Molecular dynamics-based approaches for enhanced sampling of long-time, large-scale conformational changes in biomolecules

Tamar Schlick

Address: Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA

Email: schlick@nyu.edu

Abstract

The rugged energy landscape of biomolecules together with shortcomings of traditional molecular dynamics (MD) simulations require specialized methods for capturing large-scale, long-time configurational changes along with chemical dynamics behavior. In this report, MD-based methods for biomolecules are surveyed, involving modification of the potential, simulation protocol, or algorithm as well as global reformulations. While many of these methods are successful at probing the thermally accessible configuration space at the expense of altered kinetics, more sophisticated approaches like transition path sampling or Markov chain models are required to obtain mechanistic information, reaction pathways, and/or reaction rates. Divide-and-conquer methods for sampling and for piecing together reaction rate information are especially suitable for readily available computer cluster networks. Successful applications to biomolecules remain a challenge.

Introduction and context

It has been said that, while the 19th century belonged to chemistry and the 20th century to physics, the 21st century is and will continue to be dominated by ground-breaking science in biology. However, it is clear that biology without the tools of mathematics, physics, chemistry - not to speak of computer science and engineering innovations - would be severely limited. In fact, computer modeling and simulation offer a modern tool for simulating conformational events in a variety of molecular systems and subsequently extracting related mechanistic, thermodynamic, and kinetic information.

One spectacular example of the modern marriage between technology and biology involves the theory and practice of molecular dynamics (MD). MD, with variations and extensions, has become a universal tool for probing relationships among biomolecular structure, flexibility, and function. Since successful applications were reported in the 1970s in protein dynamics [1], MD has become a universal tool, 'as if it were the differential calculus' [2].

Of course, MD is simply statistical mechanics by numbers, or Pierre-Simon de Laplace's vision of using physical laws to simulate time-dependent events [3], on modern supercomputers. And impressive progress in the development of biomolecular force fields, coupled to spectacular computer technology advances, has now made it possible to overcome the difficulty noted by Paul Dirac of solving the complex equations of motion [4].

Now, sophisticated dynamics programs, like NAMD, Desmond, or GROMACS, adapted to parallel and massively parallel computer architectures, have made simulations of biomolecular systems of one microsecond and longer feasible in several weeks of computing [5-8]. However, the hunger for longer simulations as an end by itself may not be justified given that such simulations lead to as many questions as answers, and underscore the notion that computer power alone is not likely to solve 'folding' problems. Indeed, many approximations (for example, in force field functional form as well as numerical values) and sampling limitations still

persist. And besides potential function inaccuracies, sampling the vast conformational space remains a grand challenge because standard MD methods cannot sample the vast thermally accessible configuration space of biomolecules to capture large-scale rearrangements, including reaction pathways, mechanisms, and rates. Rather, MD applications are valuable in probing local system fluctuations (Figure 1, bottom left) for different single-residue variants of DNA polymerase λ [9].

An enormous range of methods has been developed and applied to molecular systems in an attempt to enhance coverage of the thermally accessible conformational space. The efficacy/ability of each approach depends on the goals of the computation and the available computing resource. No review or presentation could possibly list them all. Instead, I will present some general concepts and highlight selected interesting applications in this brief perspective. See [10,11] for recent reviews.

Figure 1. Illustrations of various sampling methods
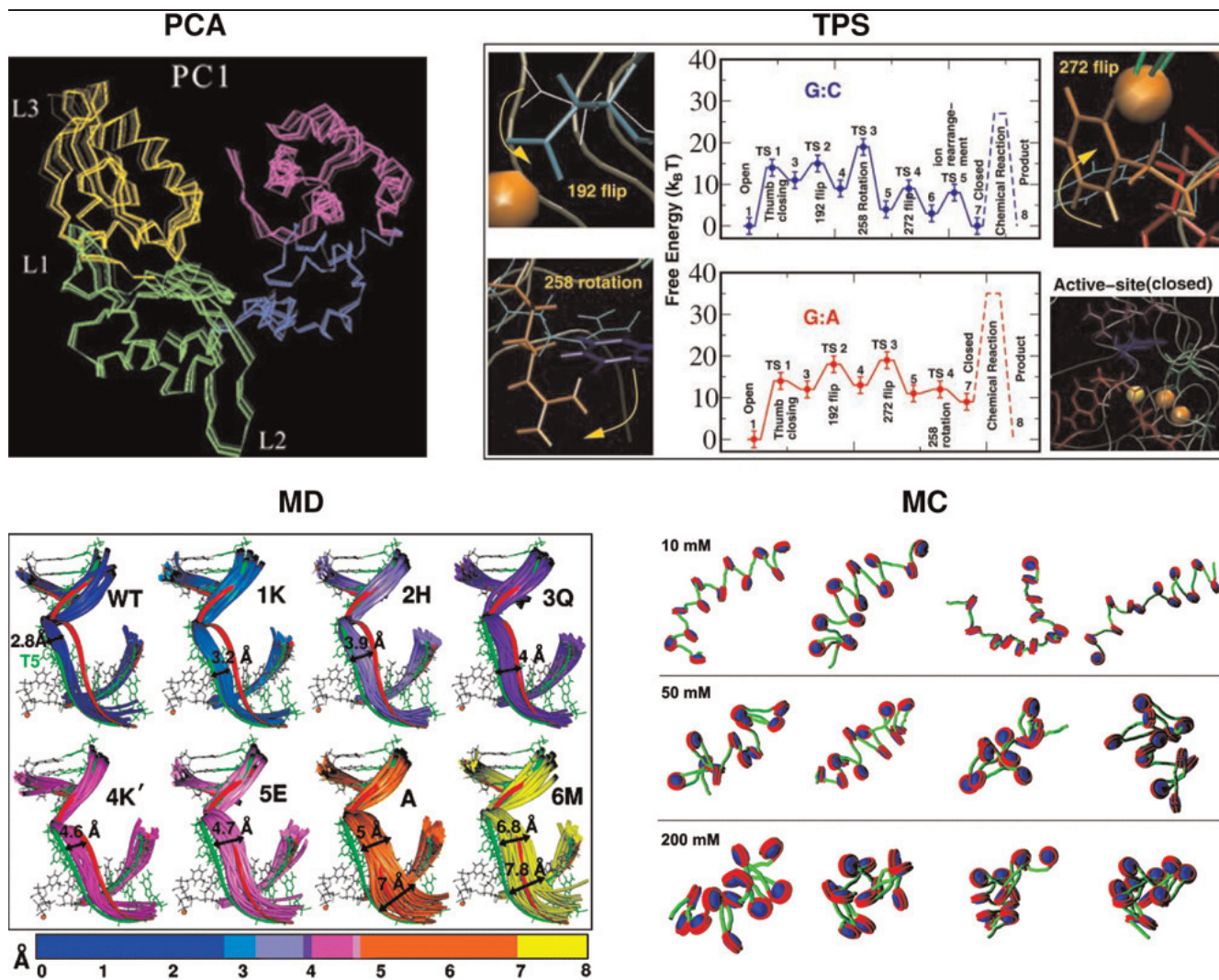
(Top left) Principal component analysis (PCA) of a polymerase β/DNA complex during the conformational closing rearrangement following nucleotide binding revealed dominant motions of the thumb subdomain, correlated to other regions of the protein [13]. Only the protein is shown. (Top right) Transition path sampling (TPS) of the same complex for the closing conformational change for both correct and incorrect base pairs reveals the free energy profiles and major barriers involved [58,59]. (Bottom right) Monte Carlo (MC) simulations of a coarse-grained multiscale model of the chromatin fiber showed the influence on the monovalent salt concentration on the oligonucleosome configurations [14]. (Bottom left) Molecular dynamics (MD) simulations of seven single-residue variants of the polymerase λ/DNA complex at residue 517 showed the importance of this residue in stabilizing the complex, as evident by the wide range of fluctuations observed [9]. PCA image reprinted with permission from *Biophysical Journal* [13]. Copyright 2004, Elsevier.

In the first review of this series [12], Monte Carlo (MC) approaches (simulated annealing, hybrid MC, and parallel tempering), harmonic approximations (methods based on spectral decompositions like principal component analysis [PCA] and essential dynamics), and coarse-graining formulations (multiscale models for protein and supramolecular systems) were surveyed; methods like PCA provide information on dominant functional motions, such as subdomain movements in a polymerase β/DNA complex [13] (Figure 1, top left), and MC simulations are especially effective for coarse-grained models, such as chromatin [14] (Figure 1, bottom right).

Here, methods based on MD are considered. These can be as simple as modifying the potential or the basic protocol or can involve more complex modifications of the algorithms, including global reformulations. Methods in all categories (for example, replica exchange MD, or REMD) are generally effective for gaining insights into broader structural and thermodynamic questions than possible in typical MD, although a rigorous assessment of their performance is generally lacking. However, increasingly sophisticated methods - grouped as global reformulations - are needed for describing mechanisms, free energy profiles, and/or detailed kinetics. These include transition path sampling (TPS), forward flux simulation (FFS), and Markov chain models, all of which pay special attention to preserving microscopic reversibility and thermal distributions of the systems.

Major recent advances
Modifying the potential energy function
The simplest and crudest approach for enhancing the sampling is to manipulate the energy function used as a basis for MD by using restrained potentials, as in targeted MD (TMD) or umbrella sampling (US) by using various experimentally based biases or guides.

In TMD, an artificial restraint potential term is added like a Lagrange multiplier with a parameter λ (0 at the initial state and 1 at the target state) so as to force the system to evolve toward a certain state (for example, closed and open state of an enzyme) in a specified number of steps. However, the actual trajectory obtained using this restraint leads to unphysical dynamics since incorrect transition states are often achieved. Nonetheless, TMD can be used to explore space in a preliminary manner.

In 2002, Kong *et al.* [15] reported a TMD application to the opening of an *Escherichia coli* membrane channel whose open and closed states were modeled. The simulations suggested a four-stage process in the opening, unlike simpler paths suggested previously. An *F1000* commentator writing on this work considered TMD to be

'one of the most promising methods to suggest possible pathways for large-scale conformational changes' [16].

