ANNUAL REVIEWS Further

Click here for quick links to Annual Reviews content online, including:

- Other articles in this volume
- Top cited articles
- Top downloaded articles
- Our comprehensive search

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

Biomolecular Simulation: A Computational Microscope for Molecular Biology

Ron O. Dror,[1] Robert M. Dirks,[1] J.P. Grossman,[1] Huafeng Xu,[1] and David E. Shaw[1,2]

[1]D. E. Shaw Research, New York, New York 10036; email: Ron.Dror@DEShawResearch.com; David.Shaw@DEShawResearch.com

[2]Center for Computational Biology and Bioinformatics, Columbia University, New York, New York 10032

Annu. Rev. Biophys. 2012. 41:429–52

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

This article's doi: 10.1146/annurev-biophys-042910-155245

Copyright © 2012 by Annual Reviews. All rights reserved

1936-122X/12/0609-0429$20.00

Keywords

molecular dynamics, millisecond timescale, Anton, protein folding, drug binding

Abstract

Molecular dynamics simulations capture the behavior of biological macromolecules in full atomic detail, but their computational demands, combined with the challenge of appropriately modeling the relevant physics, have historically restricted their length and accuracy. Dramatic recent improvements in achievable simulation speed and the underlying physical models have enabled atomic-level simulations on timescales as long as milliseconds that capture key biochemical processes such as protein folding, drug binding, membrane transport, and the conformational changes critical to protein function. Such simulation may serve as a computational microscope, revealing biomolecular mechanisms at spatial and temporal scales that are difficult to observe experimentally. We describe the rapidly evolving state of the art for atomic-level biomolecular simulation, illustrate the types of biological discoveries that can now be made through simulation, and discuss challenges motivating continued innovation in this field.

Contents

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

INTRODUCTION

Over the past half-century, breakthroughs in structural biology have provided atomic-resolution models of many of the molecules that are essential to life, including proteins and nucleic acids. Although static structures determined through crystallography and other techniques are tremendously useful, the molecules they represent are, in reality, highly dynamic, and their motions are often critical to their function (**Figure 1**). Proteins, for example, undergo a variety of conformational changes that allow them to act as signaling molecules, transporters, catalysts, sensors, and mechanical effectors. Likewise, they interact dynamically with hormones, drugs, and one another. Static structural information might be likened to a photograph of a football game; to understand more readily how the game is played, we want a video recording.

Conformational change: a transition between two alternative structures of a flexible biomolecule such as a protein

Molecular dynamics (MD) simulation: a simulation in which the positions and velocities of atoms are computed using Newton's laws of motion

A variety of experimental techniques can provide information about the dynamics of proteins and other biomolecules, but they are generally limited in their spatial and temporal resolution, and most report ensemble average properties rather than the motion of individual molecules (**Figure 2**). An attractive alternative, in principle, is to model atomic-level motions computationally, based on first-principles physics. Although such simulations have been an active area of research for decades (55), their computational expense, combined with the challenge of developing appropriate physical models, has placed restrictions on both their length and their accuracy. The past few years have seen great progress in addressing these limitations, making simulations a much more powerful tool for the study of biomolecular dynamics. This review describes several important recent advances in simulation methodology and offers an overview of what is currently possible with biomolecular simulation.

The quantum mechanical behavior of molecules at a subatomic level is described by the time-dependent Schrödinger equation, but a direct solution to this equation is in practice computationally infeasible for biological macromolecules. The standard method for simulating the motions of such molecules is a technique known as all-atom molecular dynamics (MD) simulation, in which

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

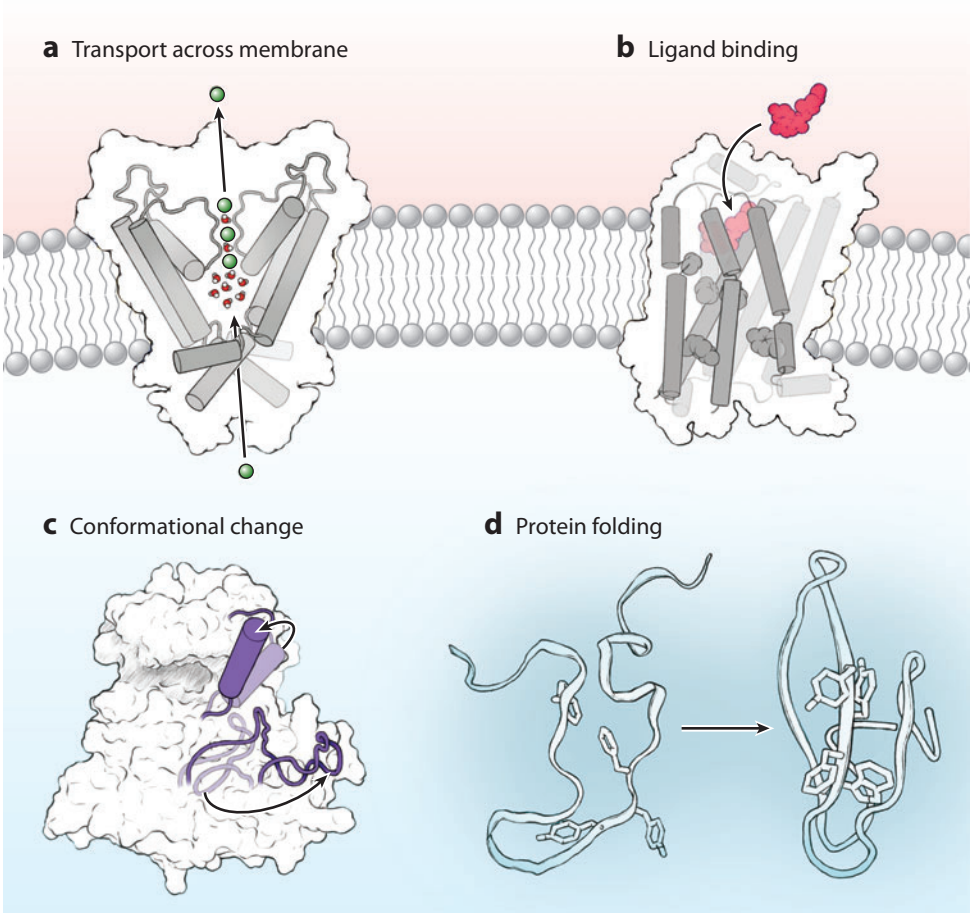

Figure 1

Examples of biomolecular processes that have been examined using molecular dynamics (MD) simulations. (*a*) Transport of small molecules across the cell membrane. (*b*) Binding of drugs to their target proteins. (*c*) Conformational transitions in proteins. (*d*) Protein folding.

the positions and velocities of particles representing every atom in the system evolve according to the laws of classical physics. The forces acting on these particles are computed using a model known as a force field, which is typically designed based on a combination of first-principles physics and parameter fitting to quantum mechanical computations and experimental data. Although MD simulation does not model the underlying physics exactly, it can provide a sufficiently close approximation to capture a wide range of critical biochemical processes. The popularity of such simulations is illustrated by the fact that they account for a majority of the computer time devoted to biomedical research at National Science Foundation supercomputer centers. Although we briefly touch on approaches that represent molecules at a coarser or finer level of detail, or that evolve positions in a nonphysical manner, all-atom MD simulations constitute the principal focus of this review.

Force field: energy function used to compute the forces acting on atoms (due to interatomic interactions) during an MD simulation

Historically, the timescales accessible to MD simulation have been shorter than those on which most biomolecular events of interest take place, thus limiting the applicability of these

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

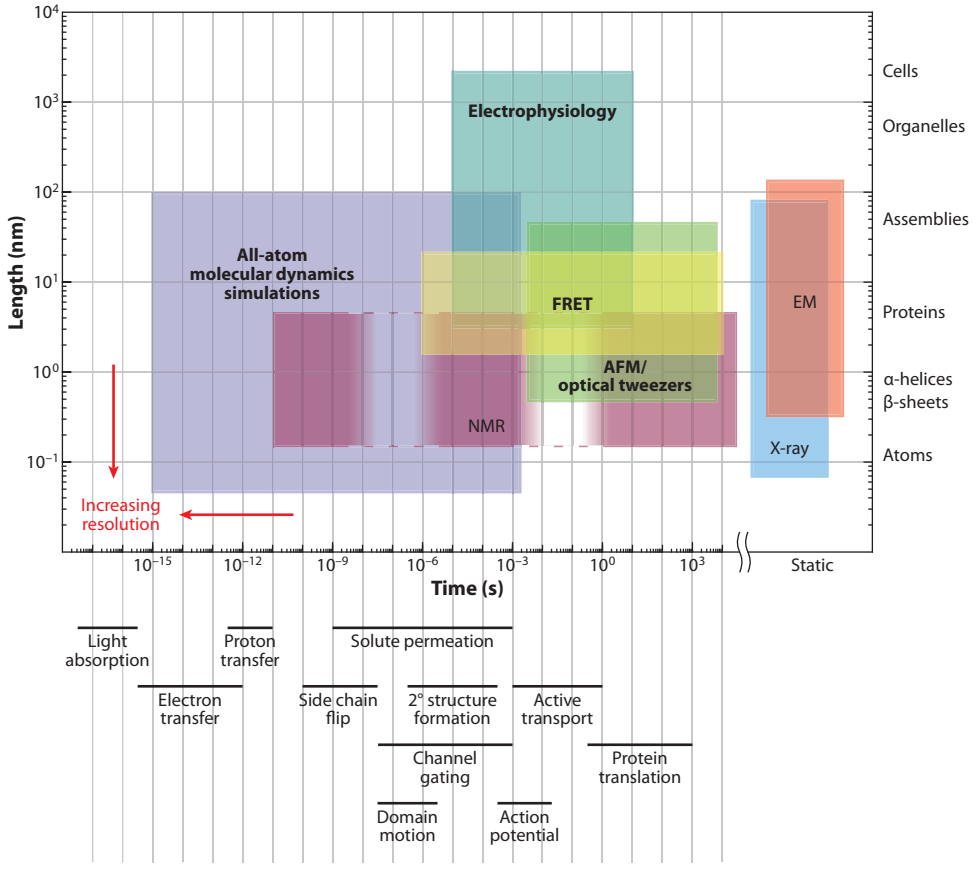

Figure 2

Spatiotemporal resolution of various biophysical techniques. The temporal (abscissa) and spatial (ordinate) resolutions of each technique are indicated by colored boxes. Techniques capable of yielding data on single molecules (as opposed to only on ensembles) are in boldface. NMR methods can probe a wide range of timescales, but they provide limited information on motion at certain intermediate timescales, as indicated by the lighter shading and dashed lines. The timescales of some fundamental molecular processes, as well as composite physiological processes, are indicated below the abscissa. The spatial resolution needed to resolve certain objects is shown at the right. Adapted from Reference 19. Abbreviations: AFM, atomic force microscopy; EM, electron microscopy; FRET, Förster resonance energy transfer; NMR, nuclear magnetic resonance.

simulations. Events such as protein folding, protein–drug binding, and major conformational changes essential to protein function typically take place on timescales of microseconds to milliseconds (**Figure 2**). MD simulations, by contrast, were until recently generally limited in practice to nanosecond timescales. Simulations of even a few microseconds required months on the most powerful supercomputers available, and longer simulations had never been performed. Recent advances in hardware, software, and algorithms have increased the timescales accessible to simulation by several orders of magnitude, enabling the first millisecond-scale simulations and allowing MD to capture many critical biochemical processes for the first time.

