A Test Set for Molecular Dynamics Algorithms

Eric Barth[1], Benedict Leimkuhler[2], and Sebastian Reich[3]

[1] Department of Mathematics and Computer Science, Kalamazoo College, Kalamazoo, MI 49006, U.S.A. (barth@kzoo.edu).
[2] Department of Mathematics and Computer Science, University of Leicester, Leicester, LE1 7RH, UK (bl12@mcs.le.ac.uk).
[3] Department of Mathematics, Imperial College, London, SW7 2BZ, UK (s.reich@ic.ac.uk).

Abstract. This article describes a collection of model problems for aiding numerical analysts, code developers and others in the design of computational methods for molecular dynamics (MD) simulation. Common types of calculations and desirable features of algorithms are surveyed, and these are used to guide selection of representative models. By including essential features of certain classes of molecular systems, but otherwise limiting the physical and quantitative details, it is hoped that the test set can help to facilitate cross-disciplinary algorithm and code development efforts.

1 Introduction and Background

Over the past two decades, computational scientists have turned increasing attention to the field of molecular modeling. Important advances have been made in the design of efficient algorithms for problems such as fast summation methods for computing non-bonded atomic interactions [1,2], long-time numerical integration of equations of motion[3–5], non-Newtonian dynamical formulations for simulation in a variety of statistical mechanical ensembles [6–8], and optimization algorithms. These projects have often involved mathematicians and/or computer scientists who, though adept at algorithm and software development, may possess limited or no physical and chemical knowledge.

An important step in the development of a computational method is testing on model problems which possess characteristic properties yet are as simple as possible so that numerical experiments are not too time consuming. The purpose is generally to assess the performance of a particular method in connection with certain targeted model features. As increasingly complex computational approaches appear in the literature, it is becoming evident that the choice of proper test systems can greatly ease code comparisons and algorithm development.

Compilations of test problems have become well established in other fields, for example in optimization [9–11] and differential equations [12,13]. The CASP project [14] (Critical Assessment of techniques for protein Structure Prediction) provides a testbed for objective testing and comparison of methods for identifying protein structure from sequence. To our knowledge, no similar collection exists for design and evaluation of molecular dynamics methods.

We have assembled here a few elementary models drawn primarily from the molecular simulation literature. Our goal is to provide adequate detail so that an applied mathematician, with little access to MD codes and virtually no chemical/physical understanding, could—based only this paper—perform simulations that resemble, in relevant ways, the more complex problems of current research interest. Our first step is to introduce a basic vocabulary for molecular dynamics and to recall some of the standard types of calculations being undertaken in the scientific literature. We then focus on certain types of problems, including a Lennard-Jones liquid, a water model, and some small conformational models. In each case, we elucidate the role of the problem in the literature and any special features that it illustrates. The article is self-contained, in the sense that all mathematical details and parameters are provided either in the paper or on the test set web site. The ongoing progress of the test set project, with further models and accompanying software, will be reported on the internet at the web site `http://www.mmc.le.ac.uk/mdt`.

1.1 Common features of molecular dynamics models

For model problems to be useful, they must possess characteristic properties representative of the challenges encountered in molecular modeling. Here we begin by briefly discussing the common (classical) molecular dynamics model and the difficulties such a model presents for simulation.

Purely collisional "hard sphere" systems were the first to be studied by molecular simulation [15], and work has continued in this area, for example extending to hard ellipsoids and to hard sphere systems in soft potentials [16–19]. In this preliminary version of the test set, we concern ourselves only with smooth potential models such as the monatomic fluids which were first discussed by Rahman [20] and Verlet [21]. These early computer experiments demonstrated that the equilibrium state of argon could be well described by the two-body Lennard-Jones potential

$$V(r) = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) \tag{1.1}$$

with r being the distance between a pair of particles, ϵ giving well depth, and σ being proportional to the equilibrium atomic separation. A representative Lennard-Jones potential is shown in Fig. 1. N particles are arranged in a cubic box with edges of length L, with periodic boundary conditions imposed so that at any instant a particle with coordinates x, y, z inside the real box is associated to an infinite number of periodic images with coordinates obtained by adding or subtracting multiples of L from each coordinate. The Cartesian positions $\mathbf{r}_i = [x_i\ y_i\ z_i]^T$ of the N particles can be assembled into a collective

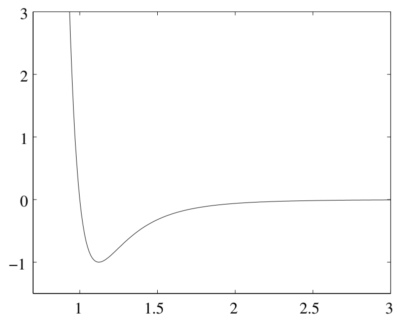


Fig. 1. The Lennard-Jones potential with $\sigma = \epsilon = 1$.

Cartesian position vector

$$\begin{aligned}\mathbf{r} &= [x_1\ y_1\ z_1\ x_2\ y_2\ z_2\ \ldots\ x_N\ y_N\ z_N]^T \\ &= [\mathbf{r}_1^T\ \mathbf{r}_2^T\ \ \ldots\ \mathbf{r}_N^T]^T,\end{aligned}$$

where the superscript T denotes a matrix transpose. The potential energy of the Lennard-Jones system is

$$V(\mathbf{r}) = \sum_{i<j}^{N} 4\epsilon \left(\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6} \right),$$

where the sum is taken over pairs of atoms (i, j) with separations defined by

$$r_{ij} = ||\mathbf{r}_i - \mathbf{r}_j||_2.$$

The Newtonian equations of motion can then be expressed as

$$\mathbf{M}\ddot{\mathbf{r}} = \mathbf{F}(\mathbf{r}) := -\nabla_{\mathbf{r}} V(\mathbf{r}),$$

where $\mathbf{M}$ is a diagonal mass matrix with diagonal

$$[m_1\ m_1\ m_1\ m_2\ m_2\ m_2\ \ldots\ m_N\ m_N\ m_N],$$

and m_i the mass of the ith particle. The gradient $\nabla_{\mathbf{r}} V$ is the column vector of all partial derivatives with respect to particle positions. It is easily verified that the total energy

$$E = \frac{\dot{\mathbf{r}}^T \mathbf{M} \dot{\mathbf{r}}}{2} + V(\mathbf{r})$$

is an integral of motion: $dE/dt = 0$. An alternative description of the system is obtained by expressing the energy as a Hamiltonian function H of positions $\mathbf{q}$ and momenta $\mathbf{p}$, where $\mathbf{p} = \mathbf{M}\dot{\mathbf{q}}$; the equations of motion then become

$$\begin{aligned}\dot{\mathbf{q}} &= \nabla_{\mathbf{p}} H(\mathbf{q}, \mathbf{p}) \\ \dot{\mathbf{p}} &= -\nabla_{\mathbf{q}} H(\mathbf{q}, \mathbf{p}).\end{aligned}$$

The system is typically simulated from given positions and velocities $\mathbf{r}(t_0) = \mathbf{r}_0, \dot{\mathbf{r}}(t_0) = \dot{\mathbf{r}}_0$ often chosen randomly in accordance with some appropriate statistical ensemble.

There are number of computational challenges to be addressed in the simulation of a Lennard-Jones fluid. First, the bulk of the computational work is spent evaluating the forces between the $\frac{1}{2}N(N-1)$ pairs of atoms, even though the interaction between many (distant) pairs is very weak. This difficulty was remedied by Verlet by imposing *distance cutoffs* by which the potential was approximated by zero for all atomic separations greater than a certain cutoff value r_c (2.5σ–3.3σ). The now-famous Verlet table was introduced, by which all distances are occasionally computed, and those pairs within distance r_M, $r_M > r_c$, are recorded in a table. Successive force evaluations include only the entries in the table. Provided that the "skin" $r_M - r_c$ is sufficiently thick, no particle pairs can move undetected into the cutoff range between table updates.

A second issue is the choice of time discretization scheme. The method proposed by Verlet propagated positions by

$$\mathbf{r}_i^{n+1} = -\mathbf{r}_i^{n-1} + 2\mathbf{r}_i^n + h^2\mathbf{F}_i^n, \tag{1.2}$$

and velocities using

$$\mathbf{v}_i^n = \left[\mathbf{r}_i^{n+1} - \mathbf{r}_i^{n-1}\right]/2h. \tag{1.3}$$

Here the superscripts denote the indices of timesteps, each of which is of size h, so

$$\mathbf{r}_i^n \approx \mathbf{r}_i(t_0 + nh),$$

and $\mathbf{F}_i^n = -\nabla_{\mathbf{r}_i} V(\mathbf{r}^n)$ is a Cartesian vector which gives the sum of forces acting on particle i due to interaction with all other particles (within the cutoff range), evaluated at the point $\mathbf{r}^n$.

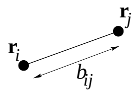
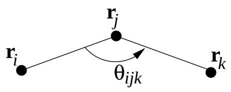
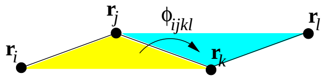
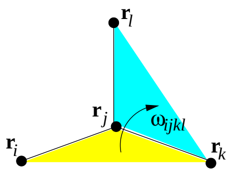
In his ground-breaking paper, Verlet reported "...small irregularities in the total energy...but the error is of no consequence." The discretization method (1.2)–(1.3) is obviously *time-reversible*, which has sometimes been mentioned as the reason for its excellent stability. In fact, though, it is now known that it is rather the *symplectic* conservation property [22] of the Verlet method, viewed as an appropriate mapping of positions and momenta, that confers its excellent long-term energy stability [23–25]. The Verlet method is now regarded as the gold standard for timestepping schemes in molecular dynamics.