Similar in spirit is the simplified multiple-basin Hamiltonian funnel-based potential model [17]. Developed for very large molecular complexes with known endpoint structures, such a model can be used, as in TMD, to simulate very large-scale motions, such as transitions between ligand-bound and unbound states.

However, over the past few years, we have come a long way. Now, we have more mathematically rigorous methods to sweep conformational space, deduce mechanisms, and compute reaction rates as well as free energies by a variety of innovative methods. Still, because the applications of those more sophisticated methods are far from routine for biomolecules, there is room for the simpler, albeit more artificial, approaches.

US focuses on specific regions of phase space by using a restraining potential. In a recent application, Mills and Andricioaei [18] developed a 'guided US' for biomolecules in which experimental restraints [for example, end-to-end distance determined by fluorescent resonance energy transfer (FRET) experiments] are used to formulate potential restraints. They showed dramatic improvement in efficiency (that is, faster convergence and better final distances for the guided versus unguided US).

Another interesting example of using experimental restraints was reported by Paci *et al.* [19]. They constructed transition state ensembles (TSEs) of simple two-state (fast) folding proteins by restraining the conformational energy to states that satisfy experimentally known contact values of ϕ (ratio of the number of native contacts present in the protein transition state to that in the native state). Interestingly, their computed TSEs for three fast-folding proteins revealed that members of these TSEs have approximately 70% native-like folds even though rms distances, relative to the folded state, can be large.

Besides various constraints or restraints, the energy can be alternatively modified by using other types of added energy terms. For example, McCammon's group [20] recently used accelerated MD in which a term is simply added to the potential energy 'to boost' the potential to allow escape from minima. For numerical considerations, the addition is implemented so that the gradient is continuous. Moreover, to be effective, the boost can be applied to one term only (for example, torsional terms) so as not to destroy liquid water structure around the biomolecule.

Such a boost approach is based on many other proposals of adding biasing potentials to overcome barrier crossing reported previously. These include Grubmüller's conformation flooding in which the bias is determined based on a coarse-grained simulation for the conformational space density [21], Voter's hyper-MD in which the biasing potential is constructed based on the smallest eigenvalue of the Hessian matrix [22], the simple local boost method based on the total potential energy [23], and a bond boost method based on bond-length deviations [24].

Modifying the simulation protocol

The MD simulation protocol rather than the energy can also be manipulated. In this subclass, we have the popular approach LES (locally enhanced sampling), in which several configurations are generated in a single energy evaluation [25], and REMD [26], based on parallel tempering MC [27], which itself is based on Metropolis-coupled Markov chain MC.

The idea in REMD is to simulate multiple, non-interacting copies (replicas) of identical systems at different temperatures. Periodically, the configurations of the replicas are exchanged (and thus the thermodynamic, or temperature, states are exchanged) with a transition probability P that maintains each temperature's equilibrium ensemble distribution in the canonical ensemble so as to retain those systems that are making better progress:

$$P = \min\{1, \exp[(\beta_{\text{new}} - \beta_{\text{old}})(U_{\text{new}} - U_{\text{old}})]\}$$

where $\beta_i = 1/(k_B T_i)$ and U_i are the internal energies of states i (new and old). These temperature-ensemble exchanges are meant to accelerate crossings of energy barriers. Though obeying detailed balance for an extended ensemble of the canonical states, the state-exchange probability destroys real kinetic properties. However, REMD has been successfully used to simulate folding/unfolding equilibria of biomolecules.

REMD has been increasing in popularity for applications ranging from small peptides to complex biological systems. However, its success has largely been empirical. For example, among the numerous applications to biomolecules reported, the folding of a small solvated RNA hairpin from an extended state [28] revealed a diverse ensemble of conformations with differing stacking and base pair arrangements; applications to solvated protein A [29] also revealed the folding as the temperature decreases from about 600 to 280 K (see posted animations [30]). REMD villin folding simulations of 200 nanoseconds in length consistently folded within 1.78 Å of the native state [31].

More recently, practitioners of REMD have emphasized the need to formulate careful configurational swapping protocols (temperature ladders and exchange/acceptance ratios) and other ways to increase the conformational sampling efficiency (see [32,33], for example). Other practical considerations have been discussed. For example, because efficient sampling in REMD requires a large number of replicas to enhance the swapping rate, the user must have a large number of concurrent processors available for the REMD simulation; this may not be the case for the average user, and the REMD variant distributed replica sampling [34] may be preferable. Kamberaj and van der Vaart [35] describe an efficient multiple scaling REMD in which a Tsallis biasing potential increases sampling efficiency because barrier crossings are enhanced so the number of replicas can be reduced. For a note on the likely computational advantage of REMD for systems with relatively high energy barriers, see [36].

