7 Molecular Dynamics Simulation: From "Ab Initio" to "Coarse Grained"

Chris Lorenz[1] · *Nikos L. Doltsinis*[1,2]
[1] Department of Physics, King's College London, London, UK
[2] Institute for Solid State Theory, Department of Physics, University of Münster, Münster, Germany

Abstract: This chapter provides an overview of different hierarchical levels of molecular dynamics (MD) simulations spanning a wide range of time and length scales – from first principles approaches via classical atomistic methods to coarse graining techniques. The theoretical background of the most widely used methods and algorithms is briefly reviewed and practical instructions are given on the choice of input parameters for an actual computer simulation. In addition, important postprocessing procedures such as data analysis and visualization are discussed.

Introduction

Molecular dynamics (MD) simulations in their different flavors are widely used in a large variety of research areas of Computational Physics and Chemistry. They represent a powerful tool to study the motion of atoms in molecules, liquids, and solids. The term MD typically refers to the propagation of point particles – atomic nuclei or effective particles combining several nuclei – according to the laws of classical mechanics. In particular, the forces acting on the particles are calculated "on the fly" only at discrete points along the trajectory. Following this definition, we discuss in this chapter Ab Initio MD (AIMD), i.e., the atomic forces are calculated from first principles, classical atomistic MD using analytical empirical interaction potentials (force-fields), which sometimes is referred to as force-field molecular dynamics, and coarse grain MD using analytical empirical potentials between effective particles representing groups of atoms. We exclude methods which go beyond classical nuclei, such as path integral MD (Tuckerman 2002; Tuckerman and Hughes 1998; Tuckerman et al. 1993) and wavepacket dynamics (Balint-Kurti 2008; Worth et al. 2008), or beyond the Born–Oppenheimer approximation (Doltsinis and Marx 2002a, b). This overview, furthermore, leaves out the vast area of semi-empirical methods (see for instance Bredow and Jug [2005] for a recent review) including self-consistent charge density functional tight-binding (SCC-DFTB) (Elstner et al. 1998) and empirical valence-bond (EVB) theory (Aqvist and Warshel 1993; Shurki and Warshel 2003; Warshel 1991, 2003).

The aim of this chapter is to offer practical guidance on how to choose the appropriate technique for a particular physical problem, how to set up a simulation, and how to analyze and visualize the output. In addition it should provide the theoretical background required to become a competent user of the available simulation software packages.

Choosing the Right Method

When choosing which type of molecular dynamics simulations to perform, it is important to understand the capabilities of each technique. The differences in the various methods are basically dependent on the detail with which each one models a physical system.

The most detailed molecular dynamics simulation technique is the ab-initio (quantum) molecular dynamics simulation approach that explicitly models the electrons of the particles within the system. Whereas, force-field molecular dynamics simulations model the nuclear interactions of the particles within the system, and therefore do not explicitly model each electron. Then the method that incorporates the least amount of detail is that of coarse grain molecular dynamics models where multiple particles are grouped together before being represented by a single interaction "bead."

Therefore, quantum molecular dynamics simulations will generate the most detailed modeling of interatomic interactions as electrons are the basis of all such interactions. Quantum simulations allow for certain phenomena like electron transport within a system to be modeled, which cannot be modeled in force-field or coarse grain molecular dynamics simulations because they do not explicitly model electrons. Also, in order to model chemical reactions, quantum simulations are the most accurate approach (Note: there have been force-field and coarse-grain molecular dynamics simulations that have modeled the formation and breaking of bonds, but some a priori knowledge must then be included in the model to allow for the reaction to take place). The major limitations of quantum simulations is that the simulations are very computationally intensive, which results in the capability to model only small system sizes ($\sim 10^2$ particles) and time ($\sim 10^{-12}$ s). Thus the systems that can be modeled are limited to small molecules or portions of larger molecules (i.e., specific amino acids within a protein).

Force-field molecular dynamics simulations offer the ability to model molecules at the particle level. Often, information from quantum simulations is used to develop the empirical equations (force-field) that are used to govern the interactions between particles. Because force-field molecular dynamics simulations use less detail than the quantum simulations, they are able to model systems that are significantly larger in size ($\sim 10^6$ particles) for a longer period of time ($< 10^{-6}$ s). Therefore, measuring the structural, mechanical, and/or transport properties of medium to large sized systems (i.e., proteins, functionalized nanoparticles, . . .) is possible.

Finally, coarse grain molecular dynamics simulations reduce the number of degrees of freedom within the simulated system even further by grouping several atoms into one interaction bead. Therefore, even larger system sizes and times (on the order of seconds) are accessible via these simulations. Several of the same properties measured via force-field molecular dynamics simulations can be measured with coarse grain molecular dynamics simulations (i.e., structural, mechanical, and transport properties). However, due to the reduced detail in the models of the molecules, it is not possible to investigate specific chemical interactions within a system, such as hydrogen bonding.

Once you have chosen the appropriate method for the particular system and property to be investigated, the next choice is what simulation package to use. For classical MD simulations, there are several free molecular dynamics packages that can be found on the web including DL_POLY (Smith et al. 2002; Todorov and Smith 2009), GROMACS (van der Spoel et al. 2005a, b), HOOMD (Anderson et al. 2008; HOOMD 2009), LAMMPS (LAMMPS 2010; Plimpton 1995), MOLDY (Refson 2000, 2001), and NAMD (Bhandarkar et al. 2009; Phillips et al. 2005b), and there are also commercial packages including AMBER (Case et al. 2005, 2008), CHARMM (Brooks et al. 2009; CHARMM 2009), and GROMOS (GROMOS 2007; Scott et al. 1999). Generally, these codes can be divided into those that are mostly used for simulations of biological systems (AMBER, CHARMM, GROMACS, GROMOS, NAMD) and those that are more general simulation packages (HOOMD, LAMMPS, MOLDY). When choosing between these options, an important criterion is to choose a code that you feel comfortable using. Outside of comfort, another aspect to take into consideration is that packages will differ in the features they offer and the additional tools to perform analysis (usually lists of analysis tools can be found in the packages' documentation).

For AIMD simulations, the user may choose from a large number of codes, for instance, ABINIT (2010; Aulbur et al. 2000), CASTEP (2009; Clark et al. 2005; Segall et al. 2002), CONQUEST (2009; Bowler et al. 2006), CP2K (Hutter et al. 2009; VandeVondele et al. 2005, 2006), CPMD (Marx and Hutter 2000, 2009; Parrinello et al. 2008), CP-PAW (2006; Blochl 1994; Blochl et al. 2003), DACAPO (2006), FHI98md (2002; Bockstedte et al. 1997), NWChem

(2008; Kendall et al. 2000), ONETEP (2005; Skylaris et al. 2005), PINY (2005), PWscf (2009; Giannozzi et al. 2009), QuantumEspresso (2009; Giannozzi et al. 2009), SIESTA (2010; Artacho et al. 2008; Soler et al. 2002), S/PHI/nX (2009; Boeck 2009), or VASP (2009; Kresse and Furthmüller 1996).

Theoretical Background

Born–Oppenheimer Approximation

Let us begin by introducing our nomenclature and by reviewing some well-known basic relations within the Schrödinger formulation of quantum mechanics. A complete, nonrelativistic, description of a dynamic system of N atoms having the positions $\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_I, \ldots, \mathbf{R}_N\}$ with n electrons located at $\mathbf{r} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_n\}$ would involve solving the time-dependent Schrödinger equation

$$\mathcal{H}\Phi(\mathbf{r}, \mathbf{R}; t) = i\hbar\frac{\partial}{\partial t}\Phi(\mathbf{r}, \mathbf{R}; t), \tag{7.1}$$

with the total Hamiltonian

$$\mathcal{H}(\mathbf{r}, \mathbf{R}) = \mathcal{T}(\mathbf{R}) + \mathcal{T}(\mathbf{r}) + \mathcal{V}_{nn}(\mathbf{R}) + \mathcal{V}_{ne}(\mathbf{r}, \mathbf{R}) + \mathcal{V}_{ee}(\mathbf{r}), \tag{7.2}$$

being the sum of kinetic energy of the atomic nuclei,

$$\mathcal{T}(\mathbf{R}) = -\frac{\hbar^2}{2}\sum_{I=1}^{N}\frac{\nabla_I^2}{M_I}, \tag{7.3}$$

kinetic energy of the electrons,

$$\mathcal{T}(\mathbf{r}) = -\frac{\hbar^2}{2m_e}\sum_{i=1}^{n}\nabla_i^2, \tag{7.4}$$

internuclear repulsion,

$$\mathcal{V}_{nn}(\mathbf{R}) = \frac{e^2}{4\pi\epsilon_0}\sum_{I=1}^{N-1}\sum_{J>I}^{N}\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}, \tag{7.5}$$

electronic–nuclear attraction,

$$\mathcal{V}_{ne}(\mathbf{r}, \mathbf{R}) = -\frac{e^2}{4\pi\epsilon_0}\sum_{I=1}^{N}\sum_{i=1}^{n}\frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}, \tag{7.6}$$

and interelectronic repulsion,

$$\mathcal{V}_{ee}(\mathbf{r}) = \frac{e^2}{4\pi\epsilon_0}\sum_{i=1}^{n-1}\sum_{j>i}^{n}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{7.7}$$

Here, M_I and Z_I denote the mass and atomic number of nucleus I; m_e and e are the electronic mass and elementary charge, and ϵ_0 is the permittivity of vacuum. The nabla operators ∇_I and ∇_i act on the coordinates of nucleus I and electron i, respectively. The total wavefunction $\Phi(\mathbf{r}, \mathbf{R}; t)$ simultaneously describes the motion of both electrons and nuclei.

The Born–Oppenheimer approximation (Doltsinis and Marx 2002b; Kołos 1970; Kutzelnigg 1997) separates nuclear and electronic motion based on the assumption that the much faster electrons adjust their positions instantaneously to the comparatively slow changes in

nuclear positions. The electronic problem is then reduced to the time-independent (electronic) Schrödinger equation for clamped nuclei,

$$\mathcal{H}_{el}(\mathbf{r}; \mathbf{R}) \Psi_k(\mathbf{r}; \mathbf{R}) = E_k(\mathbf{R}) \Psi_k(\mathbf{r}; \mathbf{R}), \tag{7.8}$$

where $\mathcal{H}_{el}(\mathbf{r}; \mathbf{R})$ is the electronic hamiltonian,

$$\mathcal{H}_{el}(\mathbf{r}, \mathbf{R}) = \mathcal{T}(\mathbf{r}) + \mathcal{V}_{nn}(\mathbf{R}) + \mathcal{V}_{ne}(\mathbf{r}, \mathbf{R}) + \mathcal{V}_{ee}(\mathbf{r}), \tag{7.9}$$

and $\Psi_k(\mathbf{r}; \mathbf{R})$ is the electronic wavefunction of state k. Meanwhile, nuclear motion is described by

$$\left[\mathcal{T}(\mathbf{R}) + E_k(\mathbf{R}) \right] \chi_k = i\hbar \frac{\partial}{\partial t} \chi_k \tag{7.10}$$

with the nuclear wavefunction $\chi_k(\mathbf{R}, t)$ evolving on the potential energy surface $E_k(\mathbf{R})$ of the electronic state k. The total wavefunction is then the direct product of the electronic and the nuclear wavefunction,

$$\Phi(\mathbf{r}, \mathbf{R}; t) = \Psi_k(\mathbf{r}, \mathbf{R}) \chi_k(\mathbf{R}, t) \tag{7.11}$$

In the classical limit (Doltsinis and Marx 2002b), the nuclear wave equation (❯ 7.10) is replaced by Newton's equation of motion

$$M_I \ddot{\mathbf{R}}_I = -\nabla_I E_k \tag{7.12}$$

For a great number of physical situations, the Born–Oppenheimer approximation can be safely applied. On the other hand, there are many important chemical phenomena such as charge transfer and photoisomerization reactions, whose very existence is due to the inseparability of electronic and nuclear motion. Inclusion of nonadiabatic effects is beyond the scope of this chapter and the reader is referred to the literature (e.g., Doltsinis 2006; Doltsinis and Marx 2002b) for more details.

The above approximations form the basis of conventional molecular dynamics, ❯ Eqs. 7.12 together with ❯ 7.8 being the working equations. Thus, in principle, a classical trajectory calculation merely amounts to integrating Newton's equations of motion (❯ 7.12). In practice, however, this deceptively simple task is complicated by the fact that the stationary Schrödinger equation (❯ 7.8) cannot be solved exactly for any many-electron system. The potential energy surface therefore has to be approximated using ab initio electronic structure methods or empirical interaction potentials (so-called force-field molecular dynamics Sutmann [2002] and Allen and Tildesley [1987]). The former approach, usually referred to as ab initio molecular dynamics (AIMD), will be the subject of section "Ab Initio Molecular Dynamics," while the latter – force-field molecular dynamics – will be discussed in section "Classical Molecular Dynamics."

Ab Initio Molecular Dynamics

In the following, we shall focus on first principles molecular dynamics methods. Due to the high computational cost associated with ab initio electronic structure calculations of large molecules, computation of the entire potential energy surface prior to the molecular dynamics simulation is best avoided. A more efficient alternative is the evaluation of electronic energy and nuclear forces "on the fly" at each step along the trajectory.

Born–Oppenheimer Molecular Dynamics

In the so-called Born–Oppenheimer implementation of such a scheme (Marx and Hutter 2000), the nuclei are propagated by integration of ❯ Eq. 7.12, where the exact energy E_k is replaced with the eigenvalue, $\tilde{E}_k$, of some approximate electronic Hamiltonian, $\tilde{\mathcal{H}}_{el}$, which is calculated at each time step. For the electronic ground state, i.e., $k = 0$, the use of Kohn–Sham (KS) density functional theory (Dreizler and Gross 1990; Parr and Yang 1989) has become increasingly popular.

Car–Parrinello Molecular Dynamics

In order to further increase computational efficiency, Car and Parrinello have introduced a technique to bypass the need for wavefunction optimization at each molecular dynamics step (Car and Parrinello 1985; Marx and Hutter 2000). Instead, the molecular wavefunction is dynamically propagated along with the atomic nuclei according to the equations of motion

$$M_I \ddot{\mathbf{R}}_I = -\nabla_I \langle \Psi_k | \tilde{\mathcal{H}}_{el} | \Psi_k \rangle \tag{7.13}$$

$$\mu_i \ddot{\psi}_i = -\frac{\delta}{\delta \psi_i^\star} \langle \Psi_k | \tilde{\mathcal{H}}_{el} | \Psi_k \rangle + \sum_j \lambda_{ij} \psi_j, \tag{7.14}$$

where the KS one-electron orbitals ψ_i are kept orthonormal by the Lagrange multipliers λ_{ij}. These are the Euler–Lagrange equations

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathcal{L}}{\partial \dot{q}} = \frac{\partial \mathcal{L}}{\partial q}, \quad (q = \mathbf{R}_I, \ \psi_i^\star) \tag{7.15}$$

for the Car–Parrinello Lagrangian (Car and Parrinello 1985)

$$\mathcal{L} = \sum_I \frac{1}{2} M_I \dot{\mathbf{R}}_I^2 + \sum_i \frac{1}{2} \mu_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle - \langle \Psi_k | \tilde{\mathcal{H}}_{el} | \Psi_k \rangle + \sum_{ij} \lambda_{ij} (\langle \psi_i | \psi_j \rangle - \delta_{ij}) \tag{7.16}$$

that is formulated here for an arbitrary electronic state Ψ_k, an arbitrary electronic Hamiltonian $\tilde{\mathcal{H}}_{el}$, and an arbitrary basis (i.e., without invoking the Hellmann–Feynman theorem).

Classical Molecular Dynamics

While first-principles molecular dynamics simulations deal with the electrons in a system, this results in a large number of particles that must be considered and therefore the calculations become significantly time-consuming. Classical molecular dynamics ignore electronic motions and calculate the energy of a system as a function of the nuclear positions only, and therefore are used to simulate larger, less detailed systems for larger timescales. The successive configurations of the system are generated by solving the differential equations that constitute Newton's second law (❯ Eq. 7.12):

$$\frac{d^2 X_I}{dt^2} = \frac{F_{X_I}}{M_I} \tag{7.17}$$

This equation describes the motion of a particle of mass M_I along one dimension (X_I), where F_{X_I} is the force on the particle in that dimension. The solution of these differential equations

results in a trajectory that specifies how the positions and velocities of the particles in the system vary with time.

In realistic models of intermolecular interactions, the force on particle I changes whenever particle I changes its position or whenever another atom with which particle I interacts changes its position. Therefore the motions of all the particles are coupled together, which results in a many-body problem that cannot be solved analytically. Therefore finite difference methods are used to integrate the equations of motion.

Generally, the integration of ❯ Eq. 7.17 is broken into consecutive steps that are conducted at different times t that are separated by increments of δt, which is generally referred to as the time step. First, the total force on each particle in the system at time t is calculated as the vector sum of its interactions with other particles.

