Preface

Modelers to the Rescue

"The purpose of models is not to fit the data but to sharpen the questions."
–Samuel Karlin, 1983.

We are in an era of data deluge. Merchants know all about what we eat, drink, wear, and read; doctors subject us to numerous tests and procedures to extract numbers and images that are intended to capture our state of health; and large-scale genome sequencing companies are producing so much data that computer disks via FedEx, rather than Internet downloads, are the only viable mode of data transfer![1] No wonder cartoonist Chris Wildt drew men and women of science pouring over paper files, computer screens, and endless diagrams before giving up: "Well, we've seen all the data and crunched the numbers... heads or tails?"

Sifting through biological and chemical data to extract knowledge, and eventually an understanding, is a task attributed today to bioinformaticians and other computational scientists through data mining approaches and computer modeling and simulation. As recently surveyed in a field perspective,[2] the trajectory of biomolecular modeling and simulation has evolved from initial euphoria and unrealistic expectations in the 1980s to a more mature, cautiously optimistic viewpoint bolstered by demonstrated successes (Figure P.1). Clearly, biomolecular modeling is not likely to supplant experimentation in the near future for predicting and understanding biomolecular structure and function; however, iterative program design between experiment and theory is becoming a reality for advancing the field. The problem of early unrealistic expectation and hype was not so much due to the inherent imperfections, approximations,

RSC Biomolecular Sciences No. 23
Innovations in Biomolecular Modeling and Simulations: Volume 1
Edited by Tamar Schlick
© Royal Society of Chemistry 2012
Published by the Royal Society of Chemistry, www.rsc.org