The other major factor limiting the applicability of MD has been the accuracy of the force field models that underlie the simulations. A number of improved force fields have been introduced

over the past several years, and the longer timescales now accessible to MD simulations have allowed more extensive validation of these force fields against experimental data.

We begin by summarizing the fundamentals of MD simulation and certain recent methodological and technological advances that have expanded its applicability. We then review the state of the art in terms of the types of biological discoveries one can make through simulation, providing a number of recent illustrative examples. Finally, we discuss several classes of important problems that MD could potentially address in the coming years and the methodological advances that may help solve them.

Parallel computation: using multiple cooperating processors to perform a computation faster than would be possible with a single processor

Moore's law: a trend dating back to 1960 in which the logic density on computer chips doubles approximately every two years

RECENT ADVANCES IN SIMULATION METHODOLOGY

Although the speed and accuracy of all-atom MD simulations has improved substantially over the past few years, the basic form of such simulations has endured (1). Each atom in the system—for example, a protein and the water surrounding it—is represented by a particle (or, in certain cases, multiple particles). The simulation steps through time, alternately computing the forces acting on each atom and using Newton's laws of motion to update the positions and velocities of all the atoms. Commonly used biomolecular force fields express the total force on an atom as a sum of three components: (*a*) bonded forces, which involve interactions between small groups of atoms connected by one or more covalent bonds; (*b*) van der Waals forces, which involve interactions between all pairs of atoms in the system but which fall off quickly with distance and are generally evaluated only for nearby pairs of atoms; and (*c*) electrostatic forces, which involve interactions between all pairs of atoms and fall off slowly with distance. Electrostatic interactions are computed explicitly between nearby pairs of atoms, whereas long-range electrostatic interactions are typically handled via one of several approximate methods that are more efficient than explicitly computing interactions between all distant pairs of atoms.

Accessing Longer Timescales

MD simulations are computationally demanding for two reasons. First, the force calculation at each time step requires substantial computation—roughly one billion arithmetic operations for a system with one hundred thousand atoms. Second, the force calculation must be repeated many times. Individual steps are limited to a few femtoseconds by fast atomic vibrations, so simulating a millisecond of physical time requires nearly one trillion time steps. On a single high-end processor core, such a simulation would take thousands of years to complete.

To make matters worse, using many computer processors in parallel to accelerate a simulation is challenging. Parallelizing a force calculation across multiple processors requires that those processors communicate with one another, and the amount of communication required increases with the number of processors. Beyond some point, adding more processors actually slows down the calculation. Furthermore, the force calculations in a single simulation must be performed sequentially, because the atom positions that serve as input for one force calculation depend on the results of the previous one. The difficulty of parallelizing an MD simulation implies that the transistor density improvements predicted by Moore's law do not automatically lead to improved MD performance, because in recent years higher transistor density has translated to more processor cores per chip, not faster individual processor cores.

In spite of these challenges, the recent performance improvements in MD simulations have far outpaced Moore's law. In half a decade, the raw performance of state-of-the-art simulations increased by over three orders of magnitude (**Figure 3**). As recently as 2007, the longest published all-atom MD simulation of a protein was 2 μs; in 2009, the first millisecond-long simulation was

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

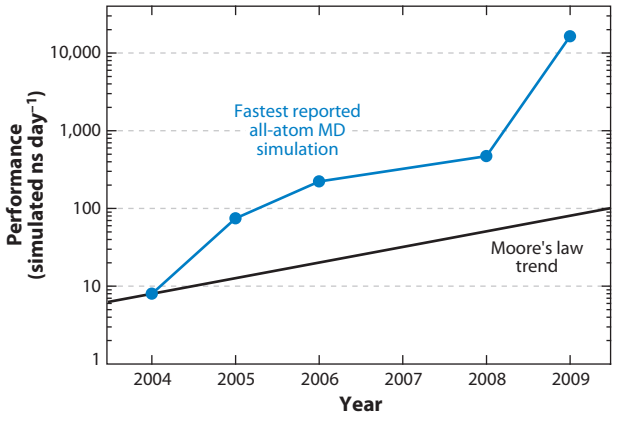

Figure 3

Fastest reported all-atom molecular dynamics (MD) simulations from 2004 to 2009 (*blue line*). The simulated systems ranged from 14,000 to 92,000 atoms, and different simulations were performed using different parameters, so this data is not intended to be a direct comparison of MD hardware and software systems. Nonetheless, an overall performance trend is evident, substantially exceeding the Moore's law growth trend in processing power (*black line*). The leftmost data point is from a 512-processor simulation using NAMD (65); the rightmost data point is from a 512-chip simulation on Anton (82). The remaining data points are from simulations run using Blue Gene/L (25) and Desmond (9, 14).

published (**Table 1**). These improvements are attributable to a variety of hardware, software, and algorithm innovations, which we discuss below.

Parallelization across general-purpose computer chips. Software that parallelizes MD force calculations across multiple computer processors has existed for two decades (69) but has become much more efficient and scalable in the past several years. IBM's Blue Matter code for its Blue Gene/L general-purpose supercomputer has been scaled up to 32,768 cores (25). The widely used MD codes NAMD (65), GROMACS (31), and AMBER (12) have all substantially improved their parallel performance in recent years. These packages can now deliver performance of over 100 ns day^{-1} on commodity computer clusters, with the number of processors required to achieve this performance scaling roughly linearly with the number of atoms in the system. Desmond, a software package for commodity clusters developed at D. E. Shaw Research, allowed simulations nearly an order of magnitude faster than previously possible on the same hardware (9) and has achieved performance approaching 500 ns day^{-1} (14).

These improvements in parallel scalability and efficiency have been possible thanks to a number of algorithmic innovations, particularly in methods for reducing the communication requirements

Table 1 Longest reported all-atom molecular dynamics simulations from 2006 to 2009

Year	Length (μs)	Protein	Platform	Reference
2006	2	Rhodopsin	Blue Gene/L[a]	54
2007	2	Villin HP-35	GROMACS[b]	22
2008	10	WW domain	NAMD[b]	27
2009	1,031	BPTI	Anton	82

[a]This simulation used IBM's Blue Matter software.

[b]These simulations were performed on a commodity computer cluster with the specified software.

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

between chips during a simulation. The class of neutral territory methods (9, 10, 25, 81, 86), for example, substantially reduces the amount of data that must be exchanged between processors in order to compute range-limited particle interactions.

Graphics processing units. Originally designed specifically to accelerate the rendering of three-dimensional graphics, graphics processing units (GPUs) have become increasingly popular for general-purpose scientific computation thanks to their ability to perform large numbers of identical computations in parallel on a single chip. Several MD implementations have been ported to GPUs (2, 30, 29, 66), and a simulation on one or a few GPUs often rivals the performance of a simulation on a small- to moderate-sized computer cluster. Unfortunately, efficiently parallelizing across many GPUs is difficult, because communication between GPUs remains slower than communication between general-purpose processors; as a result, clusters of GPUs have been unable to match the performance of large standard clusters. GPUs offer an excellent price-to-performance ratio, however, enabling reasonably fast simulations at a cost substantially lower than that for a cluster of general-purpose processors.

Special-purpose parallel architectures. By far the greatest recent speedups have been achieved through the use of special-purpose chips, in combination with new parallelization algorithms and software. In particular, a recently developed machine named Anton can perform all-atom MD simulations at up to 20 μs day^{-1}, approximately two orders of magnitude faster than the simulation rates generally achieved by any other hardware/software combination (82).