Then, assuming the force is constant over the course of the time step, the accelerations of the particles are calculated, which are then combined with positions and velocities of the particles at time t to determine the positions and velocities at time $t + \delta t$. Finally, the forces on the particles in their new positions are determined, and then new accelerations, positions, and velocities are determined at $t + 2\delta t$ and so on.

A common approach in the various finite difference methods used to integrate the equations of motions for classical molecular dynamics simulations is that it is assumed that the positions, velocities, and accelerations (as well as all other dynamic properties) can be approximated using Taylor series expansions:

$$\mathbf{R}(t + \delta t) = \mathbf{R}(t) + \delta t \mathbf{V}(t) + \frac{1}{2}\delta t^2 \mathbf{A}(t) + \frac{1}{6}\delta t^3 \mathbf{B}(t) + \frac{1}{24}\delta t^4 \mathbf{C}(t) + \dots \quad (7.18)$$

$$\mathbf{V}(t + \delta t) = \mathbf{V}(t) + \delta t \mathbf{A}(t) + \frac{1}{2}\delta t^2 \mathbf{B}(t) + \frac{1}{6}\delta t^3 \mathbf{C}(t) + \dots \quad (7.19)$$

$$\mathbf{A}(t + \delta t) = \mathbf{A}(t) + \delta t \mathbf{B}(t) + \frac{1}{2}\delta t^2 \mathbf{C}(t) + \dots \quad (7.20)$$

where $\mathbf{R}$ is the position, $\mathbf{V}$ is the velocity, $\mathbf{A}$ is the acceleration, and $\mathbf{B}$ and $\mathbf{C}$ are the third and fourth derivatives of the positions with respect to time, respectively.

Verlet Algorithm

One of the most widely used finite difference methods in classical molecular dynamics simulations is the Verlet algorithm (Verlet 1967). In the Verlet algorithm, the positions and accelerations at time t and the positions from the previous time step $\mathbf{R}(t - \delta t)$ are used to calculate the updated positions $\mathbf{R}(t + \delta t)$ using the equation:

$$\mathbf{R}(t + \delta t) = 2\mathbf{R}(t) - \mathbf{R}(t - \delta t) + \delta t^2 \mathbf{A}(t). \quad (7.21)$$

While the velocities do not explicitly appear in ❯ Eq. 7.21, they can be calculated from the difference in position over the entire time step:

$$\mathbf{V}(t) = \frac{|\mathbf{R}(t + \delta t) - \mathbf{R}(t - \delta t)|}{2\delta t} \quad (7.22)$$

or the difference in position over a half time step $(t + \frac{1}{2}\delta t)$:

$$\mathbf{V}(t + \frac{1}{2}\delta t) = \frac{|\mathbf{R}(t + \delta t) - \mathbf{R}(t)|}{\delta t} \quad (7.23)$$

The fact that the velocities are not explicitly represented in the Verlet algorithm is one of the drawbacks to this method in that no velocities are available until the positions have been determined at the next time step. Also, in order to calculate the position of particles at $t = \delta t$, it is necessary to determine the positions at $t = -\delta t$ since the algorithm requires the position at time $t - \delta t$ to calculate the position at time $t + \delta t$. Often, this drawback is overcome by using the Taylor series to calculate $\mathbf{R}(-\delta t) = \mathbf{R}(0) - \delta t \mathbf{V}(0) + \frac{1}{2} \delta t^2 \mathbf{A}(t)| + \ldots$. A final drawback of the Verlet algorithm is that there may be a loss of precision in the resulting trajectories that result from the fact that the positions are calculated by adding a small term $(\delta t^2 \mathbf{A}(t)$, to the difference of two larger terms $(2\mathbf{R}(t)$ and $\mathbf{R}(t - \delta t))$ in ❯ Eq. 7.21.

"Leap-Frog" Algorithm

In an attempt to improve upon the original Verlet algorithm, several variations have been developed. The leap-frog algorithm (Hockney 1970) is one of the variations that uses the following equations to update the positions:

$$\mathbf{R}(t + \delta t) = \mathbf{R}(t) + \delta t \mathbf{V}(t + \frac{1}{2}\delta t), \tag{7.24}$$

and the velocities:

$$\mathbf{V}(t + \frac{1}{2}\delta t) = \mathbf{V}(t - \frac{1}{2}\delta t) + \delta t \mathbf{A}(t). \tag{7.25}$$

