach the same goals. The authors highlight the potential of polarizable force fields, and suggest that further integration of statistical and machine learning tools into force field development could lead to more robust force fields.

While approaches based on molecular simulations with physical force fields may be appealing for prediction of protein-ligand binding, they are still relatively computationally expensive, considering that both conformation space and the chemical space of the ligands need to be explored. Therefore, as **Colwell** writes, it is necessary to develop rapid screening methods that are sufficiently fast that large ligand libraries can be handled. She describes recent efforts to use deep learning and neural networks to predict which ligands will bind to a given protein target. While recent results using these methods have been very promising, a major challenge is obtaining robust results beyond the test set used for fitting. Continuing on this subject, **De Fabritiis and co-authors** discuss how simulation methods can be used together with experiments to train machine learning methods. In particular, they suggest that the ever-increasing amounts of simulation data can be useful not only to answer system-specific questions, but also as input to train more general models. Examples discussed include improving the physical descriptions of molecules using quantum simulations, and improved predictions of binding affinities using molecular mechanics simulations.

In the last ten years there has been an increased focus on the role that ligand binding and unbinding kinetics may play in pharmacology. This has in turn spurred the development of computational methods to predict or rationalize kinetic properties, and **Wade and co-authors** provides a comprehensive overview both of the different methods and recent applications. The methods discussed include both biased simulation methods, as well as methods such as Markov state models or the weighted-ensemble path approach that extract kinetics from analyses of unbiased simulations.

Markov state models can also be used as a tool for enhanced sampling for problems where the time scales are inaccessible to conventional molecular dynamics simulation. An excellent example of such an approach is provided by **Huang and co-authors**, who have

applied it to the mechanism of RNA polymerase II transcription-elongation. The authors generate trial paths between different states and “shoot” trajectories from various points along the trial path to obtain extensive MD sampling, which is then combined via an MSM to provide a kinetic and mechanistic picture of each transition. The authors also describe how experimental data can be used to fill in missing rates from the overall kinetic scheme, by fitting the relaxation rate from the simulation model to experimental observations via the unknown rate parameters. The authors conclude with a perspective on the application of the kinetic models to the prediction of transcription error rates genome-wide.

Translation is another key biological process that has been studied extensively by simulations, as described in two different reviews. Going through the different stages of ribosome function, initiation, decoding, peptide bond formation, co-translational folding and termination **Grubmüller and co-workers** describe how a range of simulation methods have been used to provide molecular and mechanistic insights. These include QM studies of the process of peptide bond formation to atomistic or experimentally-driven simulations of the nascent polypeptide. On a related topic, **Sharma and O’Brien** focus on the co-translational folding of proteins as they are synthesized by the ribosome. They emphasize the inherently non-equilibrium nature of the situation, since in many cases the rates of protein synthesis may be comparable to the rate of folding. Indeed, there are several examples mentioned by the authors in which the details of the translation rate matter to the yield of the correctly folded protein. They discuss how kinetic models can be used to predict the influence of synonymous codon usage on yield of folded protein. Coarse-grained molecular simulations indicate that the positions in the sequence where the folding is furthest from equilibrium are most closely correlated with those positions where synonymous codon substitutions have the largest effect.

One potential strategy to help alleviate insufficient sampling and force field inaccuracies is to combine experiments and simulations. Such studies lie at the interface between structural biology and molecular simulations, and **MacCallum and co-workers** describe recent theoretical and algorithmic advances that have put the integration of experiments and simulations on a more secure footing. In particular, they describe the statistical foundation of the different methods that are in use, and highlight how the additional information brought by increasingly accurate physical models can aid in increasing both the accuracy and precision of conformational ensembles.

Small- and wide-angle X-ray scattering in solution is one of the experimental techniques that benefit greatly from the integration with simulations. In his review, **Hub** describes some of the particular challenges associated with interpreting X-ray scattering data, including their low information content and the difficulty in modelling accurately the effects of the solvation layer and protein dynamics on the experimental data. The review outlines some of the many different approaches developed to predict small- and wide-angle X-ray scattering data from biomolecular structures, and highlights how molecular simulation techniques, together with Bayesian methods, have the potential to unify the calculation of scattering data and structure determination.

Similar challenges are faced by cryo-electron microscopy which yields a set of single-particle images for biomolecules or biomolecular complexes which are averaged over both different molecular conformations as well as different orientations. **Cossio and Hummer** review how likelihood-based methods have been used to address this heterogeneity, starting with the class-averaging procedures typically used to infer 3D structural models based on maximum likelihood. They discuss problems (and their solutions) arising from multiple minima in the fitting function such that the derived model may be only a local minimum, dependent on the starting conditions for fitting. For many biomolecules, a small set of conformational classes is unlikely to describe the conformational distribution (e.g. molecules linked by flexible linkers). In this case, an alternative approach, based on reweighting an existing simulation ensemble (ensemble fitting) can provide a practical solution.

Tiana and Giorgetti review progress on another problem where experiments, theory and simulations have been integrated tightly, namely in studies of the structure and dynamics of mammalian chromosomes. Chromosome conformation capture (3C) experiments, and its variations 4C, 5C and Hi-C, utilize sequencing methods to provide coarse experimental data that report on long-range structure in chromatin. Using Bayesian and maximum entropy methods such data can in turn be used together with coarse-grained simulations to provide structural models of chromatin conformation within self-interacting regions termed topologically associating domains.

As these different reviews demonstrate, experiment, theory and simulations can often be fruitfully combined to study complex biological phenomena at the molecular scale. It is increasingly common to integrate multiple sources of simulation and experimental data, and improved statistical and theoretical approaches, as well as force fields and sampling algorithms, have been developed to facilitate such work. We are therefore enthusiastic about future uses of theory and simulations to aid in answering a diverse set of questions in biology and biophysics.

Biography

Robert Best completed his undergraduate degree in Chemistry at University of Cape Town and his Ph.D. at the University of Cambridge in 2003. After a post-doc at the National Institutes of Health with Bill Eaton, he took up a Royal Society Fellowship at the University of Cambridge in 2007. He returned to the NIH in 2012 and is now a Senior Investigator in the Laboratory of Chemical Physics in the National Institute of Diabetes and Digestive and Kidney Diseases. His research interests are in computational studies of protein folding, misfolding and dynamics and intrinsically disordered proteins, as well as computational sequence design.

Kresten Lindorff-Larsen trained as a biochemist at the University of Copenhagen and Carlsberg Laboratory, and completed his Ph.D. at the University of Cambridge in 2004. He then moved on to become an assistant professor in Copenhagen before joining D. E. Shaw Research in New York in 2007. He returned to Copenhagen in 2011, where he now serves as a Professor of Computational Protein Biophysics. His current research interests include

developing and applying computational methods for integrative structural biology, and the integration of biophysics and genomics research.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript