



Published as: *Chem Rev.* 2006 May ; 106(5): 1589–1615.

## Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins

Stewart A. Adcock and J. Andrew McCammon\*

*NSF Center for Theoretical Biological Physics, Department of Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0365*

### 1. Introduction

The term molecular mechanics (MM) refers to the use of simple potential-energy functions (e.g., harmonic oscillator or Coulombic potentials) to model molecular systems. Molecular mechanics approaches are widely applied in molecular structure refinement, molecular dynamics (MD) simulations, Monte Carlo (MC) simulations, and ligand-docking simulations.

Typically, molecular mechanics models consist of spherical atoms connected by springs which represent bonds. Internal forces experienced in the model structure are described using simple mathematical functions. For example, Hooke's law is commonly used to describe bonded interactions, and the nonbonded atoms might be treated as inelastic hard spheres or may interact according to a Lennard-Jones potential. Using these simple models, a molecular dynamics simulation numerically solves Newton's equations of motion, thus allowing structural fluctuations to be observed with respect to time.

Dynamic simulation methods are widely used to obtain information on the time evolution of conformations of proteins and other biological macromolecules<sup>1–4</sup> and also kinetic and thermodynamic information. Simulations can provide fine detail concerning the motions of individual particles as a function of time. They can be utilized to quantify the properties of a system at a precision and on a time scale that is otherwise inaccessible, and simulation is, therefore, a valuable tool in extending our understanding of model systems. Theoretical consideration of a system additionally allows one to investigate the specific contributions to a property through "computational alchemy",<sup>5</sup> that is, modifying the simulation in a way that is nonphysical but nonetheless allows a model's characteristics to be probed. One particular example is the artificial conversion of the energy function from that representing one system to that of another during a simulation. This is an important technique in free-energy calculations.<sup>6</sup> Thus, molecular dynamics simulations, along with a range of complementary computational approaches, have become valuable tools for investigating the basis of protein structure and function.

This review offers an outline of the origin of molecular dynamics simulation for protein systems and how it has developed into a robust and trusted tool. This review then covers more recent advances in theory and an illustrative selection of practical studies in which it played a central role. The range of studies in which MD has played a considerable or pivotal role is immense, and this review cannot do justice to them; MD simulations of biomedical importance were recently reviewed.<sup>4</sup> Particular emphasis will be placed on the study of dynamic aspects of protein recognition, an area where molecular dynamics has scope to provide broad and far-ranging insights. This review concludes with a brief discussion of the future potential offered

\*To whom correspondence should be addressed. E-mail: jmccammon@ucsd.edu.

to advancement of the biological and biochemical sciences and the remaining issues that must be overcome to allow the full extent of this potential to be realized.

### 1.1. Historical Background

MD methods were originally conceived within the theoretical physics community during the 1950s. In 1957, Alder and Wainwright<sup>7</sup> performed the earliest MD simulation using the so-called hard-sphere model, in which the atoms interacted only through perfect collisions. Rahman<sup>8</sup> subsequently applied a smooth, continuous potential to mimic real atomic interactions. During the 1970s, as computers became more widespread, MD simulations were developed for more complex systems, culminating in 1976 with the first simulation of a protein<sup>9,10</sup> using an empirical energy function constructed using physics-based first-principles assumptions. MD simulations are now widely and routinely applied and especially popular in the fields of materials science<sup>11,12</sup> and biophysics.

As will be discussed later in this review, a variety of experimental conditions may be simulated with modern theories and algorithms. The initial simulations only considered single molecules in vacuo. Over time, more realistic or at least biologically relevant simulations could be performed. This trend is continuing today.

The initial protein MD simulation, of the small bovine pancreatic trypsin inhibitor (BPTI), covered only 9.2 ps of simulation time. Modern simulations routinely have so-called equilibration periods much longer than that, and production simulations of tens of nanoseconds are routine, with the first microsecond MD simulation being reported in 1998.<sup>13</sup> In addition, the original BPTI simulation included only about 500 atoms rather than the  $10^4$ - $10^6$  atoms that are common today. While much of this advancement results from an immense increase in availability of computing power, major theoretical and methodological developments also contribute significantly.

The number of publications regarding MD theory and application of MD to biological systems is growing at an extraordinary pace. A single review cannot do justice to the recent applications of MD. Using data from ISI Web of Science, the authors estimate that during 2005 at least 800 articles will be published that discuss molecular dynamics and proteins. The historical counts are shown in Figure 1.

### 1.2. Protein Dynamics

The various dynamic processes that can be characterized for proteins have time scales ranging from femtoseconds to hours. They also cover an extensive range of amplitudes and energies. Many of these motions have critical roles in biochemical functions.<sup>14</sup> Rapid and localized motions may play a role in enzymatic reactions. Slower motions that occur on the scale of whole proteins include allosteric coupling<sup>15</sup> and folding transitions. Subunit associations occur over even longer distances.

Simulations of the longer time scale folding events are covered elsewhere in this issue.<sup>16</sup> Characteristic time scales for protein motions are shown in Table 1.

## 2. Application of Molecular Dynamics in the Study of Biomolecular Phenomena

Molecular dynamics can now be routinely applied in the investigation of a wide range of dynamic properties and processes by researchers in numerous fields, including structural biochemistry, biophysics, enzymology, molecular biology, pharmaceutical chemistry, and biotechnology. Using MD simulations, one is able to study thermodynamic properties and

time-dependent (i.e., kinetic) phenomena. This enables an understanding to be developed of various dynamic aspects of biomolecular structure, recognition, and function. However, when used alone, MD is of limited utility. An MD trajectory (i.e., the progress of simulated structure with respect to time) generally provides data only at the level of atomic positions, velocities, and single-point energies. To obtain the macroscopic properties in which one is usually interested requires the application of *statistical mechanics*, which connects microscopic simulations and macroscopic observables.

Statistical mechanics provides a rigorous framework of mathematical expressions that relate the distributions and motions of atoms and molecules to macroscopic observables such as pressure, heat capacity, and free energies.<sup>17,18</sup> Extraction of these macroscopic observables is therefore possible from the microscopic data, and one can predict, for instance, changes in the binding free energy of a particular drug candidate or the mechanisms and energetic consequences of conformational change in a particular protein.

Specific aspects of biomolecular structure, kinetics, and thermodynamics that may be investigated via MD include, for example, macromolecular stability,<sup>19</sup> conformational and allosteric properties,<sup>20</sup> the role of dynamics in enzyme activity,<sup>21,22</sup> molecular recognition and the properties of complexes,<sup>21,23</sup> ion and small molecule transport,<sup>24,25</sup> protein association,<sup>26</sup> protein folding,<sup>27,16</sup> and protein hydration.<sup>28</sup>

MD, therefore, provides the opportunity to perform a variety of studies including molecular design (drug design<sup>29</sup> and protein design<sup>30</sup>) and structure determination and refinement (X-ray,<sup>31</sup> NMR,<sup>32</sup> and modeling<sup>33</sup>).

### 3. Molecular Dynamics Methods and Theory

Given the structure of a biomolecular system, that is, the relative coordinates of the constituent atoms, there are various computational methods that can be used to investigate and study the dynamics of that system. In the present section, a number of such methods are described and discussed. The majority of important dynamics methodologies are highly dependent upon the availability of a suitable potential-energy function to describe the energy landscape of the system with respect to the aforementioned atomic coordinates. This critical aspect is, therefore, introduced first.

#### 3.1. Potential Functions and the Energy Landscape

Choice of an appropriate energy function for describing the intermolecular and intramolecular interactions is critical to a successful (i.e., valid yet tractable) molecular dynamics simulation. In conventional MD simulations, the energy function for nonbonded interactions tends to be a simple pairwise additive function (for computational reasons) of nuclear coordinates only. This use of a single nuclear coordinate to represent atoms is justified in terms of the Born-Oppenheimer approximation.<sup>34</sup> For bonded groups of atoms, that is those that form covalent bonds, bond angles, or dihedral angles, simple two-body, three-body, and four-body terms are used, as described below.

The energy functions usually consist of a large number of parametrized terms. These parameters are chiefly obtained from experimental and/or quantum mechanical studies of small molecules or fragments, and it is assumed that such parameters may be transferred to the larger molecule of interest. The set of functions along with the associated set of parameters is termed a force field. A variety of force fields have been developed specifically for simulation of proteins.

There are notable exceptions, but it is usual for a force field to be purely additive. For instance, bond lengths are not considered to be dependent on the bond angles, and atomic partial charges are fixed in magnitude. This is generally considered to give a reasonable, although not flawless,<sup>35</sup> approximation of the potential-energy landscape. Alternative methods for probing dynamics might demand additional restrictions or properties to be satisfied by the potential functions, and these are detailed in the appropriate sections of this review.

Some force fields, often termed class II force fields, do incorporate cross terms or higher order terms.<sup>36–38</sup> These class II force fields were typically developed to reproduce vibrational spectra accurately or treat structures with geometries far from their equilibrium values.

One fairly typical and widely applied force field is the CHARMM22 force field.<sup>39–41</sup> Like all widely applied force fields, it consists of several discrete terms. Each of these terms possesses a simple functional form and describes an intermolecular or intramolecular force exhibited within the system given the set of relative atomic coordinates

$$\begin{aligned}
 V(\vec{R}) = & \sum_{\text{bonds}} K_d(d - d_0)^2 + \sum_{\text{Urey-Bradley}} K_{UB}(S - S_0)^2 + \\
 & \sum_{\text{angle}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\chi(1 + \cos(n\chi - \delta)) + \\
 & \sum_{\text{impropers}} K_\phi(\phi - \phi_0)^2 + \\
 & \sum_{\text{nonbond}} \left\{ \epsilon_{ij} \left[ \left( \frac{R_{ij}^{\text{min}}}{r_{ij}} \right)^{12} - \left( \frac{R_{ij}^{\text{min}}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon_{ij}} \right\}
 \end{aligned} \tag{1}$$

where  $K_d$ ,  $K_{UB}$ ,  $K_\theta$ ,  $K_\chi$ , and  $K_\phi$  are the bond length, Urey-Bradley (1–3 bond length), bond angle, dihedral angle, and improper dihedral angle force constants, respectively. Likewise,  $d$ ,  $S$ ,  $\theta$ ,  $\chi$ , and  $\phi$  are the bond length, Urey-Bradley (1–3 bond length), bond angle, dihedral angle, and improper dihedral angle values exhibited in the current configuration, and the zero subscript represents the reference, or equilibrium, values for each of those. These terms represent the bonded interactions. The final term in the function represents the nonbonded interactions, incorporating Coulombic and Lennard-Jones interactions.  $\epsilon_{ij}$  relates to the Lennard-Jones well depth,  $R_{ij}^{\text{min}}$  is the distance at which the Lennard-Jones potential is zero,  $q_i$  is the partial atomic charge of atom  $i$ ,  $\epsilon_1$  is the effective dielectric constant, and  $r_{ij}$  is the distance between atoms  $i$  and  $j$ . The Lorentz-Berthelodt combination rules<sup>42</sup> are used to obtain the necessary Lennard-Jones parameters for each pair of different atoms;  $\epsilon_{ij}$  values are the geometric mean of the  $\epsilon_{ii}$  and  $\epsilon_{jj}$  values, while  $R_{ij}^{\text{min}}$  values are the arithmetic mean of the  $R_{ii}^{\text{min}}$  and  $R_{jj}^{\text{min}}$  values. Values for the atomic partial charges,  $q_i$ , are determined from a template-based scheme, with charges often modified to reproduce dielectric shielding effects (i.e., to mimic some of the effects of shielding from a high dielectric constant solvent). This  $\epsilon_1$  is usually set to unity for simulations incorporating explicit solvent representations.

Using eq 1 one may evaluate the potential energy,  $V(r)$ , of the system from a single set of atomic coordinates since the relevant distances and angles are easily determined. The energy is that of a single instance, termed a snapshot, of the system.

$V(r)$  includes contributions from every bond angle and dihedral angle in the system; however, it might be noted that many Urey-Bradley terms and improper dihedral angles are not used. Only those that are required for fitting computational results to observable vibrational spectra are utilized. In addition, the linear term in the original Urey-Bradley function is not incorporated at all. This is a reasonable approximation because it has been shown that, when Cartesian coordinates are used, only the quadratic term is required for determining vibrational frequencies.<sup>43</sup> The nonbonded terms are applied to all atoms except those attached through one or two covalent bonds. In certain, specific, cases the Lennard-Jones term is adjusted for

atoms connected through three covalent bonds in order to accurately reproduce experimentally observed structures. An example of such a case is the nitrogen and oxygen atoms of amides.

For the purposes of MD, it is advantageous for the force field to have efficiently accessible first and second derivatives with respect to atomic position (which correspond to the physical characteristics of atomic force and force gradients, respectively), and this is one of the more notable reasons for the very simple mathematical forms generally chosen.

The CHARMM force fields have been separately parametrized for proteins,<sup>39</sup> nucleic acids,<sup>44</sup> lipids,<sup>45</sup> and carbohydrates<sup>46,47</sup> with the goal of consistency between these sets, allowing for simulation of heterogeneous systems. Different force fields exist for small organic molecules<sup>48-51</sup> and nonbiological macromolecules such as zeolites.

### 3.2. Energy Minimization

Although not strictly a dynamics method, energy minimization is a fundamental concept upon which much of the theory discussed in this review is built.

Given a set of  $N$  independent variables,  $r$ , where  $r = (r_1, r_2, r_3, \dots, r_N)$ , the task is to find the values for each of those variables, termed  $r_{\min}$ , for which a particular function,  $V(r)$ , has its global minimum. In the case of a molecular mechanics protein model,  $N$  is typically three times the number of atoms (resulting from three degrees of freedom per atomic coordinate),  $r$  encodes the atomic coordinates (e.g., the Cartesian coordinates), and  $V$  is the potential energy as given by an equation such as eq 1. It may be seen that computationally this task is a nonlinear optimization problem.

Numerous algorithms exist for solving such nonlinear optimization problems, and a small selection of these are widely applied in molecular mechanics modeling of proteins. Relevant algorithms are reviewed elsewhere.<sup>52,53</sup>

It is extremely difficult to locate the global minimum of a general nonlinear function with more than a dozen independent variables. Typical biomolecular systems with as few as a hundred atoms will be described with on the order of 300 variables; thus, it is usually impossible to provably locate the global minimum. Also, while energy minimization methods may be used to efficiently refine molecular structures, they are totally inadequate for sampling conformational space. Given an unrefined molecular structure with bond angles and lengths distorted from their respective minima or with steric clashes between atoms, energy minimization methods can be very useful for correcting these flaws and are therefore routinely applied to protein systems. The most popular methods include those that use derivatives of various orders, including the first-order (i.e., utilizes first-order derivatives) steepest descent and conjugate gradient methods and the second-order (i.e., utilizes second-order derivatives) Newton-Raphson method.

The steepest descent method is one of several first-order iterative descent methods. These all utilize the gradient of the potential-energy surface, which directly relates to forces in the MM description of molecular systems, to guide a search path toward the nearest energy minimum. Because this corresponds to reducing the potential energy by moving atoms in response to the force applied on them by the remainder of the system, this method is attractive as it may be considered to have a behavior that is physically meaningful. Formally, the force vector is defined as  $F(r) = -d/dr V(r)$  where  $r$  is the vector of atomic coordinates.

In all of the iterative descent methods, a succession of atomic configurations are generated by applying, for iteration  $k$ , the relationship  $x(k) = x(k-1) + \lambda(k)F(k)$ , where the vector  $x$  represents the  $3N$  dimensional configuration,  $\lambda(k)$  is a step size, and  $F(k)$  is the force vector. The step size

for the first iteration is usually selected arbitrarily or else by some simple heuristic. After every iteration this step size is adjusted according to whether the overall potential energy of the system was reduced or increased by that step. If the energy increased, it is assumed that the step size was sufficiently large to jump over the local minimum along the search direction, and accordingly the step size is reduced by some multiplicative factor, typically 0.5. In the event that the energy was indeed reduced, the step size is increased by some factor, typically around 1.2. This continual adjustment of the step size keeps it roughly appropriate for the particular curvature of the potential-energy function in the region of interest. While the steepest descent method is highly inefficient for multidimensional problems with irregular potential surfaces with multiple local minima, it is robust in locating the closest local minimum. Consequently, the global motions required to locate the global energy minimum will not be observed. Nonetheless, it is very effective in removing steric conflicts and relaxing bond lengths and bond angles to their canonical values.

The Newton-Raphson method is a popular second-derivative method, although it requires some simple modifications before it is suited to typical biomolecular MM systems. The basic method relies on the assumption that, at least in the region of the minima, the potential energy is quadratically related to the individual variables.  $V(x_i) \simeq a + bx_i + cx_i^2$ , where  $a$ ,  $b$ , and  $c$  are constants. This leads to first and second derivatives of

$$\frac{dV(x)}{dx} = b + 2cx \quad \text{and} \quad \frac{d^2V(x)}{dx^2} = 2c \quad (2)$$

At the minimum,  $dV(x_{\min})/dx = 0$ , so  $x_{\min}$  may be calculated using

$$x_{\min} = \frac{-b}{2c} = \frac{\left(2cx - \frac{dV(x)}{dx}\right)}{\left(\frac{d^2V(x)}{dx^2}\right)} = x - \frac{\left(\frac{dV(x)}{dx}\right)}{\left(\frac{d^2V(x)}{dx^2}\right)} \quad (3)$$

For quadratic surfaces, no iterative searching is necessary since the exact minimum may be determined from the current configuration and the derivatives at that configuration. Unfortunately, biomolecular MM systems tend to be extremely nonquadratic and also contain many local minima. These characteristics render the basic Newton-Raphson method less useful. However, it has found widespread use as a method for efficiently completing the optimization performed via an alternative method. One modified form of the method, adopted basis set Newton-Raphson (ABNR), is very effective for large biomolecular systems.<sup>41</sup>

### 3.3. Adiabatic Mapping

The simplest approach to studying motion in proteins is the characterization of low-energy paths for specific motions. This approach is termed “adiabatic mapping”.<sup>54</sup>

The protocol typically followed involves forcing specific atoms to move along a predetermined path to cause a structural change of interest. The remaining atoms are allowed to move freely, subject to the potential-energy landscape, to reduce (or minimize) the overall potential energy at each point along the path. It is assumed that, since shifts in atomic coordinates will roughly correspond to the structural fluctuations required to allow the motion, these energies approximate the change in energies that should be observed during the associated real, spontaneous, motion.

Adiabatic mapping is computationally inexpensive and has therefore been applied to study many structural changes of various magnitudes or scales. No direct information on the time scales of dynamic mechanisms is obtained, although some approximate results can be derived

from the relaxed energies in analytical models of the dynamics (e.g., the Langevin equation, below).

The major flaw in this method is the dependence of the results on the path selected to represent, or drive, the entire motion. If the motion actually proceeds along a different path, then misleading results will be obtained. Quantitative errors are also to be expected due to any incomplete conformational relaxation, chiefly overestimation of enthalpy barriers caused by incomplete relaxation of delocalized strain. Energy minimization algorithms tend to be inefficient with respect to nonlocal strain (i.e., that which may be driving large domain motions). In addition, this approach ignores certain important thermal effects. For instance, neither the entropy nor the temperature dependence of the enthalpy is ordinarily obtained, despite these being important factors in the kinetics of structural motions.

### 3.4. Molecular Dynamics

In simple terms, molecular dynamics simulations involve the iterative numerical calculation of instantaneous forces present in a MM system and the consequential movements in that system. The MM system consists of a set of particles that move in response to their interactions according to the equations of motion defined in classical (i.e., Newtonian) mechanics. Classical MD is much more efficient than might be expected from full consideration of the physics of biomolecular systems due to the number of substantial approximations. Notably, quantum dynamical effects are usually ignored. Instead, each particle (typically a single atom, but sometimes a rigid set of atoms) is considered to be a point mass. This approximation is justified in terms of the Born-Oppenheimer approximation<sup>34</sup> (i.e., only the nuclear displacements need to be considered). This section provides a brief overview of the concepts upon which molecular dynamics simulations are justified and implemented.

For an atom,  $i$ , with mass  $m_i$ , and position indicated by the 3-dimensional vector  $\mathbf{r}_i$ , the relationship between the atom's velocity and momentum,  $p_i$ , is

$$\frac{dr_i}{dt} = \frac{p_i}{m_i} \quad (4)$$

The net force,  $F_i$ , exerted on the atom  $i$  by the remainder of the system is given by the negative gradient of the potential-energy function with respect to the position of atom  $i$

$$F_i = - \frac{dV}{dr_i} \quad (5)$$

The Newtonian equation of motion for atom  $i$  is

$$\frac{dp_i}{dt} = F_i \quad (6)$$

Given the position with respect to a single component of vector  $\mathbf{r}_i$ , (that is the position along a single dimension,  $x$ ) at a specific time,  $t$ , then the position after a short and finite interval, denoted  $\Delta t$ , is given by a standard Taylor series

$$x(t+\Delta t) = x(t) + \frac{dx(t)}{dt} \Delta t + \frac{d^2x(t)}{dt^2} \frac{\Delta t^2}{2} + \dots \quad (7)$$

The position  $x(t)$ , the velocity  $dx(t)/dt$ , and the acceleration  $d^2x(t)/dt^2$  are sufficient for numerical solution to the equations of motion if some approximation to account for higher

order terms in the Taylor series can be made. For this single dimension, Newton's second law describes the acceleration

$$\frac{d^2x(t)}{dt^2} = \frac{F_x}{m} \quad (8)$$

where  $F_x$  is the component of the net force acting on the atom parallel to the direction of  $x$ .

