

CSE 2011

Biomolecular Modeling and Simulation: The Productive Trajectory of a Field

By Tamar Schlick and Rosana Colleparado-Guevara

The great thing in the world is not so much where we stand, as in what direction we are moving.—Oliver Wendell Holmes

As the field of biomolecular modeling has increased in visibility among both scientists and the general public, it has also drawn scrutiny. In particular, biologists and funding agencies have been asking pertinent questions: What impact have computer technology and mathematical algorithms had on the field's development? What are future prospects for the field [8, 10, 17]? The focus of an invited talk* at this year's SIAM Conference on Computational Science and Engineering was the survey we conducted to address these questions, including a description of key success stories, examples of failure, and a projection of future prospects. This study began as a project in a Special Topics course at New York University and resulted in the appearance of a perspective on the field in *Quarterly Reviews of Biophysics* [15].

Our examination of progress in biomolecular modeling and simulation emphasized structure prediction and molecular dynamics simulations. Overall, the study showed that, despite unrealistic initial expectations, dramatic increases in computational technology along with steady improvements in force fields and sampling algorithms have led to considerable success in many areas with important applications to biology and medicine. Moreover, the establishment of reliable protocols and awareness of caveats have propelled the field well on its way to becoming a full partner with experiment and eventually a field in its own right.

Historical Perspective

Molecular modeling is a relatively young field, with origins in the late 1960s. It was then that force field pioneers developed simple empirical approaches based on physiochemical potential functions to represent molecular structures and interactions by basic physical models (atoms connected by springs) and model systems (e.g., peptides, nucleotides, hydrocarbon systems). Such potential functions were soon being used with energy-minimization techniques to determine favorable molecular conformations and, shortly afterward, as a refinement tool for experimental structures, such as those determined by X-ray crystallography. Rapid advances in the development of water force fields, concepts in enzyme energetics, and molecular dynamics techniques quickly led to simulations of biological systems in the late 1970s.

It was only with the advent of supercomputers in the 1980s, however, that the field gained momentum and applications to pharmacology and medicine took center stage. Indeed, the late 1980s saw an inflation of expectations, with unrealistic predictions that computer simulations might soon replace experiments in the laboratory (see [15] for detailed quotes). Expectations were particularly high for applications to drug design and other medically relevant applications, as computation appeared poised to supplant laborious experimental or technological procedures.

It soon became evident that the predictions of quick success were too optimistic. Imperfections in force field parameters cause uncertainties, and sampling the vast thermally accessible conformation space of biomolecules requires vast computer power to bridge modeling and experimental time frames. With typical molecular dynamics timesteps on the order of 1 fs, for example, 10^9 steps, each requiring a costly force evaluation, are needed to simulate 1 μ s, which is approximately the folding time for some of the smallest and fastest folding proteins; biologically interesting processes occur on timescales many orders of magnitude longer. Force field problems, such as overstabilization of α -helices, inability to model divalent ions, and instability of RNA base pairs, also became apparent.

Open discussions of these problems around the beginning of the new millennium led to clever solutions, along with fresh ideas from the mathematical, physical, and biological sciences. Improved algorithms for numerical integration [16], long-range electrostatic computations [16], and enhanced conformational sampling [13,14] emerged from these ideas. These advances, when combined with modern computer technology—loosely coupled networks of processors, efficient open-source software for biomolecular simulation, graphics processing units [20], and specialized hardware for molecular dynamics simulations—have led the field through important milestones: from the 1- μ s protein folding simulation in 1998, requiring four months of dedicated Cray computing [3], to 10- μ s simulations by aggregate dynamics a couple of years later [19] and continuous dynamics a decade after that [4], to a 1-ms simulation on a custom-built machine in 2010 [18].

This historical perspective led us to propose the “field expectation” curve shown in Figure 1, emphasizing the transition from initially unrealistic high hopes to the relatively swift progress of the past decade with a more practical viewpoint and many good ideas. (Computer scientist James Bezdek