Anton is able to achieve this speed because it is designed specifically for MD simulations. The entire computation is performed on special-purpose chips developed at D. E. Shaw Research, which are directly connected to one another in a custom network (**Figure 4**). Each Anton chip contains an array of arithmetic units hardwired for computing particle interactions, enabling a single Anton chip to compute interactions hundreds of times faster than a commodity processor core (82). Each chip also contains a dozen programmable processor cores, customized for MD, which accelerate the remainder of the computation. In addition, specialized data-movement

GPU: graphics processing unit

Special-purpose parallel architectures: computer architectures designed for a specific task, often allowing such computers to complete that task much faster than general-purpose computers

Anton: a special-purpose parallel supercomputer designed by D. E. Shaw Research to enable fast MD simulations

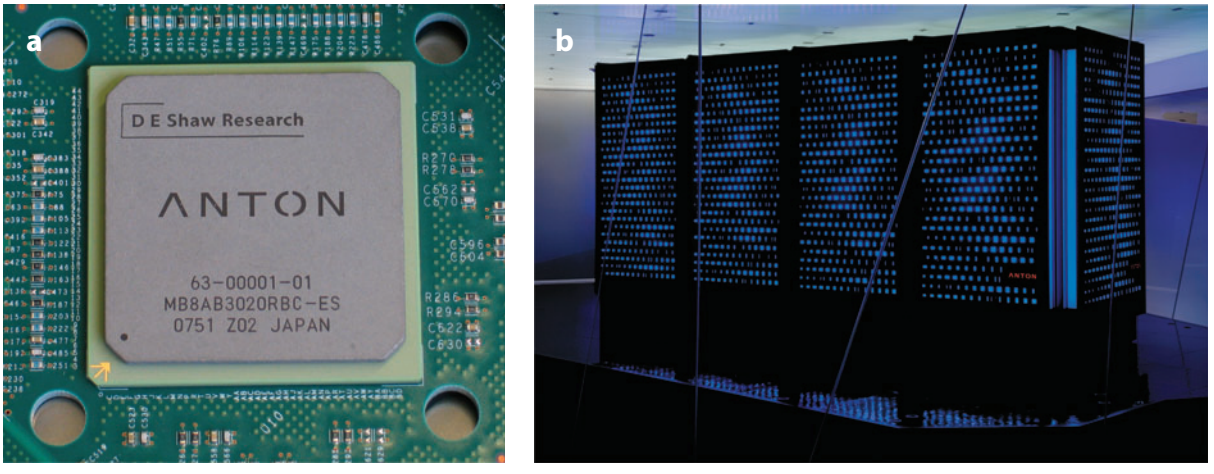

Figure 4

Anton, a special-purpose computer for molecular dynamics designed by D. E. Shaw Research, has performed all-atom protein simulations over one hundred times longer than any published previously. (*a*) A single Anton chip. (*b*) The first Anton machine, comprising 512 Anton chips connected through a specially designed network.

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

operations support fast communication between small on-chip memories, eliminating the memory cache hierarchy that typically consumes the majority of the area on commodity chips.

Enhanced sampling: algorithms devised to speed up the exploration of molecular conformations by altering the physics of the system

Several algorithmic advances also contribute to Anton's performance. A specific neutral territory method (81) was designed for Anton and is directly implemented within its specialized particle interaction hardware. Anton computes long-range electrostatic forces using the Gaussian split Ewald method (79) rather than the more commonly used particle mesh Ewald method, allowing a significant portion of the long-range electrostatics computation to be performed by the same specialized hardware that handles particle interactions. Finally, the communication patterns in Anton's MD software, which differ significantly from those in other parallel MD software packages, are designed to take advantage of Anton's specialized low-latency mechanisms for communication between and within chips (18).

Several previous projects, including FASTRUN (24), MD Engine (92), and MDGRAPE (90), have built special-purpose hardware to accelerate the most computationally expensive elements of an MD simulation. Although such hardware reduces the effective cost of simulating a given period of biological time, the speedup achieved through parallelization across many such chips is limited by the remainder of the computation as well as the communication required, precluding individual simulations on multi-microsecond timescales.

Anton has enabled all-atom MD simulations of proteins of more than a millisecond in length, over 100 times longer than any such simulation reported using other hardware. With Anton it becomes possible, for the first time, to directly observe in simulation various important biochemical processes that occur on timescales greater than a few microseconds.

Enhanced Sampling and Coarse-Graining

A variety of methods can be used in combination with MD simulation to investigate events that occur on timescales that remain inaccessible to direct all-atom MD simulation. Zuckerman (106) provided a thorough review of such enhanced sampling methods in Volume 40 of the *Annual Review of Biophysics*, in which he concluded that "No other method can routinely and reliably outperform MD by a significant amount." These methods prove essential in certain situations, however, as illustrated by several of the case studies in the following section.

One can also sidestep communication bottlenecks by performing many short simulations in parallel. A particularly impressive example is the Folding@Home project (5), which has obtained significant scientific results by using over 400,000 personal computers (PCs) to simulate many separate molecular trajectories, each limited to the timescale accessible on a single PC. The solution of many important problems, however, is greatly facilitated by the availability of long individual trajectories.

Finally, one can often substantially accelerate simulations at the cost of reduced accuracy by employing simplified system representations, such as coarse-grained models, in which each simulated particle represents several atoms, or implicit solvent models in which water atoms are replaced by a continuum representation. Models of both types have seen substantial development in recent years (13, 58).

Improving Force Field Accuracy

Long-timescale simulations place more stringent demands on force fields; a force field that proves sufficient for short-timescale simulations may not be sufficient at longer timescales. Fortunately, the past several years have seen substantial improvements in force fields for biomolecular simulation. Historically, force field parameters were determined using quantum mechanical calculations

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

and condensed phase experimental data for small molecular fragments. Recently, force field development has come to increasingly rely on experimental data for proteins and other biological macromolecules, as improvements in both simulation speed and experimental methods have led to an overlap in the timescales accessible through the two approaches.

Although the functional forms of the most widely used force fields have remained largely unchanged, their parameters have recently undergone a number of adjustments. The Amber force field, for example, incorporated changes to parameters associated with torsional angles of the protein backbone, first to improve fits to quantum calculations (32) and then to achieve better agreement between secondary structure preferences observed in long MD simulations of polypeptides and corresponding NMR measurements (7). Amber protein side chain torsions were also adjusted to better match both quantum calculations and NMR data (51). Adjustments to backbone and side chain torsions were also incorporated into the CHARMM force field (53, 67), as were modifications to the charge distributions of ionizable amino acid residues (67). Recent studies have also resulted in improved parameters for lipids in CHARMM (42) and for small drug-like molecules in the CHARMM, Amber, and OPLS-AA force fields (4, 96, 98).

G-protein-coupled receptors (GPCRs): a family of transmembrane proteins that transmit signals into cells and represent the largest class of drug targets

A recent study exploited long-timescale MD simulations on Anton to evaluate a number of protein force fields through a systematic comparison with experimental data (49). Criteria included the ability of each force field to fold small proteins to their native structures, to predict the secondary structure propensities of polypeptides, and to reproduce NMR data reporting on the structure and dynamics of folded proteins. The results indicated that the force fields examined have consistently improved over the past decade, and that the most recent versions provide an accurate description of many structural and dynamic properties of proteins. The study also highlighted certain shortcomings: None of the force fields, for example, were able to accurately capture the temperature dependency of the secondary structure propensities. It is an open question whether the ongoing parameterization of existing functional forms will be sufficient to further improve force fields. Substantial efforts are under way to develop force fields with more sophisticated functional forms, including polarizable force fields (36, 70), which capture the redistribution of electrons around each atom in response to changes in environment.

SIMULATION AS A TOOL FOR MOLECULAR BIOLOGY

In this section, we illustrate the utility of modern MD simulation as a biological research tool through a number of recent case studies, many of which are drawn from our own work, involving conformational changes in proteins, transport across membranes, protein folding, and the binding of ligands to proteins (**Figure 1**).

Conformational Changes

Under physiological conditions, proteins and other biomacromolecules constantly move from one structural state to another, and their function and regulation depend on these conformational changes. MD simulations are often used to identify novel conformations, to capture the transitional pathways between conformations, to determine equilibrium distributions among different conformations, and to characterize changes in conformational distribution as a result of mutation or ligand binding. We provide several examples involving G-protein-coupled receptors (GPCRs) and kinases.

GPCRs represent the largest class of drug targets: One-third of all marketed drugs act by binding to a GPCR and either triggering or preventing receptor activation, which involves a transition from an inactive receptor conformation to an active conformation that causes

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

G-protein-mediated signaling. The past few years have witnessed the determination of the first several crystal structures of ligand-activated GPCRs, beginning with the β_2-adrenergic receptor (β_2AR). MD has addressed several key questions raised by these structures about the conformations of inactive states and the mechanism of receptor activation (16, 17, 52, 62, 72, 73, 95).

β_2-adrenergic receptor (β_2AR): an archetypal GPCR and a target of beta blockers and beta agonists

An MD simulation study by Dror et al. (16) identified a previously unobserved inactive conformation of β_2AR, resolving an apparent contradiction between experimental results: A network of salt bridges known as the ionic lock, suggested by biochemical experiments to stabilize the inactive state of β_2AR and other GPCRs, was disrupted in the inactive state crystal structures (46). In microsecond-timescale simulations of inactive β_2AR, the receptor transitions repeatedly between two conformational states, one with the lock broken and one with it formed. In simulations of wild-type β_2AR, the lock-formed conformation predominates, in agreement with the biochemical data, but in simulations of the modified protein used for crystallographic structure determination, the lock-broken conformation predominates, explaining the crystallographic result. Both lock-broken and lock-formed conformations were subsequently observed in a crystal structure of a closely related receptor (60), lending support to these computational predictions.

More recent simulations on Anton (73) captured spontaneous transitions of β_2AR from an active to an inactive conformation, addressing a puzzle posed by two recent crystallographic structures of β_2AR bound to two different agonists (ligands that cause activation). One of these structures, which also has a G-protein-mimetic nanobody bound to its intracellular surface, appears to represent an active conformation (71). The other, which is bound to an agonist but lacks the nanobody, is almost identical to the previously solved inactive structure (73). Is this surprising structural difference due to differences between the agonists or the crystallized receptor constructs, or might it be due to the presence or absence of the G-protein-mimetic partner? In multi-microsecond simulations of β_2AR initiated from the nanobody-bound active structure, but with the nanobody removed, the agonist-bound receptor spontaneously transitioned to a conformation that closely matched the inactive crystal structure. Taken together, these simulations and the crystal structures suggest that, even with an agonist bound, the majority of the β_2AR population remains in an inactive-like conformation until a G-protein or G-protein-mimetic nanobody binds, stabilizing the active conformation.

These simulations—which represent the first in which a GPCR transitions spontaneously between crystallographic conformations representing functionally distinct states—also served to characterize the atomic-level activation mechanism of β_2AR (17). They showed that the extracellular drug-binding site is connected to the intracellular G-protein-binding site via a loosely coupled allosteric network, comprising three regions that can switch individually between distinct conformations. The simulations revealed a key intermediate conformation on the activation pathway and suggested—somewhat counterintuitively—that the first structural changes during activation often take place on the intracellular side of the receptor, far from the drug-binding site. These results may provide a foundation for the design of drugs that control receptor signaling more precisely by stabilizing specific receptor conformations.