In the leap-frog algorithm, the velocities $\mathbf{V}(t + \frac{1}{2}\delta t)$ are first calculated from the velocities at time $t - \frac{1}{2}\delta t$ and the accelerations at time t using ❯ Eq. 7.24. Then the positions $\mathbf{R}(t + \delta t)$ are calculated from the velocities $\mathbf{V}(t + \frac{1}{2}\delta t)$ and the positions $\mathbf{R}(t)$ using ❯ Eq. 7.25. The algorithm gets its name from the fact that the velocities are calculated in manner such that they "leap-frog" over the positions to give their values $t - \frac{1}{2}\delta t$. Then the positions are calculated such that they "leap-frog" over the velocities, and then the algorithm continues.

The "leap-frog" algorithm improves upon the standard Verlet algorithm in that the velocity is explicitly included in the calculations and also the "leap-frog" algorithm does not require the calculation of the differences of large numbers so the precision of the calculation should be improved. However, the fact that the calculated velocities and positions are not synchronized in time results in the fact that the kinetic energy contribution to the total energy cannot be calculated for the time at which the positions are defined. In response to this shortcoming in the "leap-frog" algorithm, a formalism to calculate the velocities at time t has been developed that follows

$$\mathbf{V}(t) = \frac{[\mathbf{V}(t + \frac{\delta t}{2}) + \mathbf{V}(t - \frac{\delta t}{2})]}{2} \tag{7.26}$$

Velocity Verlet Algorithm

The velocity Verlet method (Swope et al. 1982), which is a variation of the standard Verlet method, calculates the positions, velocities, and accelerations at the same time by using the following equations:

$$\mathbf{R}(t + \delta t) = \mathbf{R}(t) + \delta t \mathbf{V}(t) + \frac{1}{2} \delta t^2 \mathbf{A}(t) \tag{7.27}$$

$$\mathbf{V}(t + \delta t) = \mathbf{V}(t) + \frac{1}{2}\delta t [\mathbf{A}(t) + \mathbf{A}(t + \delta t)]. \tag{7.28}$$

The velocity Verlet method is a three-stage algorithm because the calculation of the new velocities (❯ Eq. 7.28) requires both the acceleration at time t and at time $t + \delta t$. Therefore, first, the positions at $t + \delta t$ are calculated using ❯ Eq. 7.27 and the velocities and accelerations at time t. The velocities at time $t + \frac{1}{2}\delta t$ are then calculated using

$$\mathbf{V}(t + \frac{1}{2}\delta t) = \mathbf{V}(t) + \frac{1}{2}\delta t \mathbf{A}(t). \tag{7.29}$$

Then the forces are computed from the current positions, which results in being able to calculate $\mathbf{A}(t + \delta t)$. Then the final step consists of calculating the velocities at time $t + \delta t$ using

$$\mathbf{V}(t + \delta t) = \mathbf{V}(t + \frac{1}{2}\delta t) + \frac{1}{2}\delta t \mathbf{A}(t + \delta t). \tag{7.30}$$

Therefore, the velocity Verlet allows for the velocities and positions to be calculated in a time-synchronized manner, and thus allows for the kinetic energy contribution of the total energy. Also, the precision of the results will be improved upon those from the standard Verlet algorithm as there are no differences of large numbers within the formalism of the method.

The selection of the best time integration method for a given problem and the size of the time step to use will be discussed in section "Setting the Time Step."

Hybrid Quantum/Classical (QM/MM) Molecular Dynamics

The ab initio and classical simulation techniques discussed in the previous sections can be viewed as complementary. While AIMD is capable of dealing with electronic processes such as chemical reactions, charge transfer, and electronic excitations, its applicability is limited to systems of modest size, precluding its use in complex, large-scale biochemical simulations. Classical MD, on the other hand, can describe much larger systems on longer timescales, but misses any of the above-mentioned electronic effects, e.g., bond breaking and formation. The basic idea of the QM/MM approach is to combine the strengths of the two methods treating a chemically active region at the quantum level and the environment using molecular mechanics (i.e., a force-field). There are several excellent review articles on the QM/MM method in the literature (Senn and Thiel 2009; Thiel 2009).

Partitioning Schemes

The entire system, **S**, is partitioned into a chemically active inner region, **I**, and a chemically inert outer region, **O**. If the border between these regions cuts through chemical bonds, so-called link atoms, **L**, are usually introduced to cap the inner region (see section "Bonds Across the QM/MM Boundary").