In the 1970s, interest developed in applying molecular dynamics methods to more complicated molecular fluids such as water [26], molecular fluids with internal degrees of freedom [27] and large flexible molecules [28]. The MD pair potential requires many modifications in these cases. The interaction between polyatomic molecules is modeled by pair potentials, both Lennard-Jones and coulombic, between *all constituent atoms* as opposed to general molecule-molecule interactions. Also, the model potential must maintain intramolecular geometries by including the "bonded" terms: bond lengths, bond angles, and dihedral angles. The result is the molecular modeling potential function which generally is of the form:

$$V(\mathbf{r}) = V_b + V_a + V_d + V_i + V_{LJ} + V_C, \tag{1.4}$$

where V_b, V_a, V_d and V_i represent bonds, angles, dihedral angles (torsions), and improper torsions, respectively. These consist of sums over various pairs, triples and quadruples of spatially localized bonded groups of atoms. V_{LJ} is the sum of Lennard-Jones contributions for all pairs of atoms, and V_C is the Coulombic potential due to charge-charge interactions. Representative terms of each sum are shown in Table 1. The various potential terms are described in more detail in section 3.

Table 1. Functional forms for the terms in the molecular dynamics potential .

bonded terms		
V_b		$\frac{\kappa_b}{2}(\\|\mathbf{r}_i - \mathbf{r}_j\\| - b_{ij}^{\mathrm{eq}})^2$
V_a		$\frac{\kappa_\theta}{2}(\theta(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) - \theta_{ijk}^{\mathrm{eq}})^2$
V_d		$\kappa_\phi[1 - \cos(n_\phi(\phi(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k, \mathbf{r}_l) - \phi_{ijkl}^{\mathrm{eq}}))]$
V_i		$\frac{\kappa_\omega}{2}(\omega(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k, \mathbf{r}_l) - \omega_{ijkl}^{\mathrm{eq}})^2$
nonbonded terms		
V_{LJ}		$4\epsilon\left(\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right)$
V_C		$\frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$

While Cartesian atomic coordinates are generally used, work on internal coordinate representations has been ongoing [29–31]. The many parameters

κ_b, κ_θ, κ_ϕ, ..., and equilibrium values b^{eq}, θ^{eq}, ϕ^{eq}, ..., in table 1 are determined empirically for interactions involving each possible combination of atoms. Many potential parameterizations are available, including CHARMM [32,33], AMBER [34,35] and OPLS [36]. A list of references to these and other MD potentials can be found in [37].

The MD potential is highly nonlinear, with many local minima. Minimization of the potential energy is a common task, but the nonpolynomial proliferation of local minima frustrates attempts to determine lowest energy states for modeled systems [38]. Also, the finite-time dynamics of a nonlinear multiple-minima system can become trapped in one potential energy well, which impedes conformational sampling. The terms in the potential represent interactions on a wide range of spatial scales (from bonds of length 1 Å$=10^{-10}$ m, to coulombic interactions which extend throughout the modeled system) and timescales (the fastest bonds have a period of 10 fs$=10^{-14}$ s, while large scale conformational interconversions may occur on the scale of seconds). Timestepping algorithms such as the Verlet method (1.2)–(1.3) require a timestep which is sufficiently short (0.5–1.0 fs) to resolve the fastest bonded motion, meaning that a computed trajectory which spans a time interval of one nanosecond (10^{-9} s) requires one million dynamics steps. As with the earliest molecular dynamics simulations, the great majority of the computational work is expended in computing the forces of interaction. Unlike the Lennard-Jones case, where the forces decay very rapidly at moderate atomic separations, the Coulombic $1/r^2$ forces are non-negligible even at large separations, making distance cutoffs unphysical and undesirable. Intense activity is ongoing on the problems of efficient timestepping and fast evaluation of non-bonded forces without distance cutoffs.

The MD potential function has several characteristics which have an impact on the performance of numerical methods: multiple minima, wide range of time and space scales, and long-range interactions between many particles. It is these features which the model problems presented below are designed to capture.

2 Newtonian and Non-Newtonian Dynamics

One can view the simplest "molecular system" as consisting of two atoms with masses m_1 and m_2 and some pairwise interaction potential $V(r)$,

$$r = ||\mathbf{r}_1 - \mathbf{r}_2||_2 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

The associated Newtonian equations of motion are easily identified as

$$m_1 \ddot{\mathbf{r}}_1 = -\frac{V'(r)}{r}(\mathbf{r}_1 - \mathbf{r}_2), \qquad m_2 \ddot{\mathbf{r}}_2 = -\frac{V'(r)}{r}(\mathbf{r}_2 - \mathbf{r}_1). \tag{2.5}$$

The interaction potential $V(r)$ can, for example, be taken as the Lennard-Jones potential (1.1) or one of the common models for a chemical bond:

the harmonic bond potential

$$V(r) = \frac{\kappa_b}{2}(r - r_0)^2 \tag{2.6}$$

or the Morse oscillator bond potential [39,40]

$$V(r) = D(1 - \exp(-\alpha(r))). \tag{2.7}$$

In case of (1.1), bounded motion is observed if the total energy

$$E = \frac{m_1}{2}||\dot{\mathbf{r}}_1||^2 + \frac{m_2}{2}||\dot{\mathbf{r}}_2||^2 + V(\|\mathbf{r}_1 - \mathbf{r}_2\|)$$

is negative and collisional dynamics if $E > 0$. Since these are the elementary interactions occurring in any molecular simulation, they provide a first simple test case for any numerical integration method.

With the mass matrix $\mathbf{M}$ defined in the introduction, the Newtonian equations (2.5) are equivalent to the system of first order equations

$$\dot{\mathbf{r}} = \mathbf{v}, \tag{2.8}$$

$$\dot{\mathbf{v}} = -\mathbf{M}^{-1}\nabla V_{\mathbf{r}}(\mathbf{r}). \tag{2.9}$$

This system possesses two fundamental geometric structures: (i) it is Hamiltonian, and (ii) it respects a time-reversal symmetry. The preservation of these geometric structures has been shown to be important for long-term stability of numerical simulations [22,41,42]. In practice, numerical methods which preserve these properties exhibit excellent conservation behavior along computed trajectories.

Note that, besides total energy, linear momentum

$$\mathbf{P} = \sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i$$

and angular momentum

$$\mathbf{L} = \sum_{i=1}^{N} m_i \mathbf{r}_i \times \dot{\mathbf{r}}_i$$

are integrals of motion for a system of N particles (in the absence of boundary conditions). These quantities should be well conserved by integration methods suitable for molecular simulations.

2.1 Constraints

A common modeling strategy in molecular dynamics is to maintain atoms at fixed separations by the use of constraint relations in Cartesian coordinates. This approach can be generalized to freeze other relationships among the variables, as well.

The positions $\mathbf{r}$ and velocities $\mathbf{v}$ of the system evolve according to the constrained Euler-Lagrange equations,

$$\dot{\mathbf{r}} = \mathbf{v}, \tag{2.10}$$
$$\mathbf{M}\dot{\mathbf{v}} = -\nabla V_{\mathbf{r}}(\mathbf{r}) + \mathbf{g}'(\mathbf{r})^T \lambda \tag{2.11}$$
$$\mathbf{0} = \mathbf{g}(\mathbf{r}). \tag{2.12}$$

Here the constraint relations $g_i(\mathbf{r}) = 0$, $i = 1, \ldots, m$, are written compactly as $\mathbf{g}(\mathbf{r}) = \mathbf{0}$, with $\mathbf{g} = (g_1, g_2, \ldots, g_m)^T$ and $\mathbf{g}'$ denotes the matrix of partial derivatives of $\mathbf{g}$ with respect to the atomic coordinates. A popular integrator for (2.10)–(2.12) is the Rattle discretization [43]. For implementation details, see [44,45].

Applied to the diatomic system (2.5), the constraint in (2.12) which maintains the two atoms at a separation of L_{eq} is given by

$$g(\mathbf{r}) = \frac{1}{2}\left((x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 - L_{\text{eq}}^2\right),$$

and the matrix of partial derivatives is given by

$$g'(\mathbf{r}) = [x_1 - x_2,\ \ y_1 - y_2,\ \ z_1 - z_2,\ \ x_2 - x_1,\ \ y_2 - y_1,\ \ z_2 - z_1]^T.$$

The dynamics of the constrained diatomic system consists entirely of rigid body rotation and translation, because all interatomic motion has been eliminated by the constraints.

An extended chain of N atoms connected by constraints, shown in Fig. 2, is a useful test system. Here the constraints take the form

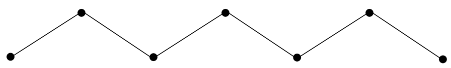


Fig. 2. An extended constraint chain.

$$g_i(\mathbf{r}) = \frac{1}{2}(\|\mathbf{r}_i - \mathbf{r}_{i+1}\|^2 - L_i^2) = 0$$

for $i = 1, \ldots, N-1$. If one end of the chain is to be fixed at position L_0, the additional constraint would be $g_0(\mathbf{r}) = \frac{1}{2}(\|\mathbf{r}_0\|^2 - L_0^2) = 0$. Let us number the multipliers $\lambda_0, \lambda_1, \ldots, \lambda_{N-1}$. Then the equations of motion are, for any of the internal nodes of the chain ($i = 2, \ldots, N-1$),

$$\dot{\mathbf{r}}_i = \mathbf{v}_i$$
$$m_i\dot{\mathbf{v}}_i = \mathbf{F}_i - \lambda_{i-1}(\mathbf{r}_i - \mathbf{r}_{i-1}) - \lambda_i(\mathbf{r}_i - \mathbf{r}_{i+1}),$$

whereas the last node obeys

$$\begin{aligned}\dot{\mathbf{r}}_N &= \mathbf{v}_N \\ m_N\dot{\mathbf{v}}_N &= \mathbf{F}_N - \lambda_{N-1}(\mathbf{r}_N - \mathbf{r}_{N-1}).\end{aligned}$$

If the first node is not maintained at a fixed position, the differential equations are

$$\begin{aligned}\dot{\mathbf{r}}_1 &= \mathbf{v}_1 \\ m_1\dot{\mathbf{v}}_1 &= \mathbf{F}_1 - \lambda_1(\mathbf{r}_1 - \mathbf{r}_2).\end{aligned}$$

If the first node is fixed, then we have

$$\begin{aligned}\dot{\mathbf{r}}_1 &= \mathbf{v}_1 \\ m_1\dot{\mathbf{v}}_1 &= \mathbf{F}_1 - \lambda_0\mathbf{r}_1 - \lambda_1(\mathbf{r}_1 - \mathbf{r}_2).\end{aligned}$$

The inter-particle forces $\mathbf{F}_i$ could be chosen to suit a particular modeling objective. With a Lennard-Jones interaction potential, the resulting chain is fully flexible as in certain polymer models. We could impose potentials for angles, dihedrals, and Coulombic interaction of non-bonded pairs, making this a model for a protein chain, which we discuss in section 3.4.

