Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Pharmacology

Reaching biological timescales with all-atom molecular dynamics simulations

Matthew C Zwier and Lillian T Chong

Molecular dynamics (MD) simulations can provide atomically detailed views of protein motions, sampling multiple timescales ranging from femtoseconds to nanoseconds on typical computing resources. The 'reach' of these computer simulations toward biologically relevant timescales (microseconds and beyond) has been improving with advances in hardware and software, as well as the development of enhanced sampling techniques. This review outlines these advances, focusing on techniques that also provide realistic, unperturbed kinetics. These longer-timescale MD simulations can provide detailed insights into the mechanisms of biological events, potentially aiding the design of pharmaceuticals.

Address
Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260, United States

Corresponding author: Chong, Lillian T (ltchong@pitt.edu)

Current Opinion in Pharmacology 2010, **10**:745–752

This review comes from a themed issue on
New technologies – the importance of protein dynamics
Edited by Dennis R Livesay

Available online 8th October 2010

1471-4892/$ – see front matter

DOI 10.1016/j.coph.2010.09.008

Introduction

Many biological processes — including enzyme catalysis, signal transduction, and protein–protein binding — involve protein motions that occur on multiple timescales [1]. As illustrated in Figure 1, these motions include ps–ns dynamics of side chains, ns–μs relative motions of protein domains, and μs–ms allosteric transitions [2]. Furthermore, the shorter-timescale dynamics can influence and be influenced by longer timescale motions [3,4]. The flexibility of proteins and the associated ensemble of alternate conformational states are important for many pharmaceutically relevant species [3,5,6].

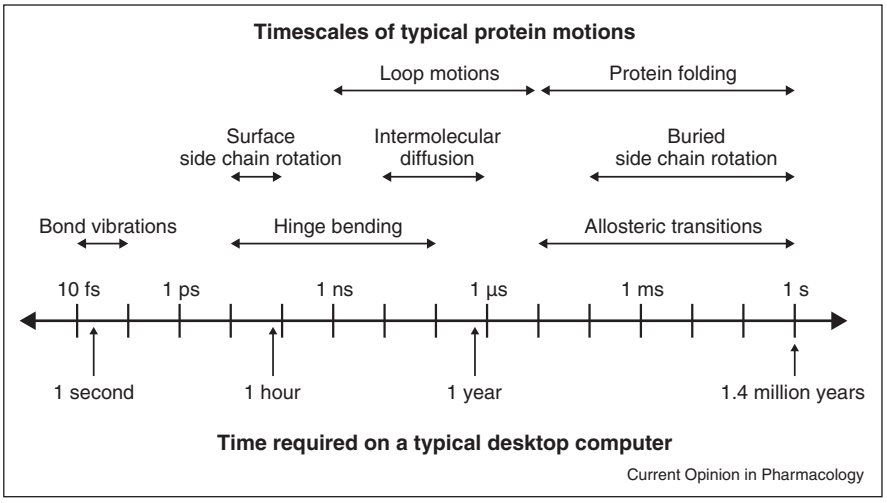
Although X-ray crystallography or NMR spectroscopy can provide ensemble-averaged structures of certain conformational states, they cannot always characterize short-lived or unstructured species, such as transient binding pockets [7], alternative conformations of active sites [8], or proteins with intrinsically disordered regions [9,10], and it may be precisely those species that could lead to new classes of pharmaceuticals [6,11]. Molecular dynamics (MD) simulations can complement experiments by providing the time resolution and atomic detail necessary for monitoring the step-by-step progression of conformational changes (e.g. the opening and closing of active-site protein 'flaps'). Given sufficient computational resources, such simulations can span multiple timescales, revealing how fast, local fluctuations (ps–ns) might facilitate slower, functionally relevant collective motions of the protein ($\geq$μs), providing detailed views of the mechanisms of conformational transitions. However, typical computing resources limit these simulations, which ideally include explicit water molecules, to the nanosecond timescale. Thus, direct 'brute force' simulations — simply running simulations for a sufficiently long time (i.e. many times longer than the slowest event of interest) — have limited use in capturing biologically relevant motions (e.g. induced-fit binding [12]). As illustrated in Figure 1, many biologically relevant motions (fs–ns) are readily accessible to modern computers, but many motions which may be of interest (μs and beyond) are far out of reach.

The desire to access biological timescales with MD simulations has driven the development of innovative enhanced sampling techniques. These techniques invariably increase computational throughput at the cost of introducing additional assumptions, e.g. the system under study is strictly at equilibrium, or that initial and final states of a transition can be unambiguously identified and rigorously defined. Common to all these enhanced sampling techniques is the assumption of a separation of timescales, where a process has a long characteristic time not because the transitions involved are slow, but rather because the transitions are rare, with long waiting times between otherwise fast events. It is the elimination of this waiting time that allows these techniques to access biologically relevant timescales.

Here, we discuss recent developments in both brute force simulation and enhanced sampling techniques. Because of space constraints, we have restricted our discussion to methods which involve motion on a single, unmodified free-energy surface.[a] The distinct advantage of such

[a] The free energy surface of a chemical system — free energy as a function of atomic coordinates — completely defines its dynamics; it depends on the microscopic interactions between atoms and macroscopic thermodynamic variables such as temperature, volume, and pressure.

Figure 1



Timescales of typical protein motions and estimated computational time to simulate them. Motions and their corresponding timescales are indicated above the axis. Below the axis is a rough estimate of the amount of 'wallclock' time required to perform a molecular dynamics (MD) simulation of a typically sized protein–protein complex solvated in explicit water (~45,000 atoms) on a typical (2.6 GHz dual-core) desktop computer. To capture the fastest motions of proteins (i.e. bond vibrations), MD simulations must employ femtosecond time steps; a large number of simulation steps are therefore required to reach biological timescales, making MD simulations of protein systems very computationally expensive. For the computer described above, tens of nanoseconds of dynamics are accessible within weeks, but to reach the millisecond timescale would require millennia (timescales are from [2,75], except for intermolecular diffusion, which is derived from [75] assuming 1 mM concentration of the diffusing species).

approaches is that the dynamics of the system under study are completely unperturbed, and furthermore, realistic kinetic information may be readily extracted from simulations; this information (e.g. kinetic rates and timescales of motion) provides another means of validating simulation with experiment. On the other hand, this admittedly limited scope precludes us from discussing in detail other promising enhanced sampling methods [2,13], including those that under some circumstances can yield realistic kinetic rates [14•,15•]. In preparing this review, we found it helpful to summarize the similarities and differences among each of the many techniques discussed below. As shown in Figure 2, the techniques discussed here can be grouped according to whether they provide continuous, atomically detailed pathways of transitions between two states or not, and also according to the amount of *a priori* information required to construct a simulation. Generally, methods that require more information about a system involve more assumptions, but can generate pathways (or trajectories) of the biological event of interest with greater efficiency.

Brute force dynamics

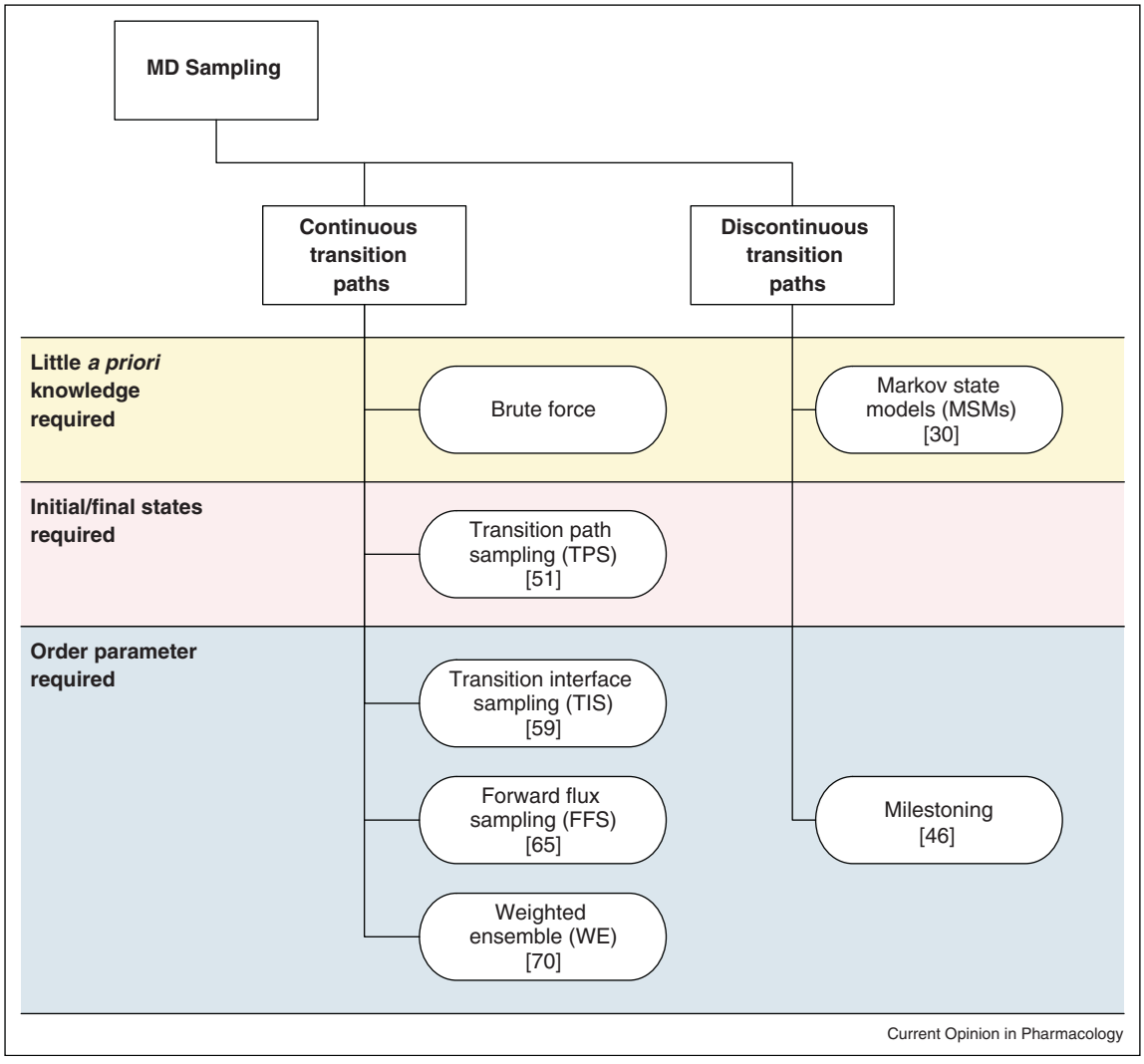
When feasible, brute force simulations provide the greatest possible detail with the fewest possible assumptions relative to other MD-based sampling techniques. As highlighted in Figure 2, very little *a priori* knowledge is required to run a brute force dynamics simulation, generally only a force field and a representative initial structure. These are not trivial concerns, particularly as the force field completely defines the thermodynamics of a system [16], and slight imbalances in force field parameterization can significantly alter the results of simulations [17], particularly for long timescales [18]; however, agreement with experiment is generally acceptable for the biological questions being addressed [16,19].