Cooke and Schmidler [37] recently presented a thorough analysis of REMD, explaining that the formulation of REMD by extension of parallel tempering MC to MD has introduced sampling and ergodicity problems stemming from the failure of the underlying constant-temperature (canonical) MD integrators to preserve certain variants. This is because, while the MC analog is based on Markov chains, REMD algorithms cannot use the symplectic leap-frog integrator popular for microcanonical (constant energy) MD and resort to isothermal (constant temperature) integrators like the Berendsen heat-bath algorithm or the Nosé-Hoover thermostat method, which are not rigorously ergodic; this can affect the dynamics of even small systems. As a remedy, combining REMD with hybrid MC to ensure ergodicity is suggested [37], thereby returning to the original advantages of parallel tempering MC. Recently developed entropy-preserving constant-temperature integrators can also solve the ergodicity problem in REMD [38].

Many other REMD combinations have been reported. Chebaro *et al.* [39] combined REMD with a coarse-grained implicit solvent model of proteins (OPEP) to determine equilibrium ensembles of proteins. Bolhuis [40] combined REMD with transition interface sampling (TIS) to compute the free energy profile and rate constants when high barriers are involved. These combinations emphasize how different sampling methods can be combined in clever ways to suit the problem and enhance sampling further.

Modifying the MD algorithm

Besides altering the energy function as in TMD and simulation protocol as in REMD, the MD algorithm itself can be altered.

Many integrators have been developed to address stability and resonance limitations of MD integrators due to the high-frequency vibrational modes and the intricate coupling among the vibrational modes of a biomolecule spanning a wide spectrum. See [41], for example, for a discussion of these issues and various solutions; most successful approaches to both stability and resonance problems consist of multiple-timestep integrators combined with stochastic dynamics.

Langevin dynamics and Brownian dynamics (BD), in which solvent collisions and thermal fluctuations are incorporated in an average sense, have long been used as a way to allow larger timesteps and hence larger timespans in dynamics simulations. For stochastic dynamics methods, it is also easier to prove ergodicity.

An interesting approach in which the algorithm is modified was recently reported by Sweet *et al.* [42]. Their approach, called 'NML', is an extension of a LIN [43] developed as a way to increase the timestep in standard MD integration. The idea in LIN was to use implicit integration for the low-frequency modes and normal-mode analysis (NMA) for the high-frequency modes; Langevin, rather than Newtonian, dynamics was also used to dampen resonance effects, which limit the timestep to around 3.3 femtoseconds when all light-atom motions are considered [41]. Sweet *et al.* [42] replace NMA by the less costly BD for the high-frequency modes and therefore keep the fast oscillations around their equilibrium values; the low-frequency modes are propagated by Langevin dynamics as in LIN.

An interesting recent algorithm for sampling the canonical distribution is based on the Nosé-Hoover formulation in the context of Markov chain MC [38]; this solution to the lack of ergodicity in the context of Nosé-Hoover integrators is similar to the idea of Cooke and Schmidler [37] for REMD. Systematic discretization errors are also eliminated by this hybrid MC approach for canonical sampling.

In this altered MD algorithm category, a noteworthy and unusual example is interactive MD (IMD) developed by the Schulten group to probe mechanisms of biological reactions using computer graphics and simulation (VMD + NAMD) [44,45]. The user steers the system to the desired state by 'feeling' the potential force and applying any desired magnitude (to 'pull' the system). IMD is particularly appropriate for studies of induced-fit mechanisms, in which substrate/solvent/protein interplay are crucial for selectivity and biological function. It requires, however, a sophisticated device on the user's desktop.

Note that MD in internal-coordinate space has been used sporadically with the rationale that the fewer degrees of freedom compared with Cartesian coordinates will allow longer integration timesteps and hence greater sampling. Indeed, peptide folding and refinement with dihedral-angle MD demonstrated a computational advantage of several orders of magnitude compared to Cartesian analogs [46]. However, internal-coordinate MD is not generally used, possibly due to the significant effect of vibrational coupling on biomolecular dynamics and the computational difficulties involved in transforming internal coordinates to/from Cartesian coordinates; it is nonetheless worth more consideration in the context of simplified models, such as macroscopic united-residue models of proteins and nucleic acids.

Global reformulations
Many innovative methods have been developed by mathematicians, physicists, and chemists in this category of algorithmic reformulations. (Note: citations below often refer to a recent work rather than the original paper, which can be found in the recent article.)

Examples of methods that aim to explore the free energy landscape include: (a) canonical adiabatic free energy sampling (CAFES) [47], which propagates dynamics of decoupled solute and solvent systems so that the former follows adiabatic dynamics of the free energy surface created by the latter at increased temperatures; (b) reference potential spatial warping algorithm (REPSWA) [48], which introduces a variable transformation in the classical partition function that increases attraction to basins; (c) metadynamics [49], which involves exploring the free energy surface by following non-Markovian dynamics determined by rewriting the equations of motion in terms of a few collective variables so that key regions of space are identified as simulation time increases; and (d) a sweep method that combines temperature-accelerated MD (TAMD) with a free energy reconstruction from the mean force using radial basis functions via optimization [50].

Methods that deduce mechanisms include: (a) TPS by Chandler and colleagues, which follows an MC protocol in MD-trajectory space to locate key transition states (as recently reviewed in [51]); (b) the nudged elastic band (NEB) method [52] that optimizes minimum energy pathways; (c) the max-flux approach [53], which constructs variationally optimized reaction pathways based on the max flux for diffusive dynamics; and (d) the string method [54], which prunes MD trajectories to retain reactive segments.