MD simulations have also shed light on the function and regulation of kinases, a class of enzymes that are actively pursued as therapeutic targets for cancer and autoimmune diseases. The activity of a typical kinase is regulated by changes to its preference for the active and various inactive conformations. Mutations to kinases often favor the "wrong" conformations, leading to aberrant signaling and consequently disease. A number of studies have used MD simulations to characterize conformational changes in kinases (6, 23, 80, 100), yielding predictions that were in agreement with subsequent experimental measurements (80) and insights that led to the design of new experimental methods (76).

Faraldo-Gómez & Roux (23) used MD simulations to characterize the regulation of Src family tyrosine kinases, which depends on an inactivation process in which the auxiliary domains (known

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

as SH2 and SH3) of a kinase assemble onto its catalytic domain, preventing catalysis. What makes such assembly robust and fast, so that kinases can be reliably and quickly turned off? To answer these questions, the authors used an enhanced sampling technique known as umbrella sampling (44) to characterize the relative free energies of various conformations connecting the disassembled and assembled states. The simulations indicated that the SH2–SH3 construct has an intrinsic propensity to adopt conformations primed for association with the catalytic domain, thus favoring and accelerating formation of the assembled (inhibitory) state. Their results also suggested that the SH2–SH3 connector is more than a passive link between the domains; rather, it is responsible for their propensity toward the assembly-ready conformation, explaining the experimental observation that mutations in the connector region increase the constitutive activity of the kinase.

Membrane transport: the movement of molecules across a cell membrane, usually facilitated by a transmembrane protein

Membrane Transport

Transport of various substrates across the cell membrane is vital both to maintaining a cell's constitution and to transmitting biochemical signals. The transport efficiency and substrate selectivity of carrier proteins often depend critically on the detailed spatial configuration of the atoms along the transport pathway as well as their subtle movement during the transport process. MD simulation, with its unique ability to simulate and record the movement of individual atoms at very fine temporal and spatial resolutions, lends itself naturally to the study of transport processes. In the past decade, MD simulations have been applied to investigate a number of transporters and channels, including aquaporins (35, 91), ion channels (8, 34, 63), and active transporters (3, 21). These studies have shed light on many mechanistic questions: How do the channels achieve a fast rate of substrate permeation? How do the transporters affect selectivity for their substrates? How is transport regulated in response to various stimuli?

Potassium channels, which allow potassium ions to move passively through the cell membrane, are essential for the transmission of nerve impulses and represent an important target for the treatment of diseases ranging from Alzheimer's to diabetes. A longstanding puzzle about these channels is why they let potassium ions, but not smaller sodium ions, pass through. Crystal structures suggest that the narrowest region of a potassium channel, known as the selectivity filter, has a geometry that snugly fits potassium ions, but the difference between the radii of potassium and sodium ions (0.38 Å) is smaller than the thermal fluctuations of the atomic positions in the selectivity filter (0.75 Å). Noskov et al. (63) and Bostick & Brooks (8) addressed this question by using MD simulations to examine several hypothetical variants of the real selectivity filter. In both studies, the authors computed the difference in the binding free energies of sodium and potassium ions to the selectivity filter and explored how this difference varied when they artificially adjusted the physical properties of the filter. Both studies suggested that selectivity was a robust feature of the filter that did not depend on its precise geometry. Instead, selectivity was a consequence of the dipole moment of the carbonyls coordinating the ions in the selectivity filter, the coordination number, and the thermal fluctuations in the filter.

Recent advances in simulation speed have allowed the first direct, atomic-resolution observations of ion permeation and pore domain closure in a voltage-gated potassium channel (**Figure 5**). Using unbiased microsecond-timescale MD simulations at various transmembrane voltages, Jensen et al. (34) followed the permeation of hundreds of potassium ions through the channel. The authors identified the transitions between microscopic states that underlie the permeation of an individual ion, thereby supplying atomistic detail of the long-postulated "knock-on" conduction mechanism, in which translocation of two selectivity-filter-bound ions is driven by a third, incoming ion.

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

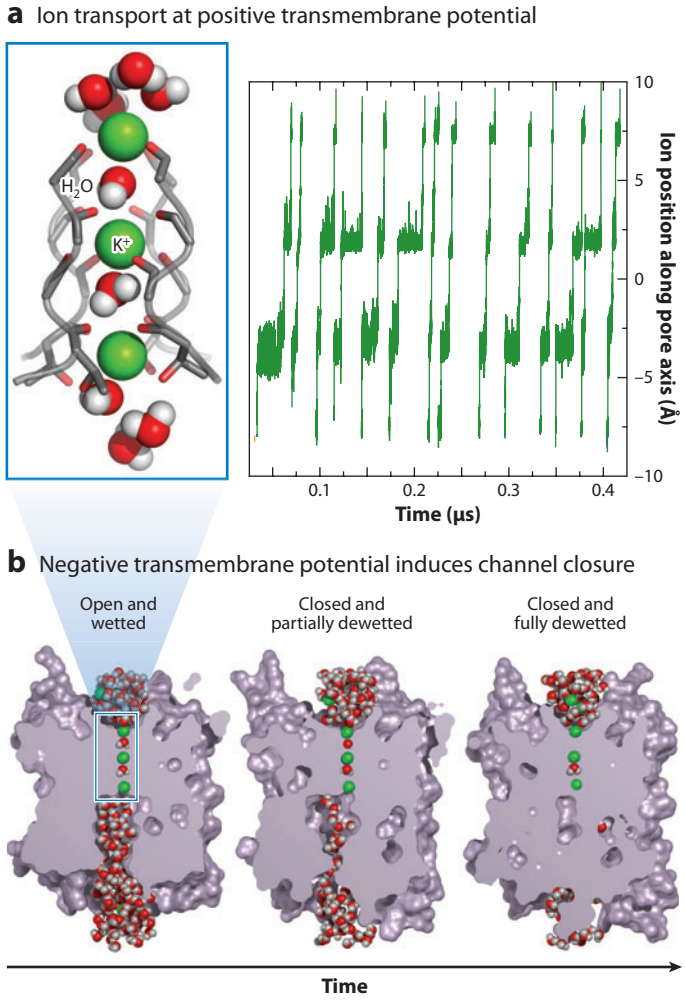

Figure 5

Simulation of ion permeation and gating in a potassium channel. (*a*) Potassium ions permeated outward (in the figure, upward) through the selectivity filter when the transmembrane potential was positive. Individual ions paused at well-defined sites within the filter, as shown by the representative traces in green. (*b*) When the transmembrane voltage was reversed, the hydrophobic cavity dehydrated, causing it to collapse and thus close the channel to conduction. Figure adapted from Reference 34.

Moreover, Jensen et al. observed channel closure—gating of the potassium channel pore domain—at negative voltages (**Figure 5**). Closure took place by means of a previously hypothesized, but unobserved (for ion channels) mechanism, called hydrophobic gating, in which the hydrophobic cavity adjacent to the selectivity filter dehydrated, causing the open pore domain to collapse into a closed conformation. This mechanism provides a molecular explanation for the experimental observation that the channel conductance is sensitive to the osmotic pressure. In particular, the change in volume upon channel closure has been measured experimentally, and it corresponds to the volume of 40–50 water molecules, closely matching the number of water molecules expelled from the pore cavity upon channel closure in the simulations (105).

MD simulations have also been used to deduce the mechanism of the sodium proton antiporter, NhaA (3), a transporter that moves sodium ions and protons in opposite directions across the cell membrane. Arkin et al. (3) performed a series of simulations in which they systematically varied the initial position of the sodium ion, as well as the protonation states of two aspartate residues—Asp163 and Asp164—critical for antiporting function. These simulations suggested that Asp164 serves as the binding site of sodium ion, with its protonation state determining whether the ion will remain bound or be released into the membrane, and that Aps163 acts as an accessibility control site, determining whether the ion will be released to the inside or outside of the cell. Although the simulations (≤100 ns each) were much shorter than the complete antiporting cycle (~1 ms), they allowed formulation of a complete transport mechanism, which was substantiated through experiments on NhaA mutants.

Folding pathway: a sequence of intermediate structures visited by a protein as it transitions from a disordered state to its native state

Protein Folding

Protein folding actually represents two challenges: Given only a protein's amino acid sequence, (*a*) determine the native structure of the protein, and (*b*) elucidate the pathways by which it folds to that structure. MD can potentially address both challenges (87), but it is particularly well suited for revealing folding pathways. Many computational (43) and experimental methods directly predict or determine protein structure, but few techniques allow direct observation of the dynamics of a folding event in atomic detail. Given an accurate force field and sufficient simulation time, MD can produce atomic-level trajectories of spontaneous folding events (22, 28, 37, 47, 83–85, 97), as well as unfolding events (93). Such a microscopic view can shed light not only on the structure and stability of the folded state, but also on the heterogeneity of folding pathways, the rate-limiting steps on these pathways, the nature of misfolded states, and other complex features of the protein folding process.

Improvements in both simulation speed and force field accuracy recently enabled Lindorff-Larsen et al. (50) to simulate repeated folding and unfolding events for a structurally diverse set of 12 small, fast-folding proteins, using a single force field. All 12 proteins folded to structures closely resembling those determined experimentally (**Figure 6**). The ability of simulations to identify the native structures is itself noteworthy, suggesting that MD may eventually serve as a viable method for predicting or refining the structures of arbitrary proteins. The most immediate utility of these simulations, however, is in allowing direct observation of the protein folding process.

Comparing the behavior of these 12 proteins suggested unifying principles for protein folding, at least for small, fast-folding proteins, and allowed the authors to address several long-standing questions regarding the mechanisms of protein folding (88). Most of the proteins studied, for example, consistently fold along a single dominant route, with local structures forming in an order that largely corresponds to the stability of those structures in the unfolded ensemble. In addition, a few long-range contacts typically form early in the folding process and help establish a nucleus to guide formation of the rest of the structure.

MD can also help guide wet-lab protein folding experiments (45). Piana et al. (68), for example, used insights gained from long simulations of a WW domain to suggest a triple mutation that should reduce the main energy barrier on the folding pathway and thus accelerate folding. Temperature-jump experiments confirmed this prediction, establishing this mutant as the fastest folding β-sheet protein known—a conclusion made more noteworthy because substantial effort had previously been dedicated to maximizing the folding rate of this WW domain through mutagenesis (61).