This just leaves the unspecified approximation for the infinite series of higher terms from the Taylor expansion to be devised. The simplest approach is to assume that the higher terms sum to zero, effectively truncating the Taylor expansion at the second derivative, the acceleration. In the general case, this is a very poor approximation as highlighted by consideration of Newton's third law. The net force acting in the entire system should be zero, resulting in conservation of the total energy (i.e., kinetic plus potential energies) and conservation of the total momentum. With the simple approximation suggested, significant fluctuations and drifting over time occur in the total energy of the system as a simulation progresses. A wide range of improvements to this simple approximation are used in modern molecular dynamics software, many of which are described later in this review.

Numerous algorithms exist for integrating the equations of motion.<sup>55–58</sup> Many of these are finite difference methods in which the integration is partitioned into small steps, each separated in time by a specific period  $\Delta t$  because the continuous potentials describing atomic interaction preclude an analytical solution. The simple Verlet algorithm<sup>55</sup> uses the atomic positions and accelerations at time  $t$  and the positions from the prior step,  $x(t - \Delta t)$ , to determine the new positions at  $t + \Delta t$

$$x(t+\Delta t) = 2x(t) - x(t - \Delta t) + \frac{d^2x(t)}{dt^2} \Delta t^2 \quad (9)$$

A slight modification of this, known as the leapfrog algorithm,<sup>59</sup> is popular. The leapfrog algorithm uses the positions at time  $t$  and the velocities at time  $t - (\Delta t/2)$  for the update of both positions and velocities via the calculated forces,  $F(t)$ , acting on the atoms at time  $t$

$$x(t+\Delta t) = x(t) + \frac{dx(t)}{dt} \left( t + \frac{\Delta t}{2} \right) \Delta t \quad (10)$$

$$\frac{dx(t)}{dt} \left( t + \frac{\Delta t}{2} \right) = \frac{dx(t)}{dt} \left( t - \frac{\Delta t}{2} \right) + \frac{d^2x(t)}{dt^2} \Delta t \quad (11)$$

Alternative finite difference method integrators include the velocity Verlet method<sup>57</sup> and the Beeman algorithm.<sup>56</sup> The velocity Verlet method synchronizes the calculation of positions, velocities, and accelerations without sacrificing precision. The Beeman algorithm exhibits improved energy conservation characteristics due to its more accurate expression for velocities.

All of these commonly used integrators are time reversible. This means the direction of simulation in time is arbitrary. If the velocities of all atoms were swapped in sign, the simulation would run in exactly the reverse direction.

The computational expense of using any particular integration scheme is important, but for practical simulations of proteins another consideration becomes even more critical. The computational demands of the integration method are insignificant compared to the calculation of all the forces acting within the system. It is therefore advantageous to limit the number of force calculations required during the simulation. One method for doing this is to select an



integrator that allows longer time steps without deviating significantly from the path of an exact, analytical trajectory.

The degree to which the Taylor series expansion is important in determining the accuracy of each integrator depends on which terms they include. The largest term in the Taylor expansion which is not considered in a given integration scheme defines the, so-called, *order* of that method. The Verlet algorithm, for example, is a fourth-order method with terms beyond  $\Delta t^4$  truncated.

A family of integration algorithms which are correct to a selected error order are known as the predictor-corrector methods.<sup>58</sup> These methods initially estimate the positions, velocities, accelerations, and any desired higher order terms of the Taylor expansion. Next, forces are calculated with these estimated positions giving new accelerations at time  $(t + \Delta t)$ . The two sets of accelerations are compared, and a correction step adjusts the originally estimated positions, velocities, etc.

**3.4.1. Simulated Environment**—A range of experimental conditions can be simulated by MD. The earliest protein simulations<sup>9,10,60</sup> considered the molecules as isolated entities, effectively in a vacuum. Later simulations included explicit water and neighboring protein molecules as in a crystal environment. It is now conventional to duplicate the system periodically in all directions to represent an essentially infinite system. Typically, a cubic lattice is used for replication of the central cubic box. The atoms outside the central box are simply images of the atoms simulated in that box. So-called *periodic boundary conditions* ensure that all simulated atoms are surrounded by neighboring atoms, whether those neighbors are images or not. The so-called *minimum image convention* guarantees that duplicate interactions between atoms are not included by calculating only one pairwise interaction for each pair of atoms. For atoms  $i$  and  $j$ , the interaction is that between the original atom  $i$  and whichever copy of atom  $j$ , original or image, is closest to atom  $i$ .

Periodic boundary conditions are not restricted to cubic systems. Other geometries are used including the rhombic dodecahedron<sup>61</sup> and the truncated octahedron.<sup>62</sup> These can significantly reduce the number of solvent atoms required in the system, leading to a corresponding reduction in the computational requirements. The range of possible geometries suitable for periodic systems is limited, but stochastic boundary conditions<sup>63</sup> can be utilized, in the absence of periodicity, with any system geometry.

Stochastic boundary conditions are particularly useful when investigating only a particular region such as the binding site in a ligand-binding study. This enables much of the system that would otherwise be simulated to be excluded, thus saving considerable computational resources. The region of interest is enclosed within a shell, usually spherical. The atoms belonging to this shell region are subject to stochastic dynamics, for example, evaluated using the Langevin equation. The stochastic shell region itself is enclosed in a bath region in which the atoms are stationary. This outer region forms a barrier that maintains the overall structure of the system, while the shell region accommodates any local fluctuations in conformation, density, or energy that occur in the central region where standard MD is performed. This approach has been applied in the study of proteins,<sup>64</sup> but the restrictive boundaries in the simplest models are known to introduce artificial density fluctuations and can alter the structure of solvents such as water.<sup>65</sup> More recent models have improved characteristics.<sup>66–68</sup>

**3.4.2. SHAKE**—From a fixed amount of computation, the length of a simulation is determined by a number of factors including the cost of evaluating interactions, number of interactions that need to be evaluated at each time step, period of that time step, and number of degrees of freedom that need to be propagated. To increase the efficiency of a computer simulation, any

of those four aspects might be improved upon. Increasing the time step period is, therefore, a simple approach for extending tractable simulation time scales, but a number of factors limit the step size.<sup>69</sup> The number of interactions to evaluate may be reduced via the use of implicit solvent models, discussed later in this review, or by a reduced representation of the biomolecular structure, also discussed later.

Numerous algorithmic improvements can be applied to enhance the stability and increase the efficiency of MD simulations. The use of integrators with good stability properties such as velocity Verlet<sup>57</sup> and extensions such as the reversible reference system propagator algorithm method (RESPA)<sup>70</sup> are typical and allow extended time steps to be utilized. Improvements in efficiency are often obtained through freezing the fastest modes of vibration by constraining the bonds to hydrogen atoms to fixed lengths using algorithms such as SHAKE,<sup>60,71</sup> RATTLE,<sup>72</sup> and LINCS.<sup>73</sup> Specifically, the use of RESPA and fixing of bond lengths involving hydrogen atoms with SHAKE, RATTLE, or LINCS allow the use of larger time-step ( $\Delta t$ ) sizes without any significant amount of degradation in the quality of the trajectory (or in the accuracy of the simulation).

The SHAKE algorithm (otherwise known as the constrained Verlet method) is a straightforward modification of the Verlet algorithm to impose constraints on the internal coordinates such as bond lengths and bond angles. The length of the time step is restricted by the requirement that  $\Delta t$  is small compared to the period of the highest frequency motions being simulated. For the biomolecular systems of interest, the highest frequency motions are the bond stretching vibrations, yet these vibrations are generally of minimal interest in the study of biomolecular structure and function. Thus, algorithms, such as SHAKE, that constrain the bonds to their equilibrium lengths are useful. In essence, they may be considered as averaging out the highest frequency vibrations.

In the SHAKE algorithm all constraints are imposed through fixed interatomic distances. In the case of bond lengths, single interatomic distances are sufficient. To constrain bond angles the fact that three constituent atomic coordinates are related through three interatomic distances is relied on.

If constraint  $k$  is on the distance between atoms  $i$  and  $j$ , then it may be expressed as

$$r_{ij}^2 - d_{ij}^2 = 0 \quad (12)$$

where  $r_{ij}$  is the vector from atom  $i$  to atom  $j$  ( $r_{ij} = r_j - r_i$ ) and  $d_{ij}$  is the desired distance. At any given step during the practical numerical simulation, the constraint is said to be satisfied whenever the deviation is less than some threshold. In the case of SHAKE, the constraint is satisfied when  $r_{ij}^2 - d_{ij}^2 < \epsilon/d_k^2$ , where  $\epsilon$  is a constant and  $d_k^2$  is the equilibrium bond length.

All except the highest frequency motions of proteins are not noticeably affected, although constraints are not recommended for valence bond angles except for those in the rigid water models.<sup>74</sup> This is chiefly due to coupling between the bond-angle motions and dihedral motions. It is also usually applied only to the bonds with the fastest vibrations, namely, those involving a hydrogen atom. Nonetheless, in practice the time step can typically be increased by a factor of 3 compared to simulations with the original Verlet algorithm.

The LINCS constraint method directly resets the constraints rather than the derivatives of the constraints (i.e., resets the constrained distances rather than the velocities), therefore avoiding drift inherent in the SHAKE method. It is also reported to produce a further speed up of about four times.<sup>73</sup>

Other improved variations of SHAKE have been proposed, including MSHAKE,<sup>75</sup> which performs matrix inversion to solve the constraint equations. Such variants are suited to systems with a limited number of interdependent constraints (e.g., water) where the cost of inverting the matrix is favorable compared to performing a large number of iterations. This method is also useful when a high level of accuracy in the application of constraints is desired. The QSHAKE method<sup>76</sup> introduces quaternion dynamics for rigid fragments. The total number of holonomic constraints is reduced, thus obtaining convergence within fewer iterations and increasing stability under larger time steps. Generalized SHAKE<sup>77</sup> adds support for general nonholonomic constraints, and no numerical drift is observed for large numbers of constraints with this approach.

**3.4.3. Experimental Conditions**—Typically, it is important to accurately simulate the experimental conditions to be replicated. Various values for physical conditions, such as pressure and temperature, may be readily considered in the simulations.

**Ensembles:** An ensemble is a collection of all possible systems that have differing microscopic states but belong to a single macroscopic or thermodynamic state.<sup>17</sup> Various different formal ensembles with differing characteristics exist. The most widely simulated are as follows. (1) The canonical ensemble (NVT): This is the collection of all systems whose thermodynamic state is characterized by a fixed number of atoms,  $N$ , fixed volume,  $V$ , and fixed temperature,  $T$ . (2) The isobaric-isoenthalpic ensemble (NPH): An ensemble with a fixed number of atoms,  $N$ , fixed pressure,  $P$ , and fixed enthalpy,  $H$ . (3) The isobaric-isothermal ensemble (NPT): An ensemble with a fixed number of atoms,  $N$ , fixed pressure,  $P$ , and fixed temperature,  $T$ . (4) The grand canonical ensemble ( $\mu$ VT): A thermodynamic state characterized by a fixed chemical potential,  $\mu$ , fixed volume,  $V$ , and fixed temperature,  $T$ . (5) The microcanonical ensemble (NVE): A thermodynamic state characterized by a fixed number of atoms,  $N$ , fixed volume,  $V$ , and fixed energy,  $E$ . This corresponds to a closed (i.e., isolated) system since energy is conserved.

Most early simulations corresponded to the microcanonical ensemble under so-called free dynamics. However, experiments are usually performed at constant temperature and volume (i.e., the canonical ensemble) or constant pressure and temperature (i.e., the isobaric-isothermal ensemble), so it is often desirable to simulate these conditions instead or mimic these conditions or those expected under physiological conditions. During a simulation at constant energy, the temperature will be observed to fluctuate due to the spontaneous interconversion of the kinetic and potential components of the total energy. The instantaneous temperature may be evaluated from the atomic velocities using

$$3k_{\text{B}}T = \frac{\sum_{i=1}^N m_i v_i^2}{N} \quad (13)$$

where  $k_{\text{B}}$  is Boltzmann's constant,  $m_i$  and  $v_i$  are the mass and velocity of atom  $i$ , respectively, and  $N$  is the total number of atoms. If desired, the atomic velocities can be rescaled or otherwise modified to keep the temperature constant during the course of a simulation. It is worth mentioning that eq 13 must be corrected when constraint algorithms, such as SHAKE, are used.<sup>78</sup>

To maintain a constant pressure during a simulation, the volume needs to be allowed to fluctuate by adjusting the dimensions of the periodic box and rescaling the atomic positions accordingly. Numerous methods exist for running MD simulations at a constant pressure properly. Methods

include the *extended system algorithm*,<sup>79</sup> the *constraint algorithm*,<sup>80</sup> *weak coupling to an external bath*,<sup>81</sup> the *hybrid method*,<sup>82</sup> and the *Langevin piston method*.<sup>83</sup>

**3.4.4. Solvation**—A reasonable representation of a protein's environment is important for characterizing its properties through simulations. At today's levels of understanding and computational resources, it is not currently possible to fully consider the full physiological environment of any typical protein. Certain specific cases can be considered more fully, but the general case is too complicated. The nonphysiological cases that involve proteins in crystal-packing arrangements or in vacuo (such as in the original BPTI simulation<sup>9</sup>) are comparatively trivial. Even simulations of in vitro systems have particular issues to consider as outlined in this section. Nonetheless, such systems are tractable, so an aqueous solvent is selected as the environment for the vast majority simulations. Great strides have been made for simulations in more specific environments such as those for transmembrane proteins.<sup>24,25,84,85</sup>

**Implicit Solvation:** Most proteins exist, at least partially, within an aqueous environment. Justified by this fact, it is common to assume that a protein is fully solvated in pure or ion-containing water during simulations. However, a considerable portion of the computation time could be spent evaluating the solvent-solvent interactions. It is therefore desirable to avoid using explicit water when possible. However, solvent effects are important and cannot be totally disregarded. Consequently, numerous implicit solvent models have been developed.<sup>86–89</sup>

In addition to the dielectric screening effects, an explicit solvent contributes specific interactions that are often important for mediating protein structure or function. Thus, explicit solvents play an important role in simulations for the accurate consideration of electrostatic effects<sup>90</sup> and for the valid decomposition of free energies, for example. Conversely, implicit models of solvation allow for better direct estimations of free energies of solvation than explicit solvation models.<sup>91</sup> The statistical mechanical characteristics and properties of implicit solvation models have been rigorously examined.<sup>92</sup>

One implicit solvent model is the generalized Born (GB) model<sup>93,88</sup> which, especially on parallel computer systems, can be used to run significantly faster MD simulations than can explicit solvent models.<sup>94</sup> Like all implicit models, GB is known to be unable to reproduce certain microscopic solvent features.<sup>87</sup> Moreover, implicit solvent models are known to facilitate modified conformational dynamics of protein molecules when compared to explicit models,<sup>95</sup> which is usually undesirable. A hybrid method which incorporates explicit solvent molecules in a defined region of the system, such as a binding site or a channel, is the generalized solvent potential method.<sup>96</sup> In this method a static solvent-shielded field from the biomolecular solute is calculated using a finite-difference Poisson-Boltzmann method. This field is used to impose a solvent reaction field, and the specific region of interest for explicit solvent is hydrated.

Apolar solvation models, using a cavity potential plus dispersion potential decomposition, such as the analytical generalized Born and nonpolar (AGBNP) solvent model,<sup>97</sup> have been shown to be very effective. The apolar component is likely to be necessary in the exploration of larger conformational changes.<sup>98</sup> This overcomes the poor correlation often found between the apolar forces from explicit solvent and implicit solvent simulations.<sup>99</sup>

**Explicit Water Models:** For cases when explicit consideration of the solvent is desirable, or necessary, there is a wide range of explicit water models available. The most popular of these models include TIP3P, TIP4P,<sup>100</sup> TIP5P,<sup>101</sup> SPC, and SPC/E.<sup>102</sup>

Commonly, the parameters in water models are adjusted such that the enthalpy of vaporization and the density of water are reproduced in simulations. All of the above models have a dipole moment of about 2.3 D instead of the experimental gas-phase value of 1.85 D. The temperature dependence of the density of water is not described well by any of these models perhaps except the TIP5P model.

The most popular models for water are all consistent with the SHAKE approximation, discussed above, since these models for water treat the molecules as completely rigid.

MD simulations with only a thin layer of water around the protein can overcome some of the problems of a purely implicit solvent.<sup>103–105</sup> The restrained water droplet model applies a weak harmonic restraining force to a 5 Å shell of water.<sup>106</sup>

**Electrostatics:** Long-range electrostatic interactions<sup>107</sup> play a dominant role in protein structural stability and are also crucial determinants in the initial encounter of many association processes.<sup>26</sup> Typically, the most computationally expensive portion of a MD simulation is the evaluation of these long-range electrostatic interactions.<sup>108,109</sup> As the number of charges in a system increases, the number of Coulombic interactions will grow as the square of that number, potentially resulting in a prohibitively large number of interactions to evaluate.

In earlier MD simulations, a cutoff was applied to the distance of electrostatic interactions, known as *spherical truncation*.<sup>110,111</sup> The interactions beyond that cutoff distance, for instance, at 12 Å, were ignored so that the maximum number of interactions becomes bounded, assuming a finite density. An abrupt cutoff distance introduces an energetic discontinuity into the system, and this can lead to unstable simulations, so smoothing functions are often applied instead. To further reduce the computational cost, group-based neighbor lists were introduced, but it is known that with such techniques the energy is not conserved.<sup>111</sup> The twin-range cutoff method overcomes some of these problems. The technique calculates the short-range electrostatic interactions at every time step, while the long-range interactions are only recalculated immediately after the nonbonded neighbor list is recalculated.

The Ewald summation method<sup>112</sup> offers a theoretically rigorous approach to the evaluation of electrostatic interactions in infinite periodic systems. While the original method is not well suited to efficient calculations within biomolecular MD simulations, more recent work<sup>113, 114</sup> has introduced versions with improved computational complexity ( $N \log N$  complexity), and these are widely applied. Particularly for systems with large periodic boxes and high dielectric solvents, the artifacts observed in simulations with the Ewald summation methods are insubstantial.<sup>115,116</sup> These methods were demonstrated as being relatively efficient.

For systems that are naturally two dimensional, special Ewald summation and particle mesh Ewald (PME) methods can be applied.<sup>117,118</sup> Such systems often include those in the simulation of transmembrane proteins that are typically simulated with periodicity in the plane of the membrane but with a finite length perpendicular.

The fast multipole (FM) method also offers an efficient way (computational complexity  $O(N)$ ) to handle long-range electrostatic interactions.<sup>114</sup>

An alternative to explicitly including all interactions in an infinite system, as is done by the Ewald and FM methods, but still considering those interactions unlike the spherical truncation methods, is to use a reaction field.<sup>67, 96</sup> This seeks to represent the surroundings by mimicking the response of the dielectric medium beyond the cutoff distance or boundary. While it is still an approximation, this gives stable and accurate results.<sup>119,109</sup>

For nonperiodic systems, provided that they are large, multipole expansion<sup>120</sup> and multigrid methods<sup>121,122</sup> are both efficient and useful.

### 3.5. Langevin Dynamics

Langevin dynamics incorporates stochastic terms to approximate the effects of degrees of freedom that are neglected in the simulation. It is based on use of the Langevin equation as an alternative to Newton's second law. This equation incorporates two additional terms. The first term is a frictional, or damping, function that is intended to represent the fictional drag experienced by solute molecules in a solvent that is not explicitly simulated. The second additional term is a random force that is applied to mimic the random impulses that would be expected from both the solvent and any coincident solute molecules. The Langevin equation for the motion of an atom,  $i$ , is

$$m_i \frac{d^2 r}{dt^2} = F_i(r) - \zeta_i \frac{dr}{dt} + R_i(t) \quad (14)$$

where  $F_i(r)$  is the usual term used in conventional MD,  $\zeta_i$  is the friction coefficient, and  $R_i(t)$  represents the random forces experienced by the atom. The temperature of the simulated system is maintained by a relationship between  $\zeta_i$  and  $R_i(t)$  (namely, the fluctuation-dissipation theorem). When  $\zeta_i = 0$ , Langevin dynamics is equivalent to conventional MD. When  $\zeta_i > 0$ , the random impulses felt by the system can assist in propagating barrier-crossing motions and, therefore, Langevin dynamics can offer improved conformational sampling characteristics over standard MD.

### 3.6. Brownian Dynamics

Brownian dynamics (BD) is a diffusional analogue of molecular dynamics<sup>26,123,124</sup> carried out through the numerical integration of the Langevin equation. When the solvent surrounding a molecule has high effective viscosity, the motion of that molecule can be described in terms of a random walk since the damping effect of the solvent will overcome any inertial effects. The Brownian dynamics method seeks to simulate the random walk to produce a representative diffusional trajectory. This is achieved by using a very large friction coefficient,  $\zeta_i$ , in the Langevin equation. In the case that a process of interest is diffusion controlled, Brownian dynamics is a useful and widely applied approach that is complementary to molecular dynamics. It is common, but not essential, for proteins to be treated as rigid bodies in BD simulations. As a result of the relative computational requirements of Brownian dynamics methods compared to molecular dynamics, time scales in the microsecond or millisecond range are readily accessible.<sup>26</sup>

Examples of biological processes which are amenable to study by Brownian dynamics include diffusion-controlled reactions, diffusional encounters, and ionic diffusion under the influence of an electrostatic field.