Although the cost of running these brute force simulations is high, computing resources continue to grow in size and power. Continued optimization of MD software, coupled with the decreasing cost of commodity hardware[b] — particularly multicore processors — has played a key role in reaching microsecond timescales for typical biological systems (e.g. $\sim 10^4$–10^5 atoms), timescales that were inaccessible only five years ago. Furthermore, recently developed hardware specialized for performing MD simulations is poised to generate microseconds of dynamics per day on similarly sized biological systems [19,20•].

Although the overall computing landscape has not changed dramatically in the last year, the use of general-purpose graphics processing units (GPGPUs) in the field of MD continues to grow (cf., e.g. [21–25]; AMBER 11, University of California, San Francisco; GROMACS, URL: http://www.gromacs.org; OpenMM 2.0, URL: https://simtk.org/home/openmm). While impressive gains in throughput are possible with GPGPUs for certain calculations involved in an MD simulation [26], the small

[b] That is, consumer-grade or business-grade general-purpose computing hardware, as opposed to specialized 'supercomputing' hardware.

Figure 2



Techniques capable of accessing biomolecular timescales. When feasible, large-scale brute force dynamics with explicit consideration of solvent provide the greatest possible detail with the fewest possible assumptions. Where such simulations are not possible, other enhanced sampling techniques can be used to obtain kinetic information, transition paths, or both. In general, where increased *a priori* knowledge (yellow, red, and blue regions) is required to run a simulation, greater efficiency gains compared to brute force dynamics are possible; the price of increased throughput is often a greater number of assumptions about or restrictions on the system being simulated. For convenience, references are provided in which each enhanced sampling method was first presented for the simulation of biochemical systems. In the case of Markov state models, the first discussion of using multiple shorter-timescale simulations to reach longer-timescale kinetics of biological systems is cited.

on-board memory and high cost of communication with GPGPUs has severely limited the ability of GPGPUs to accelerate MD calculations effectively [24]. This communication problem must be overcome before GPGPU-accelerated MD calculations can hope to supplant traditional large-scale parallel MD calculations, which continue to provide access to long-timescales [27] and large systems [28•].

Kinetic 'glue'

The high computational cost of accessing biological timescales with brute force simulations (see Figure 1) has led to the development of several methods that attempt to obtain long-timescale information by 'gluing together' shorter timescale simulations. Two examples with proven applicability to biological systems are Markov state models and Milestoning.

Markov state models

Markov state models (MSMs) — discrete-state kinetic models — seek to describe the equilibrium behavior of a system in terms of a finite number of metastable (relatively long-lived) states and the rates of transitions between them [29–36]. These transition networks are

generally constructed by grouping many conformations from multiple, relatively short brute force simulations such that conformational transitions within states (groups) are common but transitions between states are rare [30,32]. The requirements for this statewise decomposition of conformational space can be expressed in several ways: firstly, the timescales for intrastate transitions are short but the timescales for interstate transitions are long; secondly, the probability of having moved from some state i at time t to some other state j at time $t + dt$ depends only on the lag time dt; or thirdly, the probability of moving from state i to state j does not depend on how the system came to be in state i. It is the first property (the separation of timescales) that enables MSMs to capture long-timescale kinetic information from short-timescale dynamics, and it is the lattermost property ('memorylessness') that is perhaps most familiar as the defining property of a Markov process, hence the name Markov state model.