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

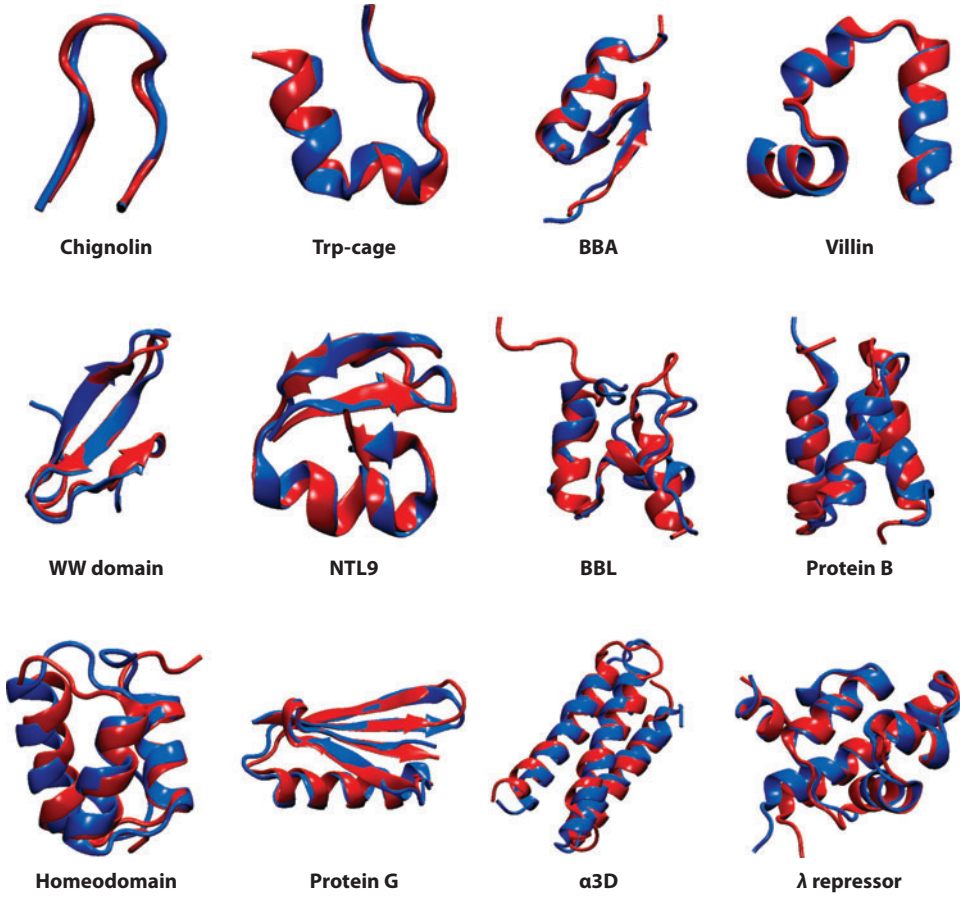

Figure 6

In simulations with a single force field, 12 structurally diverse proteins fold spontaneously to a structure (*blue*) closely resembling that determined experimentally (*red*). The simulation snapshots were chosen automatically based on a clustering analysis that did not exploit knowledge of the experimental structure. The total simulation time per protein ranged from 104 to 2,936 μs, allowing observation of at least 10 folding and 10 unfolding events for each protein. Figure adapted from Reference 50.

Ligand Binding

Interactions between proteins and small-molecule ligands play a key role in intercellular signaling and, when the ligands are drugs, in the treatment of disease. Ligands typically affect protein function by directly blocking the active site of a protein or by causing the protein to adopt a functionally altered conformational state.

Thanks to recent advances in accessible timescales, it is now possible to perform MD simulations in which ligands bind spontaneously to proteins without any prior knowledge of the binding site (20, 33, 78). In work by Shan et al. (78) on inhibitors binding to Src kinase, and by Dror et al. (20) on beta blockers and a beta agonist binding to two GPCRs, simulated drug molecules diffused extensively about the protein before discovering their binding site and binding in a location and conformation that match crystallographic observations almost exactly (**Figure 7**). These results

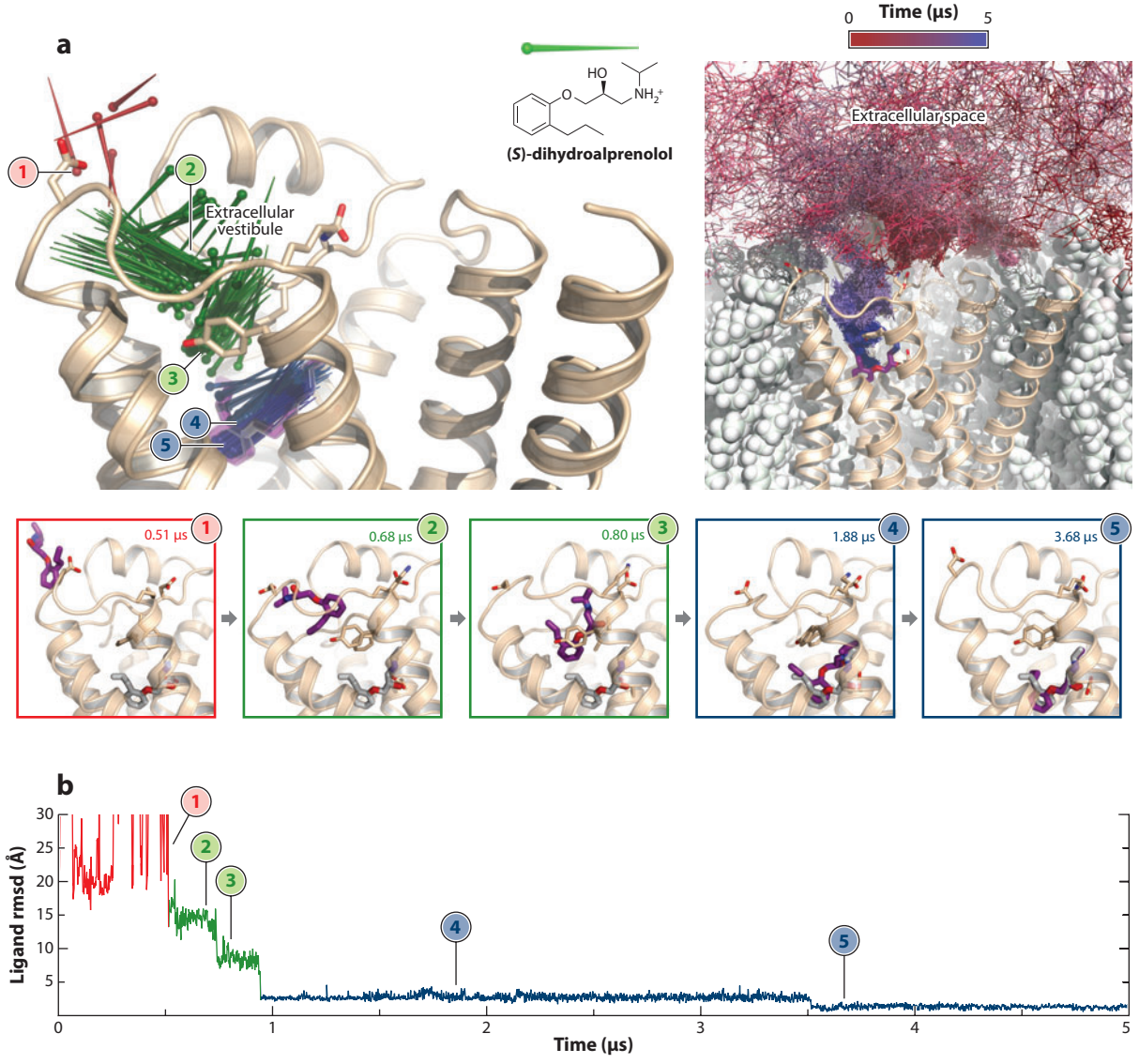

Figure 7

Beta blockers bind spontaneously to the β_2-adrenergic receptor (β_2AR) in molecular dynamics simulations, achieving the crystallographic pose and revealing several metastable intermediate states on the binding pathway. (*a*, *top left*) Pins indicate successive positions of a dihydroalprenolol molecule as it binds to β_2AR. The ligand moves from bulk solvent (pose ①), into the extracellular vestibule (poses ② and ③), and finally into the binding pocket (poses ④ and ⑤). (*a*, *bottom*) These five poses are shown in purple, with the crystallographic pose in gray. (*a*, *top right*) The path taken by the ligand as it diffuses about the receptor and then binds. (*b*) Root mean square deviation (rmsd) of the ligand in simulation from its position in the alprenolol–β_2AR crystal structure. Figure adapted from Reference 18.

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

raise the possibility of using simulation to identify novel binding sites. Indeed, both Shan et al. and Dror et al. discovered alternative binding sites, suggesting possibilities for the design of allosteric drugs with improved selectivity among kinases or GPCRs.

Such simulations also allow atomic-level characterization of the binding pathways and energetic barriers that determine binding kinetics. Dror et al. (20) found that beta blockers visit a sequence of metastable conformations en route to the binding pocket of the β_2AR (**Figure 7**). Surprisingly, they found that the largest energetic barrier on the binding pathway often occurs much earlier than receptor geometry would suggest, and appears to involve substantial dehydration that occurs as the drug associates with a particular region on the receptor surface. Shan et al. (78) also identified metastable conformations on the binding pathway, as did Buch et al. (11) in a study of an inhibitor binding to trypsin. These studies are computationally intensive: Shan et al., Dror et al., and Buch et al. performed multiple simulations totaling over 150 μs, 400 μs, and 50 μs, respectively.