Subtractive Scheme
In a subtractive scheme, the total energy, $E_{\mathrm{QM/MM}}^{\mathbf{S}}$, of the entire system,

$$E_{\mathrm{QM/MM}}^{\mathbf{S}} = E_{\mathrm{MM}}^{\mathbf{S}} + E_{\mathrm{QM}}^{\mathbf{I,L}} - E_{\mathrm{MM}}^{\mathbf{I,L}} \tag{7.31}$$

is calculated from three separate energy contributions: (1) the MM energy of the entire system, E_{MM}^{S}, (2) the QM energy of the active region (including any link atoms), $E_{QM}^{I,L}$, (3) the MM energy of the active region $E_{MM}^{I,L}$.

The role of the third term in ❯ Eq. 7.31 is to avoid double counting and to correct for any artifacts caused by the link atoms. For the latter to be effective, the force-field has to reproduce the quantum mechanical forces reasonably well in the link region.

Additive Scheme

In an additive scheme, the total energy of the system is given by

$$E_{QM/MM}^{S} = E_{MM}^{O} + E_{QM}^{I,L} + E_{QM-MM}^{I,O} \tag{7.32}$$

The difference to the subtractive scheme is that here a pure MM calculation is performed for only the outer region and the interaction between QM and MM regions is achieved by an explicit coupling term,

$$E_{QM-MM}^{I,O} = E_{QM-MM}^{bond} + E_{QM-MM}^{vdW} + E_{QM-MM}^{el} \tag{7.33}$$

where E_{QM-MM}^{bond}, E_{QM-MM}^{vdW}, E_{QM-MM}^{el}, are bonded, van der Waals, and electrostatic interaction energies, respectively.

The simplest way to treat electrostatic interactions between the **I** and **O** subsystems is to assign fixed electric charges to all **I** atoms (*mechanical embedding*). In this case the QM problem is solved for the isolated subsystem **I** without taking into account the effects of the surrounding atomic charges in **O**. The majority of implementations use an *electrostatic embedding* scheme in which the MM point charges of region **O** are incorporated in the QM Hamiltonian through a QM-MM coupling term,

$$\hat{H}_{QM-MM}^{el} = -\sum_{i}^{n} \sum_{\alpha \in \mathbf{O}} \frac{q_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \sum_{I \in \mathbf{I+L}} \sum_{\alpha \in \mathbf{O}} \frac{q_\alpha Z_I}{|\mathbf{R}_I - \mathbf{R}_\alpha|} \tag{7.34}$$

where q_α are the MM point charges at positions $\mathbf{R}_\alpha$ (all other symbols as defined in section "Born–Oppenheimer Approximation"). In this way, the electronic structure of the QM region adjusts to the moving MM charge distribution. A problem that arises when an MM point charge is in close proximity to the QM electron cloud is overpolarization of the latter, sometimes referred to as "spill-out" effect. This can be avoided by modifying the Coulomb potential in the first term of ❯ Eq. 7.34 at short range (see for instance Laio et al. 2002).

At present, in all commonly used partitioning schemes, the partitions remain *fixed* over time, i.e., an MM atom cannot turn into a QM atom and *vice versa*. This can present a serious limitation, for instance, in the case of solvent diffusion through the chemically active region. A number of *adaptive* partitioning methods have been proposed to remedy this problem (Bulo et al. 2009; Heyden et al. 2007; Hofer et al. 2005; Kerdcharoen et al. 1996; Kerdcharoen and Morokuma 2002); however the computational overhead is enormous.

Bonds Across the QM/MM Boundary

Partitioning the total system into QM and MM regions in such a way that cuts chemical bonds is best avoided. However, in many cases this is inevitable. Then one has to make sure that

any atoms participating in chemical reactions are at least three bonds away from boundary. Furthermore it is preferable to cut a bond that is unpolar and not part of a conjugated chain.

Link Atoms

Cutting a single covalent bond will create a dangling bond which must be capped by a so-called link atom; in most applications a hydrogen atom is chosen. In the QM calculation, the atoms of region **I** together with the link atoms **L** are treated as an isolated molecule in the presence of the point charges of the environment **O**. The original QM–MM bond, cut by the partitioning, is only treated at the MM level.