Construction of a Markov state model reduces many individual conformations and the detailed dynamics connecting them into a discrete set of states, their relative probabilities, and a matrix of (lag time-dependent) transition probabilities between each pair of states [30]. Each state represents a distribution of quickly interconverting conformations, which may (but need not) correspond to experimentally well-defined populations, such as those that might be identified by NMR spectroscopy. From the transition probability matrix and state populations, quantities such as the overall transition rate between two states, the set of paths connecting them, and the contribution of each path to the overall rate can be calculated [34,37]. Thus, MSMs present a coarse-grained view of both the conformational space of the system and the dynamics within it. MSMs have been used primarily to study protein folding mechanisms, showing good agreement with experiment in both structural information and folding rates [31,32,35,36,38,39,40••,41–43]. However, these models are generally applicable to the problems of identifying kinetically distinct states of proteins under given conditions (e.g. temperature and ionic strength) and determining the kinetics of slow conformational transitions. The particular strength of MSMs may in fact be their ability to indicate native state ensembles of structures, providing information complementary to that provided by X-ray crystallography and NMR spectroscopy experiments.

It should be noted that short trajectories suitable for MSM construction may be generated with replica exchange molecular dynamics (REMD) [44], a popular method for enhancing sampling of conformational space [45]. Many copies ('replicas') of a system are simulated in parallel with multiple temperatures, and configurations are occasionally swapped between temperatures (typically according to a Boltzmann criterion). Although this technique does not generally permit the extraction of reaction rates, Markov state models can be constructed from the brief trajectory segments between replica exchanges, and reaction rates may then be determined from the MSMs [14•].

Milestoning

The Milestoning approach also uses kinetic information from shorter timescale simulations to infer long-timescale kinetics [46,47••]. Unlike MSMs, from which definitions of states may be obtained, Milestoning requires *a priori* definitions of initial and final states and a one-dimensional order parameter[c] that specifies 'how far along' a simulation is in a transition between the initial and final states. This order parameter is divided by a number of surfaces ('milestones') and equilibrated ensembles of simulations are prepared at each milestone. In a second simulation phase, the constraint holding simulations to milestone surfaces is removed, and as simulations reach neighboring milestones, the time required by each simulation to reach a neighboring milestone (in either a forward or backward direction) is recorded. This simulation between milestones — rather than between initial and final states — effectively eliminates the waiting time that would otherwise be sampled by brute force simulations.

The central assumption of Milestoning is that all degrees of freedom other than the order parameter relax completely between subsequent milestones. Under this assumption, the 'incubation times' between milestones, obtained as described above, may be transformed into the global first-passage time distribution, the probability distribution of times required to reach the final state from the initial state [47••]. When a single timescale dominates a system, the first-passage time distribution is exponential and the reaction rate is simply the inverse of the mean first-passage time, but in a system where multiple timescales are important, the first-passage time distribution is capable of describing the resulting non-exponential behavior [48], as has been demonstrated explicitly for Milestoning simulations [46,47••]. This added flexibility reflects the fact that the central assumption of Milestoning (complete relaxation along all non-order-parameter coordinates) is less restrictive than that of Markov state models (where complete relaxation is assumed in *all* degrees of freedom within each state). However, the cost of this increased detail in the determination of kinetics is reduced detail in the determination of conformational states; initial and final states, the order parameter, and adequate milestones must be known before a Milestoning simulation of a system. The utility of the Milestoning approach is demonstrated

[c] This order parameter (also called a 'progress coordinate' by some practitioners) is a scalar value which varies continuously and monotonically between particular values at the initial and final states. It may, but does not necessarily, reflect a formal reaction coordinate.

well by a recent study involving the recovery stroke of myosin (a millisecond process), which provided experimentally testable mechanistic insights and a rate consistent with experiment [49].

Path sampling techniques

Path sampling approaches seek to determine the detailed dynamics of pathways between well-defined metastable states. These techniques are complementary to MSMs and Milestoning, which can provide definitions of metastable states and detailed kinetics of transitions between well-defined metastable states, respectively. The most widely used methods in recent years are transition path sampling (TPS) and its variants, such as transition interface sampling (TIS); forward flux sampling (FFS); and weighted ensemble (WE) sampling.

Transition path sampling

TPS, which is based on early work by Pratt [50], was first presented more than a decade ago [51] and has subsequently been extensively employed, refined, and reviewed [52,53,54•,55•]. TPS is a Monte Carlo sampling of MD-simulated paths between initial and final states, which (as highlighted in Figure 2) must be known *a priori*. Each path is typically generated by randomly selecting a segment of the previously sampled path, perturbing its coordinates and/or momenta, and then 'shooting off' MD trajectories both forward and backward in time from the perturbed segment [53]; thus, this scheme requires the dynamics of the system to be invariant under time reversal, i.e. the system must be at equilibrium [54]. The resulting set of paths between the initial and final states and their relative probabilities provide a detailed picture of how transitions between the initial and final states progress. TPS does not directly provide kinetic information; rather, a subsequent (computationally expensive) umbrella sampling calculation is required [53], a limitation which led directly to the development of TIS (discussed below). As with all path sampling methods, the presence of long-lived intermediate states [56] or multiple distinct transition pathways separated by substantial free-energy barriers (Zhang, *et al.*, unpublished data; URL: http://arxiv.org/abs/0902.2772) may severely reduce the effectiveness of TPS. Nonetheless, TPS is capable of describing rare transitions in biological systems. Recently, TPS was used to determine the pathways of conformational change in the activation of photoactive yellow protein (PYP), predicting experimentally detectable intermediates and suggesting experiments which can be used to validate the TPS results [57]. In a striking combination of a number of enhanced sampling techniques, TPS was used to determine the pathways of conformational change in folding and unfolding mechanisms of formin binding protein 28 (FBP28) and then map the free energy landscape of the protein; the computed unfolding barrier is in agreement with experiment [58••].