2.2 Dynamics at Constant Temperature and Pressure

Given the small physical dimension of an individual atom, it is not obvious that the simulation of a collection of particles, even many thousands or millions of atoms, can tell us much about the properties of a realistic system. Only when certain assumptions of statistical mechanics are invoked can we extrapolate from the miniscule region of space where our simulation evolves to the aggregate behavior of the much larger system. Moreover, it rarely makes sense to view an MD model as maintaining a constant total energy. The system is instead usually viewed as continually exchanging energy with a larger bath of molecules at a certain prescribed temperature, sometimes modeled stochastically, sometimes dynamically.

Statistical mechanics requires that we first identify an appropriate density or ensemble of states with respect to which all averages are computed. Such a density ρ is a steady state of the Liouville equation

$$\frac{d\rho}{dt} = \{\rho, H\} = \nabla_{\mathbf{r}}\rho \cdot \nabla_{\mathbf{p}}H - \nabla_{\mathbf{p}}\rho \cdot \nabla_{\mathbf{r}}H.$$

A typical assumption made is that ρ be a function of H. If we assume constant energy $H = E$ in a given system of N atoms constrained within a volume V, we refer to the associated distribution of states as the (N, V, E) or *microcanonical* ensemble. The corresponding density of states is

$$\rho = \delta[H - E]$$

meaning that the sampling is conducted with respect to uniformly distributed states on the energy surface $H = E$. The *canonical* or (N, V, T) ensemble has

$$\rho = \exp(-\frac{H}{k_B T}),$$

where k_B is the Boltzmann constant and T is the target temperature. Other important examples of distributions include the constant pressure, temperature and particle number (TP) ensemble and the *grand canonical ensemble* (in which both energy and particle number are allowed to fluctuate). It is important to bear in mind that the "constancy" of a quantity such as temperature does not actually mean that the *instantaneous temperature* defined by

$$T = \frac{1}{3k_B N} \sum_{i=1}^{N} m_i \|\mathbf{v}_i\|^2 \tag{2.13}$$

is held constant along trajectories (although in fact such a constraint can be imposed and certain quantities can be accurately computed from such an isokinetic model [46]).

If molecular dynamics is to be used to sample from a given ensemble, it is essential that the averages taken along trajectories reproduce the appropriate ensemble average. This requires *ergodicity*: that a given trajectory will eventually visit all states with $\rho > 0$. We can only reasonably expect our method to sample the microcanonical ensemble if we limit ourselves to a conservative (Newtonian) formulation, but it is not clear that the ergodicity assumption will hold in all cases. Indeed, for very low dimensional systems, it is common to find periodic orbits or quasiperiodic tori motion in a restricted portion of phase space; in such cases, the constant energy dynamics is not ergodic. Some modification of the dynamical system (or the introduction of a stochastic perturbation) would be necessary to make the dynamics sample the other ensembles, for example the canonical one. There are many ways to do this, but the most popular techniques are *Nosé* dynamics and *Langevin* dynamics.

In Langevin dynamics, the combination of a damping force and a specially chosen stochastic term maintains the system at a given temperature [47]. The equations of motion are given by

$$\begin{aligned} \dot{\mathbf{r}} &= \mathbf{v}, \\ \mathbf{M}\dot{\mathbf{v}} &= -\nabla V_{\mathbf{r}}(\mathbf{r}) - \gamma \mathbf{M}\mathbf{v} + \mathbf{R}(t), \end{aligned}$$

with collision parameter γ random-force vector $\mathbf{R}$ which is a stationary Gaussian process with mean and covariance given by

$$\langle \mathbf{R}(t) \rangle = 0 \qquad \langle \mathbf{R}(t\mathbf{R}(t')^T) \rangle = 2\gamma k_B T \mathbf{M} \delta(t - t'),$$

where δ denotes the Dirac δ function.

In Nosé dynamics, the Hamiltonian is extended into a larger dimensional phase space, augmented by a thermostatting variable and its canonical momentum. The Nosé extended Hamiltonian corresponding to a system with Hamiltonian $H(\mathbf{r}, \mathbf{p}) = \frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} + V(\mathbf{r})$ takes the form

$$\tilde{H} = \frac{1}{2\sigma^2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} + \frac{1}{2Q}\pi^2 + V(\mathbf{r}) + gkT\ln\sigma$$

where σ and π represent the thermostatting variable and its canonical momentum, respectively. Here Q is an artificial thermostat "mass" which must be chosen carefully to assure that the thermostat is properly coupled to dynamics of the system; see [48,49] for discussion of the choice of mass. The artificial scaling of the kinetic term impedes sampling, particularly the recovery of time-correlated averages; various reformulations of the Nosé dynamical formulation have been proposed which correct the timescale and facilitate the use of time-reversible [6] or lately symplectic [8] discretization.

One of the problems with using low-dimensional model problems is the lack of ergodicity that such systems tend to exhibit. Any of the model problems given here could be treated using Langevin dynamics or a Nosé thermostat. Moreover, it is possible to augment a thermostatted system by another thermostat (or multiple thermostats) and there is evidence that this can in some cases further improve the sampling properties [50,51].

2.3 Formulae for various computable quantities

Simulations of molecular systems can be assessed by considering a number of quantities calculated from the computed trajectories. The atomic masses, positions and velocities can be used to calculate macroscopic quantities such as temperature and internal energy. Under the ergodicity assumption, a macroscopic quantity $\mathcal{A}$, which for real systems could be observed or measured, can be thought of as a long-time average $\langle A(t)\rangle_t$ of some (instantaneous) function $A(t)$ which depends on the collective position and velocity at time t, as well as the masses. The long-time average is defined as

$$\langle A\rangle_t = \lim_{t\to\infty}\frac{1}{t}\int_0^t A(\tau)\,d\tau.$$

In practice, trajectories are computed at a large finite number of discrete times $\tau_1, \ldots, \tau_N$, in which case

$$\langle A\rangle_t = \frac{1}{\tau_N}\sum_{k=1}^{N} A(\tau_i).$$

Some macroscopic quantities of interest are collected in table 2. In this table, the Boltzmann constant is denoted k_B, and instantaneous temperature

T is proportional to kinetic energy and is defined as in (2.13). Specific heat at constant volume measures the rate of change of temperature due to a change in energy. The velocity autocorrelation function measures how the velocities at time t are related to velocities at a later time $t+\tau$. The pair correlation function, or radial distribution function for a system of particles with volume V, gives the number of particles $n(r)$ situated at a distance between r and $r+\Delta r$ from another particle ($\Delta r = 0.016\sigma$ in [52]). Experimental data is given in [53,54]. The mean square displacement $R(\tau)$ measures average atomic movement over time windows of length τ. With $t = 0$, the value of τ at which R ceases to change significantly can be understood as the time required for a simulated system to achieve equilibrium. The diffusion coefficient D is proportional to the slope of $R(\tau)$ over long times via the Einstein relation given in table 2.

More details for computing these quantities can be found in Allen and Tildesley [55], Rapaport [56], and Frenkel and Smit [57].

Table 2. Some Computable quantities

Specific Heat at constant volume

$$C_V = \left[\frac{2}{3N} - \frac{4}{9}\frac{\langle (T - \langle T\rangle_t)^2\rangle_t}{\langle T\rangle_t^2}\right]^{-1} k_B$$

Velocity autocorrelation function

$$Z(\tau) = \left\langle \frac{1}{N}\mathbf{v}(t)\cdot\mathbf{v}(t+\tau)\right\rangle_t$$

Pair correlation function (radial distribution function)

$$g(r) = \left\langle \frac{V}{N}\frac{n(r)}{(4\pi r^2 \Delta r)}\right\rangle_t$$

Mean square displacement after time τ

$$R(\tau) = \left\langle \frac{1}{N}\sum_{i=1}^{N}(r_i(t+\tau) - r_i(t))^2\right\rangle_t$$

Diffusion coefficient, D,

$$2\tau D = \frac{1}{3}R(\tau)$$

2.4 Units

The quantities of interest in a molecular simulation are extremely small in SI units (kilograms, meters, seconds). It is convenient (and numerically more feasible) to choose units that are closely associated with the modeled system. In the case of a Lennard-Jones fluid, the potential energy parameters provide a natural system of units: ϵ as the unit of energy, σ as unit length and the particle mass m as unit mass. Noting that the dimensions of energy are mass$\times$(length)^2/time^2, we can derive the unit time as $\sigma(m/\epsilon)^{1/2}$. Similarly other quantities of interest can be represented in reduced units. We have included Table 3 from Gould and Tobochnik [58] which specifies the system of units for particles in a Lennard-Jones potential.

Table 3. System of units for molecular dynamics simulation of a Lennard-Jones particle system. The quantity k_B is Boltzmann's constant, with value $k_B = 1.38 \times 10^{-23}$ J/K.

quantity	unit	value for argon
length	σ	3.4×10^{-10} m
energy	ϵ	1.65×10^{-21} J
mass	m	6.69×10^{-26} kg
time	$\sigma(m/\epsilon)^{1/2}$	2.17×10^{-12} s
velocity	$(\epsilon/m)^{1/2}$	1.57×10^{2} m/s
force	ϵ/σ	4.85×10^{-12} N
pressure	ϵ/σ^2	1.43×10^{-2} N$\cdot$m^{-1}
temperature	ϵ/k_B	120 K

As with the Lennard-Jones system, the choice of units for the general MD model is significant. Thus, if the the AKMA system is used (unit length is the angstrom, unit energy is the kilocalorie/mole, unit mass is the atomic mass unit (amu) and unit charge is taken as the electron charge e, and the unit temperature is expressed in K, as in CHARMM for example), then the unit time corresponds to 20.455 ps, Boltzmann's constant is $k_B = 1.987191\times10^{-3}$ and the coulomb constant in (1.4) is $1/(4\pi\epsilon_0) = 3.320716\times10^2$. (An adequate program will, of course, convert from and to time units in picoseconds at input and output).