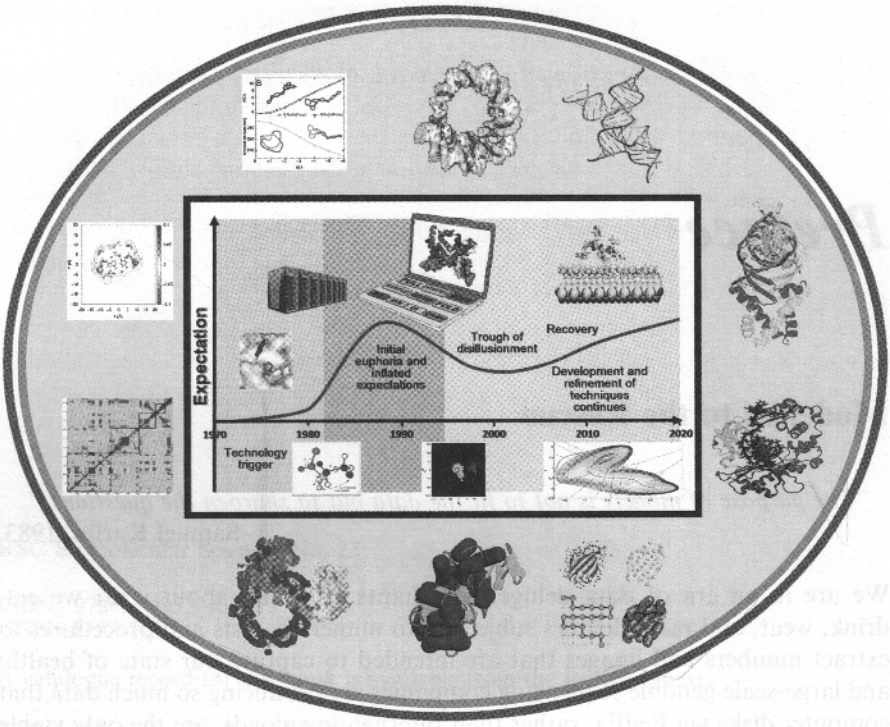

Figure P.1 **Proposed expectation curve for the field of biomolecular modeling and simulation, with approximate timeline,[2] surrounded by representative images from work collected in these volumes.** The field started when comprehensive molecular mechanics efforts began, and it took off with the increasing availability of fast workstations and later supercomputers. Following unrealistically high expectations and disappointments, the field is well on its fruitful trajectory of realistic progress and synergy between theory and experiment. The images taken from contributing chapters in this two-volume book are, *inside the box*: bottom right: contour plots connecting potential energy basins (A. E. Cardenas *et al.*); bottom middle: essential dynamics projection for HU protein/DNA junction complex (Wheatley *et al.*); top right: peptides on microarray surface (J. Feng *et al.*); and left middle: fragment grown for drug discovery applications (Z. Liu *et al.*). Outside the box, images clockwise from bottom right to top right are: proteins at various levels of resolution (N. Ceres *et al.*); coarse-grained model of the molecular motor myosin (E. R. Taylor); nucleosome chains (K. Rippe *et al.*); covariance matrix for protein/DNA/dNTP complex in DNA polymerases (M. Foley *et al.*); solvent charge density in an AT base pair plane (T. Luchko *et al.*); trefoil knot ensemble (G. Witz *et al.*); electrostatic potential of the nucleosome (R. C. Harris *et al.*); junction architecture in hairpin ribozyme (D. Lilley); structure of architectural protein HU bound to DNA (Wheatley *et al.*); and structure of DNA polymerase λ (M. Foley *et al.*).

and limitations of the models used, but rather an underestimation of the complexities of biological systems.

As we uncover the multi-layered levels of biological regulation, we are only beginning to understand what we actually do not understand. For this reason, modeling requires as much art as science in developing and appropriately applying, analyzing, and relating simulation results to the biological systems and associated experimental observations.

In this two-volume book, Innovations in Biomolecular Modeling and Simulations, twenty-six teams of leading scientists in biomolecular structure and modeling present the state-of-the-art in the field, from both a methodological and application perspective.

Volume 1 describes recent developments in force fields and electrostatics, conformational sampling and rate calculations, and coarse grained modeling and multiscale models.

Volume 2 describes atomistic simulations of nucleic acids and nucleic acid complexes; coarse-grained modeling approaches to study large-scale DNA folding, knotting, sliding and hopping; and the latest approaches and achievements in structure-based drug design.

In more detail, Volume 1 begins with Scheraga's personal perspective of the field; having worked in the field of protein structure and molecular biophysics for 65 years, Scheraga's trajectory from mainly experimental to mostly theoretical work on conformational energy calculations of polypeptides and proteins is an inspiration. The second historical perspective is contributed by writer Pollack, who tells the story of the emergence of the freely available molecular dynamics program NAMD from the dream of one immigrant, Klaus Schulten, who arrived to America in 1987 with a supercomputer in his backpack. Through interviews and studies of other pioneers, she weaves in history, science, and personal stories to present a fascinating historical account of the early days of molecular dynamics taking us to current-day applications.

Turning into the force-field advances section of Volume 1, MacKerell and co-workers describe one of the most important areas of current development, namely incorporating the polarizable component of electrostatic interactions into modern force fields; they also describe procedures for parameterizing these forms and testing them on model systems as well as highly-charged nucleic acids, where improvements can be demonstrated. Case and co-workers continue discussions of the fundamental electrostatics interactions by focusing on approaches using integral equations to model the interactions of biomolecules with water and ions and improve upon other implicit-solvation techniques such as Poisson Boltzmann approximations. Smith and co-workers share their related methodological advances using both classical molecular mechanics and quantum-mechanical techniques to study complex systems of interest to the energy bioscience field using supercomputing resources; implicit solvation and efficient scaling on large-scale machines are required to study behavior of multi-million-atom systems.

The second group of papers in Volume 1 describes the latest algorithms to survey conformational changes, associated pathways, and reactions rates from

the molecular simulations. As reviewed recently,[3,4] various clever stochastic as well as deterministic approaches have been developed and applied in recent years to increase the variety of configurations captured, encourage transitions, and extract essential quantitative information on those rare events. The methods include Milestoning and Forward Flux methods based on trajectory fragments, as described by Elber and co-workers, and conformational dynamics and Markov state models as detailed by Darve and co-workers. The alternative Monte Carlo approach for large systems is described by Mezei, who discusses special implementational considerations for macromolecules.