Transition interface sampling

The high computational cost of obtaining kinetic information with TPS inspired the development of TIS and several variants thereof [59–61]. TIS, along with FFS and WE, partitions an order parameter connecting initial and final states with several dividing surfaces ('interfaces'); this again represents an increase in the amount of information required to start a simulation (see Figure 2). In TIS, a Monte Carlo procedure highly similar to TPS — including forward and backward shooting of MD trajectories — is used to sample paths between each pair of adjacent interfaces; the reaction rate is then determined by the flux out of the initial state and the conditional probabilities of reaching each interface in turn [62]. In this way, the paths and transition rate between initial and final states are determined simultaneously. Interface-based sampling methods like TIS may suffer greatly in efficiency if significant free energy barriers exist between interfaces, particularly if the barriers must be surmounted in order to reach the next interface [63].

Forward flux sampling

The FFS method was presented as an alternative to TPS and TIS without the requirement of microscopic reversibility, thus allowing path sampling studies of nonequilibrium systems [54•,55•,64–66]. Like TIS, FFS requires well-defined initial and final states, an order parameter describing the transition between them, and partitioning of the order parameter by interfaces. Rather than using Monte Carlo techniques to sample transition paths, MD simulations — propagating forward in time only — are used to determine the paths between interfaces. When a dynamics trajectory reaches an interface, its coordinates and momenta at the interface are saved, then used to start a number of new simulations from the interface. The reaction rate is calculated in terms of a set of conditional crossing probabilities, and transition paths between initial and final states may be obtained by tracing completed paths from the final state back to the initial state [65]. As in TIS and WE, high barriers between interfaces may cause a sharp drop in sampling efficiency, as simulations can progress to the next interface only rarely [63]. A particularly interesting feature of FFS is the existence of well-defined estimates for computational efficiency as functions of FFS simulation parameters (e.g. the number of interfaces), allowing for selection of efficient parameters [67,68]. FFS has been used primarily in simplified models of various systems (cf. [54•,55•]), but it has also been applied to an all-atom folding simulation of the trp-cage mini-protein [69].

Weighted ensemble sampling

The WE sampling technique is conceptually quite similar to FFS, though it predates FFS by nearly a decade [70]. Originally conceived to accelerate sampling in Brownian dynamics simulations [70–72], WE sampling is asymptotically correct for a much broader class of stochastic simu-

lations, including MD simulations [73]. WE sampling uses independent simulations, each carrying a statistical weight, to explore conformational space. Like TIS and FFS, WE sampling requires definitions of initial and final states, an order parameter, and the partitioning of space along the order parameter into bins. Simulations are propagated for a fixed time period, after which a statistically rigorous reweighting procedure is used to keep the number of simulations in each bin constant without altering the total probability in each bin. Thus, as unoccupied bins become populated, more simulations are created with which to explore that region of phase space, and as simulations cross backwards into previously traversed bins, they will likely be eliminated, reducing oversampling. As simulations reach the destination state, their probability weights are recycled to the initial state, establishing a steady-state flow of probability from the initial state to the final state. The resulting transition paths are continuous, and the macroscopic reaction rate is obtained simultaneously as the net flow of probability into the destination state [70]. WE sampling has a theoretical and algorithmic framework that naturally supports more than one order parameter, making it an attractive option for sampling rare events in systems that cannot be described with a single order parameter [63]. Achievement of a steady-state probability flow from the initial state to the final state may be accelerated using concepts developed from nonequilibrium umbrella sampling, partially ameliorating the difficulty shared by WE, TIS, and FFS of surmounting barriers between interfaces (i.e. within bins) [63].

WE sampling in the context of a residue-based Monte Carlo simulation has recently been used to study the kinetics and conformational transitions between the *apo* and *holo* forms of calmodulin, showing excellent agreement and efficiency gains compared to brute force Monte Carlo sampling [74•]. Our own group has recently determined that WE sampling in conjunction with MD simulations is up to three orders of magnitude more efficient than brute force simulations in modeling simple molecular association events (methane/methane, methane/benzene, Na^+/Cl^-, and 18-crown-6/K^+) in explicit water, indicating that this approach is a promising one for studying protein/small molecule and protein/protein interactions.