The choice of MD versus Langevin dynamics versus BD needs to be carefully considered depending upon which contributions are thought to dominate in the motion of interest.

### 3.7. Monte Carlo

Structural and thermodynamic properties of a system can be obtained through Monte Carlo (MC) simulations, thus making these a significant alternative to molecular dynamics simulations. Monte Carlo simulations are a stochastic approach to the task of generating a set of representative configurations under given thermodynamic conditions such as temperature and volume.

One attractive aspect of conventional MC simulations is that only the potential energy is normally used in stepping through configurations; no forces need evaluation, resulting in more efficient calculations. Some biased MC approaches do utilize force data, however.

In its simplest form, the Monte Carlo algorithm is a method for numerical integration. A set of parameters are randomly selected, or randomly perturbed, and a function of these parameters is evaluated. The results of many such steps are collated, and once sufficient sampling has occurred, the probabilities of any given result occurring can be readily assessed.

Metropolis et al.<sup>125</sup> introduced a technique known as Metropolis Monte Carlo simulation. In this scheme, the problem is described in terms of a thermodynamic system at potential energy,  $V$ , and temperature,  $T$ . With a constant  $T$ , the initial configuration is perturbed and the change in energy  $dV$  is computed. If the change in energy is negative, the new configuration is accepted. If the change in energy is positive, it is accepted with a probability given by a Boltzmann factor. This process is iterated until sufficient sampling statistics for the current temperature  $T$  are achieved. This procedure simplifies the calculation of the Boltzmann average for any observable property since it is now just the mean value of this property over all samples.

There are a number of issues that seriously hamper the use of Monte Carlo simulations with large biomolecules. Importantly, efficient moves are difficult to define for macromolecules. That is, it is difficult to devise simple structural perturbations that cause changes of a sufficiently large magnitude but also avoid generating energetically infeasible configurations. Some work has eased this issue for proteins, however.<sup>126,127</sup>

Conventional MC methods are inefficient for exploring the configurational space of large biomolecules when compared to molecular dynamics.<sup>128</sup> In addition, MC methods give no information about the time evolution of structural events. Hybrid MC/MD methods might resolve both of these issues and are described in the literature.<sup>129–131</sup> A conceptually related procedure, known as the relaxed complex method, is discussed later in this review.

### 3.8. Simulated Annealing

The *simulated annealing* algorithm<sup>132</sup> is related to the MC algorithm and forms an efficient technique to find the minimum energy configuration of a system. The usual Metropolis Monte Carlo algorithm is inefficient at sampling configurations that are beyond high potential-energy barriers; thus, it is only useful when starting at a configuration that is already near the global energy minimum well. Simulated annealing overcomes this problem by initially performing Monte Carlo steps at a very high temperature. According to a periodic schedule, this simulation temperature is decreased at a logarithmic rate (or, sometimes, a linear rate) until the temperature reaches zero. This procedure is widely used in protein modeling or refinement applications. As with conventional MC methods, no information about the time evolution of structural events can be obtained.

The diffusion equation method (DEM) potential smoothing method<sup>133,134</sup> (as discussed in Enhanced Sampling section of this review) is an analytical equivalent to simulated annealing.<sup>135</sup>

Similar temperature scaling procedures utilizing MD instead of MC for generating configurations are discussed in the Enhanced Sampling section of this review, below.

### 3.9. Nondynamic Methods

**3.9.1. Conformational Sampling**—Numerous nondynamic methods besides MC and SA exist for sampling available conformational space of proteins.<sup>136</sup>

CONCOORD<sup>137</sup> is a method that does not generate a time series of configurations. However, it does generate conformations that satisfy a set of distance constraints. This is different to the above algorithms since it does not rely on a potential surface. The structures are obtained through purely geometric considerations. This allows for some conformations that might never be found in an MD simulation or other energy-based methods and is therefore complementary to the dynamical simulation methods.

**3.9.2. Principal Component Analysis**—Principal component analysis (PCA) is a method that is often used for reducing the dimensionality of a dataset. For an arbitrary dataset where there is significant correlation between the dimensions, or individual variables, the first principal component is the linear combination of these variables which gives the best-fit line through the entire dataset. In other words, it is the linear combination which describes the greatest amount of variance in the data. The second and subsequent principal components are fit to the residual variation remaining after the more significant principal components are excluded. All principal components are orthogonal.

The separation of functionally important motions from the random thermal fluctuations of a protein is one of the challenges of trajectory analysis. PCA of the covariance matrix of the atomic coordinates is termed essential dynamics (ED).<sup>138</sup> This is a powerful method for extracting the significant, large-scale, correlated motions occurring in a simulation. In this sense, the principal components are the orthogonal basis set for the trajectory's atomic coordinates. The principal components corresponding to the greatest variance can be projected onto the protein structure, either individually or in sets. All other motions, including the smaller thermal fluctuations, will be filtered out. This facilitates visualization and appreciation of the major motions that may be biologically relevant.<sup>139</sup>

### 3.10. QM/MM

Hybrid quantum mechanical/molecular mechanical (QM/MM) methods<sup>140</sup> have reached a viable state and are rapidly gaining popularity.<sup>21,141</sup>

QM/MM methods are particularly useful since they allow the study of biomolecular reaction mechanisms. This is a task for which conventional MM is unsuitable owing to their assumption that bonds are never made or broken. Conventional QM methods are also unsuitable for this task owing to their computational expense, making calculations on the scale of entire solvated proteins currently intractable. QM/MM methods are beyond the scope of this review but are covered elsewhere in this issue.<sup>142</sup>

## 4. Free-Energy Calculations

The purpose of a MD simulation is often to derive kinetic and thermodynamic data about the model system. Indeed, many thermodynamic properties can be readily extracted from sufficient sample configurations of a system. As an example, the entropy of a system is directly related to the number of different configurations that are thermally accessible to it.<sup>17</sup>

One very important thermodynamic quantity is free energy, a measure of the stability of a system. In particular, free energy of binding is a measure of the stability of a complex, a measure that is probably fundamental to all studies of biomolecular binding processes.

Rigorous techniques, including the thermodynamic cycle-perturbation method,<sup>143</sup> exist for the estimation of free energies from simulations. However, such calculations are generally only practical for small or highly constrained systems. Often the calculations are extremely expensive or the level of sampling required for reliable statistics might be beyond the feasible



limits. Methods for speeding up free-energy calculations are valuable.<sup>144</sup> Recent methods for free energy of binding estimation are briefly discussed below.

#### 4.1. Free Energy of Binding

The interactions between proteins and other molecules are critical to many biological systems and processes. Signal transduction, metabolic regulation, enzyme cooperativity, physiological response, and other processes are all dependent upon noncovalent binding. These processes may be investigated through modeling and simulation, particularly as the range of solved protein structures grows. Through MD, MC, and the various related methods described in this review, binding modes and the corresponding binding free energies<sup>6</sup> may be estimated for protein-ligand and protein-protein complexes.

Approaches available for estimating either relative or absolute binding free energies cover a broad range of accuracies and computational requirements. Free-energy perturbation (FEP) and thermodynamic integration (TI) methods are computationally expensive, but they have been successfully applied in the prediction of the binding strengths for complexes.<sup>5,145</sup> Many more or less rigorous methods<sup>146</sup> have been developed to estimate such free energies more rapidly. These include the linear interaction energy (LIE) method,<sup>147</sup> the molecular mechanics/Poisson Boltzmann surface area (MM/PBSA) method,<sup>148,149</sup> the chemical Monte Carlo/molecular dynamics (CMC/MD) method,<sup>150,151</sup> the pictorial representation of free-energy components (PRO-FEC) method,<sup>150,152</sup> the one-window free-energy grid (OWFEG) method,<sup>153,154</sup> the  $\lambda$ -dynamics method,<sup>155,156</sup> and the 4D-PMF method,<sup>157</sup> among others.

**MM/PBSA**—The molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) model is a so-called endpoint free-energy method; only the initial and final states of the system are evaluated to estimate a free-energy change. This compares to the more accurate FEP and TI methods that require an equilibrium sampling of the entire transformation path, from an initial to a final state. End-point methods are computationally efficient, and consequently, they are widely discussed and applied in the literature. Despite their simplicity, a connection between statistical thermodynamics and various end-point free-energy models has been derived.<sup>148, 158–161</sup> Certain limitations of MM/PBSA must be considered, including the fact that there is no consideration of specific water interactions, it is sensitive to the trajectory, and it is sensitive to induced fit effects.

MM/PBSA<sup>149,162,163</sup> is basically a postprocessing method to evaluate the standard free energies of molecules or the binding free energies of molecular complexes in a relatively computationally efficient manner.

The MM/PBSA method partitions the free energy into molecular mechanical (MM) energies, continuum solvation energies, and solute entropy terms as follows

$$\langle G_{\text{mol}} \rangle = \langle E_{\text{MM}} \rangle + \langle G_{\text{PBSA}} \rangle - T \langle S_{\text{MM}} \rangle \quad (15)$$

where  $\langle G_{\text{mol}} \rangle$  is the average standard free energy<sup>148</sup> of the molecule of interest, which can be the ligand, the receptor, or their complex.  $\langle G_{\text{PBSA}} \rangle$  is the molecular solvation free energy. The solute's entropy term may be estimated using a number of methods.

The average molecular mechanical energy,  $\langle E_{\text{MM}} \rangle$ , is typically defined as

$$\langle E_{\text{MM}} \rangle = \langle E_{\text{bond}} \rangle + \langle E_{\text{angle}} \rangle + \langle E_{\text{torsion}} \rangle + \langle E_{\text{vdw}} \rangle + \langle E_{\text{elec}} \rangle \quad (16)$$

where  $E_{\text{bond}}$ ,  $E_{\text{angle}}$ ,  $E_{\text{torsion}}$ ,  $E_{\text{vdw}}$ , and  $E_{\text{elec}}$  are the bond, angle, torsion, van der Waals, and electrostatics terms of intramolecular energy, respectively.

The molecular solvation free energy can be further decomposed to

$$\langle G_{\text{PBSA}} \rangle = \langle G_{\text{PB}} \rangle + \gamma \langle A \rangle \quad (17)$$

where  $\langle G_{\text{PB}} \rangle$  is the average electrostatic contribution from molecular solvation,  $\gamma$  is the surface tension of the solvent, and  $A$  is the solvent-accessible surface area (SASA). The electrostatic solvation free energy,  $G_{\text{PB}}$ , can be calculated using

$$G_{\text{PB}} = \sum_{i=1}^N [q_i(\varphi_i^{\text{aq}} - \varphi_i^{\text{g}})] \quad (18)$$

where  $N$  is the number of atoms in the molecule,  $q_i$  is the electrostatic charge of atom  $i$ , and  $\varphi_i^{\text{aq}}$  and  $\varphi_i^{\text{g}}$  are the electrostatic potentials of atom  $i$  in the aqueous and gas phase, respectively, which are usually obtained by solving the Poisson-Boltzmann equation.<sup>164,165</sup>

## 4.2. Activated Molecular Dynamics

Many biological processes are intrinsically fast but, since these processes occur infrequently, appear to have long time scales. As an example, many reactions and conformational transitions exhibit long time scales because they consist of one or more activated processes. In fact, activated processes such as local conformational changes associated with ligand binding<sup>166</sup> are widespread in biology. An activated process is one in which a high-energy barrier exists between the initial and final states and this barrier must be overcome. The actual barrier crossing is often relatively rapid, but the time required for the system's random thermal fluctuations to provide the constituent atoms with suitable momentum can be long.

Conventional MD is unsuitable for investigating activated processes in biology because the tractable simulation time scales are of the order of nanoseconds while the biological process might take milliseconds or longer. However, a procedure known as *actiVated molecular dynamics* makes the study of such processes possible provided that the primary structural changes for the process are known beforehand.

Activated molecular dynamics is a two-stage process. First, a series of simulations is performed. Each of these simulations is constrained to a successive portion of the transition pathway. The purpose of these simulations is to locate the free-energy barrier peak. The second stage involves running conventional simulations from the region of the free-energy barrier. The resulting trajectories can be run in forward and reverse to generate a set of representative barrier crossing events. Analysis of these trajectories gives useful information regarding the mechanism of the activated process.<sup>167</sup>

## 4.3. Steered Molecular Dynamics

Steered molecular dynamics (SMD) simulations introduce a time-dependent or position-dependent force. The purpose of this force is to steer systems along particular degrees of freedom. This allows one to focus on dynamic events of interest while keeping computational expense to a minimum.<sup>168,169</sup> For example, the external force could drive a particular binding or unbinding event.

SMD offers scope for interactive steering in an immersive 3D environment. Implementations of such interactive environments include one<sup>170</sup> based on SIGMA<sup>171</sup> and VMD<sup>172</sup> and another based on NAMD and VMD that utilizes a haptic feedback device.<sup>173</sup>

In many respects, SMD simulation is the computational analogue of the experimental techniques which apply external mechanical manipulations to biomolecules. These experimental techniques include atomic force microscopy (AFM),<sup>174</sup> optical tweezers,<sup>175</sup> biomembrane force probes,<sup>176</sup> and dynamic force spectroscopy<sup>177</sup> experiments.

In the limit of weak forces that only slowly change in direction, any induced structural change will be minor and SMD will be equivalent to umbrella sampling. The results of SMD simulations are, however, often more interesting when this limit is violated significantly. Such conditions would be disastrous to many applications of standard molecular dynamics, including umbrella sampling itself and also methods relying upon free-energy perturbation theory and the weighted histogram analysis. Therefore, SMD might be useful in cases where major structural changes will be experienced and correspondingly major deviations from equilibrium would occur. Examples of such nonequilibrium cases include ligand unbinding and protein unfolding as initiated by stretching of termini. Equilibrium descriptions cannot be applied in the analysis of such simulations. Therefore, the extraction of valid potentials of mean force from SMD simulations is not straightforward, but several approaches have been proposed.<sup>178–181</sup> Free-energy differences can be obtained from the exponential averages of irreversible work,<sup>182,183</sup> and this leads to the most promising approach employed in extracting free-energy profiles from SMD simulations.<sup>184,185</sup>

Jarzynski's equation<sup>182,183</sup> relates equilibrium free-energy differences and work done through nonequilibrium processes. Consider a system described by a parameter  $\lambda$  and a process that causes this parameter to evolve from  $\lambda_0$  at time zero to  $\lambda_t$  at time  $t$ . According to the second law of thermodynamics, assuming the system is quasi-static, the average work done on that system cannot be smaller than the difference between free energies of the system corresponding to  $\lambda_0$  and  $\lambda_t$

$$\Delta F = F(\lambda_0) - F(\lambda_t) \leq \langle W \rangle \quad (19)$$

In other words, a nonequilibrium process provides only an upper limit to the free-energy difference. Jarzynski<sup>182</sup> proposed an equality that is independent of the speed of the process

$$e^{-\beta\Delta F} = \langle e^{-\beta W} \rangle \quad (20)$$

where the average is over a set of trajectories of the nonequilibrium process.

This equality has been validated both experimentally<sup>186</sup> and through computational simulation.<sup>183</sup> Thus, the equality provides a method for calculating free energies from nonequilibrium processes despite conventional thermodynamic integration being invalid because  $\Delta F$  does not equate to  $\langle W \rangle$ . The major difficulty that remains, however, is that the average of exponential term of Jarzynski's equality is dominated by trajectories corresponding to small values of work. These trajectories are infrequent in the simulations, leading to inadequate sampling. Currently, practical application is limited to slow processes where the fluctuation of work is comparable to the simulation temperature.<sup>184</sup> Coupling SMD to certain enhanced sampling techniques described later in this review might extend this limit of practicality.

A related technique is targeted MD in which a force that is dependent upon the difference between the current conformation and a target conformation. The aim is to drive the evolution of the simulation toward the given target conformation. Targeted MD has been applied in the prediction of pathways between particular protein conformations<sup>187</sup> and in protein folding.<sup>188</sup>

## 5. Recent Advances in the Theoretical Aspects of Molecular Dynamics

### 5.1. Force Fields

The accuracy of the potential-energy function is of crucial importance to the validity and stability of MD simulations of proteins<sup>189,190</sup> and indeed all macromolecules. As indicated above, the form of the energy function must be simple in order to make such computations tractable. It is also important that derivatives are readily accessible to facilitate efficient minimization and efficient integrators of motion.

Most force fields utilized in MD simulations of proteins share a significant number of similarities. Harmonic terms describe bond lengths and angles, Fourier series describe torsions, and pairwise atomic interactions are described using a Lennard-Jones function and a Coulombic function. Usually, although not always, parameters are first obtained for protein systems and subsequent parameters are derived for nucleic acid, lipids, and other biological molecules in such a way that they are consistent with the protein set. The main differences between the various force fields result from the diverse approaches taken to derive the individual parameters. It is not unusual for the parameters to contain significant interrelations and compensatory components such that the final results within a full simulation system reproduce desired experimental observables. Relatively innocuous seeming differences in the way that different software packages handle technical details in the simulation, such as the treatment of long-range electrostatic effects or the treatment of interactions between atoms bonded through a small number of intermediate atoms, can lead to substantially divergent energies with an alternative energy function. A drawback of this is that parameters for a given atom type cannot be compared between the force fields. The direct transfer of parameters from one force field to another is, generally, not valid.

The consistent-valence force field (CVFF)<sup>191</sup> differs from most of these force fields in the sense that it has a more complex functional form. Most of the others differ only through minor points such as how improper torsions (i.e., out of plane dihedral angles) are treated, what scaling factors are used for nonbonded interactions, or whether hydrogen bonds are included explicitly. The van der Waals parameters of all of the force fields listed above were developed through empirical fitting to small molecule model systems in liquid or solid phases. As a result, the densities of solvated protein systems tend to be close to reality. The torsional parameters tend to be fit to a mixture of QM and empirical data. The parametrization of template partial charges for the atoms in residues is more challenging. The resultant electrostatic interactions must be balanced with the particular water models. A typical approach is to determine the gas-phase partial charges through a QM calculation of a model compound and to scale these calculated charges by a somewhat justified multiplicative factor.

While it is not clear whether alanine tetrapeptide makes a reasonable model system for proteins, a study of 20 different protein force fields using *ab initio* quantum mechanical calculations indicated that there were discrepancies in all of these force fields.<sup>192</sup> Furthermore, it was suggested that in order to yield accurate electrostatics, the force fields would need to incorporate non-atom-centered partial charges.

The AMBER95 force field<sup>193</sup> for proteins is an example of one of the several widely used force fields that are developed alongside particular MD simulation software packages. The accuracy of partial charges assigned to various atoms in a protein structure is critical. Partial charges for the AMBER force field were determined using the restrained electrostatic potential (RESP) method.<sup>194,195</sup> This method fits a quantum mechanically calculated electrostatic potential at molecular surfaces using an atom-centered point charge model. Subsequent studies were conducted to assess how well the method performed in calculating conformational energies, and it performed better than the other tested force fields.<sup>196</sup> RESP charges have been

calculated for molecules with a description of lone-pair donor sites and atom-centered polarization.<sup>197</sup> In this study the partial charges were determined self-consistently from the charges and induced dipoles to reproduce a quantum mechanical electrostatic potential. In a separate study<sup>198</sup> an automated systematic search and genetic algorithm-based approach was applied for parameter optimization. For both of these methods the error in the conformational energy was lower than with the older AMBER95 force field.

The general AMBER force field (GAFF)<sup>199</sup> for organic molecules is designed to be compatible with existing AMBER force fields. It has parameters for most pharmaceutically relevant compounds and can be applied to a range of molecules with an automated methodology. This makes it suitable for applications such as rational drug design where protein-ligand simulations are utilized and manually assigning appropriate parameters to all ligands is not practical.

In the CHARMM22 force field<sup>200</sup> the atomic charges were derived from *ab initio* quantum chemical calculations of the interactions between model compounds and water molecules.

The majority of the bond length and bond angle parameters of the OPLS-AA force field<sup>201</sup> were extracted directly from the AMBER95 force field. The torsional and nonbonded parameters were instead derived using a combination of *ab initio* molecular orbital calculations and Monte Carlo simulations. A study using a similar set of parameters for amines concluded that there is no need to consider polarizability.<sup>202</sup> The OPLS-AA force field was further improved by reparametrizing the torsional coefficients.<sup>203</sup> The deviation in the energy compared to those from *ab initio* calculations of peptide was significantly reduced.

**5.1.1. Polarizable Force Fields**—One physical characteristic of molecular entities that is currently avoided in contemporary MM force fields is the effect of fluctuations in polarization.<sup>204,205</sup> There are two schools of thought on the matter. The first would suggest that since force fields are freely parametrized, these parameters could be fixed such that the effects of polarization are implicitly incorporated, at least in an approximate or average sense. Others would argue, however, that standard empirical force fields do not include any polarization terms and only by explicitly including such terms would accurate reproduction of experimental observables be enabled.<sup>192</sup> The hypothetical polarizable force fields were for many years touted as the “next-generation” of force field.<sup>192,206–208</sup> To date the most common methodologies for incorporating explicit polarizability include induced dipole models<sup>209</sup> and fluctuating charge models.<sup>210,211</sup> Unfortunately, the practical issues involved along with uncertainties in the best way to approximate the physics have resulted in a distinct lack of usable polarizable force field for proteins and other biomolecules. Nonetheless, the simpler case of homogeneous liquids such as water has yielded to the efforts of force field developers.<sup>212–214</sup> For example, in one study which found increased accuracy in a water model with polarization compared to simpler condensed phase models, charges and dipoles were calculated by fitting to *ab initio* potentials of isolated molecules and additional polarizability parameters were fitted to a range of potentials from applying electric fields to the molecules.<sup>215</sup> A recently presented alternative is based on the Drude oscillator.<sup>216–218</sup>

More recently, some attempts toward producing a usable polarizable force field for proteins have produced some more promising results. A fluctuating charge model for protein has been demonstrated in a nanosecond time scale simulation.<sup>219,220</sup>

A promising atomic multipole method<sup>221</sup> is distributed with the Tinker MD package<sup>222</sup> (<http://dasher.wustl.edu/tinker/>).