While atomistic models are essential for describing detailed events at the fine level, coarse-grained models are required to describe events relevant to very large complex systems like the chromatin fiber or major rearrangements involving topological changes such as DNA knotting. For such problems, simplified representations are needed, and the challenge is not only to construct useful models that approximate many aspects of the systems, while detailing others, but to eventually telescope between the macroscopic and microscopic views. Examples of such approaches are described by Lavery and coworkers for proteins, by Taylor and co-workers for the molecular motor myosin, by Radhakrishnan and co-workers for protein-protein and protein-membrane networks, by Pettitt and co-workers for complexes at material surfaces, and by Fish and co-workers in the context of large-scale industrial applications.

Volume 2 begins with state-of-the-art examples of atomistic simulations of nucleic acids and nucleic acid complexes. Lankas describes the nuts and bolts of atomistic simulations of DNA in explicit solvent with a viewpoint into coarse-grained models and a focus on understanding sequence and anharmonic effects. Neidle and co-workers describe applications to quadruplex systems of nucleic acids, their complex hierarchical arrangements as studied by classical dynamics, enhanced sampling methods, and hybrid classical/quantum approaches, as well their relevance for supplementing current static structure resolution approaches. Rohs, Fenley and co-workers discuss electrostatic and structural aspects of DNA that have tailored DNA for interactions with proteins, as deciphered from atomistic studies of nucleic acids and their complexes. Schlick and co-workers describe insights into DNA polymerase repair mechanisms as deduced from atomistic dynamics and principal component analysis of several polymerase enzymes; the resulting correlated motions suggest mechanisms that are hybrid between traditional induced fit and conformational selections models. Mukerji, Beveridge and co-workers report their latest study for a structural prediction from molecular dynamics of a novel protein/DNA complex involving a four-way junction.

Continuing to RNA at atomic resolution, Sponer and co-workers describe the state of-the-art in simulating RNAs by all-atom molecular dynamics, a field that has lagged behind RNA's famous cousin due to force-field limitations, among others. Lilley then describes the fascinating and hierarchical folding of RNA junctions, important since junctions organize the global architecture of RNA.

Events that require different kinds of methods are then investigated in the second part of Volume 2. Witz and Stasiak describe studies of DNA knots and catenanes using topological frameworks explored by Monte Carlo and Brownian dynamics simulations. Rippe and co-workers describe Monte Carlo simulations of nucleosome chains using simple coarse-grained models that capture essential features of the systems. Levy and Marcovitz then describe a study of DNA sliding on the genomic level.

The book concludes with overviews of drug design approaches, from both industrial (Singh and co-workers) and academic perspectives (Amaro and co-workers, McCammon). McCammon's personal perspective focuses on the identification of two anti-viral drug targets for AIDS using a structure and dynamics-based approach.

The range of subjects and methods as reflected by these exciting contributions indicate not only the ingenuity of the methods and wide range of tools available to the molecular modeler, but also the endless questions that arise from the modeling. These questions, in turn, serve to advance the field toward an ultimate understanding of the complex interaction between biomolecular structure and function.

The data deluge is unlikely to cease in the near future. Nonetheless, our modeling approaches will undoubtedly become better and more diverse, allowing us to tackle yet more challenging problems and unravel complex relationships, patterns, and long-time events in biomolecules.

Tamar Schlick
New York

References

1. A. Pollack, DNA Sequencing Caught in Deluge of Data, 30 Nov., *The New York Times*, 2011, page C1.
2. T. Schlick, R. Collepardo-Guevara, L. A. Halvorsen, S. Jung and X. Xiao, Biomolecular Modeling and Simulation: A Field Coming of Age. *Quart. Rev. Biophys.*, 2011, **44**, 191–228.
3. T. Schlick, Monte Carlo, Harmonic Approximation, and Coarse-Graining Approaches for Enhanced Sampling of Biomolecular Structure. *F1000 Biol. Rep.*, 2009, **1**, 48.
4. T. Schlick, Molecular-Dynamics Based Approaches for Enhanced Sampling of Long-Time, Large-Scale Conformational Changes in Biomolecules. *F1000 Biol. Rep.*, 2009, **1**, 51.