Methods that compute reaction rates include (a) FFSs [55], which use interfaces to partition phase space; (b) 'milestoning' [56], which coarse-grains temporal and spatial descriptors of the system by sequential transitions; and (c) the popular Markov state models (MSMs) [57], which define transition networks to describe biomolecular kinetics and thermodynamics. Essentially, all of these methods compute rates by various divide-and-conquer strategies.

TPS developed by Chandler, Bolhuis, Dellago, and Geissler has been popular and successful in various biomolecular application contexts [51]. The idea is to run an MC in trajectory space in a way to create reactive trajectories that preserve microscopic reversibility and local thermal distributions. Full kinetic information can be extracted by clever extensions (such as efficient US, BOLAS [58], and network models). For example, an application of TPS to DNA polymerase β deduced the mechanism, free energy profile, and rate for pol β's closing conformational change [59]: a complex landscape was revealed where sequential, subtle, side-chain rearrangements lead the enzyme from open to the closed state and where Arg258 is a 'gatekeeper' for the reaction (Figure 1, top right); the overall computed rate of 10 per second corresponds to the $27k_BT$ barrier and agrees well with the experimental value of 3 to 10 per second for the overall reaction. The application of BOLAS [58], instead of US, was crucial for achieving acceptable margins of error; unlike US, BOLAS avoids large biases (which in turn lead to large error bars) by sampling each window by many short MD trajectories.

The same TPS protocol applied to the mismatch (G:A instead of G:C) suggested that the higher free energy barrier for the mismatch comes from the instability of the closed mismatched state compared with the matched base pair system [60].

Rogal and Bolhuis [61] recently extended TPS to connect several intermediate states in phase space.

The idea in the popular MSMs [57] is to compute reaction rates from different interfaces. Combinations such as by Pan and Roux [62], which employ the single-sweep method of Vanden-Eijnden and colleagues [50] with MSM, can be successful for proteins because phase space is surveyed to map dynamically important regions, from which free simulations are initiated, and then the transition matrix is constructed by piecing the information together.

The FFS approach [55] similarly enhances phase space sampling of rare events in stochastic non-equilibrium systems in which the phase space distribution is not known *a priori*. The phase space is partitioned, and an adaptive procedure is used to find kinetic 'bottleneck regions' by estimating rate constants associated with reaching subsequent interfaces. Then, FFS concentrates on sampling those bottleneck regions only. FFS is appropriate for discretely defined surfaces, like protein folding on a lattice; a recent application located the most probable transition state isosurfaces for a 48-unit protein [63].

Finally, the similar idea in milestoning [56] is to recover reaction kinetics by integral equations and global path optimization strategies. Milestoning is compared with MSM, FFS, and TIS in [64], where advantages of milestoning compared with the other methods are suggested in terms of accuracy, efficiency, and parallel-machine implementations.

Numerous other innovative mathematical methods use various constructs from physics, mathematics, and engineering such as domain decomposition, clustering techniques, or Markov models to sample and/or survey conformational space and dynamics. For example, a robust Perron cluster analysis propagates a system's essential dynamics with preservation of Markov properties [65], and a hierarchical clustering approach by a Markov model helps identify groups of configurations with common features and transitions between them [66].

Future directions

Certainly, long MD simulations are being reported more regularly now compared with several years ago as a result of efficient and parallel simulation packages like NAMD, Desmond, GROMACS, and others [5-8], and new computer systems like Anton, hard-wired for long MD simulations [67]. For example, a 1-microsecond atomic-level simulation of the voltage-modulated potassium channel Kv1.2 (120,000 atoms) recently described the opening/closing mechanism involved [8]. Thus, many problems thought to be intractable several decades ago are now being solved. But faster computer technology alone is insufficient to address challenging biomolecular problems. Algorithmic developments for enhanced sampling such as described here, especially those divide-and-conquer methods suitable for loosely coupled processor architectures that are more readily available to the average user, are needed to complement long atomistic simulations.

While simple potential modifications like TMD can provide conformational insights and REMD approaches can enhance sampling, success has been empirical rather

than rigorous; because the actual kinetics of the systems are altered, caution is warranted in biological interpretations. However, enhanced sampling protocols can be useful in specific contexts, as recently demonstrated by using accelerated MD [20] to reproduce residual dipolar coupling measurements concerning the slow modes from nuclear magnetic resonance data on a domain of the protein GB3 [68]. Furthermore, the studies pinpointed the slow motions and led to a model of the potential energy landscape of the protein.

REMD has recently received some scrutiny concerning effective ensemble exchange protocols, general computational efficiency, and ergodicity questions, and this has led to variants with stochastic elements - motivated by REMD's origin in MC parallel tempering methods - that can improve sampling (for example, [37]). Because REMD applications require concurrent processors (one per replica), the technique may not always be practical, especially for large atomistic systems. Variants like the distributed replica sampling [34] can be more viable by performing stochastic moves of independent replicas instead of pairwise exchanges of replicas.