3 The model problems

Here we present the test problems. We have organized them with the simplest models first. In each case, we cite publications in which the problems have been used. We indicate the feature of each problem which makes it relevant for molecular dynamics methods. We describe the analytical techniques which have been used to assess the performance of numerical methods.

3.1 Systems of three atoms

After considering the diatomic system in section 2, the next step is three body systems, either in bounded or collisional motion. Let us, for example, consider a simple tri-atomic molecule that resembles a single water molecule. We have two hydrogen atoms with position $\mathbf{r}_{\mathrm{H1}}$, $\mathbf{r}_{\mathrm{H2}}$ respectively, and mass $m_{\mathrm{H}} = 1\,\mathrm{amu}$, and a single oxygen atom with position $\mathbf{r}_{\mathrm{O}}$ and mass $m_{\mathrm{O}} = 16\,\mathrm{amu}$. The potential energy function is

$$V = \frac{\kappa_b}{2}(r_{\mathrm{OH1}} - r_0)^2 + \frac{\kappa_b}{2}(r_{\mathrm{OH2}} - r_0)^2 + \frac{\kappa_\theta}{2}(\theta - \theta_0)^2,$$

where θ is the bond angle between the O-H1 and the O-H2 bond defined by

$$\cos(\theta) = \frac{\mathbf{r}_{\mathrm{OH1}} \cdot \mathbf{r}_{\mathrm{OH2}}}{||\mathbf{r}_{\mathrm{OH1}}||\,||\mathbf{r}_{\mathrm{OH2}}||},$$

$\mathbf{r}_{\mathrm{OH}i} = \mathbf{r}_{\mathrm{O}} - \mathbf{r}_{\mathrm{H}i}$, $i = 1, 2$. Parameter values can be taken from table 4 [59].

It is often assumed that the bond stretching and bond bending modes can be replaced by rigid approximations

$$r_{\mathrm{OH1}} - r_0 = 0, \quad r_{\mathrm{OH2}} - r_0, \quad \theta - \theta_0 = 0$$

as we discussed in section 2. In fact, that leads to the approximation of a single water molecule as a rigid body with frozen atomic distances $r_{\mathrm{OH1}} = r_{\mathrm{OH2}} = r_0$ and $r_{\mathrm{HH}} = ||\mathbf{r}_{\mathrm{H1}} - \mathbf{r}_{\mathrm{H2}}|| = r_1$. The associated constrained equations of motion are

$$\begin{aligned} m_{\mathrm{H}}\ddot{r}_{\mathrm{H1}} &= -\lambda_1(\mathbf{r}_{\mathrm{H1}} - \mathbf{r}_{\mathrm{O}}) - \lambda_2(\mathbf{r}_{\mathrm{H1}} - \mathbf{r}_{\mathrm{H2}}), \\ m_{\mathrm{H}}\ddot{r}_{\mathrm{H2}} &= -\lambda_3(\mathbf{r}_{\mathrm{H2}} - \mathbf{r}_{\mathrm{O}}) - \lambda_2(\mathbf{r}_{\mathrm{H2}} - \mathbf{r}_{\mathrm{H1}}), \\ m_{\mathrm{O}}\ddot{r}_{\mathrm{O}} &= -\lambda_1(\mathbf{r}_{\mathrm{O}} - \mathbf{r}_{\mathrm{H1}}) - \lambda_3(\mathbf{r}_{\mathrm{O}} - \mathbf{r}_{\mathrm{H2}}), \end{aligned}$$

where the Lagrange multipliers λ_i, $i = 1, 2, 3$, are implicitly determined by the three (holonomic) constraints

$$r_{\mathrm{OH1}} = r_{\mathrm{OH2}} = r_0, \qquad r_{\mathrm{HH}} = r_1.$$

Numerical integration of these constrained equations of motion again provides a simple test problem.

3.2 Lennard-Jones Fluids

Careful description of argon simulations are given in [52] and [58]; we summarize these here. Chapter 4 of Frenkel and Smit [57] includes detailed algorithms for simulating a Lennard-Jones system and analyzing the resulting trajectory.

The atoms of liquid argon are assumed to interact with a Lennard-Jones potential (1.1). In reduced-unit computer simulation of the Lennard-Jones

Table 4. Potential energy parameters for a single water molecule. An artificial bond term is given for interaction between the two hydrogen atoms. Such a term is often included in addition to the angle term. It also provides an equilibrium value for the H-H length if the constrained formalism is used.

bonds		
	O-H	$\kappa_b = 450.0$ kcal/mol, $b^{\rm eq} = 0.957$ Å
	H-H	$\kappa_b = 450.0$ kcal/mol, $b^{\rm eq} = 1.514$ Å
angle		
	H-O-H	$\kappa_\theta = 55.0$ kcal/mol, $\theta^{\rm eq} = 104.52°$

systems, the parameters σ, ϵ and m are set to 1. Table 3 can be used to interpret the computed results in a manner which is physically meaningful for argon. An important advantage of the reduced unit system is that the same simulation could be used to study a system other than argon, substituting appropriate conversion parameters analogous to the last column of the table.

The force on particle i due to a single interaction with particle j is given by

$$\mathbf{F}_{ij} = \left[-\frac{\partial V}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial x_i} \quad -\frac{\partial V}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial y_i} \quad -\frac{\partial V}{\partial r_{ij}} \frac{\partial r_{ij}}{\partial z_i} \right]^T \tag{3.14}$$

$$= -24 \left(\left(\frac{1}{r_{ij}^7} \right) - 2 \left(\frac{1}{r_{ij}^{13}} \right) \right) \left[\frac{\partial r_{ij}}{\partial x_i} \quad \frac{\partial r_{ij}}{\partial x_i} \quad \frac{\partial r_{ij}}{\partial x_i} \right]^T \tag{3.15}$$

$$= -24 \left(\left(\frac{1}{r_{ij}^8} \right) - 2 \left(\frac{1}{r_{ij}^{14}} \right) \right) \left[(x_i - x_j) \; (y_i - y_j) \; (z_i - z_j) \right]^T \tag{3.16}$$

$$= -24 \left(\left(\frac{1}{r_{ij}^8} \right) - 2 \left(\frac{1}{r_{ij}^{14}} \right) \right) (\mathbf{r}_i - \mathbf{r}_j). \tag{3.17}$$

We have used reduced units in (3.14)–(3.17). Thus the equations of motion for a general Lennard-Jones fluid are:

$$\dot{\mathbf{r}}_i = \mathbf{v}_i \tag{3.18}$$

$$\dot{\mathbf{v}}_i = 24 \sum_{j \neq i} \frac{\mathbf{r}_i - \mathbf{r}_j}{r_{ij}^2} \left(2 \left(\frac{1}{r_{ij}} \right)^{12} - \left(\frac{1}{r_{ij}} \right)^6 \right). \tag{3.19}$$

The particles are placed in a cubic box, subdivided into q^3 smaller cubes. Initially the particles are arranged according to the "face-centered cubic" (fcc) crystal structure appropriate for argon — an atom is placed at each of the corners of a cube, plus an atom at the center of each cube face. The result is $4q^3$ atoms. The standard simulation involves a cubic box with edges of length $L = 10.229\sigma$, with 864 ($4q^3$ with $q = 6$) atoms arranged as described above. L, σ and q are chosen to achieve a physical density (1.37 g cm^{-3}).

Periodic boundary conditions are imposed by the *minimum image convention* in which an atom interacts with all others contained in a self-centered cubic box with edges of length $L/2$. To set the initial temperature, velocities are assigned according to a Gaussian (normal) distribution with mean 0 and standard deviation $\sqrt{k_B T/m}$ in each component. Trajectories can be assessed by computing average quantities discussed in section 2.3.

3.3 Water

Liquid water, due to its all-important biological role in aqueous solution, has been the focus of considerable attention in molecular simulation for 30 years [60,61]. At the heart of molecular dynamics water calculations is the effective pair potential [62] between a pair of water molecules. There are a number of models for molecular interactions in water. The simplest models, such as the 3-atom molecule "simple point charge" (SPC) model [63] and TIP3P model [62], treat the oxygen and the two hydrogen atoms as locations of electrostatic charge, with coulombic interaction V_C as in table 1. Alternatively, four-site models such as TIP4P [62] move the charge from the oxygen towards the hydrogens along the bisector of the H-O-H angle. Five-site models such as ST2 [64] and TIP5P [65] have charges on the oxygen, the two hydrogens, plus two isolated sites located so that the hydrogens and the 2 additional charge sites form a regular tetrahedron centered around the oxygen atom. The various models mentioned here are are illustrated in Fig. 3 and compared in [62,65].

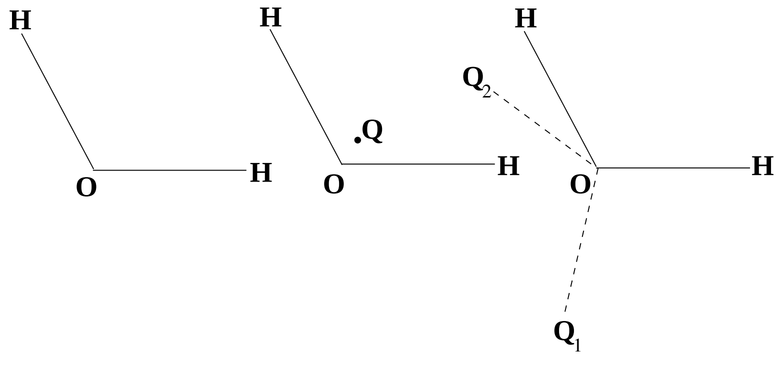


Fig. 3. From left: three-site water model with charges placed at the atomic sites; four-site model with the charge moved from the oxygen to the site Q along the bisector of the H-O-H angle; the five-site model with charges at the three atomic sites, plus two isolated positions Q_1 and Q_2 located at the same distance from the oxygen as the hydrogens, resulting in a regular tetrahedron with the oxygen atom at the center.