Conclusions

Conformational changes in biologically relevant systems span an enormous range of timescales, from picosecond dynamics of side chains through microsecond or slower dynamics of coordinated conformational transitions. All-atom MD simulations have typically been limited by computing power to microseconds of simulation time or less. Even so, with increasing computing power, brute force MD simulations continue to provide detailed views on biologically relevant conformational transitions. Additionally, a number of enhanced sampling techniques have matured to the point of reaching biological timescales with MD simulations. The most promising avenue for exploration of the dynamics and kinetics of pharmacologically relevant systems appears not to be any single MD sampling technique, but combinations of techniques that, when used together, yield far more information than any technique alone (e.g. [58••]). With advances in simulation approaches and computing power, MD simulation is becoming increasingly useful in providing detailed structural and mechanistic insight into biologically relevant events that are of pharmaceutical interest.

Acknowledgements

This work was supported by an NSF CAREER Award (MCB-0845216) to LTC and a University of Pittsburgh Arts & Sciences Fellowship to MCZ. We are grateful to John Chodera and Dan Zuckerman for insightful discussions and constructive comments on the manuscript.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Henzler-Wildman KA, Kern D: **Dynamic personalities of proteins**. *Nature* 2007, **450**:964-972.
2. Adcock SA, McCammon JA: **Molecular dynamics: survey of methods for simulating the activity of proteins**. *Chem Rev* 2006, **106**:1589-1615.
3. Hammes-Schiffer S, Benkovic SJ: **Relating protein motion to catalysis**. *Annu Rev Biochem* 2006, **75**:519-541.
4. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D: **A hierarchy of timescales in protein dynamics is linked to enzyme catalysis**. *Nature* 2007, **450**:913-916.
5. Teague SJ: **Implications of protein flexibility for drug discovery**. *Nat Rev Drug Discov* 2003, **2**:527-541.
6. Lee GM, Craik CS: **Trapping moving targets with small molecules**. *Science* 2009, **324**:213-215.
7. Wells JA, McClendon CL: **Reaching for high-hanging fruit in drug discovery at protein–protein interfaces**. *Nature* 2007, **450**:1001-1009.
8. Hornak V, Simmerling C: **Targeting structural flexibility in HIV-1 protease inhibitor binding**. *Drug Discov Today* 2007, **12**:132-138.
9. Dyson HJ, Wright PE: **Intrinsically unstructured proteins and their functions**. *Nat Rev Mol Cell Biol* 2005, **6**:197-208.
10. Uversky VN, Oldfield CJ, Dunker AK: **Intrinsically disordered proteins in human diseases: introducing the D2 concept**. *Annu Rev Biophys* 2008, **37**:215-246.
11. Boehr DD, Nussinov R, Wright PE: **The role of dynamic conformational ensembles in biomolecular recognition**. *Nat Chem Biol* 2009, **5**:789-796.
12. Chong LT, Snow CD, Rhee YM, Pande VS: **Dimerization of the p53 oligomerization domain: identification of a folding nucleus by molecular dynamics simulations**. *J Mol Biol* 2005, **345**:869-878.
13. Elber R: **Long-timescale simulation methods**. *Curr Opin Struct Biol* 2005, **15**:151-156.
14. • Buchete N, Hummer G: **Peptide folding kinetics from replica exchange molecular dynamics**. *Phys Rev E* 2008, **77**:1-4.
Markov models are used to extract realistic kinetic information from replica exchange molecular dynamics simulations, which are generally unable to provide such information.

15. Xin Y, Doshi U, Hamelberg D: **Examining the limits of time**
• **reweighting and Kramers' rate theory to obtain correct kinetics from accelerated molecular dynamics**. *J Chem Phys* 2010, **132**:224101.
A unique study of how exact, unperturbed thermodynamic averages and (in some cases) realistic macroscopic reaction rates can be extracted from perturbed dynamics.

16. Mackerell AD: **Empirical force fields for biological macromolecules: overview and issues**. *J Comput Chem* 2004, **25**:1584-1604.

17. Best RB, Hummer G: **Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides**. *J Phys Chem B* 2009, **113**:9004-9015.

18. Freddolino PL, Park S, Roux B, Schulten K: **Force field bias in protein folding simulations**. *Biophys J* 2009, **96**:3772-3780.

19. Klepeis JL, Lindorff-Larsen K, Dror RO, Shaw DE: **Long-timescale molecular dynamics simulations of protein structure and function**. *Curr Opin Struct Biol* 2009, **19**:120-127.

20. Pitera JW: **Current developments in and importance of high-**
• **performance computing in drug discovery**. *Curr Opin Drug Discov Dev* 2009, **12**:388-396.
A lucid introduction to high-performance computing as applied to biomolecular systems.

21. Eastman P, Pande VS: **Efficient nonbonded interactions for molecular dynamics on a graphics processing unit**. *J Comput Chem* 2010, **31**:1268-1272.

22. Larsson P, Lindahl E: **A high-performance parallel-generalized born implementation enabled by tabulated interaction rescaling**. *J Comput Chem* 2010, **31**:2593-2600.