## 5.2. Constant pH Molecular Dynamics

The protonation state of acidic and basic residues of a protein along with the protonation state of any substrate will be influenced by their interactions with the environment, their internal interactions, and their mutual interactions.<sup>223</sup> While it is important to correctly assign protonation states, to run realistic simulations, it is especially vital to estimate the most favorable protonation states accurately in order to reliably estimate the free energy of the system. Numerous non-MD methods are available for estimating protonation states of proteins.<sup>224–227</sup>

The general problem of allowing for protonations and deprotonations of titratable residues during MD simulations could be important for accurate representation of proteins and has been examined.<sup>228–232</sup> A number of procedures for incorporating such events have been proposed.<sup>233</sup> In some of these methods, protons are added or removed in continuous, nonintegral fashion. In other methods, the Poisson-Boltzmann equation is used to gauge the correct protonation state but not for the propagation of the MD trajectory.

Recently a method using physically realistic, integral changes in protonation and consistent potentials for both titration and propagation was presented.<sup>234</sup> This approach uses a generalized Born model for the aqueous solvent. To ensure that surface groups on the solute protein exhibit proper canonical configurational sampling, Langevin dynamics is used to propagate the solute trajectories. The interdependence of titration states and solute conformation is recognized by use of periodic Monte Carlo sampling of the protonation states of the titratable residues. In the Monte Carlo step, a titratable site and protonation state are chosen at random, and the transition energy is calculated using

$$\Delta G = k_B T (\text{pH} - \text{p}K_{a,\text{ref}}) \ln 10 + \Delta G_{\text{elec}} - \Delta G_{\text{elec,ref}} \quad (21)$$

where  $k_B$  is the Boltzmann constant,  $T$  is temperature,  $\text{pH}$  is the specified solvent  $\text{pH}$ ,  $\text{p}K_{a,\text{ref}}$  is the  $\text{p}K_a$  of the appropriate reference compound,  $\Delta G_{\text{elec}}$  is the electrostatic component of the transition energy calculated for the titratable group in the protein, and  $\Delta G_{\text{elec,ref}}$  is the electrostatic component of the transition energy for the reference compound, a solvated dipeptide amino acid. The electrostatic portion of the transition energy is determined by taking the difference between the potential calculated with the charges for the current protonation state and the potential calculated with the charges for the proposed state. There is no need for solvent equilibration because an implicit model is incorporated, so this is done in a single step. The equation can then be used to calculate the total transition energy, as all other terms are known. The total transition energy,  $\Delta G_{\text{elec}}$ , is used as the basis for applying the Metropolis acceptance criterion to determine whether this transition will be accepted. If the transition is accepted, MD is continued with the titratable group in the new protonation state. Otherwise, MD continues without change to the protonation state. Applications of the method to hen egg white lysozyme yielded agreement with the experimental  $\text{p}K_a$  to within plus or minus one unit for most titratable sites. This is comparable to methods designed specifically for  $\text{p}K_a$  prediction.

## 5.3. Advanced Sampling Techniques

Sometimes one knows two biologically relevant but distinct conformational states but knows little about the necessary dynamic events of paths that convert one state into the other. Sometimes one knows a single conformational state, often a crystallographically determined structure but also knows that there must be a conformational change to provide a particular activity of interest. In many cases, the time scales involved in the conformation changes are not accessible via conventional MD techniques. To address this issue, numerous accelerated MD variants are proposed in the literature,<sup>235,236</sup> and the more recent advances are covered here. These accelerated MD methods extend the conformational sampling characteristics,

enabling extended time scales, or effective time scales, to be accessed. These methods aim to allow rare dynamic events to be observed more readily.

Accelerated MD methods can be grouped into three broad classes. The first class alters sampling of conformational space through explicit modification of the potential surface. The second class also alters the sampling but by using non-Boltzmann sampling to increase the probability of high-energy states. The third class includes those methods that enhance the sampling of certain degrees of freedom at the expense of other, typically faster, degrees of freedom. There is some overlap between these classes.

All of these accelerated MD approaches are distinct from fundamental improvements in the core MD algorithm, such as improved integrators and parallelism. These latter types of enhancements are discussed elsewhere in this review.

The most interesting or promising enhanced sampling methods are described in the following sections, but the reader should note that this area is actively researched, and further approaches are keenly anticipated. A separate group of methods are those that apply some non-MD algorithm. Although these are not actually MD methods, they are relevant to the issue at hand and are therefore also discussed here.

There are two primary goals driving development of enhanced sampling methods. Some of these approaches aim to increase the volume of conformational space that is explored during the simulation, while others aim to drive the system to a particular conformation or to the global minimum energy conformation more rapidly.

**5.3.1. Modified Potentials**—The basic principle behind various potential-energy modification methods is to reduce the amount of time that the simulated system remains in a local energy minimum well, speeding the transitions from the region of one local minimum to another, forcing the system to sample the remainder of the available conformational space. In each of these approaches, the potential-energy function is altered to enhance sampling by reducing the propensity of energy wells to act as conformational traps. Methods that modify the potential-energy surface include the deflation method,<sup>237</sup> conformational flooding,<sup>238, 239</sup> umbrella sampling,<sup>240</sup> local elevation,<sup>241</sup> potential smoothing,<sup>135</sup> puddle-skimming<sup>242,243</sup> and puddle-jumping method,<sup>244</sup> hyperdynamics,<sup>245,246</sup> and accelerated MD.<sup>247</sup>

Another method is the well-studied diffusion equation method (DEM),<sup>133–135,248–251</sup> a type of potential smoothing. In DEM, the diffusion equation is solved analytically for the potential-energy surface, thus deforming and smoothing it. A time-reversal process is used to restore the potential to its original form as the simulation proceeds.

The smoothing or flattening of the potential-energy landscape can be applied either globally to increase the overall sampling or only along a specific, predefined, reaction coordinate to enhance the conformational evolution along the direction of a desirable transformation or enhance sampling over a particular set of conformations, depending upon the method. For example, given some prior information about the desired conformations, one widely used approach is umbrella sampling.<sup>240</sup> A compensating function, known as an umbrella potential, is added to the potential-energy function to bias the sampling. Obviously, construction of the umbrella function requires prior knowledge of the conformations of interest. Nonetheless, umbrella sampling is a powerful alternative to adiabatic mapping. In umbrella sampling, local strain may be relaxed more effectively and kinetic effects can be included to some extent. Importantly, some deviations from any specified conformational path can be tolerated.

In principle, these schemes are all applicable to methods other than MD. They can be used for enhancing sampling by MC simulations, for example.

The local elevation method<sup>241</sup> enhances sampling by adding a penalty potential to any conformations previously sampled. This resembles the widely used tabu search algorithm in that a list of previous solutions is maintained and new solutions are driven away from solutions existing in this list. This approach has not been adapted for use with proteins. It is likely that the storage overhead for configurations of such large molecules would be overwhelming.

Another approach that aims to force systems out of their sampled local minima is conformational flooding.<sup>238,239</sup> In conformational flooding the initial state of the system is destabilized by adding an extra unfavorable potential at this initial state. This method predicts the so-called essential degrees of freedom using PCA, described earlier. A Gaussian potential is added to the system to force it along these essential degrees of freedom. Unlike local elevation, this method has been demonstrated in protein simulations.

**Hyperdynamics and Accelerated MD:** The recently described *accelerated molecular dynamics*<sup>247</sup> allows for more rapid sampling of the configurational space in systems with rough energy landscapes and also allows one to calculate the correct thermodynamic properties of the system. It resembles the puddle-skimming method<sup>243</sup> except that its formulation avoids nonsmooth potential-energy surfaces that cause significant problems in MD simulations. In addition, the puddle-skimming method results in random sampling on a flat energy landscape in the regions of the minima where sampling becomes very inefficient.

In the accelerated MD method, time becomes a statistical quantity in the simulation. The effective time scale of simulations is increased by several orders of magnitude at the expense of sampling around the energy minima. Because the potential is altered analytically, the statistics of sampled configurations can be corrected to reproduce the canonical probability distribution for the original potential surface.

The original potential function  $V(r)$  is altered via

$$V'(r) = \begin{cases} V(r) & V(r) \geq E \\ V(r) + \Delta V(r) & V(r) < E \end{cases} \quad (22)$$

where  $\Delta V(r)$  has a nonnegative value as given by

$$\Delta V(r) = \frac{(E - V(r))^2}{\alpha + (E - V(r))} \quad (23)$$

The modified potential echoes the original potential, so sampling directions during the simulation are still representative of the unbiased system. Thus, the random-sampling behavior of puddle skimming in the regions of the original minima is not observed.

When applied to the dihedral angle term and the 1–4 term of the Amber potential, in a Langevin dynamics simulation using the GB implicit solvent model, substantial increases in configuration sampling could be obtained. Also, after applying a correction for the sampling bias, the method yielded the expected distribution of configurations for small peptides.<sup>247</sup>

**5.3.2. Modified Sampling**—The basic principle behind most of the second class of methods is also to reduce the amount of time that the simulated system remains in local energy minimum wells, forcing the system to sample the remainder of the conformational space available. However, these techniques use alternative methods for sampling rather than performing



conventional MD on explicitly modified energy landscapes. One obvious approach is simulated annealing (SA), as described earlier, which effectively smoothes a potential-energy surface through additional entropic contributions.<sup>132</sup> It can be shown that SA is the stochastic equivalent of the DEM method discussed above, and the relationships between these methods have been investigated.<sup>135</sup>

Modified sampling methods include high-temperature MD,<sup>252</sup> locally enhanced sampling (LES),<sup>253–256</sup> replica exchange,<sup>257,258</sup> parallel tempering,<sup>259,260</sup> self-guided MD,<sup>261</sup> targeted MD,<sup>187</sup> milestoneing,<sup>262,263</sup> repeated annealing,<sup>264</sup> additional degrees of freedom,<sup>265,266</sup> and various non-Boltzmann sampling methods.<sup>267,268</sup>

**Locally Enhanced Sampling:** In the LES method<sup>253,256</sup> a fragment of the system, maybe a side chain or a ligand molecule, is duplicated so that the simulation contains  $N$  noninteracting copies of that fragment. The remainder of the system experiences each fragment through interactions that are reduced by a factor of  $1/N$  from their original magnitudes. This use of multiple copies and the reduction of the interaction potentials significantly enhances the sampling of the conformations of the fragments.

Evaluating the free energy from LES simulations requires two additional perturbation calculations: one in transforming from the single-copy representation of the reference state to the multiple-copy representation and a second in transforming from the multiple-copy representation of the perturbed state into the single-copy representation. However, the benefits of adopting a multiple-copy representation outweigh the additional costs of introducing two more perturbation calculations.<sup>269</sup> This method has been applied to the study of  $\alpha \rightarrow \beta$  anomerization of glucose, and it was found that the free-energy calculations converged an order of magnitude quicker than with the single-copy method.<sup>270</sup>

**High-Temperature Molecular Dynamics:** Perhaps the most obvious approach for enhancing sampling of high-energy states is to raise the simulation temperature. This approach is known as high-temperature dynamics and has been evaluated by several groups.<sup>252,271,272</sup> It is suggested that high-temperature molecular dynamics is a useful aid in conformational searches, but physiologically relevant low-energy structures are not generally obtained even after minimization of the generated high-energy structures. Indeed, the generated structures often have an infeasible proportion of cis-peptides.<sup>252</sup> Consequently, this method is not without criticism. Another possibly major issue is that MM force fields, generally, have not been designed for, or validated with, temperatures much beyond the physiologically relevant 300–330 K that most MD simulations are run at. Whether or not the force field is physically correct, there may be debate about whether the use of such high temperatures leads to appropriate sampling characteristics since the entropic contribution to the free energy is significantly enhanced (and therefore over-sampled).<sup>238</sup> Although there are several problematic issues involved with using high temperatures to accelerate MD simulations, the basic principle acts as the fundamental basis for a few more complex approaches discussed below.

**Multiple-Copy Dynamics:** Several approaches that use a series of simultaneous (or parallel) MD simulations are demonstrated in the literature.<sup>255,256,273–276</sup> In these individual molecules in the separate simulations may, or may not, interact in some way. Similar approaches may be applied in simulated annealing.<sup>277</sup>

One interesting multiple-copy approach is SWARM-MD.<sup>256</sup> The basic idea behind SWARM-MD was motivated by the efficient search behavior observed of swarms of social insects. Despite the absence of any higher intelligence, whole swarms of insects often appear to exhibit significant organization and planning. The cooperative rules that lead to the swarm's overall behavior is mimicked for conformational search in a simulated swarm of molecules. In such

simulations, each molecule is subject to supplementary artificial forces to introduce the cooperative behavior. The artificial force drives the trajectory of each molecule toward the mean trajectory of the entire swarm.

It would be easy to conceive a scheme in which each molecule is driven away from the mean trajectory in order to enhance the volume of conformational space that would be sampled.

Multiple-copy approaches tend to be very well suited to trivial parallelization with communication only between the individual systems. Indeed, the well-known distributed protein-folding project, Folding@Home, uses a multiple-copy approach.<sup>278</sup>

A number of multiple-copy approaches specific to ligand binding or design have been introduced,<sup>275,279,280</sup> and these are discussed further in the MD for Ligand Docking section, below.

**Replica Exchange Molecular Dynamics:** Like the multiple-copy approaches, replica exchange molecular dynamics (REMD)<sup>281</sup> and the closely related parallel tempering method<sup>260</sup> utilize a series of simultaneous and noninteracting simulations, known as replicas. With proteins, these simulations are typically MD, but earlier work applied Monte Carlo simulations. The replicas are simulated over a range of temperatures, and at particular intervals the temperatures of these simulations may be swapped (i.e., replicas are exchanged). The methods differ, however, in the way the individual simulations are coupled. Usually these replicas may exchange temperatures according to Monte Carlo-like transition probability. Such exchanges occur through a simple swapping of the simulation temperatures via velocity reassignment. The high-temperature replicas jump from basin to basin, but the low-temperature replicas explore a single valley with sampling characteristics just like conventional MD.

While REMD is widely applied to smaller molecules, particularly in peptide and protein folding experiments,<sup>282–284</sup> it is found to be extremely computationally expensive when applied to large proteins.

In a study in which REMD was applied to a 20-residue peptide it was found that at physiologically relevant temperatures the conformational space was sampled much more efficiently than it was with conventional constant temperature MD<sup>284</sup> and with similar thermodynamic properties.

**Self-Guided Molecular Dynamics:** Self-guided molecular dynamics (SGMD)<sup>261,285,286</sup> applies an additional guiding force to drive the simulation. The guiding force is a continuously updated time average of the force of the current simulation, leading to increased search efficiency by assisting the system over energy barriers. For efficient sampling of the available conformational space, the correlation between the guiding forces and the actual physical forces must be low;<sup>287</sup> nonetheless, the algorithm produces stable dynamics.<sup>286</sup>

A related method, the momentum-enhanced hybrid Monte Carlo method (MEHMC)<sup>288</sup> overcomes some of the inherent problems observed with SGMD. The SGMD algorithm lacks reversibility because the effective potential-energy landscape is a function of the trajectory rather than a function of the coordinates. This irreversibility results in substantial errors in canonical averages from the trajectory. MEHMC differs by using average momentum instead of average force to bias the initial momentum within a hybrid MD/Monte Carlo procedure. This is believed to yield correct canonical averages.<sup>288</sup>

**5.3.3. Modified Dynamics**—This third class of enhanced sampling method encompasses those methods in which the dynamics along the “slow” degrees of freedom are accentuated

relative to the “fast” degrees of freedom. One such method is extremely widely used, to the extent that simulations not applying it (or one of its close descendants) are exceptionally rare. This is the SHAKE algorithm,<sup>71</sup> where constraints are applied to particular bond lengths to allow larger time steps to be taken without encountering excessive forces. SHAKE was described earlier in this review. Another method is the MBO-(N)D algorithm<sup>289</sup> in which groups of atoms are partitioned into substructures that are considered to be rigid during the simulation. It is possible to integrate the equations of motion faster by separating the faster components of the dynamic propagator from the slower components.<sup>290</sup> Alternatively, the faster (i.e., higher-frequency) motions may be reduced or eliminated completely. Several algorithms based on this principle exist including dynamic integration within a subspace of low-frequency eigenvectors,<sup>291</sup> generalized moment expansion,<sup>292</sup> various types of coarse-grained modeling<sup>293,294</sup> such as network models,<sup>295</sup> mode coupling theory and projection of a generalized Langevin equation onto certain degrees of freedom,<sup>296</sup> digital filtering of selected velocities,<sup>297–299</sup> large time-step dynamics using stochastic action,<sup>300,301</sup> and leap dynamics.<sup>302</sup>

**Gaussian Network Model:** A large number of coarse-grained models for proteins have been described in the literature. The Gaussian network model<sup>293,303–305</sup> is one such coarse-grained method in which the energy function evaluates the system at the level of residues rather than atoms or in some cases even more coarse fragments.<sup>306</sup> The residues interact through simple harmonic and nonharmonic terms. This model has been widely applied in MD of proteins.<sup>307,308</sup>

This simplified model has been reduced further, even to the point where sequence information is not considered.<sup>309</sup> In addition, sampling has been further enhanced through amplification of slower motions<sup>310</sup> and coupled to conventional MD to assist sampling.<sup>311</sup>

**Leap Dynamics:** In the leap-dynamics<sup>302</sup> method a combination of MD and essential dynamics is applied. Conformational “leaps” applied to the system force it over energy barriers. The structures from the leaping process are refined using MD. The method was demonstrated to correctly predict the enhanced partial flexibility of a mutant structure in comparison to native bovine pancreatic trypsin inhibitor.

A potentially significant disadvantage of many of these modified dynamics methods is that the dynamics are, of course, artificial. This may, or may not, be a problem depending upon the desired application of the results. Some methods might still be anticipated to yield a Boltzmann distribution of structures if run for a sufficiently large number of steps, but there can be no such guarantees for others.

**Digitally Filtered Molecular Dynamics:** Digitally filtered molecular dynamics (DFMD) applies the theory of digital filters to MD simulations, selectively enhancing or suppressing motions on the basis of frequency.<sup>312</sup> This method was applied to the Syrian hamster prion protein, and a high degree of selectivity and control was demonstrated in the enhancement of the rate of conformational changes.<sup>298</sup> A time reversible version of the method has also been developed.<sup>299</sup>

**Multiple Time-Step Methods:** A number of multiple time-step methods are available<sup>313</sup> with the reference system propagation algorithm (RESPA)<sup>314–316</sup> being most widely applied to biomolecular systems.<sup>317,318</sup> A reasonable MD trajectory may be generated at 35–50% of the usual computational expense. Coupled with the Langevin methods described above, speed ups of as much as 2 orders of magnitude have been reported for protein simulations.<sup>319</sup>

The basic concept in multiple time-step methods is to separate the slow motions from the fast motions and evaluate the interactions relating to the slow motions less frequently. One common assumption in such methods is that interactions involving longer distances change more slowly, and therefore, the forces due to those interactions can be reapplied over several time steps before being recalculated.

**Multiple-Body O(N) Dynamics:** Multiple-body O(N) dynamics (MBO(N)D)<sup>289</sup> is a molecular dynamics technique using a coarse-grained model that scales linearly with the number of bodies in the system. Speed ups of 5–30 times compared to conventional MD are claimed.<sup>320</sup> This method combines rigid-body dynamics with multiple time steps. The highest frequency harmonic motions are removed while retaining the low-frequency anharmonic motions. It has been demonstrated to reproduce the global essential dynamic properties of both proteins and nucleic acid systems.<sup>289</sup> One notable problem with this otherwise extremely promising method is that it relies upon the user to determine the level of granularity from an empirical study of such levels for the system of interest.

## 6. Recent Advances in the Computational Aspects of Molecular Dynamics

The practical application of molecular dynamics is fuelled, in part, by the wide availability of software and the growing availability of significant computational resources.

### 6.1. Software

There is a considerable amount of diversity in the software packages that may be chosen today. While many molecular dynamics packages aim to have a broad or comprehensive range of capabilities, each widely used package does have certain features or advantages that set it apart from the others. Few research groups restrict their simulations to a single software package. This encourages development of compatibility-oriented features and uniform benchmarks, although there is still much scope for improvement in these areas. The majority of popular MD packages can utilize force field, structure, and trajectory file formats that were originally introduced in other packages. This enables a certain amount of validation and facilitates reproduction of published results, even without the original software, an exigency for some areas of computational chemistry.

Many of the general-purpose protein modeling packages contain some kind of MD facility, although in many cases this is nothing more than an interface to one of the specialized MD software packages. Such interfaces are useful in themselves as they can provide a simple mechanism for invoking the simulations without an understanding of the underlying, and often complex, software. They, therefore, allow the nonspecialist to readily perform simple simulations.

Taking the authors' research group as an example, GROMACS, AMBER, and NAMD are all routinely used on a regular basis for MD studies as warranted by the specific aims or requirements of the current project. In addition, CHARMM, NWChem,<sup>321,322</sup> and others are used when the situation demands.