TPS, MSMs, and approaches that aim to compute reaction rates have been more rigorously grounded in theory, and successful biomolecular applications have been reported. Still, their application to biomolecules in general remains far from routine. In this goal, various combinations of methods that deduce mechanisms and compute reaction rates by divide-and-conquer approaches will undoubtedly be effective in today's readily available distributed computing resources of cluster networks, especially by enhancing them with coarse-grained models and MC elements.

Ultimately, the most successful applications reflect a combination of novel and rigorous mathematical ideas with physical intuition. And such applications can be successful not only by reproducing experimental data concerning slow motions but also in helping suggest mechanistic and energetic details. Clearly, the gap between experimental and theoretical timeframes is steadily narrowing, leading to renewed interest in MD-based modeling by theorists and experimentalists alike.

Abbreviations
BD, Brownian dynamics; CAFES, canonical adiabatic free energy sampling; FFS, forward flux simulation; FRET, fluorescent resonance energy transfer; GROMACS, GROningen MAchine for Chemical Simulation; IMD, interactive molecular dynamics; LES, locally enhanced sampling; LIN; Langevin, implicit Eular and normal-mode; MC, Monte Carlo; MD, molecular dynamics;

MSM, Markov state model; NAMD, nanoscale molecular dynamics; NEB, nudged elastic band; NMA, normal-mode analysis; NML, normal-mode Langevin; OPEP, optimized potential for efficient structure prediction; PCA, principal component analysis; REMD, replica exchange molecular dynamics; REPSWA, reference potential spatial warping algorithm; TAMD, temperature-accelerated MD; TIS, transition interface sampling; TMD, targeted molecular dynamics; TPS, transition path sampling; TSE, transition state ensemble; US, umbrella sampling; VMD, visual molecular dynamics.

Competing interests
The author declares that she has no competing interests.

References
1. McCammon JA, Gelin BR, Karplus M: **Dynamics of folded proteins.** *Nature* 1977, **267**:585.

2. Maddox J: **Statistical mechanics by numbers.** *Nature* 1988, **334**:561.

3. Laplace PS: **Théorie analytique des probabilités.** In *Oevres Complètes.* Vol. 7. Paris: Courcier Imprimeur; 1812.

4. Dirac PA: **Quantum mechanics of many-electron systems.** *Proc R Soc Lond A* 1929, **123**:714-33.

5. Pérez A, Luque FJ, Orozco M: **Dynamics of B-DNA on the microsecond time scale.** *J Am Chem Soc* 2007, **129**:14739-45.

6. Maragakis P, Lindorff-Larsen K, Eastwood MP, Dror RO, Klepeis JL, Arkin IT, Jensen MØ, Xu H, Trbovic N, Friesner RA, Iii AG, Shaw DE: **Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins.** *J Phys Chem B* 2008, **112**:6155-8.

7. Freddolino PL, Liu F, Gruebele M, Schulten K: **Ten-microsecond molecular dynamics simulation of a fast-folding WW domain.** *Biophys J* 2008, **94**:L75-L77.

8. Bjelkmar P, Niemelä PS, Vattulainen I, Lindahl E: **Conformational changes and slow dynamics through microsecond polarized atomistic molecular simulation of an integral Kv1.2 ion channel.** *PLoS Comput Biol* 2009, **5**:e1000289.

9. Foley M, Schlick T: **Simulations of DNA pol lambda R517 mutants indicate 517's crucial role in ternary complex stability and suggest DNA slippage origin.** *J Am Chem Soc* 2008, **130**:3967-77.

10. Lei H, Duan Y: **Improved sampling methods for molecular simulation.** *Curr Opin Struct Biol* 2007, **17**:187.

11. Liwo A, Czaplewski C, Oldziej S, Scheraga HA: **Computational techniques for efficient conformational sampling of proteins.** *Curr Opin Struct Biol* 2008, **18**:134-9.

12. Schlick T: **Monte Carlo, harmonic approximation, and coarse-graining approaches for enhanced sampling of biomolecular structure.** *F1000 Biology Reports* 2009, **1**:48.

13. Arora K, Schlick T: **In silico evidence for DNA polymerase-β's substrate-induced conformational change.** *Biophys J* 2004, **87**:3088-99.

14. Sun J, Zhang Q, Schlick T: **Electrostatic mechanism of nucleosomal array folding revealed by computer simulation.** *Proc Natl Acad Sci U S A* 2005, **102**:8180-5.

15. Kong Y, Shen Y, Warth TE, Ma J: **Conformational pathways in the gating of Escherichia coli mechanosensitive channel.** *Proc Natl Acad Sci U S A* 2002, **99**:5999-6004.

> F1000 Factor 3.0 *Recommended*
> Evaluated by Volkhard Helms 17 Jun 2002

16. Helms V: **Evaluation of: Kong Y, Shen Y, Warth TE, Ma J: Conformational pathways in the gating of Escherichia coli mechanosensitive channel.** *Proc Natl Acad Sci U S A* 2002, **99**:5999-6004. Faculty of 1000 Biology. 2002 Jun 17.