There is theoretical evidence that the central H-O-H bond angle remains unchanged in both ice and liquid water. Accordingly, in many water models the internal degrees of freedom (the O-H stretches and the H-O-H bend) are maintained rigidly with internal coordinates [26] or through Cartesian coordinates with constraints [27], or restrained near equilibrium values with harmonic potential terms on the bonds and angle [66]. In the rigid models described here, Lennard-Jones interactions between a pair of water molecules are included only for the oxygen atoms

There is some interest in flexible water models, in which the O-H stretch and H-O-H angle are modeled by harmonic potentials as in section 3.1. In addition, the Lennard-Jones interactions are often included for H-H and O-H pairs, and some models include polarizability through variable charges [67,68]. Two flexible models based on SPC are described in [66,69].

Once a particular water model is adopted, the practical difficulty arises that the Coulombic interaction unlike Lennard-Jones, is significant even at large separations r_{ij}. Force cutoffs, usually 8–9 Å, were used in the parametrization many of the models mentioned here. In a buffer region near the cutoff distance, the force (or potential, see below) is smoothly switched to zero by a switching function such as (3.21). There is evidence that cutoffs can cause undesirable artifacts [70,71]. A review of this issue is found in [72]. For this reason attention has turned to the development of summation schemes, such as fast multipole [73,74,1,75] and Ewald summation [76–78], for faster evaluation of electrostatic energies and forces without distance cutoffs. It is often the case that fast summation methods enjoy computational speedups, as compared to direction summation, only for sufficiently large systems. Large periodic water systems were used for testing fast electrostatic methods in [78] and are available electronically. Non-periodic water systems for testing timestepping methods are described in [79,59].

New summation techniques are generally compared, in terms of accuracy and speedup, with existing methods based on single force and/or energy calculations for a given model system. This limited testing is sometimes insufficient to predict the behavior of the new methods in molecular dynamics applications. For example, it has been demonstrated that insufficient accuracy leads to loss of smoothness in the forces from timestep to timestep, causing instabilities in high-performance molecular dynamics timestepping schemes with fast multipole evaluation of electrostatics [80,81]. It is preferable therefore to compare force evaluation methods in broader contexts. To verify the behavior of novel electrostatic treatments, energy stability throughout a dynamics simulation should be confirmed. As another measure of correctness for force evaluation methods, the short-range structure of liquid water can be ascertained as with the liquid argon simulation described earlier, by calculation of the radial distribution function (see table 2). For water, there are three such functions: g_{OO}, g_{OH} and g_{HH}, corresponding to oxygen-oxygen, oxygen-hydrogen, and hydrogen-hydrogen distances, respectively between pairs of molecules. These quantities are calculated as time averages of the radial distribution function over trajectories resulting from molecular dynamics or Monte Carlo simulations. Computed radial distribution functions are often compared with experimental results from x-ray and neutron refraction data given in tabular form in [82].

The choice of water model is based on a variety of factors related to the agreement of the computed results with experiment. The choice can have important practical implications for molecular dynamics simulations. The five-site models clearly require more computation for evaluating the electrostatic forces than the three-site models. If a flexible model is chosen, the additional high-frequency terms in the potential necessitate a smaller timestep (typically 0.5–1.0 fs) to assure the stability of the time discretization method. Time steps of 2-4 fs are common for rigid water models. An alternative to rigid water has been recently introduced which maintains the elasticity of flexible water molecules while removing the high frequency modes [83]. In addition, the presence of the O-H bond terms can produce instabilities in advanced timestepping algorithms, which we discuss next.

A system with long and short range forces. Since the Coulombic interactions are significant at large separations, systems of water molecules have been used as test systems for force-splitting numerical integration schemes [84,85]. The force is decomposed into fast and slow components using a switching function $S(r)$,

$$\mathbf{F} = \mathbf{F}_{\text{fast}} + \mathbf{F}_{\text{slow}} = -S(r)\nabla V - (1 - S(r))\nabla V, \tag{3.20}$$

where

$$S(r) = \begin{cases} 1, & r < r_c - \lambda \\ 1 + R^2(2R-3), & r_c - \lambda \le r \le r_c \\ 0, & r_c < r \end{cases} \tag{3.21}$$

with $R = (r - (r_c - \lambda))/\lambda$. Here r_c is the cutoff distance beyond which forces are considered slow, and λ is a "healing length" over which the switching function S smoothly varies between one and zero. The form of the switching function is somewhat arbitrary, though sufficient smoothness is required.

In multiple timestep (MTS) time discretization methods the short range forces, which can change rapidly in time, are updated frequently and included in the numerical dynamics with small timesteps. The long range forces are treated with larger steps in time, appropriate to the timescale on which they vary significantly. MTS methods have been developed which share the abstract geometric properties (and also favorable energy conservation) with the Verlet method [84,86].

For solvated systems, fast summation methods have been successfully combined with multiple timestep integration methods [87,88]. Variations of the method given in (3.20) include employing switching functions such as

(3.21) to intermolecular (water-water) distances rather than interatomic distances or using switching functions to split the potential, rather than the force. Procacci and Berne point out that these schemes can result in stable integration methods, however the force-splitting formulation results in an altered Hamiltonian $H = T + V^*$, where $-\nabla V^* = F_{\text{fast}} + F_{\text{slow}}$, during the propagation of the fast motion [89].

In flexible water models, the fast forces include the O-H bond and H-O-H angle terms. In multiple timestep methods, these forces are treated with an appropriately short timestep (0.5 fs), while the long-range nonbonded forces can, in principle, be updated at much longer intervals. It has become clear over the last several years, however, that to assure stability of the integration method the long timestep in MTS methods cannot exceed 5 fs. The 5 fs barrier is understood to be a resonance artifact coinciding with the half-period of bonds such as O-H. Impulses introduced into the dynamics at each large step excite the bonds and lead to catastrophic energy growth. A number of methods have been proposed to overcome the MTS timestep barrier, including averaging methods [59] which allow timesteps of up to 6 fs for flexible water, and nonconservative MTS methods [79] which are less susceptible to resonance at timesteps up to 48 fs for flexible water, but require stabilization by the addition of substantial stochastic noise and dissipation through Langevin dynamics.

3.4 Larger molecular models

We have focused to this point on systems of molecular fluids. The challenges posed by flexible water, with both local (bond and angle) and long range forces, are present to a much greater degree in flexible molecules like proteins. An important goal of molecular modeling is to provide insight into questions of protein structure and function. Proteins are characterized by their linear sequence of amino acid residues (primary structure). Cooperative arrangements between groups of residues (secondary structure) such as alpha helices and beta sheets, together with higher level organization of secondary structures, lead to a multitude of folded conformations (tertiary structure). It has long been appreciated that tertiary structure is vitally important in determining the biological activity of proteins.

With 10-20 atoms per residue, detailed molecular dynamics models of proteins can contain thousands of atoms. In addition, realistic models of biological systems must include solvent, which can add many thousands of atoms. Systems of this size limit the feasible length of computed trajectories to perhaps a few nanoseconds. Recent work has sought to address the computational burden of solvent by development of implicit solvent models [90,91]. Algorithmic advances in protein folding could have tremendous impact on our understanding of life itself. Here the need for test systems is especially clear. In this section we present several simple polymer models. They are suitable tests for sampling and global optimization because they possess many distinct low-energy structures and are easily scalable over a large range of system sizes.

Butane molecule The consideration of flexible molecules leads to an additional type of potential energy term which comes from the dihedral degrees of freedom. The associated dihedral angles are best understood for four atomic molecular systems such as butane in united atom presentation CH_3-CH_2-CH_2-CH_3. In the united atom approach, a heavy atom and attached hydrogens can be modeled as a single particle with additional mass: CH_3 and CH_2 are particles of mass 15 and 14, respectively. Let us denote the sequence of four particle positions by $\mathbf{r}_i$, $i = 1, 2, 3, 4$. Then each pair of distance vectors $(\mathbf{r}_{12}, \mathbf{r}_{32})$ and $(\mathbf{r}_{32}, \mathbf{r}_{34})$ spans a plane and the angle between these two planes is called the dihedral angle ϕ, shown in Fig. 4 and table 1. Let us give a

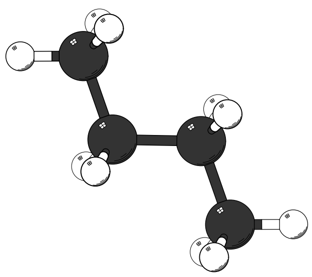


Fig. 4. The butane molecule with carbon atoms represented by large dark spheres and hydrogen by small light spheres. The image was rendered using MolScript [92].

precise mathematical definition. We introduce the two vectors

$$\mathbf{m} = \mathbf{r}_{12} \times \mathbf{r}_{32} \quad \text{and} \quad \mathbf{n} = \mathbf{r}_{32} \times \mathbf{r}_{34}. \tag{3.22}$$

Then the classical definition of the dihedral angle is given by

$$\phi := \operatorname{sign}(\mathbf{n} \cdot \mathbf{r}_{12}) \arccos\left(\frac{\mathbf{m} \cdot \mathbf{n}}{||\mathbf{m}||\,||\mathbf{n}||}\right). \tag{3.23}$$

In analogy to our discussion of LJ systems, the potential is given in internal coordinates (bond lengths, angles, etc.), and the Cartesian force vector is obtained by differentiating the potential with respect to the Cartesian coordinates by the chain rule. It is possible to derive the following elegant formulas for the forces associated with a dihedral potential energy term $V_d(\phi)$ from table 1 [93–95]:

$$\mathbf{F}_1 = -\frac{dV(\phi)}{d\phi} r_{32} \frac{\mathbf{m}}{||\mathbf{m}||},$$
$$\mathbf{F}_4 = +\frac{dV(\phi)}{d\phi} r_{32} \frac{\mathbf{n}}{||\mathbf{n}||},$$
$$\mathbf{F}_2 = -\mathbf{F}_1 + \frac{\mathbf{r}_{12} \cdot \mathbf{r}_{32}}{r_{32}^2} \mathbf{F}_1 - \frac{\mathbf{r}_{34} \cdot \mathbf{r}_{32}}{r_{32}^2} \mathbf{F}_4,$$
$$\mathbf{F}_3 = -\mathbf{F}_4 - \frac{\mathbf{r}_{12} \cdot \mathbf{r}_{32}}{r_{32}^2} \mathbf{F}_1 + \frac{\mathbf{r}_{34} \cdot \mathbf{r}_{32}}{r_{32}^2} \mathbf{F}_4.$$

An interesting test case is the simulation of butane gas in a setting similar to argon. See Schlick et al [96]. The CHARMM (version 19) united atom force field parameter values for butane are given in table 5.