While it might be argued by some that such a broad base of actively developed software fractures the field and wastes a lot of effort through duplication, it in fact fosters a great deal of friendly competition. Such competition is a sign of a healthy worldwide research effort. Other reasons, such as the independent validation or verification of theories, should not be understated either. As MD approaches and methods pass from the province of specialized experts into the wider realm of scientists, particular software packages can offer important advantages or accommodate particular needs of the diverse set of potential users.

**6.1.1. GROMACS**—GROMACS<sup>323,324</sup> (<http://www.gromacs.org/>) is advertised as a versatile package which is primarily designed to perform molecular dynamics simulations of biochemical molecules such as proteins and lipids. However, it is also claimed that since GROMACS is extremely fast at calculating the nonbonded interactions that typically dominate simulations, many groups are also using it for research on nonbiological systems such as polymers. GROMACS was initially a rewrite of the GROMOS<sup>325</sup> package (<http://www.igc.ethz.ch/gromos/>), which itself, like AMBER, was originally derived from an early version of CHARMM.

The main advantages of GROMACS are its ease of use and its exceptional performance on standard personal computers. A lot of effort was expended in optimizing the code to run efficiently on desktop computers such as those using the Intel Pentium IV and PowerPC G4 processors. The authors report that it is normally between 3 and 10 times faster than other MD programs.

One factor that many users might find attractive is its lack of a scripting engine. GROMACS is actually a suite of small command-line programs each with a simple set of options.

GROMACS file formats are somewhat interesting. All files are plain-text based, so they are, in principle, human readable. This helps them avoid the platform dependence malediction of many MD packages. These plain-text formats result in much larger file sizes than binary formats would, so GROMACS transparently utilizes standard UNIX compression tools. In addition, trajectories may be stored in a very condensed form using lossy compression. Lossy compression describes a class of data compression algorithms that achieve impressive reductions in data size by only approximating the original data. While some fidelity is lost, it generally has little practical consequence much like the lossy compression used in the well-known JPEG image format.

Most of the standard types of data analysis can be performed using the set of accompanying tools which can also produce publication-ready plots in a straightforward manner.

From a practical point-of-view, one particularly attractive reason to choose GROMACS is the fact that it is distributed as free software under the terms of the GNU General Public license (<http://www.gnu.org/>). This provides certain freedoms and significant pedagogical benefits. Source code is generally available (at least for academic groups) for all major MD packages, but the use and reuse of such code is often highly restricted.

**6.1.2. NAMD**—Whereas GROMACS is renowned for its spectacular performance on modest desktop computers, NAMD<sup>326</sup> (<http://www.ks.uiuc.edu/Research/namd/>) exhibits inspiring performance on high-end parallel computing platforms with large numbers of processors.<sup>327</sup> NAMD is able to comfortably handle system sizes which are well beyond the practical, if not absolute, limits of other MD packages. For instance, NAMD was used in the simulation of a 200 000 atom lac repressor system.<sup>328</sup> Such huge biological systems are simulated on massively parallel supercomputers and Beowulf-style clusters. NAMD scales to (i.e., runs efficiently on) large numbers, even thousands, of processors impressively.<sup>329</sup> Until the recent development of PMEMD<sup>330</sup> no other MD software came close on this point.

Besides its efficiency, another advantageous property is the level of integration with the VMD molecular visualization software.<sup>172</sup> This facilitates interactive molecular dynamics, for example.<sup>331</sup>

One specific factor that probably contributes to its efficiency is that its feature set is relatively modest compared to the other major MD packages. However, NAMD is file compatible with

packages such as AMBER and CHARMM, so system preparation and the postprocessing and analysis of results are often performed using other software.

NAMD is distributed free of charge, with full source code, but certain conditions on use apply. Its object-oriented design implemented using the C++ programming language<sup>332</sup> facilitates pedagogical use and the incorporation of new algorithms.

**6.1.3. AMBER**—Two of the more established and respected MD packages are CHARMM and AMBER<sup>333</sup> (<http://amber.scripps.edu/>). While the respective developers can trace the programs heritage to a common source, they have been independently developed since the early 1980s and adopted slightly different philosophies. AMBER consists of a suite of separate programs, each performing a specific task. The CHARMM developers have taken a more integrated approach where one single and monolithic application does everything from system preparation through simulation to analysis. There are inherent benefits to each approach, but the overall merit, more or less, comes down to personal preference.

AMBER was principally maintained in the research group of the late Professor Peter Kollman, and ongoing maintenance is now coordinated in the research group of Professor David Case. Code contributions came from a variety of locations, however. Along with CHARMM, it often incorporates new methodologies and algorithms before any other packages.

A recent addition to the AMBER suite<sup>334</sup> is PMEMD, a stripped-down and optimized version of the general MD program known as Sander. PMEMD provides scaling on massively parallel platforms that is comparable with NAMD.

The simulations executed using PMEMD are intended to replicate AMBER's Sander calculations within the limits of computational precision. However, the computation is performed much more quickly, in roughly one-half of the memory, and with significantly less overhead on larger numbers of processors. A number of benchmark cases are presented on the PMEMD website (<http://amber.scripps.edu/pmemd-get.html>).

Like CHARMM, a series of force fields are developed in conjunction with the AMBER simulation software.<sup>189,335</sup>

**6.1.4. CHARMM**—CHARMM (Chemistry at HARvard Molecular Mechanics) is described as a program for macromolecular simulations, including energy minimization, molecular dynamics, and Monte Carlo simulations.<sup>39–41</sup> It is predominantly developed within the research group of Professor Karplus at Harvard University, although, as highlighted on the CHARMM website (<http://yuri.harvard.edu/>), a network of developers in the United States and elsewhere actively contribute to its ongoing development. The CHARMM software is developed in unison with a series of force fields, as described above.

A variant of CHARMM, named CHARMMm, is widely deployed in commercial settings. While it lacks some of the more cutting-edge features, it is arguably more robust and bug free. Commercial support is provided by Accelrys (<http://www.accelrys.com/support/life/charmm/>).

## 6.2. Hardware

The great strides made in the molecular dynamics field are due in part to the phenomenal developments in computer hardware. Protein simulations make heavy demands on the available computing facilities. The length and accuracy of simulations is chiefly restricted by the availability of processor power, while memory and disk space are also important, especially

for large model systems. The most significant developments in computational infrastructure with applicability to MD are briefly discussed here with details of this applicability.

**6.2.1. Parallel Computing**—The most powerful supercomputers available today consist of arrays of processors that communicate via fast interconnects. The potential power of these systems is only utilized when the algorithm to be run can be partitioned into a number of separate processes that can individually be run on a single processor. With no or minimal communication required between the processes, the bandwidth of the interconnects will not become the bottleneck in the calculation. In addition, if the processes do not need to remain synchronized, then processing power will not be wasted as one of the processes waits for others to catch up. While this might be the case for many computational tasks, it is not for typical MD algorithms. Particularly in the case of nonbonded interactions, significant communication is required and the processes will need to be synchronized during each time step.

**Blue Gene:** Blue Gene<sup>336</sup> is a widely publicized project, with an ultimate goal of implementing and utilizing a so-called hugely parallel supercomputer architecture.

*IBM's Blue Gene Project.* (<http://researchweb.watson.ibm.com/bluegene/index.html>) is described as representing a unique opportunity to explore novel research into a number of areas, notably including biomolecular simulation. The planned scientific program will require and foster a collaborative effort across many disciplines and the participation of the worldwide scientific community to make best use of this exciting computational resource.

**Commodity Clusters:** The availability of Beowulf-style clusters built from commodity PC components is being increasingly leveraged, as a cost-effective alternative to traditional supercomputer platforms, to facilitate large-scale MD simulations. There is one caveat, however. The technical characteristics of typical clusters do not suit them to fully efficient MD simulations as detailed in the Future Prospects and Challenges section. Clusters are suitable for running many simulations in parallel, with minimal communication between the processing nodes and therefore maintaining much of the efficiency of the serial codes. Some types of simulation, including the relaxed complex method, are highly suited to clusters.

## 7. Recent Applications of Molecular Dynamics

There have been numerous and varied simulations described in the literature. As MD approaches the point in its development at which it becomes a routine tool for nonspecialist researchers, such simulations will only increase in frequency. It is impossible to give a broad overview that will do justice to any of these simulations. Consequently, a tiny set of recent examples driven by simulations of biomolecules is selected here for brief discussion.

### 7.1. Functional Mechanism of GroEL

In the highly concentrated milieu of the cell, chaperone molecules are essential to facilitate the correct folding of many proteins. For example, in *Escherichia coli* it is thought that around 10% of proteins located in the cytoplasm require an experimentally well-characterized protein, chaperonin GroEL, for correct folding. This protein is a homomeric complex of 14 subunits arranged in two heptagonal rings.<sup>337</sup> Critical to its function are large conformational changes that are regulated through cooperative binding and hydrolysis of ATP in the presence of a chaperonin GroES. This cooperativity is positive in a given ring but negative between the rings. The conformational changes occur in all subunits, converting them from a “closed” form to an “open” form upon binding of ATP and GroES.

It was impossible to determine the conformational pathway between the open and closed forms experimentally. It is believed that this conformational transition occurs on a millisecond time

scale, so conventional MD of this transition would not be tractable. However, targeted MD is applicable in such cases and was applied to determine the transition pathway between the two known conformations.<sup>338</sup> The targeted MD simulation predicts a particular intermediate conformation with ATP bound before GroES. It was also indicated that steric interactions along with salt bridges between the individual subunits mediate the pattern of positive and negative cooperativity of the ATP binding and hydrolysis. It is seen that early in the pathway ATP binding triggers a downward motion of a small intermediate domain, and this causes the larger motion of the apical and equatorial domains. Subsequent cryoelectron microscopy results support the results of the simulation, indicating that this intermediate domain plays a critical role in the conformational behavior.<sup>339</sup>

## 7.2. Simulation of Membrane Proteins

An active area where simulations play a key role is the study of ion diffusion through pores and channels and the gating mechanisms associated with such channels, topics that are frequently reviewed.<sup>25,340–342</sup> Experimentally probing the structure of transmembrane proteins is difficult, but valuable insights have been obtained through the application of MD in cases where a reasonable structure is known or can be predicted. One system that has been widely simulated is the M2 protein of influenza A. During a 4 ns MD simulation a funnel-like structure formed, but it appeared to be occluded by a particular histidine residue,<sup>343</sup> while it has also been shown that the protonation of this residue can drive channel opening.<sup>344</sup> Both of these predictions have been validated by NMR results. This transmembrane protein fragment has been used as the basis of a number of model ion channels for various viral proteins.<sup>340</sup>

## 7.3. Molecular Dynamics for Docking and Ligand Design

The interactions between proteins and substrates are critical to many biological systems and processes. Signal transduction, metabolic regulation, enzyme cooperativity, physiological response, and other processes are all dependent upon noncovalent binding. These processes may be investigated through modeling and simulation, particularly as the range of solved protein structures grows. Through MD, MC, and the various related methods described in this review, binding modes and the corresponding binding free energies may be estimated for protein-ligand<sup>29</sup> and protein-protein<sup>345</sup> complexes.

Ligand docking is the prediction of protein-ligand complexes; the use of MD is widespread in such ligand-docking studies.<sup>346,347</sup> Most ligand-docking methods are MM-based; however, the present discussion is limited to methods that actually use MD rather than just using MM-based scoring functions.

When calculating free energy of binding estimates there is a necessary balance to be found between the accuracy or reliability of these estimates and the computational cost of the calculations. It is not always essential to determine highly accurate binding constants for productive studies in drug design.

At the upper end of the accuracy versus computational speed, one factor that becomes crucial is that the ligand-binding process can lead to conformational changes in the receptor protein itself. These changes could be necessary for the receptor to accommodate the bound ligand. While it is important to explore the conformational space available to the receptor (i.e., the protein) molecule, it is often difficult to predict or represent the plasticity of the binding site.<sup>348</sup> This is particularly important when there may be multiple, allosterically connected, binding sites. While many approaches are available for considering such flexibility,<sup>349–351</sup> it is suggested that the use of multiple protein structures in the docking process is a wise approach to the task. One simple approach to generate such multiple structures is through the use of MD<sup>352,353</sup> with standard “static” docking to a series of individual snapshots from the



simulation trajectory. It may be tempting to use MD as the docking search method, but during reasonable simulation times the system is likely to be stuck in local minima, with energy barriers of more than 1–2 kT unlikely to be overcome. With some modification of the potential function to smooth the energy surface and allow further exploration of receptor conformations such simulations can become practical.<sup>354,355</sup> By coupling different degrees of freedom of the system to different temperatures, the system can be assisted in escaping local minima by varying temperatures that specifically mediate the flexibility of the ligand.<sup>356,357</sup> The concept of varying temperatures to enhance sampling for docking simulations is commonly applied in the guise of simulated annealing MD.<sup>358</sup>

**7.3.1. Advanced Molecular Dynamics-Based Methods for Drug Discovery**—It has been shown that in a study of different search algorithms, an MD-based method was the most efficient approach for large search spaces and produced the lowest (i.e., best) mean energies for the docked conformations.<sup>359,360</sup>

Further docking methods based around MD have been proposed. The dynamic pharmacophore method<sup>361</sup> and the relaxed complex method<sup>362,363</sup> are both designed to take receptor flexibility into account in the analysis of ligand-receptor binding. These are described below.

SGMD, as described earlier, would probably be suited to investigation of protein-ligand systems. This was applied to a host-guest system,<sup>364</sup> but no study for protein-ligand systems has been reported to date.

The main impediment to the use of MD in docking studies remains the computational cost of running suitably long simulations. Approaches using only short MD simulations have been shown to improve the performance of docking procedures versus methods using static structures, in certain cases.<sup>365</sup>

The filling potential method<sup>366</sup> is an MD-based approach for estimating free-energy surfaces for protein-ligand docking. This is a modified umbrella potential sampling method which enables the ligand molecule to drift out of local minima through a self-avoiding (via a tabu list) random walk consisting of an iterative cycle of local-minimum searches and transition-state searches.

Another interesting flexible docking method relies on calculation of the flexible degrees of freedom using MD simulations.<sup>367</sup> This approach allows relaxation of the protein conformation in precalculated soft flexible degrees of freedom. These soft flexible modes are extracted as principal components of motion from a conventional MD simulation.

MD is also widely used in the refinement of docked conformations from the results of approximate, generally rigid-body, ligand docking. For example, a three-stage method has been presented.<sup>368</sup> A grid-based method is used to sample the conformations of an unbound ligand in the first stage. Next, the lowest energy ligand conformers are rigidly docked into the binding site. The docked modes are refined in the third stage by molecular mechanics minimization, conformational scanning at the binding site, and a short period of MD-based simulated annealing. This procedure was applied to ligand-protein complexes with as many as 16 rotatable bonds in the ligand with final root-mean-square deviations ranging from 0.64 to 2.01 Å compared to the crystal structures.

Taking this refinement a stage further, a combined quantum mechanical/molecular mechanical (QM/MM) docking method is described.<sup>369</sup> In this method AutoDock<sup>370</sup> is used to generate initial starting points for the docked structures, semiempirical AM1 QM/MM optimization of the complex gives an improved description of the binding mode, and the electronic properties of the ligand within the environment of a flexible protein to simulate the limited structural

changes of the enzyme upon ligand binding. This method was able to reproduce the induced fit structural changes when a simple optimization was adequate for reproducing the protein's movement.

**Multiple-Copy Simultaneous Search:** Multiple-copy simultaneous search is a method for determining locations of favorable binding for functional groups to the surface of a protein.<sup>275,371</sup> A few thousand copies of a functional group are randomly placed at the protein surface and then subjected to simultaneous energy minimization and/or quenched molecular dynamics. The resulting locations of the functional group yield a functionality map of the protein receptor site, which included consideration of its flexibility. A set of these functionality maps can be used for the analysis of protein-ligand interactions and rational drug design.<sup>280</sup>

**Locally Enhanced Sampling:** In LES a fragment of the system exists as several copies in the same simulation, as described above. The individual copies do not interact with each other but do interact with the environment. In the case that the fragment is a ligand, LES becomes a method for sampling ligand conformations under the influence of a protein.<sup>372</sup>

**Dynamic Pharmacophore Method:** The dynamic pharmacophore method<sup>351,361,373</sup> requires a set of instantaneous snapshots of the fluctuating receptor molecule. These snapshots are typically extracted from MD simulations, although structures that are consistent with NMR data or other sources might be used instead. Probes corresponding to fundamental functional groups (e.g., methyl, hydroxyl, phenyl) are docked to each snapshot with the aim of detecting consensus patterns for the whole ensemble of snapshots.

Generally, receptor-based pharmacophore models are developed using a single receptor structure. These pharmacophore models based on one receptor structure could fail to identify inhibitors that bind to structures that are somewhat different from the experimental or model structure but that are still readily accessible at physiological temperatures. The dynamic pharmacophore model was developed to address this issue.

For each snapshot from an MD simulation, a pharmacophore model was constructed by identifying favorable binding sites of chemical functional groups using the multiunit search for interacting conformers (MUSIC) procedure of the BOSS program.<sup>374</sup> This identifies favorable binding sites of probe molecules by simultaneously refining the energy of a large number of probe molecules, which do not interact with each other, in the potential field of the receptor molecule. Strong binding sites tend to cluster many probe molecules in well-defined orientations and locations. Thus, strong binding sites can be selected as those which consistently appear in many snapshots rather than a few snapshots. These sites are used to form the important components in a pharmacophore model. This approach also uncovers useful binding sites that might not be readily recognized as such in the initial starting structure.

While this procedure increases the number of false positives,<sup>375</sup> it produces pharmacophore models that perform better than any single conformation model for potent inhibitors of HIV-1 integrase.<sup>373</sup>

**Relaxed Complex Method:** The relaxed complex method<sup>362,363</sup> resembles the dynamic pharmacophore method but involves docking of whole ligand molecules to the initial set of receptor snapshots with subsequent rescoring of the most favorable structures within a rigorous statistical mechanical framework.<sup>148</sup>

It is possible that ligands may bind to conformations that occur only rarely in the dynamics of the receptor and that strong binding often reflects multivalent attachment of the ligand to the receptor. Two successful experimental approaches that recognize this fact are *SAR by*

*NMR*<sup>376</sup> and the *tether method*.<sup>377</sup> The relaxed complex method was inspired by these methods and aims to reliably consider the induced fit at the binding site.

The relaxed complex method is a three-step process. The first step is to generate a series of target conformations of the receptor. This task is typically performed by selecting representative snapshots from an apo-protein MD trajectory. Alternative methods for generating the receptor conformations are possible. Such methods include the replica exchange method or even just an ensemble of short time-scale MD simulations. During the second step small libraries of candidate ligands are docked into each representative receptor structure. In the original work<sup>362,363</sup> the Autodock<sup>370</sup> software was used for this docking process, although the overall method is independent of the choice of docking software. The final stage seeks to improve the scoring of the predicted binding configurations by use of a more rigorous, but computationally expensive, method for estimating the standard free energies of binding. This rescoring process has been demonstrated with MM/PBSA<sup>363</sup> as implemented in the AMBER software with the electrostatic terms calculated using the APBS software,<sup>164</sup> but again, the overall method is not dependent upon this choice.

A double-ligand variation of the above procedure incorporates consideration of the fact that two ligands with relatively low binding affinities might be linked to form a single high-affinity ligand. Because the binding of the first ligand could introduce unfavorable interactions for the binding of the second ligand, the combination of the best-ranked ligands for respective binding sites does not necessarily produce the most favorable composite compound. Continuing from the previous single-ligand studies, the first ligand is treated as part of the receptor molecule and the docking simulations of the second ligand are repeated but limited to a region of space consistent with the allowable lengths of linkers. As before, the binding of the second ligand is subsequently recalculated using the more accurate approach.

Initial results presented from work applying the relaxed complex approach covered the binding of two ligands to the FK506 binding protein (FKBP) with conformations generated via a 2 ns MD simulation. This demonstrated that the ligand binding is sensitive to conformational fluctuations in the protein; the binding energies covered a range of 3–4 kcal mol<sup>-1</sup>, which corresponds to a 100–1000-fold difference in binding affinities.<sup>362</sup> The use of MM/PBSA free-energy evaluations allowed for correct prediction of binding modes compared to the crystallographic structures.<sup>363</sup>

**Lambda Dynamics Method:** The  $\lambda$  dynamics method<sup>155,156</sup> is another technique intended for speeding up free-energy calculations. In the  $\lambda$  dynamics method, another multiple-copy method, multiple ligands are simultaneously located in the receptor binding site. However, the interaction potential of each ligand is reduced from its full strength. The fraction,  $\lambda_i^2$ , of the interaction potential for each ligand is determined dynamically during the simulation as an additional degree of freedom. Specifically,  $\lambda_i$  is treated as a particle with a fictitious mass. Because the interaction potential of each ligand is reduced, the barriers for conformational transitions are lowered. The reduced barriers allow each ligand to further explore orientational and conformational space more readily. The ranking of the ligands can emerge rapidly during the simulation because  $\lambda_i^2$  is rapidly able to increase for the winners at the expense of the losers. Distinguishing the strong binders from the weaker binders can be much faster than by performing many individual free-energy perturbation calculations with a single ligand each time. This method has been demonstrated to efficiently identify strong benzamidine inhibitors of trypsin.<sup>378</sup>

**7.3.2. Practical Applications of Molecular Dynamics in Drug Discovery**—There are, undoubtedly, many examples where computer simulation and molecular dynamics have played a demonstrated role in the discovery or development of therapeutic drugs.<sup>4</sup> There are

particular cases where MD provided an otherwise unavailable description of the flexibility of the binding site to aid in the development of drug candidates. It is hoped that MD will help in the discovery of practical HIV-1 integrase inhibitors, for instance, as it had in the earlier discovery of widely prescribed HIV-1 protease inhibitors.<sup>379</sup> Concepts from HIV-1 integrase MD simulation have proved useful in developing promising novel antiviral compounds.<sup>380</sup> MD simulations indicated sizable fluctuations of the catalytic site.<sup>381–383</sup> MD studies predicted favorable binding of compounds that utilize a hitherto unknown additional binding trench adjacent to the catalytic site,<sup>384</sup> as shown in Figure 2, and this has been validated by experimental studies.<sup>380</sup> This has the potential to greatly reduce the likelihood of resistant strains developing.