23. Harvey MJ, Giupponi G, De Fabritiis G: **ACEMD: accelerating biomolecular dynamics in the microsecond time scale**. *J Chem Theory Comput* 2009, **5**:1632-1639.

24. Phillips JC, Stone JE, Schulten K: **Adapting a message-driven parallel application to GPU-accelerated clusters**. *2008 SC—International Conference for High Performance Computing, Networking, Storage and Analysis*. 2008 doi: 10.1109/SC.2008.5214716.

25. Harvey MJ, De Fabritiis G: **An implementation of the smooth particle Mesh Ewald method on GPU hardware**. *J Chem Theory Comput* 2009, **5**:2371-2377.

26. Friedrichs MS, Eastman P, Vaidyanathan V, Houston M, Legrand S, Beberg AL, Ensign DL, Bruns CM, Pande VS: **Accelerating molecular dynamic simulation on graphics processing units**. *J Comput Chem* 2009, **30**:864-872.

27. 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.

28. Schulz R, Lindner B, Petridis L, Smith JC: **Scaling of multimillion-**
• **atom biological molecular dynamics simulation on a petascale supercomputer**. *J Chem Theory Comput* 2009, **5**:2798-2808.
A detailed discussion of the technical complexities involved in scaling molecular dynamics simulations to large core counts.

29. Ozkan SB, Dill KA, Bahar I: **Computing the transition state populations in simple protein models**. *Biopolymers* 2003, **68**:35-46.

30. Chodera JD, Swope WC, Pitera JW, Dill KA: **Long-time protein folding dynamics from short-time molecular dynamics simulations**. *Multiscale Model Simul* 2006, **5**:1214.

31. 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.

32. Chodera JD, Singhal N, Pande VS, Dill KA, Swope WC: **Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics**. *J Chem Phys* 2007, **126**:155101.

33. Buchete N, Hummer G: **Coarse master equations for peptide folding dynamics**. *J Phys Chem B* 2008, **112**:6057-6069.

34. Metzner P, Schütte C, Vanden-Eijnden E: **Transition path theory for Markov jump processes**. *Multiscale Model Simul* 2009, **7**:1192.

35. Noé F, Fischer S: **Transition networks for modeling the kinetics of conformational change in macromolecules**. *Curr Opin Struct Biol* 2008, **18**:154-162.

36. Bowman GR, Beauchamp KA, Boxer G, Pande VS: **Progress and challenges in the automated construction of Markov state models for full protein systems**. *J Chem Phys* 2009, **131**:124101.

37. Sriraman S, Kevrekidis IG, Hummer G: **Coarse master equation from Bayesian analysis of replica molecular dynamics simulations**. *J Phys Chem B* 2005, **109**:6479-6484.

38. Huang X, Bowman GR, Bacallado S, Pande VS: **Rapid equilibrium sampling initiated from nonequilibrium data**. *Proc Natl Acad Sci U S A* 2009, **106**:19765-19769.

39. Yang S, Banavali NK, Roux B: **Mapping the conformational transition in Src activation by cumulating the information from multiple molecular dynamics trajectories**. *Proc Natl Acad Sci U S A* 2009, **106**:3776-3781.

40. Noé F, Schütte C, Vanden-Eijnden E, Reich L, Weikl TR:
•• **Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations**. *Proc Natl Acad Sci U S A* 2009, **106**:19011-19016.
The folding of a small protein (the PinWW domain) is studied using short MD trajectories in explicit solvent, Markov state models, and transition path theory; folding occurs along multiple pathways, and therefore the authors argue that a probabilistic view of protein folding mechanisms is necessary.

41. Voelz VA, Bowman GR, Beauchamp KA, Pande VS: **Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1–39)**. *J Am Chem Soc* 2010, **132**:1526-1528.

42. Bowman GR, Pande VS: **Protein folded states are kinetic hubs**. *Proc Natl Acad Sci U S A* 2010, **2010**:1-6.

43. Bowman GR, Huang X, Pande VS: **Network models for molecular kinetics and their initial applications to human health**. *Cell Res* 2010, **20**:622-630.

44. Hansmann U: **Parallel tempering algorithm for conformational studies of biological molecules**. *Chem Phys Lett* 1997, **281**:140-150.

45. Mitsutake A, Sugita Y, Okamoto Y: **Generalized-ensemble algorithms for molecular simulations of biopolymers**. *Biopolymers* 2001, **60**:96-123.

46. Faradjian AK, Elber R: **Computing time scales from reaction coordinates by milestoning**. *J Chem Phys* 2004, **120**:10880-10889.

47. West AM, Elber R, Shalloway D: **Extending molecular dynamics**
•• **time scales with milestoning: example of complex kinetics in a solvated peptide**. *J Chem Phys* 2007, **126**:145104.
An exceptionally clear and thorough description of the Milestoning procedure, with a superior balance of detailed theory and straightforward explanation, a forthright discussion of the strengths and weaknesses of the Milestoning method, and detailed comparison to similar methods.