A common application of protein–ligand simulations is to compute the binding affinity of a ligand, often a drug candidate, to a known binding site. Unbiased MD simulations of ligand binding are usually ill suited for this purpose, as precise estimation of ligand affinity would typically require seconds to hours of simulated time in order to observe sufficiently many binding and unbinding events. Fortunately, binding affinity calculations can be performed much more efficiently using methods such as free energy perturbation (107) or thermodynamic integration (40), which involve using a family of modified force fields to bias a series of simulations in ways that accelerate the forming and breaking of protein-ligand interactions. These biasing forces can be physically intuitive, such as forcibly pulling a ligand into or out of a known binding pocket (99), or more abstract, such as gradually turning off all interactions between a ligand and its surroundings (38). If the artificial energy functions are properly constructed, unbiased binding affinities can be efficiently and quantitatively derived from the biased simulations.

One compelling example of simulation-based binding affinity calculations is recent work on HIV reverse transcriptase. Starting with a weakly binding ligand that displayed no activity as a reverse transcriptase inhibitor, Zeevaart et al. (103) used Monte Carlo simulations (closely related to MD) to calculate the relative binding energies of a family of closely related molecules. By selecting variants predicted to bind more tightly, they discovered several molecules that proved experimentally active in protecting human T-cells from HIV infection.

FUTURE FRONTIERS OF BIOMOLECULAR SIMULATION

As MD simulations become faster and more accurate, they will almost certainly find applications beyond the categories we have discussed. In this section, we speculate on other areas on which biomolecular simulation may make an impact in the coming years, and highlight some of the methodological problems whose solutions may help make these possibilities reality.

Drug Design

A major goal of structural biology in general, and biomolecular simulation in particular, has long been to assist in the design of therapeutics. Simulations are already sometimes utilized as part of the drug development process. Simulation-based binding affinity calculations, for example, guided the design of HIV reverse transcriptase inhibitors mentioned above (103), as well as the subsequent design of inhibitors that maintain high potency in the presence of a drug-resistance mutation (37). The use of MD in mainstream drug discovery efforts, however, remains limited.

In the future, simulation may offer a number of opportunities for improving the drug discovery process. Simulation-based methods can compute ligand–protein affinities more accurately than

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

standard docking methods, aiding in the identification of lead compounds through virtual screening of drug candidates or through a fragment-based approach. Accurate evaluation of binding affinities may prove even more useful in the subsequent process of lead optimization, or in avoiding toxicity by ensuring that drug candidates do not bind to known antitargets. MD also has the potential to discover novel binding sites, including pockets that are not present in existing crystal structures (75, 78). In addition, simulations may allow refinement of low-resolution structural models for proteins, thus enabling structure-based drug design.

MD also allows the examination of interactions between known drugs and genetic variants of protein targets. If a disease becomes resistant to a drug, simulations of the mutated targets may elucidate the mechanism of resistance and facilitate modifications that restore drug efficacy (37). Simulations might even aid in the design of drugs or drug cocktails tailored to the genetic makeup—and thus the unique protein variants—of a particular individual.

Perhaps more importantly, the insights MD can provide into the functional mechanisms of proteins involved in disease pathways may facilitate the identification of appropriate targets and the design of drugs that target those proteins. Many drugs may prove more effective if they bind preferentially to a specific conformation of their target protein. Such conformational selectivity could allow finer control of cellular signaling by stabilizing a particular conformation of a receptor, or reduce side effects by favoring binding to a protein when it is in a particular state of activity.

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

Protein Design

By facilitating optimization of properties such as structure, ligand-binding affinity, or enzymatic activity, MD may play a role in the design of proteins for use as biosensors, industrial catalysts, or therapeutic antibodies, among other potential applications. MD has already been used to rank candidate amino acid sequences on the basis of calculated properties such as binding affinity (41, 59, 101). It may be used in the future not only to test whether a protein binds a ligand, forms a desired interface with another protein, or folds correctly, but to guide the design process in order to achieve such properties. One might even imagine a simulation during which a protein gradually evolves, favoring mutations that improve some measure of its fitness.

Modeling Nucleic Acids

Although nucleic acids are often viewed as static structures that primarily carry sequence information, experimental investigations of both RNAs and DNAs, such as those found in ribosomes, riboswitches, Holliday junctions, and nucleosomes, reveal diverse structures, functions, and dynamics. As with proteins, accurate, long-timescale simulations of nucleic acids should enrich our understanding of important biological mechanisms and may help reveal promising drug targets and binding sites.

To date, MD simulations of proteins have vastly outnumbered those of nucleic acids (56). Reasons for this disparity include both the smaller number of available nucleic acid structures and the relative immaturity of nucleic acid force fields. The Protein Data Bank (**http://www.pdb.org/**) currently contains over 75,000 solved structures, whereas the Nucleic Acid Database (**http://ndbserver.rutgers.edu/**) contains just over 5,000. Force fields for nucleic acids are still undergoing extensive refinement (15, 64, 102, 104), and it is unclear whether the high charge density of nucleic acids and commonly associated divalent cations can be accurately simulated using simple point charges; polarizability may be essential.

Simulation of Complex Biological Structures

The increasing availability of high-resolution structures for biological units of greater size and complexity—motor proteins, multiprotein complexes, ribosomes, whole organelles, and even simple organisms—offers the opportunity to investigate their dynamics through simulation. All-atom MD simulations have already been employed to model a few such units. Using simulations of more than two million atoms, Sanbonmatsu et al. (74) investigated how a ribosome selects transfer RNAs (tRNAs) with high efficiency and accuracy and found that the flexibility of tRNAs was crucial to maintaining the complementary codon–anticodon geometry. Freddolino et al. (26) conducted MD simulations of the complete satellite tobacco mosaic virus, including a capsid consisting of 60 copies of a single protein and a 1,058-base RNA genome. These simulations, with more than a million atoms, suggested that the presence of the RNA is necessary for the assembly and stability of the capsid.

Although such million-atom simulations are impressive, cellular organelles, let alone entire cells, are dramatically larger; a mitochondrion, for instance, is about half a micron in diameter and comprises over ten billion atoms. Further, the functional timescales of large protein complexes and organelles tend to be substantially longer than those of individual proteins, often extending to seconds or more. Such spatial and temporal scales are well beyond those of even the most advanced MD simulations. Fortunately, complex biological structures usually have a hierarchical and modular organization; it may thus be especially productive to develop multiscale models that use the most appropriate abstraction and representation for each temporal and spatial scale. The challenge lies in integrating all-atom MD simulations seamlessly into such multiscale models.

Enabling Longer and Cheaper Simulations

Although the millisecond-scale simulations recently performed on Anton represent a significant milestone, many biochemical events take place on timescales well above a millisecond and involve systems larger than those currently being simulated. The efficiency of a parallel MD simulation on a given hardware platform is determined roughly by the ratio of atoms to processors, so to first approximation one can roughly maintain simulation speed for larger chemical systems by using additional processors. In practice, however, cost often becomes a limiting factor. Accessing longer timescales is even more difficult: Because of interprocessor communication limits, one cannot increase the speed of today's fastest simulations by simply adding more processors. Innovation in some combination of algorithms, computer architecture, and enhanced sampling methods is thus necessary to reach longer timescales and to reduce the cost of large simulations. Such work is currently under way, both in our group and elsewhere.

More Accurate Biochemistry

Despite recent improvements in force fields, MD simulations still face a number of sources of error. First, force fields remain imperfect. Second, most contemporary MD simulations do not fully capture the detailed molecular composition of biological systems, in which many different types of molecules are present, both in the aqueous phase and in lipid bilayers. In the past, force field inaccuracies often overshadowed the effects of unrealistic composition, but as force fields improve, accurately modeling molecular composition will become more important.

Third, classical MD simulations treat covalent bonds as unchanging. Chemical reactions, in which covalent bonds break or form, are typically simulated using techniques such as quantum mechanics/molecular mechanics simulations, in which the bulk of the system is simulated as

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

in classical MD, but a small part is evaluated using more computationally intensive quantum mechanical approaches (39, 77). Several methods are under development to handle chemical reactions directly within an MD framework. Some of these concentrate on capturing changes to the protonation states of ionizable amino acid residues (57, 89). Reactive force fields, which allow covalent bonds between arbitrary pairs of atoms to break or form, have thus far been limited to simple inorganic and organic molecules (94) but may eventually capture more general enzymatic reactions.

A promising avenue to improve the accuracy of MD simulations is to incorporate experimental data directly into the simulations. NMR data, for example, has been used to restrain MD simulations, biasing the protein conformations toward those compatible with the experimental measurements (48). A general framework allowing incorporation of biophysical, biochemical, and even evolutionary data into MD simulations may prove useful both in interpreting experimental data and in making simulation-based predictions.

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

SUMMARY POINTS

1. MD simulation can serve as a computational microscope, revealing the workings of biomolecular systems at a spatial and temporal resolution that is often difficult to access experimentally.
2. Until recently, even the longest atomic-level MD simulations fell short of the microsecond and millisecond timescales on which biochemical events such as protein folding, protein–drug interactions, and major conformational changes typically take place. The speed of the fastest MD simulations has increased 1,000-fold over a period of several years, however, due to the development of specialized hardware and better parallelization algorithms. All-atom simulations of proteins can now reach timescales in excess of a millisecond.
3. These developments, combined with the improvements to the force field models that underlie MD simulations, have allowed MD to capture in atomistic detail processes such as the conformational transitions essential to protein function, the folding of proteins to their native structures, the transport of small molecules across cell membranes, and the binding of drugs to their targets.

FUTURE ISSUES

1. Many biochemical events still take place on long timescales that are inaccessible to atomic-level MD simulations, or on large spatial scales that make atomic-level simulation inordinately expensive. Further improvements in algorithms and computer architectures are needed to make simulations faster and more cost-effective.
2. Multiscale models and enhanced sampling methods will likely also play an essential role in capturing events at larger temporal and spatial scales.
3. Force fields require further improvement and validation, particularly for the modeling of nucleic acids, certain ions, and some types of ligands.
4. Classical MD does not capture breaking and formation of covalent bonds, but it may be possible to handle such reactive chemistry within a generalized MD framework.

5. Application of MD simulation to the design of drugs and proteins remains fertile ground for future research.

DISCLOSURE STATEMENT

David E. Shaw is the beneficial owner of D. E. Shaw Research and serves as its Chief Scientist.