## 8. Future Prospects and Challenges

### 8.1. Efficiency and Stability

Standard MD methods often fail to explore configurational space adequately for the accurate evaluation of thermodynamic and kinetic properties for proteins. This is partly because such systems typically have enthalpic and entropic barriers that are significantly higher than the thermal energy at physiologically relevant temperatures. When systems are trapped in local regions of configurational space over the time scale of a simulation, due to high free-energy barriers, they appear nonergodic.<sup>385</sup> That is, for these systems the time averages of observable characteristics do not equal the corresponding ensemble averages. The simple fact that the low-frequency motions of proteins typically correspond to the larger conformational changes, and these are often the more interesting motions, aggravates the issue. Such motions sometimes do not involve crossing of a very high energy barrier but may have a slow, diffusional character. Thus, the problem is just a matter of sampling for an inadequate length of time. Many different enhanced sampling methods have been introduced in the literature, as described above, to reduce this problem. However, no perfect solution has been devised to date. Indeed, certain approaches are better suited to specific systems or observable characteristics than others. Future progress toward resolving this issue will be of great interest.

Besides the development of improved sampling protocols, simply enhancing the efficiency of MD routines will increase its practical scope. For example, improvements to integrators might allow larger time steps to be used. Likewise, improved methods for long-range force evaluation, particularly in terms of computational parallelization, would lead to more efficient simulations. Each of these putative improvements is ripe for exploration.

Fairly fundamental incremental improvements to the underlying MD algorithms are still being made (e.g., a fast and readily parallelized pair list construction algorithm for nonbonded interactions was described recently<sup>386</sup>).

### 8.2. Electrostatics

Generalized Born treatments of electrostatics have had the important advantage of lower computational requirements over the more rigorous Poisson-Boltzmann treatments. With the recent development of Poisson-Boltzmann solvers that are fast enough to conduct simulations for proteins,<sup>164,387,388</sup> it would be desirable to choose Poisson-Boltzmann methods when the drawbacks of GB might affect the conclusions being drawn from a simulation. While analogous methods that use Poisson-Boltzmann rather than generalized Born treatments of the electrostatics require no conceptual jumps, those methods cannot be implemented as efficiently as GB at this time. The necessary theory for determining forces and MD trajectories from the Poisson-Boltzmann formalism exists<sup>387,389,390</sup> but can be prohibitively expensive to calculate at present. Analytical methods for higher order derivatives in such methods are desired.

### 8.3. Solvation and Solvent Models

Improvements to the speed and accuracy of calculations regarding solvent will be particularly beneficial. Speed increases will be useful because a major portion of a typical simulation will consist of solvent. Accuracy improvements are important because the solvent often mediates important aspects of protein structure, dynamics, and function. Replicating explicit solvent effects with new implicit solvent models will help to advance the understanding of such aspects. Currently, specific interactions, solvation shells, and long-range order are ignored in most implicit models, although some hybrid methods do seek to resolve this. In particular, the poor correlation between apolar solvation forces exhibited in explicit solvent simulations and implicit solvent simulations needs to be addressed.<sup>99,97</sup>

### 8.4. MD in Ligand Docking and Molecular Design Studies

While protein flexibility undoubtedly plays a critical role in determining molecular recognition, most drug design and modeling efforts disregard these effects since they are computationally expensive to include. With the rapid progress in algorithms and computational resources, as discussed in this review, it is becoming feasible to consider these effects in a wider range of drug discovery tasks. The more demanding but rigorous free-energy calculation methods can often be used in the later stages of lead optimization. The more rapid but approximate methods, those relying on single reference states, can be used to quickly identify favorable and unfavorable features of a putative lead compound in molecular recognition. The identification and classification of these features can help to suggest modifications of the compound to improve binding affinities. These features also aid in construction of pharmacophore models for locating possible alternative lead compounds from chemical libraries or databases. Even the generation of improved libraries can be assisted via the use of pharmacophore-based constraints in their design. Free-energy methods that may be considered intermediate to the two extremes, for example, MM/PBSA and the semiempirical linear response approach, can be utilized to further screen out less promising compounds suggested by the single reference state models before more rigorous free-energy calculations are performed. Improved implicit or hybrid implicit/explicit solvent models also have a role in facilitating rapid conformational sampling, allowing protein flexibility to be fully accounted for in the early stages of a rational drug design process.

### 8.5. Benchmarks

Benchmarks are inherently subjective; it is impossible to objectively compare all MD algorithms or software on equal terms when they have widely differing capabilities and aims. However, there are certain common example simulations, or analysis tasks, that are amenable to benchmarking for both numerical accuracy and computational cost. The MD community would benefit greatly from a diverse set of well-conceived and publicly available benchmark tests based on these.

There are few direct comparisons of the various accelerated sampling techniques in the literature, for example. A standard set of benchmark tests would alleviate the need for researchers to setup and run unfamiliar methods and software to fairly compare these against newly developed approaches. This would be of particular value when methods are specifically developed, or optimized, for computer systems unavailable to the researchers who would otherwise be performing the comparisons.

### 8.6. Computing Facilities

MD continues to benefit immensely from improvements in computer technology. As computers become faster, it has become possible to handle larger molecules and explore their dynamics for longer time scales. Moreover, the recent advent of Beowulf-style clusters has

resulted in a considerable increase in the number of research groups able to undertake biomolecular simulations. Currently, a typical simulation might have a system size of around  $10^5$ - $10^6$  atoms, and a multiple nanosecond simulation will probably require of the order of  $10^6$ - $10^7$  time steps. Such a simulation could be expected to take a couple of weeks on between 8 and 32 processors (obviously this is highly dependent on the nature of the particular protocols employed and on the efficiency of the simulation software). During this period it could generate gigabytes of data for subsequent analysis and visualization. Computer resources with the capability of handling and storing such large quantities of data are now widely available, but meaningful visualization is increasingly becoming a challenging task. The tools and techniques developed for large-scale data-mining efforts will also be increasingly valuable in the study of MD trajectories.

### 8.7. BioSimGrid

As just indicated, a typical and routine large-scale MD simulation might produce several gigabytes of raw data that needs to be processed when complete. An additional unresolved issue faced by the developers and users of modern molecular dynamics technology is the archival, indexing, and dissemination of this output data. Building upon the current efforts toward Grid computing, BioSimGrid (<http://www.biosimgrid.org/>) may provide the solution or insight toward future solutions. BioSimGrid is a collaborative project between several of the leading U.K. research groups in the field of molecular simulation.<sup>391</sup>

The BioSimGrid project seeks to build, using Grid technology, a publicly accessible database of biomolecular simulation data. The data will include the raw simulation output, information about the generic properties of that output and the corresponding software configuration data, and information derived from analysis of the raw data. One valuable outcome of such a database might be integration of the simulation data with experimental and bioinformatic data, opening a wide range of data-mining possibilities.

## 9. Summary

MD simulations of proteins have provided many insights into the internal motions of these biomolecules. Simulation of *in silico* models aids in the interpretation and reconciliation of experimental data.

With ongoing advances in both methodology and computational resources, molecular dynamics simulations are being extended to larger systems and longer time scales. This enables investigation of motions and conformational changes that have functional implications and yields information that is not available through any other means. Today's results suggest that (subject to the continuing utilization of synergies between experiment and simulation) the applications of molecular dynamics will command an increasingly critical role in our understanding of biological systems.

Investigation of the structural and functional characteristics of intriguing biochemical systems is being made possible by computer simulation with techniques such as molecular dynamics.

## Acknowledgements

The authors acknowledge support from Accelrys, the Howard Hughes Medical Institute, the NIH, the NSF, the National Biomedical Computation Resource, the San Diego Supercomputer Center, and the NSF Center for Theoretical Biological Physics through funding and provision of resources.

## Biographies

Stewart A. Adcock was a recent postdoctoral researcher in J. Andrew McCammon's research group at the University of California, San Diego (UCSD). He was born in Norwich, England, in 1975. He received his MChem degree in Chemistry from the University of Sheffield and his D.Phil. degree from Oxford University. His doctoral research concerned the structural prediction and simulation of transmembrane proteins, under the guidance of W. G. Richards. Between 2001 and 2004 he was a UCSD research scholar, during which time he developed protocols and software for protein structure modeling and prediction. Currently, he is a scientific software consultant with Tessella Support Services PLC and has interest in high-performance scientific software and computing platforms.



J. Andrew McCammon is an Investigator of the Howard Hughes Medical Institute and holds the Joseph E. Mayer Chair of Theoretical Chemistry at the University of California, San Diego (UCSD). He was born in Lafayette, Indiana, in 1947. He received his B.A. degree from Pomona College and his Ph.D. degree in Chemical Physics from Harvard University, where he worked with John Deutch on biological applications of statistical mechanics and hydrodynamics. In 1976–78 he was a research fellow at Harvard, where he developed the computer simulation approach to protein dynamics in collaboration with Martin Karplus. He joined the faculty of the University of Houston as Assistant Professor of Chemistry in 1978 and was appointed to the M. D. Anderson Chair at Houston in 1981. He moved to his current positions at UCSD in 1995. His industrial consulting started in 1987 with the establishment of the computer-aided drug discovery program of Agouron Pharmaceuticals. About 50 of his former students hold tenured or tenure-track positions at leading universities or research institutes. His awards include the George Herbert Hitchings Award for Innovative Methods in Drug Design from the Burroughs Wellcome Fund and the Smithsonian Institution's Award for Breakthrough Computational Science, sponsored by Cray Research. New directions in his research group include theoretical efforts to understand how supramolecular and cellular activity emerge from the molecular level, particularly in neural synapses.



## 11. Abbreviations

**AGBNP**

analytic generalized Born and nonpolar

**AFM**

	atomic force microscopy
<b>BD</b>	Brownian dynamics
<b>BPTI</b>	bovine pancreatic trypsin inhibitor
<b>DEM</b>	diffusion equation method
<b>ED</b>	essential dynamics
<b>FEP</b>	free-energy perturbation
<b>FM</b>	fast multipole
<b>GB</b>	generalized Born model
<b>IMD</b>	interactive molecular dynamics
<b>LES</b>	locally enhanced sampling
<b>LIE</b>	linear interaction energy
<b>LINCS</b>	linear constraint solver
<b>MBO(N)D</b>	multiple-body O(N) dynamics
<b>MC</b>	Monte Carlo
<b>MD</b>	molecular dynamics
<b>MEHMC</b>	momentum-enhanced hybrid Monte Carlo method
<b>MM</b>	molecular mechanics
<b>MM/GBSA</b>	molecular mechanics/generalized Born-surface area method
<b>MM/PBSA</b>	molecular mechanics/Poisson Boltzmann-surface area method
<b>PCA</b>	principal component analysis



<b>PMEMD</b>	particle mesh Ewald molecular dynamics
<b>PMF</b>	potential of mean force
<b>QM/MM</b>	hybrid quantum mechanics/molecular mechanics
<b>REMD</b>	replica exchange molecular dynamics
<b>RESPA</b>	reversible reference system propagator algorithms
<b>SA</b>	simulated annealing
<b>SASA</b>	solvent-accessible surface area
<b>SGMD</b>	self-guided molecular dynamics
<b>SMD</b>	steered molecular dynamics
<b>TI</b>	thermodynamic integration

## References

1. Cheatham TE, Kollman PA. *Ann Rev Phys Chem* 2000;51:435. [PubMed: 11031289]
2. Karplus M, McCammon JA. *Nat Struct Biol* 2002;9:646. [PubMed: 12198485]
3. Norberg J, Nilsson L. *Acc Chem Res* 2002;35:465. [PubMed: 12069632]
4. Moraitakis G, Purkiss AG, Goodfellow JM. *Rep Prog Phys* 2003;66:383.
5. Straatsma TP, McCammon JA. *Ann Rev Phys Chem* 1992;43:407.
6. Simonson T, Archontis G, Karplus M. *Acc Chem Res* 2002;35:430. [PubMed: 12069628]
7. Alder BJ, Wainwright TE. *J Chem Phys* 1957;27:1208.
8. Rahman A. *Phys Rev A* 1964;136:405.
9. McCammon JA. *Molecular Dynamics study of the bovine pancreatic trypsin inhibitor. Models for Protein Dynamics; CECAM Orsay France* 1976:137.
10. McCammon JA, Gelin BR, Karplus M. *Nature* 1977;267:585. [PubMed: 301613]
11. Kremer K. *Macromol Chem Phys* 2003;204:257.
12. Rickman JM, LeSar R. *Ann Rev Mater Res* 2002;32:195.
13. Duan Y, Kollman PA. *Science* 1998;282:740. [PubMed: 9784131]
14. Berendsen HJ, Hayward S. *Curr Opin Struct Biol* 2000;10:165. [PubMed: 10753809]
15. Kern D, Zuiderweg ERP. *Curr Opin Struct Biol* 2003;13:748. [PubMed: 14675554]
16. Daggett V. *Chem Rev* 2006;106:1898. [PubMed: 16683760]
17. McQuarrie, DA. *Statistical Mechanics*. Harper and Row; New York: 1976.
18. Callen, HB. *Thermodynamics and an Introduction to Thermostatistics*. Wiley; New York: 1985.
19. Tiana G, Simona F, De Mori GMS, Broglia RA, Colombo G. *Protein Sci* 2004;13:113. [PubMed: 14691227]

20. Kim HW, Shen TJ, Sun DP, Ho NT, Madrid M, Tam MF, Zou M, Cottam PF, Ho C. *Proc Natl Acad Sci, USA* 1994;91:11547. [PubMed: 7972099]
21. Wang W, Donini O, Reyes CM, Kollman PA. *Ann Rev Biophys Biomol Struct* 2001;30:211. [PubMed: 11340059]
22. Warshel A. *Ann Rev Biophys Biomol Struct* 2003;32:425. [PubMed: 12574064]
23. Brooijmans N, Kuntz ID. *Ann Rev Biophys Biomol Struct* 2003;32:355.
24. Roux B. *Acc Chem Res* 2004;35:366. [PubMed: 12069621]
25. Bond PJ, Sansom MSP. *Mol Membr Biol* 2004;21:151. [PubMed: 15204623]
26. Elcock AH. *Numerical Comput Methods, D: Methods Enzymol* 2004;383:166.
27. Day R, Daggett V. *Protein Simul Rev Protein Chem* 2003;66:373.
28. Pettitt BM, Makarov VA, Andrews BK. *Curr Opin Struct Biol* 1998;8:218. [PubMed: 9631296]
29. Wong CF, McCammon JA. *Rev Protein Chem* 2003;66:87.
30. Koehl P, Levitt M. *J Mol Biol* 1999;293:1161. [PubMed: 10547293]
31. Brunger AT, Adams PD. *Acc Chem Res* 2002;35:404. [PubMed: 12069625]
32. Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M. *Proteins: Struct, Funct Genet* 2003;50:496. [PubMed: 12557191]
33. Fan H, Mark AE. *Protein Sci* 2004;13:211. [PubMed: 14691236]
34. Born M, Oppenheimer JR. *Ann Phys* 1927;84:457.
35. Cheatham TE. *Curr Opin Struct Biol* 2004;14:360. [PubMed: 15193317]
36. Ewig CS, Berry R, Dinur U, Hill JR, Hwang MJ, Li H, Liang C, Maple J, Peng Z, Stockfisch TP, Thacher TS, Yan L, Ni X, Hagler AT. *J Comput Chem* 2001;22:1782. [PubMed: 12116411]
37. Sun H. *J Phys Chem B* 1998;102:7338.
38. Palmo K, Mannfors B, Mirkin NG, Krimm S. *Biopolymers* 2003;68:383. [PubMed: 12601797]
39. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. *J Phys Chem B* 1998;102:3586.
40. MacKerell, AD., Jr; Brooks, B.; Brooks, CL., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. *The Encyclopedia of Computational Chemistry*. Schleyer, PvR, et al., editors. 1. John Wiley & Sons; Chichester: 1998. p. 271
41. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. *J Comput Chem* 1983;4:187.
42. Reed, TM.; Gubbins, KE. *Statistical Mechanics*. McGraw; New York: 1973.
43. Pettitt BM, Karplus M. *J Am Chem Soc* 1985;107:1166.
44. MacKerell AD, Wiorkiewicz-kuczera J, Karplus M. *J Comput Chem* 1995;117:11946.
45. Feller SE, MacKerell AD. *J Phys Chem B* 2000;104:7510.
46. Kuttel M, Brady JW, Naidoo KJ. *J Comput Chem* 2002;23:1236. [PubMed: 12210149]
47. Reiling S, Schlenkrich M, Brickmann J. *J Comput Chem* 1996;17:450.
48. Cournia Z, Vaiana AC, Ullmann GM, Smith JC. *Pure Appl Chem* 2004;76:189.
49. Halgren TA. *J Comput Chem* 1996;17:514.
50. Halgren TA. *J Comput Chem* 1996;17:520.
51. Halgren TA. *J Comput Chem* 1996;17:553.
52. McCammon, JA.; Harvey, SC. *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press; Cambridge: 1987.
53. Leach, AR. *Molecular Modelling: Principles and Applications*. Pearson Education; London: 2001.
54. Bolhuis PG, Chandler D, Dellago C, Geissler PL. *Ann Rev Phys Chem* 2002;53:291. [PubMed: 11972010]
55. Verlet J. *Phys Rev* 1967;159:98.
56. Beeman D. *J Comput Phys* 1976;20:130.
57. Swope WC, Andersen HC, Berens PH, Wilson KR. *J Chem Phys* 1982;76:637–649.

58. Gear, CW. Numerical Initial Value Problems in Ordinary Differential Equations. Prentice; New York: 1971.
59. Hockney RW. Methods Comput Phys 1970;9:136.
60. van Gunsteren WF, Berendsen HJC. Mol Phys 1977;34:1311.
61. Wang SS, Krumhansi JA. J Chem Phys 1972;56:4287.
62. Adams DJ. Chem Phys Lett 1979;62:329.
63. Berkowitz M, McCammon JA. Chem Phys Lett 1982;90:215.
64. Brooks CL, Brunger A, Karplus M. Biopolymers 1985;24:843. [PubMed: 2410050]
65. Belch AC, Berkowitz M. Chem Phys Lett 1985;113:278.
66. Beglov D, Roux B. J Phys Chem 1994;100:9050.
67. Banavali NK, Im W, Roux B. J Chem Phys 2002;117:7381.
68. Roux, B.; Beglov, D.; Im, W. Generalized Solvent Boundary Potentials for Computer Simulations, Santa Fe Workshop. In: Pratt, LR.; Hummer, G., editors. Treatments of electrostatics interactions in computer simulations of condensed media. 492. 1999. p. 473
69. Schlick T, Barth E, Mandziuk M. Ann Rev Biophys Biomol Struct 1997;26:181. [PubMed: 9241418]
70. Tuckerman M, Berne BJ, Martyna GJ. J Chem Phys 1992;97:1990–2001.
71. Ryckaert JP, Ciccotti G, Berendsen HJC. J Comput Phys 1977;23:327.
72. Andersen HC. J Comput Phys 1983;52:24–34.
73. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. J Comput Chem 1997;18:1463.
74. van Gunsteren WF, Karplus M. Macromolecules 1982;15:1528.
75. Krautler V, van Gunsteren WF, Hunenberger PH. J Comput Chem 2001;22:501.
76. Forester TR, Smith W. J Comput Chem 1998;19:102.
77. Kutteh R. J Chem Phys 1999;111:1394.
78. Allen, MP.; Tildesley, DJ. Computer Simulation of Liquids. Oxford University Press; Oxford: 1987.
79. Andersen HC. J Chem Phys 1980;72:8205.
80. Evans DJ, Morris GP. Chem Phys 1983;77:63.
81. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. J Chem Phys 1984;81:3684.
82. Martyna GJ, Tobias DJ, Klein ML. J Chem Phys 1994;101:4177.
83. Feller SE, Zhang Y, Pastor RW, Brooks BR. J Chem Phys 1995;103:4613.
84. Bond PJ, Sansom MSP. J Mol Biol 2003;329:1035. [PubMed: 12798692]
85. Gumbart J, Wang Y, Aksimentiev A, Tajkhorshid E, Schulten K. Curr Opin Struct Biol 2005;15:423. [PubMed: 16043343]
86. Cramer CJ, Truhlar DG. Chem Rev 1999;99:2161. [PubMed: 11849023]
87. Simonson T. Curr Opin Struct Biol 2001;11:243. [PubMed: 11297935]
88. Feig M, Brooks CL. Curr Opin Struct Biol 2004;14:217. [PubMed: 15093837]
89. Baker NA. Curr Opin Struct Biol 2005;15:137. [PubMed: 15837170]
90. Levy RM, Gallicchio E. Ann Rev Phys Chem 1998;49:531. [PubMed: 9933909]
91. Lazardis T, Karplus M. Biophys Chem 2003;100:367. [PubMed: 12646378]
92. Roux B, Simonson T. Biophys Chem 1999;78:1. [PubMed: 17030302]
93. Bashford D, Case DA. Ann Rev Phys Chem 2000;51:129. [PubMed: 11031278]
94. Sosa CP, Hewitt T, Lee MR, Case DA. J Mol Struct: Theochem 2001;549:193.
95. Zhang LY, Gallicchio E, Friesner RA, Levy RM. J Comput Chem 2001;22:591.
96. Im W, Berneche S, Roux B. J Chem Phys 2001;114:2924.
97. Gallicchio E, Levy RM. J Comput Chem 2004;25:479. [PubMed: 14735568]
98. Felts AK, Harano Y, Gallicchio E, Levy RM. Proteins: Struct, Funct, Bioinform 2004;56:310.
99. Wagoner J, Baker NA. J Comput Chem 2004;25:1623. [PubMed: 15264256]
100. Jorgensen WL, Chandrasekhar J, Madura J, Impey RW, Klein ML. J Chem Phys 1983;79:926.
101. Mahoney MW, Jorgensen WL. J Chem Phys 2000;112:8910.
102. Berendsen HJC, Grigera JR, Straatsma TP. J Phys Chem 1987;91:6269.