17. Okazaki K, Koga N, Takada S, Onuchic JN, Wolynes PG: **Multiple-basin energy landscapes for large-amplitude conformational motions of proteins: structure-based molecular dynamics simulations.** *Proc Natl Acad Sci U S A* 2006, **103**:11844-9.

> F1000 Factor 6.0 *Must Read*
> Evaluated by David Osguthorpe 01 Sep 2006

18. Mills M, Andricioaei I: **An experimentally guided umbrella sampling protocol for biomolecules.** *J Chem Phys* 2008, **129**:114101.

19. Paci E, Lindorff-Larsen K, Dobson CM, Karplus, Vendruscolo M: **Transition state contact orders correlate with protein folding rates.** *J Mol Biol* 2005, **352**:495-500.

20. Hamelberg D, de Oliveira CA, McCammon JA: **Sampling of slow diffusive conformational transitions with accelerated molecular dynamics.** *J Chem Phys* 2007, **127**:155102.

21. Grubmüller H: **Predicting slow structural transitions in macromolecular systems: Conformational flooding.** *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 1995, **52**:2893-906.

22. Voter AF: **A method for accelerating the molecular dynamics simulation of infrequent events.** *J Chem Phys* 1997, **106**:4665-77.

23. Wang JC, Pal S, Fichthorn KA: **Accelerated molecular dynamics of rare events using the local boost method.** *Phys Rev B* 2001, **63**:085403.

24. Miron RA, Fichthorn KA: **Accelerated molecular dynamics with the bond-boost method.** *J Chem Phys* 2003, **119**:6210-6.

25. Roitberg A, Elber R: **Modeling of side chains in peptides and proteins: application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations.** *J Chem Phys* 1991, **95**:9277-87.

26. Sugita Y, Okamoto Y: **Replica-exchange molecular dynamics methods for protein folding.** *Chem Phys Lett* 1999, **314**:141-51.

27. Hansmann UE: **Parallel-tempering algorithm for conformational studies of biological molecules.** *Chem Phys Lett* 1997, **281**:140-50.

28. Garcia AE, Pascheck D: **Simulation of the pressure and temperature folding/unfolding equilibrium of a small RNA hairpin.** *J Am Chem Soc* 2008, **130**:815-7.

29. Garcia AE, Onuchic JN: **Folding a protein in a computer: an atomic description of the folding/unfolding of protein A.** *Proc Natl Acad Sci U S A* 2003, **100**:13898-903.

> F1000 Factor 4.8 *Must Read*
> Evaluated by Victor Munoz 26 Jan 2004, Martin Gruebele 27 Jan 2004

30. **Index of /~angel/movies.** [http://www.rpi.edu/~angel/movies]

31. Lei H, Wu C, Liu H, Duan Y: **Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations.** *Proc Natl Acad Sci U S A* 2007, **104**:4925-30.

> F1000 Factor 3.0 *Recommended*
> Evaluated by Richard Henchman 20 Apr 2007

32. Sindhikara D, Meng Y, Roitberg AE: **Exchange frequency in replica exchange molecular dynamics.** *J Chem Phys* 2008, **128**:024103.

33. Roitberg AE, Okur A, Simmerling C: **Coupling of replica exchange simulations to a non-Boltzmann structure reservoir.** *J Phys Chem B* 2007, **111**:2415-8.

34. Rodinger T, Howell PL, Pomès R: **Distributed replica sampling.** *J Chem Theory Comput* 2006, **2**:725-31.

35. Kamberaj H, van der Vaart A: **Multiple scaling replica exchange for the conformational sampling of biomolecules in explicit water.** *J Chem Phys* 2007, **127**:234102.

36. Zuckerman DM, Lyman E: **Second Look at Canonical Sampling of Biomolecules using Replica Exchange Simulation.** *J Chem Theory Comput* 2006, **2**:12001202.

37. Cooke B, Schmidler SC: **Preserving the Boltzmann ensemble in replica-exchange molecular dynamics.** *J Chem Phys* 2008, **129**:164112.

38. Leimkuhler B, Reich S: **A Metropolis adjusted Nose Hoover thermostat.** *Math Mod Num Anal,* in press.

39. Chebaro Y, Dong X, Laghaei R, Derreumaux P, Mousseau N: **Replica exchange molecular dynamics simulations of coarse-grained proteins in implicit solvent.** *J Phys Chem B* 2009, **113**:267-74.

40. Bolhuis PG: **Rare events via multiple reaction channels sampled by path replica exchange.** *J Chem Phys* 2008, **129**:114108.

41. Schlick T: *Molecular Modeling and Simulation: An Interdisciplinary Guide.* New York: Springer-Verlag New York, Inc; 2002.

42. Sweet CR, Petrine P, Pande VS, Izaguirre JA: **Normal mode partitioning of Langevin dynamics for biomolecules.** *J Chem Phys* 2008, **128**:145101.

> F1000 Factor 3.0 *Recommended*
> Evaluated by Nathan Baker 20 May 2008

43. Zhang G, Schlick T: **LIN: a new algorithm combining implicit integration and normal mode techniques for molecular dynamics.** *J Comput Chem* 1993, **4**:1212-33.