ACKNOWLEDGMENTS

We thank Anton Arkhipov, David Borhani, Michael Eastwood, Morten Jensen, John Klepeis, Kresten Lindorff-Larsen, Venkatesh Mysore, Albert Pan, Stefano Piana, and Yibing Shan for stimulating discussions, helpful comments, and assistance with figures, and Mollie Kirk for editorial assistance.

LITERATURE CITED

1. Adcock SA, McCammon JA. 2006 . Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem. Rev.* 106:1589–615
2. Anderson JA, Lorenz CD, Travesset A. 2008. General purpose molecular dynamics simulations fully implemented on graphics processing units. *J. Comp. Phys.* 227:5342–59
3. Arkin IT, Xu H, Jensen MØ, Arbely E, Bennett ER, et al. 2007. Mechanism of Na^+/H^+ antiporting. *Science* 317:799–803
4. Banks JL, Beard HS, Cao Y, Cho AE, Damm W, et al. 2005. Integrated modeling program, applied chemical theory (IMPACT). *J. Comp. Chem.* 26:1752–80
5. **Beberg AL, Ensign D, Jayachandran G, Khaliq S, Pande VS. 2009. Folding@home: lessons from eight years of volunteer distributed computing. *Proc. 2009 IEEE Intl. Symp. Parallel Distributed Processing* (*IPDPS 09*). Washington, DC: IEEE Comp. Soc.**
6. Berteotti A, Cavalli A, Branduardi D, Gervasio FL, Recanatini M, Parrinello M. 2009. Protein conformational transitions: the closure mechanism of a kinase explored by atomistic simulations. *J. Am. Chem. Soc.* 131:244–50
7. Best RB, Hummer G. 2009. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* 113:9004–15
8. Bostick DL, Brooks CL III. 2007. Selectivity in K^+ channels is due to topological control of the permeant ion's coordination state. *Proc. Natl. Acad. Sci. USA* 104:9260–65
9. Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, et al. 2006. Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proc. Intl. Conf. for High Performance Computing, Networking, Storage and Analysis* (*SC06*). New York: IEEE
10. Bowers KJ, Dror RO, Shaw DE. 2007. Zonal methods for the parallel execution of range-limited N-body simulations. *J. Comp. Phys.* 221:303–29
11. Buch I, Giorgino T, De Fabritiis G. 2011. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* 108:10184–89
12. Case DA, Cheatham TE III, Darden T, Gohlke H, Luo R, et al. 2005. The Amber biomolecular simulation programs. *J. Comp. Chem.* 26:1668–88
13. Chen J, Brooks CL III, Khandogin J. 2008. Recent advances in implicit solvent-based methods for biomolecular simulations. *Curr. Opin. Struct. Biol.* 18:140–48
14. Chow E, Rendleman CA, Bowers KJ, Dror RO, Hughes DH, et al. 2008. Desmond performance on a cluster of multicore processors. *D. E. Shaw Research Technical Report DESRES/TR—2008-01*, D. E. Shaw Research, New York, NY

5. Folding@home uses spare cycles on over 400,000 processors distributed across the internet to perform MD simulations.

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

15. Denning EJ, Priyakumar DU, Nilsson L, Mackerell AD Jr. 2011. Impact of 2-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA. *J. Comput. Chem.* 32:1929–43
16. Dror RO, Arlow DH, Borhani DW, Jensen MØ, Piana S, Shaw DE. 2009. Identification of two distinct inactive conformations of the β_2-adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci. USA* 106:4689–94
17. Dror RO, Arlow DH, Maragakis P, Mildorf TJ, Pan AC, et al. 2011. Activation mechanism of the β_2-adrenergic receptor. *Proc. Natl. Acad. Sci. USA* 108:18684–89
18. Dror RO, Grossman JP, Mackenzie KM, Towles B, Chow E, et al. 2010. Exploiting 162-nanosecond end-to-end communication latency on Anton. *Proc. Conf. for High Performance Computing, Networking, Storage and Analysis* (*SC10*). New York: IEEE
19. Dror RO, Jensen MØ, Borhani DW, Shaw DE. 2010. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *J. Gen. Physiol.* 135:555–62
20. Dror RO, Pan AC, Arlow DH, Borhani DW, Maragakis P, et al. 2011. Pathway and mechanism of drug binding to G protein–coupled receptors. *Proc. Natl. Acad. Sci. USA* 108:13118–23
21. Enkavi G, Tajkhorshid E. 2010. Simulation of spontaneous substrate binding revealing the binding pathway and mechanism and initial conformational response to GlpT. *Biochemistry* 49:1105–14
22. Ensign DL, Kasson PM, Pande VS. 2007. Heterogeneity even at the speed limit of folding: large-scale molecular dynamics study of a fast-folding variant of the villin headpiece. *J. Mol. Biol.* 274:806–16
23. Faraldo-Gómez J-D, Roux B. 2007. On the importance of a funneled energy landscape for the assembly and regulation of multidomain Src tyrosine kinase. *Proc. Natl. Acad. Sci. USA* 104:13643–48
24. Fine R, Dimmler G, Levinthal C. 1991. FASTRUN: a special purpose, hardwired computer for molecular simulation. *Proteins* 11:242–53
25. Fitch BG, Rayshubskiy A, Eleftheriou M, Ward TJC, Giampapa M, et al. 2006. Blue Matter: strong scaling of molecular dynamics on Blue Gene/L. *Comp. Sci.* 3992:846–54
26. Freddolino PL, Arkhipov AS, Larson SB, McPherson A, Schulten K. 2006. Molecular dynamics simulations of the complete satellite Tobacco mosaic virus. *Structure* 14:437–49
27. Freddolino PL, Liu F, Gruebele M, Schulten K. 2008. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys. J.* 94:L75–77
28. Freddolino PL, Schulten K. 2009. Common structural transitions in explicit-solvent simulations of villin headpiece folding. *Biophys. J.* 97:2338–47
29. Friedrichs MS, Eastman P, Vaidyanathan V, Houston M, Legrand S, et al. 2009. Accelerating molecular dynamic simulation on graphics processing units. *J. Comput. Chem.* 30:864–72
30. Harvey MJ, Giupponi G, De Fabritiis G. 2009. ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.* 5:1632–39
31. Hess B, Kutzner C, van der Spoel D, Lindahl D. 2008. GROMACS 4: algorithms for highly efficient, load-balanced and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–47
32. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65:712–25
33. Hurst DP, Grossfield A, Lynch DL, Feller S, Romo TD, et al. 2010. A lipid pathway for ligand binding is necessary for a cannabinoid G protein-coupled receptor. *J. Biol. Chem.* 285:17954–64
34. Jensen MØ, Borhani DW, Lindorff-Larsen K, Maragakis P, Jogini V, et al. 2010. Principles of conduction and hydrophobic gating in K^+ channels. *Proc. Natl. Acad. Sci. USA* 107:5833–38
35. Jensen MØ, Dror RO, Xu H, Borhani DW, Arkin IT, et al. 2008. Dynamic control of slow water transport by aquaporin 0: implications for hydration and junction stability in the eye lens. *Proc. Natl. Acad. Sci. USA* 105:14430–35
36. Jiang W, Hardy DJ, Phillips JC, Mackerell AD Jr, Schulten K, Roux B. 2011. High-performance scalable molecular dynamics simulations of a polarizable force field based on classical Drude oscillators in NAMD. *J. Phys. Chem. Lett.* 2:87–92
37. Jorgensen WL, Bollini M, Thakur VV, Domaoal RA, Spasov KA, Anderson KS. 2011. Efficient discovery of potent anti-HIV agents targeting the Tyr181Cys variant of HIV reverse transcriptase. *J. Am. Chem. Soc.* 133:15686–96

17. Transitions between conformational states in unbiased simulations of a GPCR suggest a mechanism for drug-induced activation.

20. Drugs bind spontaneously in simulation, revealing a sequence of intermediate states en route to the crystallographic binding pose.

34. Simulations of a potassium channel elucidate the mechanisms of ion permeation and voltage gating.

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

38. Jorgensen WL, Buckner JK, Boudon S, Tirado-Rives J. 1988. Efficient computation of absolute free energies of binding by computer simulations: application to the methane dimer in water. *J. Chem. Phys.* 89:3742–46
39. Kamerlin SCL, Warshel A. 2011. The empirical valence bond model: theory and applications. *Wiley Interdisciplinary Rev. Comput. Mol. Sci.* 1:30–45
40. Kirkwood JG. 1935. Statistical mechanics of fluid mixtures. *J. Chem. Phys.* 3:300–13
41. Kiss G, Röthlisberger D, Baker D, Houk KN. 2010. Evaluation and ranking of enzyme designs. *Protein Sci.* 19:1760–73
42. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, et al. 2010. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J. Phys. Chem. B* 144:7830–43
43. Kryshtafovych A, Fidelis K, Moult J. 2011. CASP9 results compared to those of previous CASP experiments. *Prot. Struct. Funct. Bioinform.* 79(Suppl. 10):196–207
44. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13:1011–21
45. Ladurner AG, Itzhaki LS, Daggett V, Fersht AR. 1998. Synergy between simulation and experiment in describing the energy landscape of protein folding. *Proc. Natl. Acad. Sci. USA* 95:8473–78
46. Lefkowitz RJ, Sun J-P, Shukla AK. 2008. A crystal clear view of the β2-adrenergic receptor. *Nat. Biotechnol.* 26:189–91
47. Lei H, Wu C, Liu H, Duan Y. 2007. Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* 104:4925–30
48. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M. 2005. Simultaneous determination of protein structure and dynamics. *Nature* 433:128–32
49. Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE. 2012. Systematic validation of protein force fields against experimental data. *PLoS ONE.* 7:e32131
50. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. 2011. How fast-folding proteins fold. *Science* 334:517–20
51. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, et al. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 78:1950–58
52. Lyman E, Higgs C, Kim B, Lupyan D, Shelley JC, et al. A role for a specific cholesterol interaction in stabilizing the Apo configuration of the human A_{2A} adenosine receptor. *Structure* 17:1660–68
53. Mackerell AD Jr, Feig M, Brooks CL III. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–15
54. Martinez-Mayorga K, Pitman MC, Grossfield A, Feller SE, Brown SA. 2006. Retinal counterion switch mechanism in vision evaluated by molecular simulations. *J. Am. Chem. Soc.* 128:16502–3
55. McCammon JA, Gelin BR, Karplus M. 1977. Dynamics of folded proteins. *Nature* 267:585–90
56. McDowell SE, Spackova N, Sponer J, Walter NG. 2007. Molecular dynamics simulations of RNA: an in silico single molecule approach. *Biopolymers* 85:169–84
57. Mongan J, Case DA. 2005. Biomolecular simulations at constant pH. *Curr. Opin. Struct. Biol.* 15:157–63
58. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink S-J. 2008. The MARTINI coarse grained force field: extension to proteins. *J. Chem. Theory Comput.* 4:819–34
59. Morra G, Baragli C, Colombo G. 2010. Selecting sequences that fold into a defined 3D structure: a new approach for protein design based on molecular dynamics and energetics. *Biophys. Chem.* 146:76–84
60. Moukhametzianov R, Warne T, Edwards PC, Serrano-Vega MJ, Leslie AG, et al. 2011. Two distinct conformations of helix 6 observed in antagonist-bound structures of a β_1-adrenergic receptor. *Proc. Natl. Acad. Sci. USA* 108:8228–32
61. Nguyen H, Jäger M, Kelly J, Gruebele M. 2005. Engineering a β-sheet protein towards the folding speed limit. *J. Phys. Chem. B* 109:15182–86
62. Niesen MJ, Bhattacharya S, Vaidehi N. 2011. The role of conformational ensembles in ligand recognition in G-protein coupled receptors. *J. Am. Chem. Soc.* 133:13197–204
63. Noskov SY, Bernèche S, Roux B. 2004. Control of ion selectivity in potassium channels by electrostatic and dynamic properties of carbonyl ligands. *Nature* 431:830–34