103. Guenot J, Kolman PA. *Protein Sci* 1992;1:1185. [PubMed: 1304396]
104. Steinback PJ, Brooks BR. *Proc Natl Acad Sci, USA* 1993;90:9135. [PubMed: 8415667]
105. Sen S, Nilsson L. *Biophys J* 1999;77:1782. [PubMed: 10512803]
106. Sankaramakrishnan R, Konvicka K, Mehler EL. *Int J Quantum Chem* 2000;77:174.
107. Gilson MK. *Curr Opin Struct Biol* 1995;5:216. [PubMed: 7648324]
108. Norberg J, Nilsson L. *Biophys J* 2000;79:1537. [PubMed: 10969015]
109. Walser R, Hunenberger PH, van Gunsteren WF. *Proteins: Struct, Funct Genet* 2001;43:509. [PubMed: 11340666]
110. Brooks CL III, Pettitt BM, Karplus M. *J Chem Phys* 1985;83:5897.
111. Steinbach PJ, Brooks BR. *J Comput Chem* 1994;15:667.
112. Ewald PP. *Ann Phys* 1921;64:253.
113. Toukhami AY, Board JA. *Comput Phys Commun* 1996;95:73.
114. Sagui C, Darden TA. *Ann Rev Biophys Biomol Struct* 1999;28:155. [PubMed: 10410799]
115. Hunenberger PH, McCammon JA. *J Chem Phys* 1999;110:1856.
116. Weber W, Hunenberger PH, McCammon JA. *J Phys Chem B* 2000;104:3668.
117. Yeh IC, Berkowitz ML. *J Chem Phys* 1999;111:3155.
118. Kawata M, Nagashima U. *Chem Phys Lett* 2001;240:165.
119. Rozanska X, Chipot C. *J Chem Phys* 2000;112:9691.
120. Chao SD, Kress JD, Redondo A. *J Chem Phys* 2004;120:5558. [PubMed: 15267431]
121. Holst M, Kozack RE, Saied F, Subramaniam S. *J Biomol Struct Dyn* 1994;11:1437. [PubMed: 7946084]
122. Skeel RD, Tezcan I, Hardy DJ. *J Comput Chem* 2002;23:673. [PubMed: 11939600]
123. Ermak DL, McCammon JA. *J Chem Phys* 1978;69:1352.
124. Gabdoulline RR, Wade RC. *Curr Opin Struct Biol* 2002;12:204. [PubMed: 11959498]
125. Metropolis N, Rosenbluth A, Rosenbluth R, Teller A, Teller E. *J Chem Phys* 1953;21:1087.
126. Dinner AR. *J Comput Chem* 2000;21:1132.
127. Ulmschneider JP, Jorgensen WL. *J Chem Phys* 2003;118:4261.
128. Northrup SH, McCammon JA. *Biopolymers* 1980;19:1001. [PubMed: 7378544]
129. La Berge LJ, Tully JC. *Chem Phys* 2000;260:183.
130. Faller R, de Pablo JJ. *J Chem Phys* 2002;116:55.
131. Hampton SS, Izaguirre JA. *J Comput Phys* 2004;200:581–604.
132. Kirkpatrick S, Gelatt CD, Vecchi MP. *Science* 1983;220:671. [PubMed: 17813860]
133. Kostrowicki J, Scheraga HA. *J Phys Chem* 1992;96:7442.
134. Huber T, Torda A, van Gunsteren W. *J Phys Chem* 1997;101:5926.
135. Hart RK, Pappu RV, Ponder JW. *J Comput Chem* 2000;20:531.
136. Tai K. *Biophys Chem* 2004;107:213. [PubMed: 14967236]
137. De Groot B, van Aalten D, Scheek R, Amadei A, Vriend G. *Proteins: Struct, Funct Genet* 1997;29:240. [PubMed: 9329088]
138. Garcia AE. *Phys Rev Lett* 1992;68:2696. [PubMed: 10045464]
139. Mongan J. *J Comput-Aided Mol Des* 2004;18:433. [PubMed: 15663003]
140. Warshel A, Levitt M. *J Mol Biol* 1976;103:227–249. [PubMed: 985660]
141. Shurki A, Warshel A. *Rev Protein Chem* 2003;66:249.
142. Warshel A. *Chem Rev* 2006;106:1737. [PubMed: 16683752]
143. Lybrand TP, McCammon JA, Wipff G. *Proc Natl Acad Sci, USA* 1986;83:833. [PubMed: 3456569]
144. Roderinger T, Pomes R. *Curr Opin Struct Biol* 2005;15:164. [PubMed: 15837174]
145. Kollman P. *Chem Rev* 1993;93:2395.
146. Ajay A, Murcko MA. *J Med Chem* 1995;38:4953. [PubMed: 8544170]
147. Almlof M, Brandsdal BO, Aqvist J. *J Comput Chem* 2004;25:1242. [PubMed: 15139037]
148. Gilson MK, Given JA, Bush BL, McCammon JA. *Biophys J* 1997;72:1047. [PubMed: 9138555]

149. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA. *J Am Chem Soc* 1998;120:9401.
150. Eriksson MA, Pitera J, Kollman PA. *J Med Chem* 1999;42:868. [PubMed: 10072684]
151. Pitera JW, Kollman PA. *Proteins: Struct, Funct Genet* 2000;41:385. [PubMed: 11025549]
152. Randmer RJ, Kollman PA. *J Comput-Aided Mol Des* 1998;12:215. [PubMed: 9749366]
153. Pearlman DA. *J Med Chem* 1999;42:4313. [PubMed: 10543875]
154. Pearlman DA, Charifson PS. *J Med Chem* 2001;44:502. [PubMed: 11170640]
155. Banba S, Guo ZY, Brooks CL III. *J Phys Chem B* 2000;104:6903.
156. Damodaran KV, Banba S, Brooks CL III. *J Phys Chem B* 2001;105:9316.
157. Rodiger T, Howell PL, Pomes R. *J Chem Phys* 2005;123:034104.
158. Lazaridis T, Masunov A, Gandolfo F. *Proteins: Struct, Funct Genet* 2002;47:194. [PubMed: 11933066]
159. Luo HB, Gilson MK. *J Am Chem Soc* 2000;122:2934.
160. Luo HB, Sharp K. *Proc Natl Acad Sci, USA* 2002;99:10399. [PubMed: 12149474]
161. Swanson JMJ, Henschman RH, McCammon JA. *Biophys J* 2004;86:67. [PubMed: 14695250]
162. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. *Acc Chem Res* 2000;33:889. [PubMed: 11123888]
163. Lee MR, Duan Y, Kollman PA. *Proteins: Struct, Funct Genet* 2000;39:309. [PubMed: 10813813]
164. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. *Proc Natl Acad Sci, USA* 2001;98:10037. [PubMed: 11517324]
165. Honig B, Nicholls A. *Science* 1995;268:1144. [PubMed: 7761829]
166. Debrunner PG, Frauenfelder H. *Ann Rev Phys Chem* 1982;33:283.
167. McCammon JA, Lee CY, Northrup SH. *J Am Chem Soc* 1983;105:2232.
168. Isralewitz B, Gao M, Schulten K. *Curr Opin Struct Biol* 2001;11:224. [PubMed: 11297932]
169. Isralewitz B, Baudry J, Gullingsrud J, Kosztin D, Schulten K. *J Mol Graphics Modelling* 2001;19:13.
170. Prins JF, Hermans J, Mann G, Nyland G, Simons M. *Future Generation Comput Syst* 1999;15:485.
171. Hermans, J.; Yun, RH.; Leech, J.; Cavanaugh, D. University of North Carolina; 1994. *SIGMA documentation*, 8080/HERMANS/software/SIGMA/;. <http://hekto.med.unc.edu>
172. Humphrey W, Dalke A, Schulten K. *J Mol Graphics* 1996;14:33.
173. Grayson P, Tajkhorshid E, Schulten K. *Biophys J* 2003;85:36. [PubMed: 12829462]
174. Binning G, Quate CF, Gerber G. *Phys Rev Lett* 1986;56:930. [PubMed: 10033323]
175. Ashkin A, Dziedzic JM. *Science* 1987;235:1517. [PubMed: 3547653]
176. Evans E, Ritchie K, Merkel R. *Biophys J* 1995;68:2580. [PubMed: 7647261]
177. Evans E. *Ann ReV Biophys Biomol Struct* 2001;30:105. [PubMed: 11340054]
178. Balsara M, Stepaniants S, Izrailev S, Oono Y, Schulten K. *Biophys J* 1997;73:1281. [PubMed: 9284296]
179. Evans E, Ritchie K. *Biophys J* 1997;72:1541. [PubMed: 9083660]
180. Gullingsrud JR, Braun R, Schulten K. *J Comput Phys* 1999;151:190.
181. Park S, Schulten K. *J Chem Phys* 2004;120:5946. [PubMed: 15267476]
182. Jarzynski C. *Phys Rev Lett* 1997;78:2690.
183. Jarzynski C. *Phys Rev E* 1997;56:5018.
184. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K. *J Chem Phys* 2003;119:3559.
185. Jensen MO, Park S, Tajkhorshid E, Schulten K. *Proc Natl Acad Sci, USA* 2002;99:6731. [PubMed: 11997475]
186. Liphardt J, Dumont S, Smith SB, Tinoco I Jr, Bustamante C. *Science* 2002;296:1832. [PubMed: 12052949]
187. Kruger P, Verheyden S, Declerck PJ, Engelborghs Y. *Protein Sci* 2001;10:798. [PubMed: 11274471]
188. Ferrara P, Apostolakis J, Caflisch A. *Proteins: Struct, Funct Genet* 2000;39:252. [PubMed: 10737947]
189. Ponder JW, Case DA. *Rev Protein Chem* 2003;66:27.

190. MacKerell AD Jr. *J Comput Chem* 2004;25:1584. [PubMed: 15264253]
191. Ewig CS, Thacher TS, Hagler AT. *J Phys Chem B* 1999;103:6998.
192. Beachy M, Chassman D, Murphy R, Halgren T, Friesner R. *J Am Chem Soc* 1997;119:5908.
193. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. *J Am Chem Soc* 1995;117:5179.
194. Bayly CI, Cieplak P, Cornell WD, Kollman PA. *J Phys Chem* 1993;97:10269.
195. Cornell WD, Cieplak P, Bayly CI, Kollman PA. *J Am Chem Soc* 1993;115:9620.
196. Wang J, Cieplak P, Kollman PA. *J Comput Chem* 2000;21:1049.
197. Cieplak P, Caldwell J, Kollman P. *J Comput Chem* 2001;22:1048.
198. Wang J, Kollman PA. *J Comput Chem* 2001;22:1219.
199. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. *J Comput Chem* 2004;25:1157. [PubMed: 15116359]
200. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. *J Phys Chem B* 1998;102:3586.
201. Jorgensen WL, Maxwell DS, Tirado-Rives J. *J Am Chem Soc* 1996;118:11225.
202. Rizzo RC, Jorgensen WL. *J Am Chem Soc* 1999;121:4827.
203. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. *J Phys Chem B* 2001;105:6474.
204. Halgren TA, Damm W. *Curr Opin Struct Biol* 2001;11:236. [PubMed: 11297934]
205. Rick SW, Stuart SJ. *Rev Comput Chem* 2002;18:89.
206. Stern HA, Kaminski GA, Banks JL, Zhou R, Berne BJ, Friesner RA. *J Phys Chem B* 1999;103:4730.
207. Ridard J, Levy B. *J Comput Chem* 1999;20:473.
208. Banks JL, Kaminski GA, Zhou R, Mainz DT, Berne BJ, Friesner RA. *J Chem Phys* 1999;110:741.
209. Dang LX. *J Phys Chem B* 1998;102:620.
210. Yoshii N, Miyauchi R, Niura S, Okazaki S. *Chem Phys Lett* 2000;317:414.
211. Llanta E, Ando K, Rey R. *J Phys Chem B* 2001;105:7783.
212. Rick SW. *J Chem Phys* 2001;114:2276.
213. Guillot B, Guissani Y. *J Chem Phys* 2001;114:6720.
214. Mahoney MW, Jorgensen WL. *J Chem Phys* 2001;114:9337.
215. Mannfors B, Palmo K, Krimm S. *J Mol Struct* 2000;556:1.
216. Lamoureux G, MacKerell AD Jr, Roux B. *J Chem Phys* 2003;119:5185.
217. Lamoureux G, Roux B. *J Chem Phys* 2003;119:3025.
218. van Maaren PJ, van der Spoel D. *J Phys Chem B* 2001;105:2618.
219. Patel S, Brooks CL III. *J Comput Chem* 2004;25:1. [PubMed: 14634989]
220. Patel S, Brooks CL III. *J Comput Chem* 2004;25:1504. [PubMed: 15224394]
221. Grossfield A, Ren PY, Ponder JW. *J Am Chem Soc* 2003;125:15671. [PubMed: 14664617]
222. Ponder JW, Richards FM. *J Comput Chem* 1987;8:1016.
223. Garcia-Moreno B, Fitch CA. *Methods Enzymol* 2004;380:20. [PubMed: 15051331]
224. Nielsen JE, McCammon JA. *Protein Sci* 2003;12:313. [PubMed: 12538895]
225. Kuhn B, Kollman PA, Stahl M. *J Comput Chem* 2004;2:1865. [PubMed: 15376253]
226. Antosiewicz J, McCammon JA, Gilson MK. *J Mol Biol* 1994;238:415. [PubMed: 8176733]
227. Oberoi H, Allewell NM. *Biophys J* 1993;65:48. [PubMed: 8369451]
228. Baptista AM, Martel PJ, Petersen SB. *Proteins: Struct, Funct Genet* 1997;27:523. [PubMed: 9141133]
229. Börjesson U, Hünenberger PH. *J Chem Phys* 2001;114:9706.
230. Burgi R, Kollman PA, Van Gunsteren WF. *Proteins: Struct, Funct Genet* 2002;47:469. [PubMed: 12001225]
231. Baptista AM, Teixeira VH, Soares CM. *J Chem Phys* 2002;117:4184.
232. Walczak AM, Antosiewicz JM. *Phys Rev E* 2002;66:051911.

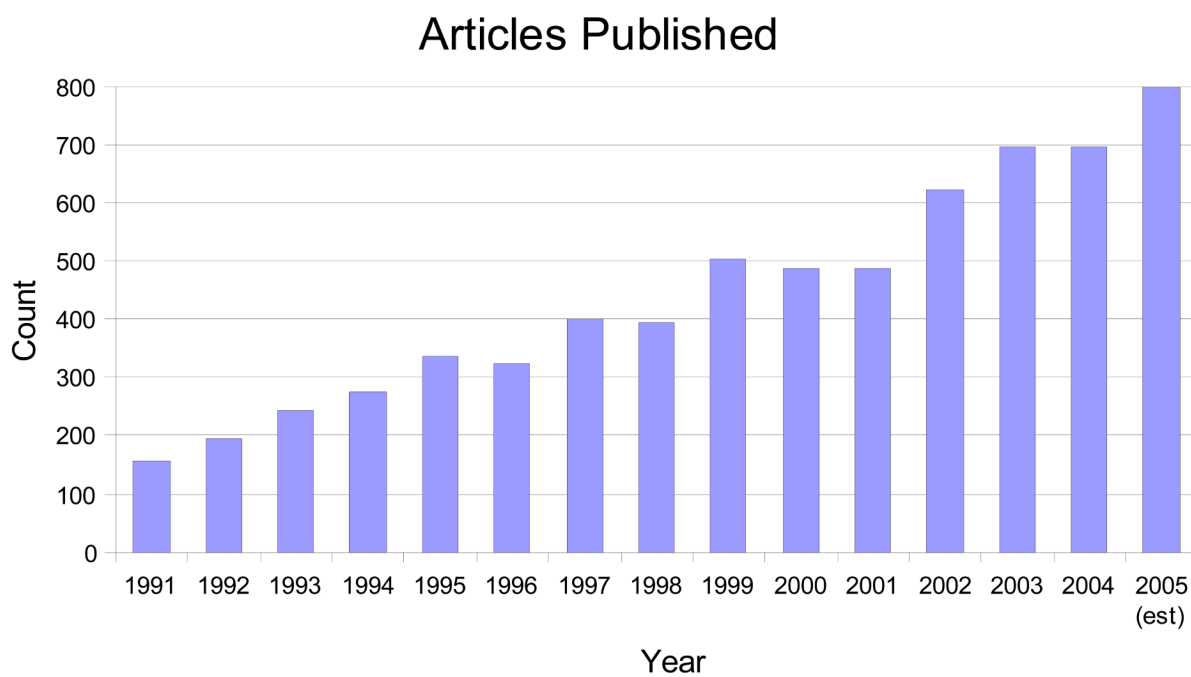
233. Mongan J, Case DA. *Curr Opin Struct Biol* 2005;15:157. [PubMed: 15837173]
234. Mongan J, Case DA, McCammon JA. *J Comput Chem* 2004;25:2038. [PubMed: 15481090]
235. Berne BJ, Straub JE. *Curr Opin Struct Biol* 1997;7:181. [PubMed: 9094324]
236. Elber R. *Curr Opin Struct Biol* 2005;15:151. [PubMed: 15837172]
237. Crippen GM, Scheraga H. *Proc Natl Acad Sci, USA* 1969;64:42. [PubMed: 5263023]
238. Grubmuller H. *Phys Rev E* 1995;52:2893.
239. Schulze BG, Grubmuller H, Evanseck JD. *J Am Chem Soc* 2000;122:8700.
240. Torrie GM, Valleau JP. *J Comput Phys* 1977;23:187.
241. Huber T, Torda AE, Van Gunsteren WF. *J Comput-Aided Mol Des* 1994;8:695. [PubMed: 7738605]
242. Steiner MM, Genilloud PA, Wilkins JW. *Phys Rev B* 1998;57:10236.
243. Rahman JA, Tully JC. *J Chem Phys* 2002;116:8750.
244. Rahman JA, Tully JC. *Chem Phys* 2002;285:277.
245. Voter AF. *Phys Rev Lett* 1997;78:3908.
246. Voter AF. *J Chem Phys* 1997;106:4665.
247. Hamelberg D, Mongan J, McCammon JA. *J Chem Phys* 2004;120:11919. [PubMed: 15268227]
248. Piela L, Kostrowicki J, Scheraga H. *J Phys Chem* 1989;93:3339.
249. Kostrowicki J, Piela L, Cherayil B, Scheraga H. *J Phys Chem* 1991;95:4113.
250. Wawak R, Wimmer M, Scheraga H. *J Phys Chem* 1992;96:5138.
251. Pappu RV, Hart RK, Ponder JW. *J Phys Chem B* 1998;102:9725.
252. Brucoleri RE, Karplus M. *Biopolymers* 1990;29:1847. [PubMed: 2207289]
253. Czerminski R, Elber R. *Proteins: Struct, Funct Genet* 1991;10:70. [PubMed: 2062829]
254. Ulitsky A, Elber R. *J Phys Chem* 1994;98:1034.
255. Simmerling C, Miller JL, Kollman PA. *J Am Chem Soc* 1998;120:7149.
256. Huber T, van Gunsteren WF. *J Phys Chem A* 1998;102:5937.
257. Mitsutake A, Sugita Y, Okamoto Y. *Biopolymers* 2001;60:96. [PubMed: 11455545]
258. Sanbonmatsu KJ, Garcia AE. *Proteins: Struct, Funct Genet* 2002;46:225. [PubMed: 11807951]
259. Frantz DD, Freeman DL, Doll JD. *J Chem Phys* 1990;93:2769.
260. Hansmann UH. *Chem Phys Lett* 1997;281:140.
261. Wu XW, Wang SM. *J Phys Chem B* 1998;102:7238.
262. Faradjian AK, Elber R. *J Chem Phys* 2004;120:10880. [PubMed: 15268118]
263. Moroni D, Bolhuis PG, van Erp TS. *J Chem Phys* 2004;120:4055. [PubMed: 15268572]
264. Kamiya N, Hugo J. *J Comput Chem* 2001;22:1098.
265. Crippen GM. *J Comput Chem* 1982;3:471.
266. Van Schaik R, Berendsen H, Torda A, van Gunsteren W. *J Mol Biol* 1993;234:751. [PubMed: 8254671]
267. Krivov SV, Chekmarev SF, Karplus M. *Phys Rev Lett* 2002;88:038101. [PubMed: 11801090]
268. Andricioaei I, Straub JE. *J Chem Phys* 1997;107:9117.
269. Verkhivker G, Elber R, Nowak W. *J Phys Chem* 1992;97:7838.
270. Simmerling C, Fox T, Kollman PA. *J Am Chem Soc* 1998;120:5771.
271. Garel T, Niel JC, Orland H, Smith J, Velikson B. *J Chim Phys Phys-Chim Biol* 1991;88:2479.
272. Luo ZW, Butcher DJ, Huang ZW. *Protein Eng* 1997;10:1039. [PubMed: 9464567]
273. Keasar C, Elber R. *J Phys Chem* 1995;99:11550.
274. Keasar C, Elber R, Skolnick J. *Folding Des* 1997;2:247.
275. Stultz CM, Karplus M. *Proteins: Struct, Funct Genet* 1999;37:512. [PubMed: 10651268]
276. Shirts MR, Pande VS. *Phys Rev Lett* 2001;86:4983. [PubMed: 11384401]
277. Huber GA, McCammon JA. *Phys Rev E* 1997;55:4822.
278. Snow CD, Nguyen VS, Pande M, Gruebele M. *Nature* 2002;420:102. [PubMed: 12422224]
279. Miranker A, Karplus M. *Proteins: Struct, Funct Genet* 1991;11:29. [PubMed: 1961699]
280. Joseph-McCarthy D, Alvarez JC. *Proteins: Struct, Funct Genet* 2003;51:189. [PubMed: 12660988]