Table 5. A suitable set of united atom parameters for the butane. (From CHARMM.)

bonds	
CH_2-CH_2	$\kappa_b = 225.0$ kcal/mol, $b^{\mathrm{eq}} = 1.52$ Å
CH_2-CH_3	$\kappa_b = 225.0$ kcal/mol, $b^{\mathrm{eq}} = 1.54$ Å
angles	
CH_3-CH_2-CH_2	$\kappa_\theta = 45.0$ kcal/mol, $\theta^{\mathrm{eq}} = 110.0°$
dihedrals	
CH_3-CH_2-CH_2-CH_3	$\kappa_\phi = 1.6$ kcal/mol, $\phi^{\mathrm{eq}} = 0.0°$, $n_\phi = 3$
Lennard-Jones	
CH_2	$\epsilon = 0.1142$ kcal/mol, $\sigma = 2.235$ Å
CH_3	$\epsilon = 0.1811$ kcal/mol, $\sigma = 2.165$ Å

A small flexible molecule A frequently-used test problem in molecular dynamics simulation is the so-called alanine dipeptide (more properly N-Acetylalanyl-N'-Methylamide) [32,97,47,98]. Small in size (22 atoms), alanine dipeptide is a good model system because it exhibits conformational flexibility in the 2-dimensional $\phi - \psi$ dihedral space. The dihedral angles ϕ and ψ are illustrated in Fig. 5. The changes in structure seen through the dynamics of the backbone dihedral angles make alanine dipeptide an attractive prototype for understanding the structure of folding proteins. Figure 6 shows three distinct conformations at major minima: C_5: $\phi \approx -180°$, $\psi \approx 180°$; C_7 equatorial: $\phi \approx -120°$, $\psi \approx 60°$; and C_7 axial: $\phi \approx 60°$, $\psi \approx -60°$. The exact location of the minima varies with the choice of potential function. A Ramachandran plot, which gives the conformational energy distribution as a function of the two dihedral angles is given in Fig. 7. The energies were calculated by restraining the dihedral angles at 10° increments and minimizing all other degrees of freedom (no other constraints were imposed). The process is described in [47]. Alanine dipeptide parameters from the CHARMM potential are provided on the web site `http://www.mmc.le.ac.uk/mdt`.

In many molecules (including peptides and nucleic acids), certain arrangements of four or more atoms along the peptide backbone have a naturally planar equilibrium arrangement — the so-called improper torsion angles. For these, deformations from planarity may be described with quadratic potential energy terms in the dihedral angles relating sets of four atoms (V_i in table 1). In alanine dipeptide, the left and right "arms", shown by shaded polygons in Fig. 5, are each formed by six atoms in a plane. In the left arm (darker rectangle) the planar configuration is maintained by two improper torsions, CH_3-C-O-N and C-N-H-CH and a dihedral angle CH_3-C-N-CH. Similarly the right arm is maintained in a plane by improper torsion angle terms for CH-C-O-N and CH_3-N-H-C and the dihedral angle CH-C-N-CH_3.

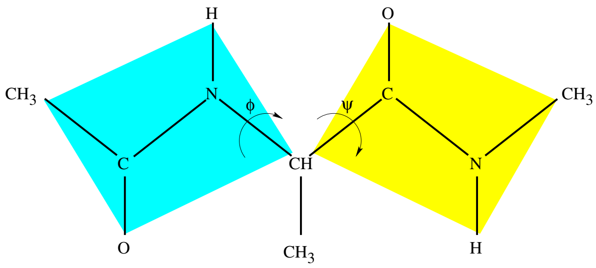


Fig. 5. Alanine Dipeptide.

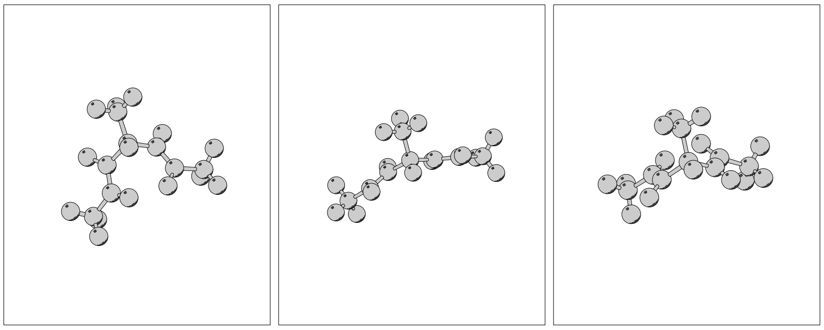


Fig. 6. Alanine Dipeptide in three conformations. From left: C_7 equatorial, C_5 and C_7 axial. The images were rendered using MolScript [92].

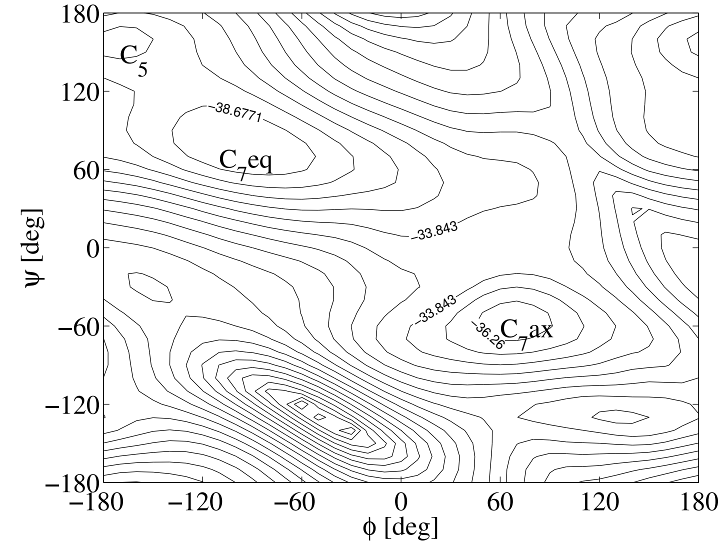


Fig. 7. Ramachandran plot for alanine dipeptide — potential energy as a function of the two dihedral angles ϕ and ψ indicated in Fig. 5. Energy contours are labeled in units of kcal/mol.

As with the flexible water models, the fast oscillatory force components present in alanine dipeptide restrict the allowable timestep for dynamics simulations to about 0.5 fs. The bond lengths in alanine dipeptide can be constrained to fixed lengths using the SHAKE algorithm [27]. A commonly used procedure is to constrain the bonds involving hydrogens (the fastest of the bonds) which allows a timestep as large as 2 fs. MTS methods have been applied to unconstrained dynamics simulations of this system with the short range forces consisting of the local terms (bonds, angles, dihedrals) and sometimes the Lennard-Jones interactions between close pairs of atoms [98]. Because there are not any truly long range forces in a single small molecule, the MTS time step for the slower forces is in the range 2–3 fs.

Methods for dynamics and sampling applied to alanine dipeptide can be assessed by calculating observable quantities as time series over a computed trajectory and time averages and variances. Examples of such quantities are total energy, individual components of potential energy (especially bond energy, where MTS resonance artifacts are first observed), and mean square displacements from the starting structure (see table 2). The backbone dihedral angles ϕ and ψ provide a compact way to monitor sampling of conformation space [98] and transitions between conformations [47]. The regions of conformational space available for a particular MD potential function can be ascertained from a calculation like the one used to produce the Ramachandran plot in Fig. 7.

"Minimal models" for proteins. Grubmüller and Tavan [99] have proposed a simplified protein model consisting of a chain of 100 identical particles. They use the same CH_2 united atom model and parameters given for butane in section 3.4 (see table 5).

The potential energy used in this model is a subset of the CHARMM potential (1.4), including bonds, angles, Lennard-Jones and electrostatics. The CH_2 parameters are given in Table 6. In order to impose a heterogeneous primary structure, the model is divided into five electrostatic regions, three positive and two negative, by assignment of partial charges given approximately by $q_n = \frac{1}{2}\cos(n/8)e$ where n is the index of a carbon atom, numbered from the end of the chain, as shown in Fig. 8, and e is the unit electrostatic charge.

Table 6. Suitable parameters for the minimal protein model potential.

bonds	$\kappa_b = 225.0$ kcal/mol, $b^{\text{eq}} = 1.52$ Å
angles	$\kappa_\theta = 45.0$ kcal/mol, $\theta^{\text{eq}} = 110.0°$
Lennard-Jones	$\epsilon = 0.1142$ kcal/mol, $\sigma = 4.470$ Å
Coulomb	$q_n = \frac{1}{2}\cos(n/8)e$

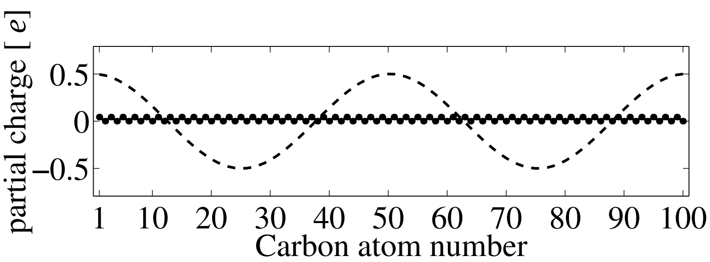


Fig. 8. The simplified protein, with partial charge (in units of electron charge) assigned along the length of the chain.