50. Simulations of 12 structurally diverse proteins reproduce experimentally observed structures and provide insight into folding mechanisms.

63. Simulations suggest the physical mechanisms underlying ion selectivity in potassium channels.

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

64. Pérez A, Marchán I, Svozil D, Sponer J, Cheatham TE, et al. 2007. Refinement of the AMBER force field for nucleic acids improving the description of α/γ conformers. *Biophys. J.* 92:3817–29
65. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. 2005. Scalable molecular dynamics with NAMD. *J. Comp. Chem.* 26:1781–802
66. Phillips JC, Stone JE, Schulten K. 2008. Adapting a message-driven parallel application to GPU-accelerated clusters. *Proc. Int. Conf. High Performance Computing, Networking, Storage and Analysis* (*SC08*). New York: IEEE
67. Piana S, Lindorff-Larsen K, Shaw DE. 2011. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* 100:L47–49
68. Piana S, Sarkar K, Lindorff-Larsen K, Guo M, Gruebele M, Shaw DE. 2011. Computational design and experimental testing of the fastest-folding β-sheet protein. *J. Mol. Biol.* 405:43–48
69. Plimpton S. 1995. Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.* 117:1–19
70. Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, et al. 2010. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B* 114:2549–64
71. Rasmussen SGF, Choi H-J, Fung JJ, Pardon E, Casarosa P, et al. 2011. Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature* 469:175–80
72. Romo TD, Grossfield A, Pitman MC. 2010. Concerted interconversion between ionic lock substates of the β_2 adrenergic receptor revealed by microsecond timescale molecular dynamics. *Biophys. J.* 98:76–84
73. Rosenbaum DM, Zhang C, Lyons JA, Holl R, Aragao D, et al. 2011. Structure and function of an irreversible agonist-β_2 adrenoceptor complex. *Nature* 469:236–40
74. Sanbonmatsu KY, Joseph S, Tung CS. 2005. Simulating movement of tRNA into the ribosome during decoding. *Proc. Natl. Acad. Sci. USA* 102:15854–59
75. Schames J, Henchman RH, Siegel JS, Sotriffer CA, Ni H, McCammon JA. 2004. Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* 47:1879–81
76. Seeliger MA, Ranjitkar P, Kasap C, Shan Y, Shaw DE, et al. 2009. Equally potent inhibition of c-Src and Abl by compounds that recognize inactive kinase conformations. *Cancer Res.* 69:2384–92
77. Senn HM, Thiel W. 2009. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed. Engl.* 48:1198–229
78. Shan Y, Kim ET, Eastwood MP, Dror RO, Seeliger MA, Shaw DE. 2011. How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* 133:9181–83
79. Shan Y, Klepeis JL, Eastwood MP, Dror RO, Shaw DE. 2005. Gaussian split Ewald: a fast Ewald mesh method for molecular simulation. *J. Chem. Phys.* 122:054101
80. Shan Y, Seeliger MA, Eastwood MP, Frank F, Xu H, et al. 2009. A conserved protonation-dependent switch controls drug binding in the Abl kinase. *Proc. Natl. Acad. Sci. USA* 106:139–44
81. Shaw DE. 2005. A fast, scalable method for the parallel evaluation of distance-limited pairwise particle interactions. *J. Comp. Chem.* 26:1318–28
82. Shaw DE, Dror RO, Salmon JK, Grossman JP, Mackenzie KM, et al. 2009. Millisecond-scale molecular dynamics simulations on Anton. *Proc. Conf. High Performance Computing, Networking, Storage and Analysis* (SC09). New York: ACM
83. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, et al. 2010. Atomic-level characterization of the structural dynamics of proteins. *Science* 330:341–46
84. Seibert MM, Patriksson A, Hess B, van der Spoel D. 2005. Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. *J. Mol. Biol.* 354:173–83
85. Simmerling S, Strockbine B, Roitberg AE. 2002. All-atom structure prediction and folding simulations of a stable protein. *J. Am. Chem. Soc.* 124:11258–59
86. Snir M. 2004. A note on N-body computations with cutoffs. *Theor. Comput. Sys.* 37:295–318
87. Snow CD, Sorin EJ, Rhee YM, Pande VS. 2005. How well can simulation predict protein folding kinetics and thermodynamics? *Annu. Rev. Biophys. Biomol. Struct.* 34:43–69
88. Sosnick TR, Barrick D. 2011. The folding of single domain proteins—Have we reached a consensus? *Curr. Opin. Struct. Biol.* 21:12–24
89. Stern HA. 2007. Molecular simulation with variable protonation states at constant pH. *J. Chem. Phys.* 126:164112

66. Multiple GPUs are used to accelerate the NAMD parallel MD software.

78. Drug molecules diffuse from solution into their binding site in simulation, spontaneously adopting the crystallographically observed conformation.

82. Reports the performance of Anton, a special-purpose parallel supercomputer, and completion of the first millisecond-long MD simulation.

83. MD simulations as long as a millisecond capture conformational changes, folding, and unfolding of several proteins.

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

90. Taiji M, Narumi T, Ohno Y, Futatsugi N, Suenaga A, et al. 2003. Protein explorer: a petaflops special-purpose computer system for molecular dynamics simulations. *Proc. ACM/IEEE Conf. on Supercomputing* (*SC03*). New York: ACM
91. Tajkhorshid E, Nollert P, Jensen MØ, Miercke LJW, O'Connell J, et al. 2002. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* 296:525–30
92. Toyoda S, Miyagawa H, Kitamura K, Amisaki T, Hashimoto E, et al. 1999. Development of MD engine: high-speed accelerator with parallel processor design for molecular dynamics simulations. *J. Comp. Chem.* 20:185–99
93. van der Kamp MW, Schaeffer RD, Jonsson AL, Scouras AD, Simms AM, et al. 2010. Dynameomics: a comprehensive database of protein dynamics. *Structure* 18:423–35
94. van Duin ACT, Dasgupta S, Lorant F, Goddard WA III. 2001. ReaxFF: a reactive force field for hydrocarbons. *J. Phys. Chem. A* 105:9396
95. Vanni S, Neri M, Tavernelli I, Rothlisberger U. 2009. Observation of "ionic lock" formation in molecular dynamics simulations of wild-type β1 and β2 adrenergic receptors. *Biochemistry* 48:4789–97
96. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, et al. 2010. CHARMM general force field (CGenFF): a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comp. Chem.* 31:671–90
97. Voelz VA, Bowman GR, Beauchamp K, Pande VS. 2011. Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1–39). *J. Am. Chem. Soc.* 132:1526–28
98. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. 2004. Development and testing of a general Amber force field. *J. Comp. Chem.* 25:1157–74
99. Woo H-J, Roux B. 2005. Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. USA* 102:6825–30
100. Yang S, Banavali NK, Roux B. 2009. Mapping the conformational transition in Src activation by cumulating the information from multiple molecular dynamics trajectories. *Proc. Natl. Acad. Sci. USA* 106:3776–81
101. Yang W, Pan Y, Zheng F, Cho H, Tai HH, Zhan CG. 2009. Free-energy perturbation simulation on transition states and redesign of butyrylcholinesterase. *Biophys. J.* 96:1931–38
102. Yildirim I, Stern HA, Tubbs JD, Kennedy SD, Turner DH. 2011. Benchmarking AMBER force fields for RNA: Comparisons to NMR spectra for single-stranded r(GACC) are improved by revised χ torsions. *J. Phys. Chem. B* 115:9261–70
103. Zeevaart JG, Wang L, Thakur VV, Leung CS, Tirado-Rives J, et al. 2008. Optimization of azoles as anti-human immunodeficiency virus agents guided by free-energy calculations. *J. Am. Chem. Soc.* 130:9492–99
104. Zgarbova M, Otyepka M, Sponer J, Mladek A, Banas P, et al. 2011. Refinement of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* 7:2886–902
105. Zimmerberg J, Bezanilla F, Parsegian VA. 1990. Solute inaccessible aqueous volume changes during opening of the potassium channel of the squid giant axon. *Biophys. J.* 57:1049–64
106. Zuckerman DM. 2011. Equilibrium sampling in biomolecular simulations. *Annu. Rev. Biophys.* 40:41–62
107. Zwanzig RW. 1954. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* 22:1420–26

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

Annual Review of Biophysics
Volume 41, 2012

Contents

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.

Annu. Rev. Biophys. 2012.41:429-452. Downloaded from www.annualreviews.org
Access provided by Aalto University on 01/27/19. For personal use only.