48. Hänggi P, Talkner P, Borkovec M: **Reaction-rate theory: fifty years after Kramers**. *Rev Mod Phys* 1990, **62**:251-341.

49. Elber R, West AM: **Atomically detailed simulation of the recovery stroke in myosin by Milestoning**. *Proc Natl Acad Sci U S A* 2010, **107**:5001-5005.

50. Pratt LR: **A statistical method for identifying transition states in high dimensional problems**. *J Chem Phys* 1986, **85**:5045-5048.

51. Dellago C, Bolhuis PG, Csajka F, Chandler D: **Transition path sampling and the calculation of rate constants**. *J Chem Phys* 1998, **108**:1964.

52. Bolhuis PG, Chandler D, Dellago C, Geissler PL: **Transition path sampling: throwing ropes over rough mountain passes, in the dark**. *Annu Rev Phys Chem* 2002, **53**:291-318.

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

54. Escobedo FA, Borrero EE, Araque JC: **Transition path sampling**
• **and forward flux sampling. Applications to biological systems**. *J Phys Condens Matter* 2009, **21**:333101.
A concise review of enhanced sampling techniques, focusing on biological applications.

55. Allen RJ, Valeriani C, ten Wolde PR: **Forward flux sampling for**
• **rare event simulations**. *J Phys Condens Matter* 2009, **21**:463102.
Another review of enhanced sampling techniques, with a particular focus on forward flux sampling (FFS).

56. Grünwald M, Dellago C, Geissler PL: **Precision shooting: sampling long transition pathways**. *J Chem Phys* 2008, **129**:194101.

57. Vreede J, Juraszek J, Bolhuis PG: **Predicting the reaction coordinates of millisecond light-induced conformational changes in photoactive yellow protein**. *Proc Natl Acad Sci U S A* 2010, **107**:2397-2402.

58. Juraszek J, Bolhuis PG: **(Un)Folding mechanisms of the FBP28**
•• **WW domain in explicit solvent revealed by multiple rare event simulation methods**. *Biophys J* 2010, **98**:646-656.
A model study demonstrating the great utility of combining multiple enhanced sampling techniques. The authors use three techniques, each according to its own particular strength, to elucidate the folding mechanism of a small protein (the FBP28 WW domain).

59. van Erp TS, Moroni D, Bolhuis PG: **A novel path sampling method for the calculation of rate constants**. *J Chem Phys* 2003, **118**:7762.

60. Moroni D, Bolhuis PG, van Erp TS: **Rate constants for diffusive processes by partial path sampling**. *J Chem Phys* 2004, **120**:4055-4065.

61. van Erp TS: **Reaction rate calculation by parallel path swapping**. *Phys Rev Lett* 2007, **98**:1-4.

62. Vanerp T, Bolhuis PG: **Elaborating transition interface sampling methods**. *J Comput Phys* 2005, **205**:157-181.

63. Bhatt D, Zhang BW, Zuckerman DM: **Steady-state simulations using weighted ensemble path sampling**. *J Chem Phys* 2010, **133**:014110.

64. Allen RJ, Warren P, ten Wolde PR: **Sampling rare switching events in biochemical networks**. *Phys Rev Lett* 2005, **94**:018104.

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

66. Dickson A, Dinner AR: **Enhanced sampling of nonequilibrium steady states**. *Annu Rev Phys Chem* 2010, **61**:441-459.

67. Allen RJ, Frenkel D, ten Wolde PR: **Forward flux sampling-type schemes for simulating rare events: efficiency analysis**. *J Chem Phys* 2006, **124**:194111.

68. 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.

69. Juraszek J, Bolhuis PG: **Rate constant and reaction coordinate of Trp-cage folding in explicit water**. *Biophys J* 2008, **95**:4246-4257.

70. Huber GA, Kim S: **Weighted-ensemble Brownian dynamics simulations for protein association reactions**. *Biophys J* 1996, **70**:97-110.

71. Rojnuckarin A, Kim S, Subramaniam S: **Brownian dynamics simulations of protein folding: access to milliseconds time scale and beyond**. *Proc Natl Acad Sci U S A* 1998, **95**:4288-4292.

72. Rojnuckarin A, Livesay DR, Subramaniam S: **Bimolecular reaction simulation using weighted ensemble Brownian dynamics and the University of Houston Brownian dynamics program**. *Biophys J* 2000, **79**:686-693.

73. Zhang BW, Jasnow D, Zuckerman DM: **The 'weighted ensemble' path sampling method is statistically exact for a broad class of stochastic processes and binning procedures**. *J Chem Phys* 2010, **132**:054107.

74. Zhang BW, Jasnow D, Zuckerman DM: **Efficient and verified**
• **simulation of a path ensemble for conformational change in a united-residue model of calmodulin**. *Proc Natl Acad Sci U S A* 2007, **104**:18043-18048.
The first application of weighted ensemble sampling to large-scale conformational transitions of a protein switch (*apo* and *holo* forms of calmodulin).

75. Fersht AR: *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding.*W.H. Freeman and Company; 1999.