281. Sugita Y, Okamoto Y. *Chem Phys Lett* 1999;314:141.
282. Zhou R, Berne BJ, Germain R. *Proc Natl Acad Sci, USA* 2001;98:14931. [PubMed: 11752441]
283. Fieg M, MacKerell AD Jr, Brooks CL III. *J Phys Chem B* 1999;107:2831.
284. Rao F, Caflisch A. *J Chem Phys* 2003;119:4025.
285. Wu X, Wang S. *J Chem Phys* 1999;110:9401.
286. Shinoda W, Mikami M. *Chem Phys Lett* 2001;335:265.
287. Lahiri A, Nilsson L, Laaksonen A. *J Chem Phys* 2001;114:5993.
288. Andricioaei I, Dinner AR, Karplus M. *J Chem Phys* 2003;118:1074.
289. Chun HM, Padilla CE, Chin DN, Watanabe M, Karlov VI, Alper HE, Soosaar K, Blair KB, Becker OM, Caves LSD, Nagle R, Haney DN, Farmer BL. *J Comput Chem* 2000;21:159.
290. Tuckerman ME, Martyna GJ, Berne BJ. *J Chem Phys* 1990;93:1287.
291. Space B, Rabitz H, Askar A. *J Chem Phys* 1993;99:9070.
292. Nadler W, Schulten K. *J Chem Phys* 1985;82:151.
293. Tirion MM. *Phys Rev Lett* 1996;77:1905. [PubMed: 10063201]
294. Tozzini V. *Curr Opin Struct Biol* 2005;15:144. [PubMed: 15837171]
295. Kundu S, Sorensen DC, Phillips GN Jr. *Proteins: Struct, Funct Genet* 2004;57:726.
296. Kostov KS, Freed KF. *J Chem Phys* 1997;106:771.
297. Dauber-Osguthorpe P, Maunder CM, Osguthorpe DJ. *J Comput-Aided Mol Des* 1996;10:177. [PubMed: 8808735]
298. Phillips SC, Essex JW, Edge CM. *J Chem Phys* 2000;112:2586.
299. Phillips SC, Swain MT, Wiley AP, Essex JC, Edge CM. *J Chem Phys* 2003;107:2098.
300. Elber R, Meller J, Olender R. *J Phys Chem B* 1999;103:899.
301. Elber R, Ghosh A, Cardenas A. *Acc Chem Res* 2002;35:396. [PubMed: 12069624]
302. Kleinjung J, Bayley P, Fraternali F. *FEBS Lett* 2000;470:257. [PubMed: 10745078]
303. Bahar I, Atilgan AR, Erman B. *Folding Des* 1997;2:173.
304. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. *Biophys J* 2001;80:505. [PubMed: 11159421]
305. Doruker P, Jernigan RL, Bahar I. *J Comput Chem* 2002;23:119. [PubMed: 11913377]
306. Isin B, Doruker P, Bahar I. *Biophys J* 2002;82:569. [PubMed: 11806902]
307. Bahar I, Jernigan RL. *Biochemistry* 1999;38:3478. [PubMed: 10090734]
308. Bahar I, Erman B, Jernigan RL, Atilgan AR, Covell DG. *J Mol Biol* 1999;285:1023. [PubMed: 9887265]
309. Ming D, Kong Y, Wakil SJ, Brink J, Ma J. *Proc Natl Acad Sci, USA* 2002;99:7895. [PubMed: 12060737]
310. Zhang Z, Shi Y, Liu Y. *Biophys J* 2003;84:3583. [PubMed: 12770868]
311. He J, Zhang Z, Shi Y, Liu H. *J Chem Phys* 2003;119:4005.
312. Sessions RB, Dauber-Osguthorpe P, Osguthorpe DJ. *J Mol Biol* 1989;210:617. [PubMed: 2614836]
313. Schlick T. *Structure* 2001;9:R45. [PubMed: 11525172]
314. Marchi M, Procacci P. *J Chem Phys* 1998;109:5194.
315. Samuelson S, Martyna GJ. *J Phys Chem B* 1998;103:1752.
316. Zhou R, Harder E, Xu H, Berne BJ. *J Chem Phys* 2001;115:2348.
317. Grubmuller H, Tavan P. *J Comput Chem* 1998;19:1534.
318. Cheng A, Merz KM Jr. *J Phys Chem B* 1999;103:5396.
319. Eastman P, Doniach S. *Proteins: Struct, Funct Genet* 1998;30:215. [PubMed: 9517537]
320. Elezgaray J, Marcou G, Sanejouand YH. *Phys Rev E* 2002;66:031908.Part 1
321. Straatsma TP, Philippopoulos M, McCammon JA. *Comput Phys Commun* 2000;128:377.
322. Straatsma TP, McCammon JA. *IBM Syst J* 2001;40:328.
323. Berendsen HJC, van der Spoel D, van Drunen R. *Comput Phys Commun* 1995;91:43.
324. Lindahl E, Hess B, van der Spoel D. *J Mol Modelling* 2001;7:306.

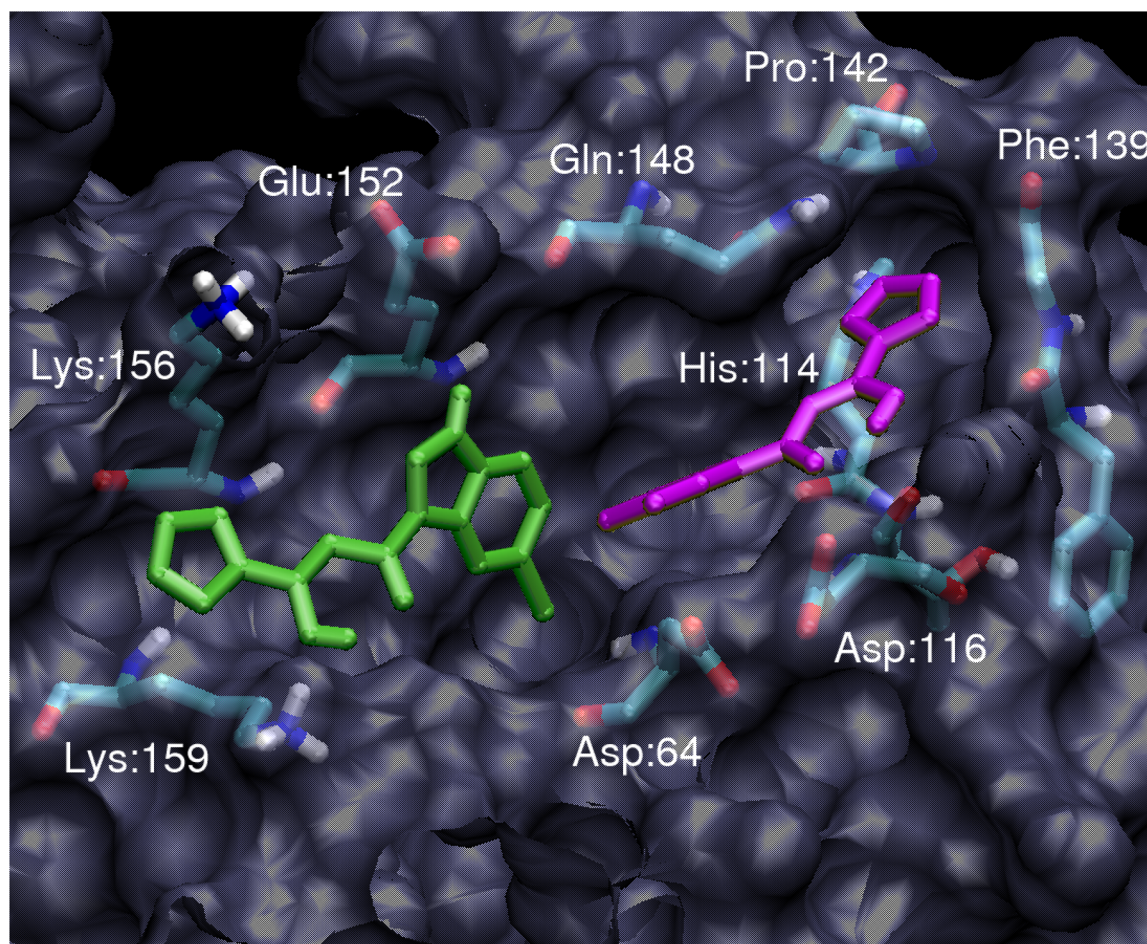


325. Scott WRP, Hunenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Kruger P, van Gunsteren WF. *J Phys Chem A* 1999;103:3596.
326. Kalé L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. *J Comput Phys* 1999;151:283.
327. Kalé LV, Kumar S, Zheng GB, Lee CW. *Computer Sciences ICCS 2003, PT IV, Proceedings lecture notes in computer science*. 2003;2660:23.
328. Balaeff A, Mahadevan L, Schulten K. *Struct* 2004;12:123.
329. Phillips JC, Zheng G, Kumar S, Kalé LV. *NAMD: Biomolecular Simulation on Thousands of Processors. Proceedings of SC2002*. 2002
330. Duke, RE.; Pedersen, LG. *PMEMD 3.1*. University of North Carolinas Chapel Hill; Chapel Hill, NC: 2003.
331. Stone, J.; Gullingsrud, J.; Grayson, P.; Schulten, K. 2001 ACM Symposium on InteractiVe 3D Graphics. Hughes, JF.; Sequin, CH., editors. ACM SIGGRAPH; New York: 2001. p. 191
332. Nelson MT, Humphrey W, Gursoy A, Dalke A, Kale LV, Skeel RD, Schulten K. *Int J Supercomput Appl High Perform Comput* 1996;10:251.
333. Pearlman DA, Case DA, Caldwell JW, Ross WR, Cheatham TE III, DeBolt S, Ferguson D, Seibel G, Kollman PA. *Comput Phys Commun* 1995;91:1.
334. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM Jr, Onufriev A, Simmerling C, Wang B, Woods R. *J Comput Chem* 2005;26:1668. [PubMed: 16200636]
335. Cheatham TE, Young MA. *Biopolymers* 2001;56:232. [PubMed: 11754338]
336. Bhanot G, Chen D, Gara A, Vranas P. *Nucl Phys B-Proceed Suppl* 2003;119:114.
337. Xu A, Horwich AL, Sigler PB. *Nature* 1997;388:741. [PubMed: 9285585]
338. Ma JP, Sigler PB, Xu ZH, Karplus M. *J Mol Biol* 2000;302:303. [PubMed: 10970735]
339. Ranson NA, Farr GW, Roseman AM, Gower B, Fenton WA, Horwich HL, Saibil HR. *Cell* 2001;107:869. [PubMed: 11779463]
340. Sansom MSP, Shrivastava IH, Ranatunga KM, Smith GR. *Trends Biochem Sci* 2000;25:368. [PubMed: 10916155]
341. Sansom MSP, Shrivastava IH, Bright JN, Tate J, Capener CE, Biggin PC. *Biochim Biophys Acta-Biomembr* 2002;1565:294.
342. Roux B, Allen T, Berneche S, Im W. *Q Rev Biophys* 2004;37:15. [PubMed: 17390604]
343. Zhong QF, Husslein T, Moore PB, Newns DM, Pattnaik P, Klein MI. *FEBS Lett* 1998;434:265. [PubMed: 9742936]
344. Newns D, Zhong Q, Moore P, Husslein T, Pattnaik P, Klein M. *Parallel Comput* 2000;26:965.
345. Elcock AH, Sept D, McCammon JA. *J Phys Chem B* 2001;105:1504.
346. Brooijmans N, Kuntz ID. *Ann Rev Biophys Biomol Struct* 2003;32:335. [PubMed: 12574069]
347. Halperin I, Ma BY, Wolfson H, Nussinov R. *Proteins: Struct, Funct Genet* 2002;47:409. [PubMed: 12001221]
348. Wong CF, McCammon JA. *Ann Rev Pharmacol Toxicol* 2003;43:31. [PubMed: 12142469]
349. Teodoro ML, Kaviraki LE. *Curr Pharm Des* 2003;20:1635. [PubMed: 12871062]
350. Carlson HA. *Curr Pharm Des* 2002;8:1571. [PubMed: 12052201]
351. Carlson HA, McCammon JA. *Mol Pharmacol* 2000;57:213. [PubMed: 10648630]
352. Kua J, Zhang YK, McCammon JA. *J Am Chem Soc* 2002;124:8260. [PubMed: 12105904]
353. Kua J, Zhang YK, Eslami AC, Butler JR, McCammon JA. *Protein Sci* 2003;12:2675. [PubMed: 14627729]
354. Nakajima N, Higo J, Kidera A, Nakamura H. *Chem Phys* 1997;278:297.
355. Pak Y, Wang S. *J Phys Chem B* 2000;104:354.
356. Di Nola A, Roccatano D, Berendsen HJC. *Proteins: Struct, Funct Genet* 1994;19:174. [PubMed: 7937732]
357. Mangoni M, Roccatano D, Di Nola A. *Proteins: Struct, Funct Genet* 1999;35:153. [PubMed: 10223288]
358. Goodsell DS, Morris GM, Olson AJ. *J Mol Recognit* 1996;9:1. [PubMed: 8723313]

359. Vieth M, Hirst JD, Dominy BN, Daigler H, Brooks CL. *J Comput Chem* 1998;19:1623.
360. Vieth M, Hirst JD, Kolinski A, Brooks CL. *J Comput Chem* 1998;19:1612.
361. Carlson HA, Masukawa KM, McCammon JA. *J Phys Chem* 1999;103:10213.
362. Lin JH, Perryman AL, Schames JR, McCammon JA. *J Am Chem Soc* 2002;124:5632. [PubMed: 12010024]
363. Lin JH, Perryman AL, Schames JR, McCammon JA. *Biopolymers* 2003;68:47. [PubMed: 12579579]
364. Varady J, Wu X, Wang S. *J Phys Chem B* 2002;106:4863.
365. Broughton HB. *J Mol Graphics Modelling* 2000;18:247.
366. Fukunishi Y, Mikami Y, Nakamura H. *J Phys Chem B* 2003;107:13201.
367. Zacharias M. *Proteins: Struct, Funct Genet* 2004;54:759. [PubMed: 14997571]
368. Wang J, Kollman PA, Kuntz ID. *Proteins: Struct, Funct Genet* 1999;36:1. [PubMed: 10373002]
369. Beierlein F, Lanig H, Schurer G, Horn AHC, Clark T. *Mol Phys* 2003;101:2469.
370. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. *J Comput Chem* 1998;19:1639.
371. Miranker A, Karplus M. *Proteins: Struct, Funct Genet* 1991;11:29. [PubMed: 1961699]
372. Das B, Helms V, Lounnas V, Wade RC. *J Inorg Biochem* 2000;81:121. [PubMed: 11051557]
373. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, Lins RD, Briggs JM, McCammon JA. *J Med Chem* 2000;43:2100. [PubMed: 10841789]
374. Jorgensen, WL. BOSS. Yale University; New Haven, CT: 1997.
375. Mustata GI, Briggs JM. *J Comput-Aided Mol Des* 2002;16:935. [PubMed: 12825624]
376. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. *Science* 1996;274:1531. [PubMed: 8929414]
377. Erlanson DA, Braisted AC, Raphael DR, Randal M, Stroud RM, Gorden EM, Wells JA. *Proc Natl Acad Sci, USA* 2000;97:9367. [PubMed: 10944209]
378. Guo ZY, Brooks CL III. *J Am Chem Soc* 1998;120:1920.
379. Hodge, CN.; Straatsma, TP.; McCammon, JA.; Wlodawer, A. *Structural Biology of Viruses*. In: Chiu, W.; Burnett, RM.; Garcea, R., editors. *Rational Design of HIV Protease Inhibitors*. Oxford University Press; Oxford: 1997. p. 451-473.
380. Hazuda DJ, Anthony NJ, Gomez RP, Jolly SM, Wai JS, Zhuang L, Fisher TE, Embrey M, Guare JP Jr, Egbertson MS, Vacca JP, Huff JR, Felock PJ, Witmer MV, Stillmock KA, Danovich R, Grobler J, Miller MD, Espeseth AS, Jin L, Chen IW, Lin JH, Kassahun K, Ellis JD, Wong BL, Xu W, Pearson PG, Schleif WA, Cortese R, Emini E, Summa V, Holloway MK, Young SD. *Proc Natl Acad Sci, USA* 2004;101:11233. [PubMed: 15277684]
381. Barreca ML, Lee KW, Chimirri A, Briggs JM. *Biophys J* 2003;84:1450. [PubMed: 12609852]
382. Ni H, Sotriffer CA, McCammon JA. *J Med Chem* 2001;44:3043. [PubMed: 11543671]
383. Lins RD, Briggs JM, Straatsma TP, Carlson HA, Greenwald J, Choe S, McCammon JA. *Biophys J* 1999;76:2999. [PubMed: 10354426]
384. Schames JR, Henchman RH, Siegel JS, Sotriffer CA, Ni H, McCammon JA. *J Medicinal Chem* 2004;47:1879.
385. Palmer RG. *Rev Phys* 1982;31:669.
386. Heinz TN, Hunenberger PH. *J Comput Chem* 2004;25:1474. [PubMed: 15224391]
387. Im W, Beglov D, Roux B. *Comput Phys Commun* 1998;111:59.
388. Luo R, David L, Gilson MK. *J Comput Chem* 2002;23:1244. [PubMed: 12210150]
389. Gilson MK, Davies ME, Luty BA, McCammon JA. *J Phys Chem* 1993;97:3591.
390. Baker NA. *Methods Enzymol* 2004;383:94. [PubMed: 15063648]
391. Tai K, Murdock S, Wu B, Ng HH, Johnston S, Fangohr H, Cox SJ, Jeffreys P, Essex JW, Sansom MSP. *Org Biomol Chem* 2004;2:3219. [PubMed: 15534698]
392. Goldgur Y, Craigie R, Cohen GH, Fujiwara T, Yoshinaga T, Fujishita T, Sugimoto H, Endo T, Murai H, Davies DR. *Proc Natl Acad Sci, USA* 1999;96:13040. [PubMed: 10557269]



**Figure 1.** Articles matching ISI Web of Science query “TS=(protein) AND TS=(molecular dynamics)”.



**Figure 2.** Two predicted binding conformations of a HIV-1 integrase inhibitor, 5CITEP, to a MD snapshot of the protein. The green conformation is similar to that in the crystal structure<sup>392</sup> and the purple is in a secondary predicted binding trench.

**Table 1**

## Characteristic Time Scales for Protein Motions

event	spatial extent (nm)	amplitude (nm)	time (s)	appropriate simulations
bond-length vibration	0.2–0.5	0.001–0.01	$10^{-14}$ – $10^{-13}$	QM methods
elastic vibration of globular domain	1.0–2.0	0.005–0.05	$10^{-12}$ – $10^{-11}$	conventional MD
rotation of solvent-exposed side chains	0.5–1.0	0.5–1.0	$10^{-11}$ – $10^{-10}$	conventional MD
torsional libration of buried groups	0.5–1.0	0.05	$10^{-11}$ – $10^{-9}$	conventional MD
hinge bending (relative motion of globular domains)	1.0–2.0	0.1–0.5	$10^{-11}$ – $10^{-7}$	Langevin dynamics, enhanced sampling MD methods?
rotation of buried side chains	0.5	0.5	$10^{-4}$ –1	enhanced sampling MD methods?
allosteric transitions	0.5–4.0	0.1–0.5	$10^{-5}$ –1	enhanced sampling MD methods?
local denaturation	0.5–1.0	0.5–1.0	$10^{-5}$ – $10^1$	enhanced sampling MD methods?
loop motions	1.0–5.0	1.0–5.0	$10^{-9}$ – $10^{-5}$	Brownian dynamics?
rigid-body (helix) motions		1.0–5.0	$10^{-9}$ – $10^{-6}$	enhanced sampling MD methods?
helix-coil transitions		>5.0	$10^{-7}$ – $10^4$	enhanced sampling MD methods?
protein association	>>1.0			Brownian dynamics