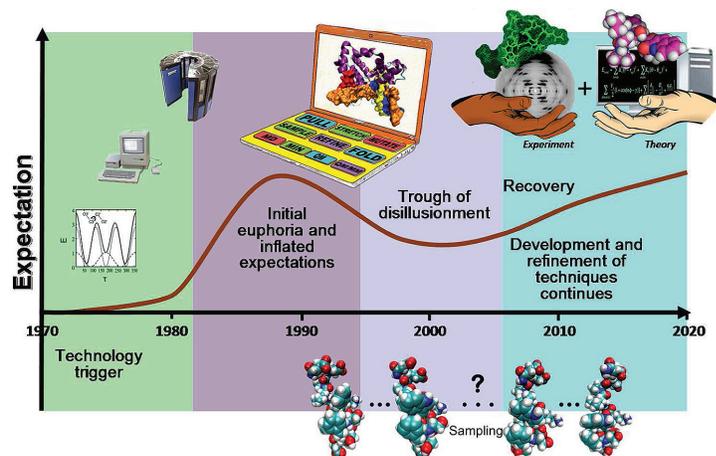


Figure 1. Proposed expectation curve for the field of biomolecular modeling and simulation, with approximate timeline. Reprinted with permission from “Biomolecular modeling and simulation: A field coming of age,” by T. Schlick, R. Colleparado-Guevara, L.A. Halvorsen, S. Jung, and X. Xiao, *Quart. Rev. Biophys.*, 44 (2011), 191–228, [15].

*Slides and synchronized audio for the talk, given by Tamar Schlick on Tuesday, March 1, can be accessed at siam.org.

first described such field expectation curves in 1993 when he founded a journal on fuzzy mathematics.)

Success Stories

The many successes of biomolecular modeling and simulation, as documented in [15], include the establishment of reliable and efficient biomolecular simulation algorithms (e.g., symplectic integrators, particle-mesh Ewald methods for electrostatic computations, Markov state models for sampling); structure predictions, especially by knowledge-based methods, that preceded experimental discovery; theories of protein folding describing experimental observations related to protein statics, thermodynamics, and kinetics; modeling-aided drug discovery and design; and interpretation of experimental structures. We elaborate here on the two last categories.

Structure-based rational drug design has always been the poster child for the field, illustrating the immense potential of molecular modeling. Indeed,

as proteins and macromolecular complexes implicated in human disease (e.g., HIV enzymes, misfolded protein fibrils) are resolved experimentally in astonishing detail, new opportunities for (and expectations of) modeling emerge: With the ability to define molecular specificity and functional interactions, it may become possible to pinpoint, and eventually design, drug targets. The HIV protease inhibitor and SARS virus inhibitors are classic examples. While the monetary cost and time required to develop a successful drug remain high, modeling has been instrumental in opening doors that might otherwise have remained closed. For example, modeling can systematically test and reveal various binding modes and possibilities.

Consider the solved complex of HIV integrase with an inhibitor [5] and a related complex for an evolutionarily similar virus [7]. Studies in both cases revealed the drug docked in one main region. Molecular dynamics simulations based on such structures suggested that inhibitors can bind in more than one orientation—namely, preferred and flipped [11,12]—as shown in Figure 2. Simulations further suggested that binding modes can be selected to exploit stronger interactions in specific regions or orientations, and that different divalent-ion arrangements are associated with these binding sites and fluctuations. Such information can be utilized in systematic exploration for the binding of similar active compounds, as additional inhibitors, to the alternative site; more than one inhibitor is often needed as drug resistance builds up quickly in the HIV enzymes.

Modeling and simulation have also been instrumental in resolving conflicting experimental information and/or adding new hypotheses that can eventually be tested. For example, two proposed models (Figure 3A and B) have long been debated for the chromatin fiber, the protein/DNA complex in eukaryotes that packages the genetic material.

Evidence from X-ray crystallography of nucleosomes and electron microscopy images of reconstituted chromatin fibers has been contradictory. Modeling, as well as single-force microscopy experiments, in which polymers like RNA or chromatin are stretched to produce curves that describe the observed extension vs. the applied force, have been utilized to address such questions. Whereas modeling is challenged by the large spatial and temporal ranges associated with chromatin rearrangements, the experiments are limited by the indirect inference of shape/structure from the force vs. extension curves.

In one recent study, a solenoid fiber model was interpreted to be consistent with the force microscopy data [9]. Our recent Monte Carlo simulations of a mesoscale chromatin model, in collaboration with a team of experimentalists, showed that, in fact, solenoid features can be present in a mostly zigzag fiber under typical physiological (divalent-ion) conditions [6]; the notion of a heteromorphic helix that merges the two features is appropriate for a floppy polymer in solution (Figure 3C).