44. Grayson P, Tajkhorshid E, Schulten K: **Mechanisms of selectivity in channels and enzymes studied with interactive molecular dynamics.** *Biophys J* 2003, **85**:36.

45. Schlick T: **Engineering teams up with computer-simulation and visualization tools to probe biomolecular mechanisms.** *Biophys J* 2003, **85**:1.

46. Chen J, Im W, Brooks CL 3rd: **Application of torsion angle molecular dynamics for efficient sampling of protein conformations.** *J Comput Chem* 2005, **26**:1565-78.

47. VandeVondele J, Rothlisberger U: **Canonical adiabatic free energy sampling (CAFES): a novel method for the exploration of free energy surface.** *J Phys Chem B* 2002, **106**:203-8.

48. Minary P, Tuckerman ME, Martyna GJ: **Dynamical spatial warping: a novel method for the conformational sampling of biophysical structure.** *SIAM J Sci Comput* 2008, **30**:2055-83.

49. Parrinello M: **Eppur si muove.** In *Physical Biology: 4D Visualization of Complexity.* Edited by Zewail AH. London: Imperial College Press; 2008:247-66.

50. Maragliano L, Vanden-Eijnden E: **Single-sweep methods for free energy calculations.** *J Chem Phys* 2008, **128**:184110.

51. Dellago C, Bolhuis PG: **Transition path sampling simulations of biological systems.** *Top Curr Chem* 2007, **268**:291-317.

52. Henkelman G, Jónsson H: **Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points.** *J Chem Phys* 2000, **113**:9978-85.

53. Huo S, Straub JE: **The MaxFlux algorithm for calculating variationally optimized reaction paths for conformational transitions in many body systems at finite temperature.** *J Chem Phys* 1997, **107**:5000-6.

54. Miller TF, Vanden-Eijnden E, Chandler D: **Solvent coarse-graining and the string method applied to the hydrophobic collapse of a hydrated chain.** *Proc Natl Acad Sci U S A* 2007, **104**:14559-64.

55. Borrero EE, Escobedo FA: **Optimizing the sampling and staging for simulations of rare events via forward flux sampling schemes.** *J Chem Phys* 2008, **129**:024115.

56. Elber R: **A milestoning study of the kinetics of an allosteric transition: atomically detailed simulations of deoxy scapharca hemoglobin.** *Biophys J* 2007, **92:**L85-L87.

57. Noé F, Fischer S: **Transition networks for modeling the kinetics of conformational change in macromolecules.** *Curr Opin Struct Biol* 2008, **18:**154-62.

58. Radhakrishnan R, Schlick: **Biomolecular free energy profiles by a shooting/umbrella sampling protocol, 'BOLAS'.** *J Chem Phys* 2004, **121:**2436.

59. Radhakrishnan R, Schlick T: **Orchestration of cooperative events in DNA synthesis and repair mechanism unraveled by transition path sampling of DNA polymerase beta's closing.** *Proc Natl Acad Sci U S A* 2004, **101:**5970-5.

60. Radhakrishnan R, Schlick T: **Fidelity discrimination in DNA polymerase β: differing closing profiles for a mismatched (G:A) versus matched (G:C) base pair.** *J Am Chem Soc* 2005, **127:**13245-52.

61. Rogal J, Bolhuis PG: **Multiple state transition path sampling.** *J Chem Phys* 2008, **129:**224107.

62. Pan AC, Roux B: **Building Markov state models along pathways to determine free energies and rates of transitions.** *J Chem Phys* 2008, **129:**064107.

63. Allen RJ, Frenkel D, ten Wolde PR: **Simulating rare events in equilibrium or nonequilibrium stochastic systems.** *J Chem Phys* 2006, **124:**024102.

64. Vanden-Eijnden E, Venturoli M, Ciccotti G, Elber R: **On the assumptions underlying milestoning.** *J Chem Phys* 2008, **129:**174102.

65. Weber M, Kube S, Wlater L, Deuflhard P: **Stable computational of probability densities for metastable dynamical systems.** *Multiscale Model Simul* 2007, **6:**396-416.

66. Noé F, Horenko I, Schütte C, Smith JC: **Hierarchical analysis of conformational dynamics in biomolecules: transition networks of metastable states.** *J Chem Phys* 2007, **126:**155102.

67. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, Eastwood MP, Gagliardo J, Grossman JP, Ho CR, Ierardi DJ, Kolossváry I, Klepeis JL, Layman T, McLeavey C, Moraes MA, Mueller R, Priest EC, Shan Y, Spengler J, Theobald M, Towles B, Wang SC: **Anton, a special-purpose machine for molecular dynamics simulation.** In *Proceedings of the 34th Annual International Symposium on Computer Architecture (ISCA): 9-13 June 2007; San Diego, California.* New York: Association for Computing Machinery; 2007.

68. Markwick PRL, Bouvignies G, Blackledge M: **Exploring multiple timescale motions in protein GB3 using accelerated molecular dynamics and NMR spectroscopy.** *J Am Chem Soc* 2007, **129:**4724-30.