Molecular dynamics simulations of this model exhibit large-scale conformational changes between distinct tertiary structures. Configurational interconversions occur on an accelerated timescale, compared with detailed protein models [99].

Inter-particle distance between neighbors along the chain is maintained near equilibrium with a harmonic bond potential, necessitating a small timestep of 1 fs, as in traditional molecular dynamics simulations applied to detailed models. A larger integration timestep could be used if rigid constraints were to be imposed on the bond lengths, most likely without substantive effect on the overall dynamics. A similar particle chain has been proposed by Honeycutt and Thirumalai [100] to mimic a protein with 46 amino acids. These are represented as particles interacting locally as in the Grubmüller model with bonds replaced by algebraic constraints, while the nonbonded interactions are modeled by the LJ potential with parameters for interaction between specific pairs chosen to reflect properties of individual residues in the amino acid sequence. Details can be found in [101].

4 Conclusions

In this paper, we have presented a number of small model systems for use in testing the comparing molecular dynamics algorithms. Beginning with an atomic fluid, we have sequentially treated systems of increasing molecular complexity which possess important features of the molecular dynamics model. We have explicitly included potential energy functional forms and suitable parameter values. We describe some of the difficulties in developing MD simulation methods and discuss theoretical and algorithmic advances for these problems. It is our hope that this paper will serve as a starting point for ongoing test set development effort which includes contributions from researchers in mathematical and chemical science. The progress of the test set will be published on the MD Test Set web site `http://www.mmc.le.ac.uk/mdt`.

References

1. L. Greengard and V. Rokhlin, *A fast algorithm for particle simulations*, J. Comp. Phys., 73 (1987), pp. 325–348
2. R. Krasny and Z.-H. Duan, *Treecode algorithms for computing nonbonded particle interactions*, this volume.
3. M. Tuckerman, G. Martyna and B.J. Berne, *Molecular Dynamics algorithms for multiple time scales: Systems with Long Range Forces*, J. Chem. Phys., 94 (1991), pp. 6811–6815
4. R.D. Skeel and J. Izaguirre *The Five Femtosecond Time Step Barrier*, in Deuflhard, P., Hermans, J., Leimkuhler, B., Mark, A., Reich, S., and Skeel, R. D., eds., *Computational Molecular Dynamics, Challenges, Methods, Ideas* (Springer-Verlag), pp. 303-318, (1998)
5. T. Schlick, R. D. Skeel, A. T. Brunger, L. V. Kalé, J. Hermans, K. Schulten, and J. A. Board, Jr., *Algorithmic Challenges in Computational Molecular Biophysics*, J. Comp. Phys., 151 (1999), pp. 9–48

6. S. Nosé, *A unified formulation of the constant temperature molecular dynamics methods*, J. Chem. Phys., 81 (1984), pp. 511–519
7. G.J. Martyna, M.E. Tuckerman, D.J. Tobias, and M.L. Klein, *Explicit reversible integration algorithms for extended systems*, Mol. Phys. 87 (1996), pp. 1117
8. S.D. Bond, B.J. Leimkuhler and B.B. Laird, *The Nose-Poincare method for constant temperature molecular dynamics*, J. Comp. Phys. 151 (1999), pp. 114–134
9. J.J. Moré, B.S. Garbow and K.E. Hillstrom, *Testing Unconstrained Optimization Software*, ACM TOMS, 7 (1981), 17–41
10. E.D. Dolan and J.J. Moré, *Benchmarking optimization software with COPS*, Mathematics and Computer Science Division, Argonne National Laboratory, Technical Report ANL/MCS-246, November 2000 (Revised November 30), `http://www-unix.mcs.any.gov/~more/cops/`
11. C.A. Floudas, P.M. Pardalos, C. Adjiman, W.R. Esposito, Z.H. Gümüs, S.T. Harding, J.L. Klepeis, C.A. Meyer, C.A. Schweiger, *Handbook of test problems in local and global optimization*, Volume 33 of *Nonconvex Optimization and Its Applications*, Kluwer Academic Publishers, Dordrecht, 1999
12. W.M. Lioen and J.J.B. de Swart, *Test set for initial value problems*, Report MAS-R 9832, Centrum voor Wiskunde en Informatica, Amsterdam `http://www.cwi.nl/cwi/projects/IVPtestset`
13. E. Hairer and G. Wanner, *Solving ordinary differential equations. Volume II*, Springer Series in Comput. Mathematics, Vol. 14, Springer-Verlag 1996, `http://www.zib.de/uwe.poehle/ode.html`
14. *Proceedings of the first meeting on the critical assessment of techniques for protein structure prediction*, Proteins: Structure, Function and Genetics, 23 (1995), see also `http://predictioncenter.llnl.gov`
15. B.J. Alder and T.E. Wainwright, *Phase transition for a hard sphere system*, J. Chem. Phys., 27 (1957), pp. 1208–1209
16. J.D. Weeks, D. Chandler, and H.C. Andersen, *Role of repulsive forces in determining the equilibrium structure of simple liquids*, J. Chem. Phys., 54 (1971), pp. 5237–5247
17. R.M. Stratt, S.L. Holmgren, and D. Chandler, *Constrained impulsive molecular dynamics*, Mol. Phys., 42 (1981), pp. 1233–1243
18. S.-H. Suh, L. Mieryteran, H.S. White, and H.T. Davis, *Molecular dynamics study of the primitive model of 1–3 electrolyte solutions*, Chem. Phys., 142 (1990), pp. 203–211
19. Y.A. Houndonougbo, B.B. Laird and B.J. Leimkuhler, *Molecular dynamics algorithms for mixed hard-core/continuous potentials*, Mol. Phys., 98 (1999), pp. 309–316
20. A. Rahman, *Correlations in the motion of atoms in liquid argon*, Phys. Rev. A, 136 (1964), pp. 405–411
21. L. Verlet, *Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones Molecules*, Phys. Rev., 159 (1967), pp. 98–103
22. R.D. Ruth, *A canonical integration technique*, IEEE Trans. Nucl. Sci., 30 (1983), p. 2669-2671
23. F.M. Lasagni, *Integration methods for Hamiltonian differential equations*, Unpublished manuscript, (1990).

24. E. Hairer, *Backward analysis of numerical integrators and symplectic methods*, in K. Burrage, C. Baker, P. v.d. Houwen, Z. Jackiewicz, and P. Sharp, editors, *Scientific Computation and Differential Equations*, volume 1 of Annals of Numer. Math., pp. 107–132, Amsterdam, 1994. J.C. Baltzer. Proceedings of the SCADE'93 conference, Auckland, New-Zealand, January 1993
25. G. Benettin and A. Giorgilli, *On the Hamiltonian Interpolation of Near to the Identity Symplectic Mappings*, J. Statist. Phys., 74 (1994), pp. 1117–1143
26. A. Rahman and F.H. Stillinger, *Molecular dynamics study of liquid water*, J. Chem. Phys., 55 (1971), pp. 3336–3359
27. J.P. Ryckaert, G. Ciccotti and H.J.C. Berendsen, *Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*, J. Comp. Phys., 23 (1977), pp. 327–341
28. J.A. McCammon, B.R. Gelin and M. Karplus, *Dynamics of folded proteins*, Nature, 267 (1977), pp. 585–590
29. A.K. Mazur, V.E. Dorofeev and R.A. Abagyan, *Derivation and testing of explicit equations of motion for polymers described by internal coordinates*, J. Comp. Phys., 92 (1991), pp. 261–272
30. S. He and H.A. Scheraga, *Macromolecular conformational dynamics in torsional angle space*, J. Chem. Phys., 108 (1998), pp. 271–286
31. S.-H. Lee, K. Palmo and S. Krimm, *A new formalism for molecular dynamics in internal coordinates*, J. Chem. Phys., to appear
32. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations*, J. Comp. Chem, 4 (1983), pp. 187–217
33. A.D. MacKerell Jr., D. Bashford, M. Bellott, R.L. Dunbrack Jr., J. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher III, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *An all-atom empirical potential for molecular modeling and dynamics of proteins*, J. Phys. Chem., 102 (1998), pp. 3586–3616
34. S.J. Weiner, P.A. Kollman, D.T. Nguyen, and D.A. Case, *An all atom force field for simulations of proteins and nucleic acids*, J. Comp. Chem., 7 (1986), pp. 230–252
`http://www.amber.ucsf.edu/amber/`
35. W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, Jr, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell and P.A. Kollman, *A second generation force field for the simulation of proteins and nucleic acids*, J. Am. Chem. Soc., 117 (1995), pp. 5179-5197
36. W.L. Jorgensen and J. Tirado-Rives, *The OPLS potential functions for proteins. Energy minimization for crystals of cyclic peptides and crambin*, J. Am. Chem. Soc., 110 (1988), pp. 1657–1666
37. `http://bmbiris.bmb.uga.edu/wampler/8200/using-ff/mmrefs.html`.
38. T. Schlick, *Optimization methods in computational chemistry*, in *Reviews in Computational Chemistry*, Volume 3, Chapter 1, pages 1-71, K. B. Lipkowitz and D. B. Boyd, eds., VCH Publishers, New York (1992)
39. N.B. Slater, *Classical motion under a Morse potential*, Nature, 180 (1957), pp. 1352–1353