Modeling of fiber unfolding showed that the zigzag architecture is also consistent with the experimental data; hence, the solenoid interpretation is not unique. Our modeling of chromatin stretching also suggested

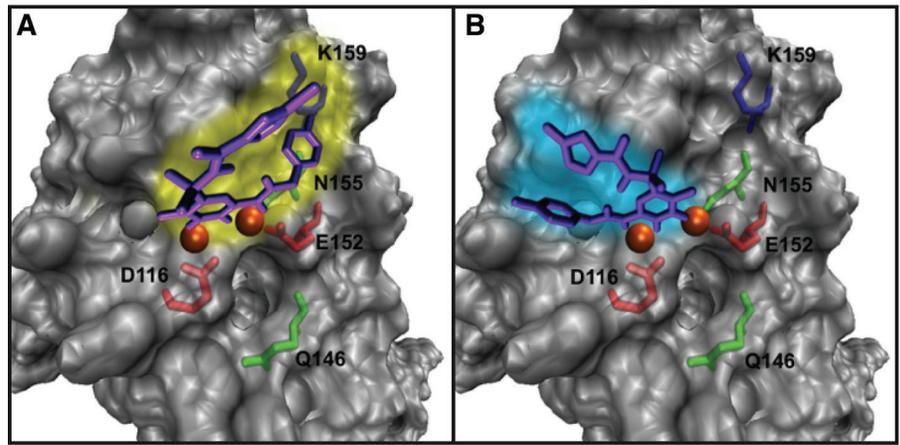


Figure 2. Drug/inhibitor binding pockets to HIV integrase as predicted by modeling [11]. A: Preferred binding (yellow shading) for the binding region of the anti-retroviral drug Raltegravir to the HIV-1 integrase catalytic core domain. B: Alternative, or flipped, binding (turquoise shading) for Raltegravir to the HIV-1 integrase catalytic core domain. The magnesium ions are shown as orange spheres.

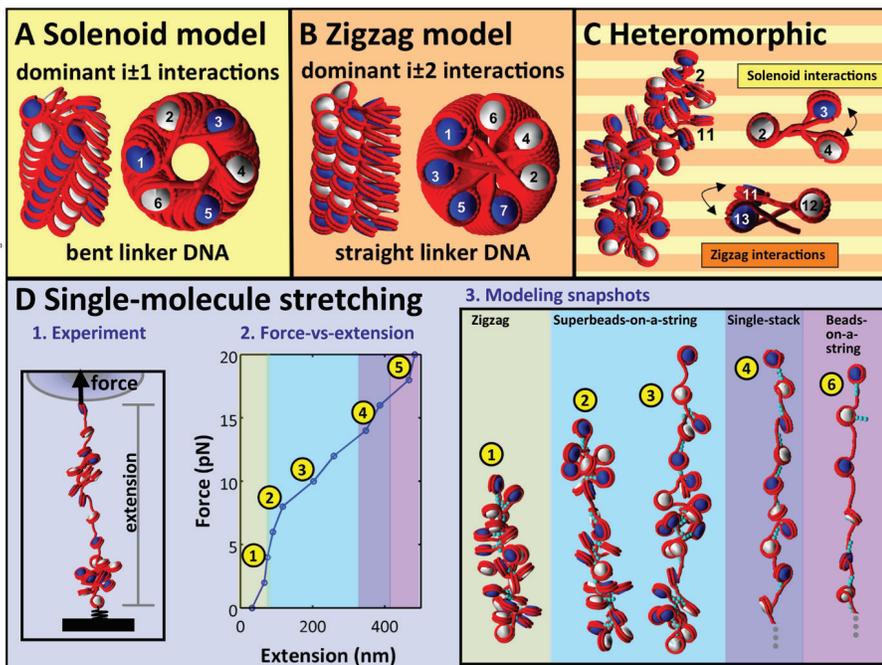


Figure 3. Chromatin organization: Ideal models and simulation-generated models and stretching profiles. A: Ideal solenoid model in which DNA linkers (the DNA segments, shown in red, connect nucleosomes) are bent and neighboring nucleosomes ($i \pm 1$) are in closest contact. B: Ideal zigzag model in which DNA linkers are straight and next-nearest neighbors ($i \pm 2$) are in closest contact. C: Heteromorphic architecture obtained by modeling [6] in divalent-ion environments with the chromatin model. In all views, connecting DNA linkers and DNA wrapping the nucleosomes are shown in red; odd and even nucleosomes are blue and white, respectively. D: Simulations mimicking single-molecule chromatin unfolding experiments with resulting force vs. extension curves and associated snapshots [1,2].

structural pathways associated with the experimental data, energy trends, and insights into the effect of different binding modes of protein linkers on resulting fiber elasticity (Figure 3D) [1,2]. Such information regarding the features that stiffen/soften the rigidity of chromatin fibers is important because chromatin is subjected to pulling forces of molecular motors in the natural cellular environment, and these molecular machines initiate various biological processes.

These are only a few of the many examples in which modeling has helped clarify, resolve, interpret, and supplement experimental data regarding biological structures, mechanisms, and interactions. Such valuable insights from careful design and execution of modeling “experiments” continue to propel the field on its productive trajectory (Figure 1) despite its many limitations. Taking Holmes’s view, perhaps being headed in the right direction will allow us to be more tolerant of these imperfections and to keep moving.

References

- [1] R. Collepardo-Guevara and T. Schlick, *Crucial role of dynamic linker histone binding for DNA accessibility and gene regulation revealed by mesoscale modeling of oligonucleosomes*, 2011, submitted.
- [2] R. Collepardo-Guevara and T. Schlick, *The effect of linker histone’s nucleosome binding affinity on chromatin unfolding mechanisms*, 2011, submitted.
- [3] Y. Duan and P.A. Kollman, *Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution*, *Science*, 282 (1998), 740–744.
- [4] P.L. Freddolino, F. Liu, M. Gruebele, and K. Schulten, *Ten-microsecond molecular dynamics simulation of a fast-folding WW domain*, *Biophys. J.*, 94 (2008), L75–L77.
- [5] Y. Goldgur, R. Craigie, G.H. Cohen, T. Fujiwara, T. Yoshinaga, T. Fujishita, H. Sugimoto, T. Endo, H. Murai, and D.R. Davies, *Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design*, *Proc. Natl. Acad. Sci. USA*, 96 (1999), 13040–13043.
- [6] S.A. Grigoryev, G. Arya, S. Correll, C.L. Woodcock, and T. Schlick, *Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions*, *Proc. Natl. Acad. Sci. USA*, 106 (2009), 13317–13322.
- [7] S. Hare, S.S. Gupta, E. Valkov, A. Engelman, and P. Cherepanov, *Retroviral intasome assembly and inhibition of DNA strand transfer*, *Nature*, 464 (2010), 232–236.
- [8] *IMAG Futures Meeting: The Impact of Modeling on Biomedical Research*, http://www.imagwiki.nibib.nih.gov/mediawiki/images/b/b5/IMAG_Futures_Report.pdf, NIH/IMAG, December 2009.
- [9] M. Kruithof, F.-T. Chen, A. Routh, C. Logie, D. Rhodes, and J. van Noort, *Single-molecule force microscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber*, *Nat. Struct. Mol. Biol.*, 16 (2009), 534–540.
- [10] *A New Biology for the 21st Century: Ensuring the United States Leads the Coming Biology Revolution*, The National Academies Press, <http://www.nap.edu/catalog/12764.html>, 2009.
- [11] A.L. Perryman, S. Forli, G.M. Morris, C. Burt, Y. Cheng, M.J. Palmer, K. Whitby, J.A. McCammon, C. Phillips, and A.J. Olson, *A dynamic model of HIV integrase inhibition and drug resistance*, *J. Mol. Biol.*, 397 (2010), 600–615.
- [12] J.R. Schames, R.H. Henchman, J.S. Siegel, C.A. Sotriffer, H. Ni, and J.A. McCammon, *Discovery of a novel binding trench in HIV integrase*, *J. Med. Chem.*, 47 (2004), 1879–1881.
- [13] T. Schlick, *Molecular-dynamics based approaches for enhanced sampling of long-time, large-scale conformational changes in biomolecules*, *F1000 Biol. Rep.*, 1:51 (2009).
- [14] T. Schlick, *Monte Carlo, harmonic approximation, and coarse-graining approaches for enhanced sampling of biomolecular structure*, *F1000 Biol. Rep.*, 1:48 (2009).
- [15] 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.*, 44 (2011), 191–228.
- [16] T. Schlick, R.D. Skeel, A.T. Brünger, L.V. Kalé, J.A. Board, Jr., J. Hermans, and K. Schulten, *Algorithmic challenges in computational molecular biophysics*, *J. Comput. Phys.*, Special Volume on Computational Biophysics, 151 (1999), 9–48.
- [17] *Scientific Grand Challenges: Opportunities in Biology at the Extreme Scale of Computing*, <https://www.cels.anl.gov/events/workshops/extremecomputing/biology/report.php>, DOE, August 2009.
- [18] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, and W. Wriggers, *Atomic-level characterization of the structural dynamics of proteins*, *Science*, 330 (2010), 341–346.
- [19] M. Shirts and V. Pande, *Screen savers of the world unite!*, *Science*, 290 (2000), 1903–1904.
- [20] J.E. Stone, D.J. Hardy, I.S. Ufimtsev, and K. Schulten, *GPU-accelerated molecular modeling coming of age*, *J. Mol. Graph. Mod.*, 29 (2010), 116–125.

Tamar Schlick (schlick@nyu.edu), a member of the Department of Chemistry and the Courant Institute of Mathematical Sciences at New York University, is a professor of chemistry, mathematics, and computer science. Rosana Collepardo-Guevara is a postdoctoral researcher and Schlumberger Faculty for the Future Scholar in the Department of Chemistry at NYU.