40. M. Mandziuk and T. Schlick, *Resonance in chemical systems simulated by the implicit midpoint method*, Chem. Phys. Lett., 237 (1995), pp. 525–535
41. J.M. Sanz-Serna and M.P. Calvo, *Numerical Hamiltonian problems*, Chapman and Hall, 1994.
42. J. Frank, W. Huang and B. Leimkuhler, *Geometric integrators for classical spin systems*, J. Comp. Phys., 133 (1997), pp. 160–172.
43. H.C. Andersen, *Rattle: a 'velocity' version of the shake algorithm for molecular dynamics calculations*, J. Comp. Phys., 52 (1983), pp. 24–34
44. B. Leimkuhler and R.D. Skeel, *Symplectic numerical integrators in constrained Hamiltonian systems*, J. Comp. Phys., 112 (1994), pp. 117–125
45. E. Barth, K. Kuczera, B. Leimkuhler and R.D. Skeel, *Algorithms for constrained molecular dynamics*, J. Comp. Chem., 16 (1995), pp. 1192–1209
46. D.J. Evans, *Computer "experiment" for nonlinear thermodynamics of Couette flow*, J. Chem. Phys., 78 (1983), pp. 3297–3302
47. R.J. Loncharich, B.R. Brooks and R.W. Pastor, *Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-Acetylalanyl-N′-Methylamide*, Biopolymers, 32 (1992), pp. 523–535
48. S. Nosé, *A molecular dynamics method for simulations in the canonical ensemble*, Mol. Phys., 52 (1984), pp. 255-268
49. J.B. Sturgeon and B.B. Laird, *Symplectic algorithm for constant-pressure molecular-dynamics using a Nose-Poincare thermostat*, J. Chem. Phys. 112 (2000), 3474
50. G.J. Martyna, M.L. Klein, and M.E. Tuckerman, *Nose-Hoover chains: The canonical ensemble via continuous dynamics*, J. Chem. Phys., 97 (1992), pp. 2635–2643
51. W.G. Hoover, C.G. Hoover, and D.J. Isbister, *Chaos, ergodic convergence, and fractal instability for a thermostatted canonical harmonic oscillator*, Phys. Rev. E, 63 (2001), 026029
52. D. Okunbor and R.D. Skeel, *Canonical numerical methods for molecular dynamics simulations*, J. Comp. Chem., 15 (1994), pp. 72–79
53. J.L. Yarnell, M.J. Katz, R.G. Wenzel and S.H. Koenig, *Structure factor and radial distribution function for liquid argon at 85° K*, Phys. Rev. A, 7 (1973), pp. 2130–2144
54. A.K. Soper, *On the determination of the pair correlation function from liquid structure factor measurements*, Chem. Phys., 107 (1986), pp. 61–74
55. M.P. Allen and D.J. Tildesley, *Computer simulation of liquids*, Oxford Science Publications, 1987
56. D.C. Rapaport, *The art of molecular dynamics simulation*, Cambridge University Press, 1995, http://uk.cambridge.org/physics/resource
57. D. Frenkel and B. Smit, *Understanding molecular simulation. From algorithms to applications*, Academic Press, 1996
58. H. Gould and J. Tobochnik, *An introduction to computer simulation methods: Applications to physical systems*, Addison-Wesley, 1988
59. J.A. Izaguirre, S. Reich and R.D. Skeel, *Longer time steps for molecular dynamics*, J. Chem. Phys, 110 (1999), pp. 9853–9864
60. F.H. Stillinger, *Theory and molecular models for water*, Adv. Chem. Phys., 31 (1975), pp. 1–101
61. F.H. Stillinger, *Water revisited*, Science, 209 (1980), pp. 451–457

62. W. Jorgensen, J. Chandrasekar, J. Madura and R. Impey and M. Klein, *Comparison of simple potential functions for simulating liquid water*, J. Chem. Phys., 79 (1983), pp. 926–935
63. H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren and J. Hermans, in *Intermolecular Forces*, B. Pullman, Editor, Reidel, Dordrecht, 1981
64. F.H. Stillinger and A. Rahman, *Improved simulation of liquid water by molecular dynamics*, J. Chem. Phys., 60 (1974), pp. 1545–1557
65. M.W. Mahoney and W.L. Jorgensen, *A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions*, J. Chem. Phys., 112 (2000), pp. 8910–8922
66. K. Toukan and A. Rahman, *Molecular dynamics study of atomic motions in water*, Phys. Rev. B., 32 (1985), pp. 2643–2648
67. H.J.C. Berendsen, J.R. Grigera, and T.P. Straatsma, *The missing term in effective pair potentials*, J. Phys. Chem., 91 (1987), pp. 6269–6271
68. S.J. Stuart and B.J. Berne, *Effects of Polarizability on the Hydration of the Chloride Ion*, J. Phys. Chem., 100 (1996), pp. 11934–11943
69. J. Anderson, J.J. Ullo and S. Yip, *Molecular dynamics simulation of dielectric properties of water*, J. Chem. Phys., 87 (1987), pp. 1726–1732
70. R.R. Gabdoulline and Chong Zheng, *Effects of the cutoff center on the mean potential and pair distribution functions in liquid water*, J. Comp. Chem., 16 (1995), pp. 1428–1433
71. M.Saito, *Molecular dynamics simulations of proteins in solution: Artifacts caused by the cutoff approximation*, J. Chem. Phys., 101 (1994), pp. 4055–4061
72. R.M. Levy and E. Gallicchio, *Computer simulations with explicit solvent: Recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects*, Annu. Rev. Phys. Chem., 49 (1998), pp. 531–567
73. A.W. Appel, *An efficient program for many-body simulations*, SIAM J. Sci Stat. Comput., 6 (1985), pp. 85–103
74. J. Barnes and P. Hut, *A hierarchical O(N log N) force calculation algorithm*, Nature, 324 (1986), pp. 446–449
75. Z.-H. Duan and R. Krasny, *An adaptive treecode for computing nonbonded potential energy in classical molecular systems*, J. Comp. Chem., 21 (2000), pp. 1–12
76. R.W. Hockney and J.W. Eastwood, *Computer simulation using particles*, McGraw-Hill, New York, 1981
77. T. Darden, D. York and L. Pedersen, *Particle mesh Ewald: an N*log(N) method for computing Ewald sums*, J. Chem. Phys., 98 (1993), pp. 10089–10092
78. Z.-H. Duan and R. Krasny, *An Ewald summation based multipole method*, J. Chem. Phys., 113 (2000), pp. 3492–3495 `http://www.math.lsa.umich.edu/~zduan/math/`
79. E. Barth and T. Schlick, *Overcoming stability limitations in biomolecular dynamics: Combining force splitting via extrapolation with Langevin dynamics in LN*, J. Chem. Phys., 109 (1998), pp. 1617-1632
80. T. Bishop, R. Skeel and K. Schulten, *Difficulties with multiple timestepping and the fast multipole algorithm in molecular dynamics*, J. Comp. Chem., 18 (1997), pp. 1785–1791
81. P. Procacci, M. Marchi and G. Martyna, *Electrostatic calculations and multiple time scales in molecular dynamics simulation of flexible molecular systems*, J. Chem. Phys., 108 (1998), pp. 8799–8803

82. A.K. Soper and M.G. Phillips, *A new determination of the structure of water at 25C*, Chem. Phys., 107 (1986), pp. 47–60
83. J. Zhou, S. Reich and B.R. Brooks, *Elastic molecular dynamics with self-consistent flexible constraints*, J. Chem. Phys., 112 (2000), pp. 7919–7929
84. M. Tuckerman, B.J. Berne and G.J. Martyna, *Reversible multiple time scale molecular dynamics*, J. Chem. Phys., 97 (1992) pp. 1990–2001
85. M. Tuckerman and B.J. Berne, *Molecular dynamics in systems with multiple timescales — systems with stiff and soft degrees of freedom and with short and long-range forces*, J. Chem. Phys., 95 (1991), pp. 8362–8364
86. H. Grubmüller, H. Heller, A. Windemuth and K. Schulten, *Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions*, Mol. Sim., 6 (1991), pp. 121–142
87. R. Zhou and B.J. Berne, *A new molecular dynamics method combining the reference system propagator algorithm with a fast multipole method for simulating proteins and other complex systems*, J. Chem. Phys., 103 (1995), pp. 9444–9459
88. P. Procacci and M. Marchi, *Taming the Ewald sum in molecular dynamics simulations of solvated proteins via a multiple time step algorithm*, J. Chem. Phys., 104 (1996), pp. 3003–3012
89. P. Procacci and B. Berne, *Computer simulation of solid C_{60} using multiple time-step algorithms*, J. Chem. Phys., 101 (1994), pp. 2421–2431
90. W.C. Still, A. Tempczyk R.C. Hawley and T. Hendrickson, *Semianalytical treatment of solvation for molecular mechanics and dynamics*, J. Am. Chem. Soc., 112 (1990), pp. 6127–6129
91. D. Qiu, P.S. Shenkin, F.P. Hollinger and W.C. Still, *The GB/SA continuum model for solvations. A fast analytical method for the calculation of approximate Born radii*, J. Phys. Chem. A, 101 (1997), pp. 3005–3014
92. P.J. Kraulis, *MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures*, J. of Appl. Cryst., 24 (1991), pp. 946-950. http://www.avatar.se/molscript/
93. H.B. Thompson, *XXXXX* J. Chem. Phys., 47 (1967), pp. 3407–XXX
94. J. Hermans, *Rationalization of Molecular Models*, Methods in Enzymology, 115 (1985), pp. 171–189
95. H. Bekker, H.J.C. Berendsen and W.F. van Gunsteren, *Force and virial of torsional-angle dependent potentials*, J. Comput. Chem., 16 (1995), 527–533
96. G. Zhang and T. Schlick, *LIN: A new algorithm combining implicit integration and normal mode techniques for molecular dynamics*, J. Comp. Chem., 14 (1993), pp. 1212-1233
97. D.J. Tobias and C.L. Brooks III, *Molecular dynamics with internal coordinate constraints*, J. Chem. Phys., 89 (1988), pp. 5115–5127
98. E. Barth, M. Mandziuk and T. Schlick, *A separating framework for increasing the timestep in molecular dynamics*, in Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, Volume 3, chapter 4, W.F. van Gunsteren, P.K. Weiner and A. J. Wilkinson, Editors, ESCOM, Leiden, The Netherlands, 1996
99. H.B. Grubmüller and Paul Tavan, *Molecular dynamics of conformational substates for a simplified protein model*, J. Chem. Phys., 101 (1994), pp. 5047–5057
100. J.D. Honeycutt and D. Thirumalai, *Metastability of the folded states of globular proteins*, Proc. Natl. Acad. Sci. USA, 87 (1990), pp. 3526-3529

101. J.-E. Shea, Y.D. Nochomovitz, Z. Guo and C.L Brooks, III, *Exploring the space of protein folding Hamiltonians: The balance of forces in a minimalist β-barrel model*, J. Chem. Phys, 109 (1998), pp. 2895–2903