Boundary Atoms

Boundary atom schemes have been developed to avoid the artifacts introduced by a link atom. The boundary atom appears as a normal MM atom in the MM calculation, while carrying QM features to saturate the QM–MM bond and to mimic the electronic properties of the MM side. The QM interactions are achieved by placing a pseudopotential at the position of the boundary atom, parameterized to reproduce electronic properties of certain chemical end group, e.g., a methyl group in the case of a cut C–C bond. Among the various flavors that have been proposed, the pseudobond method for first principles QM calculations (Zhang 2005, 2006; Zhang et al. 1999) and the pseudopotential approach for plane-wave DFT (Laio et al. 2002) are the most relevant in the present context.

Frozen Localized Orbitals

The basic idea behind the various frozen orbital methods (Amara et al. 2000; Assfeld and Rivail 1996; Assfeld et al. 1998; Day et al. 1996; Ferré et al. 2002; Fornili et al. 2003, 2006a, b; Gao et al. 1998; Garcia-Viloca and Gao 2004; Gordon et al. 2001; Grigorenko et al. 2002; Jensen et al. 1994; Jung et al. 2007; Kairys and Jensen 2000; Loos and Assfeld 2007; Monard et al. 1996; Murphy et al. 2000; Nemukhin et al. 2002, 2003; Philipp and Friesner 1999; Pu et al. 2004a, b, 2005; Sironi et al. 2007; Théry et al. 1994; Warshel and Levitt 1976) is to saturate the cut QM–MM bond by placing on either the MM or the QM atom at the boundary localized orbitals that have been determined in a prior quantum-mechanical SCF calculation on a model molecule containing the bond under consideration. To preserve the properties of the bond, the localized orbitals are then kept fixed in the subsequent QM/MM calculation. Different flavors are the Local SCF (LSCF) method (Assfeld and Rivail 1996; Assfeld et al. 1998; Ferré et al. 2002; Monard et al. 1996; Théry et al. 1994), extremely localized molecular orbitals (ELMOs) (Fornili et al. 2003, 2006b; Sironi et al. 2007), frozen core orbitals (Fornili et al. 2006a), optimized LSCF (Loos and Assfeld 2007), frozen orbitals (Murphy et al. 2000; Philipp and Friesner 1999), generalized hybrid orbitals (Amara et al. 2000; Gao et al. 1998; Garcia-Viloca and Gao 2004; Jung et al. 2007; Pu et al. 2004a, b, 2005), and effective fragment potentials (EFP) (Day et al. 1996; Gordon et al. 2001; Grigorenko et al. 2002; Jensen et al. 1994; Kairys and Jensen 2000; Nemukhin et al. 2002, 2003).

Of the three types of boundary treatment, the link atom method is the simplest both conceptually and in practice, and is hence the most widely used. The boundary atom and in particular the frozen orbital methods can potentially achieve higher accuracy but require careful a priori parametrization and bear limitations on transferability (Senn and Thiel 2009).

Coarse Grain Molecular Dynamics

A large number of important problems in fields that are often studied using molecular dynamics simulations (i.e. soft condensed matter physics, structural biology, chemistry and materials science) take place over a time span of microseconds to seconds and distances of few hundred nanometers to a few microns. However, these time and length scales are still unattainable via quantum or force-field molecular dynamics methods despite significant computational hardware advances (Mervis 2001; Reed 2003; Shirts and Pande 2000) and the development of increasingly powerful software (Lindahl et al. 2001; MacKerell et al. 1998; Phillips et al. 2005a; Wang et al. 2004). Therefore one approach that has been utilized in order to be able to study these complex problems is to reduce the computational demand of the simulation by reducing the number of atoms represented and therefore the degrees of freedom of the simulated system. This procedure of reducing the number of atoms represented in a system is done by grouping atoms together and representing them as a single interaction site and is generally referred to as "coarse graining" of the system. ❯ *Figure 7-1* shows a comparison of the atomistic, united-atom and coarse grain representation.

The "bead-spring" coarse grain model of polymer chains that was created by Kremer and Grest in 1990 has served as the foundation for many of the coarse grain models that have been developed for a wide range of phenomena (at the current date this paper has been cited over 860 times) including various studies of polymers and biomolecules including DNA solutions. Many of the more recent coarse grain models have been developed for biological macromolecules since there are many examples of interesting biophysical phenomena that occur at large length and timescales. The most widely used coarse grain models for biological systems include the generic model of Lipowsky et al. (Goetz et al. 1999; Shillcock and Lipowsky 2002), the solvent-free model of Deserno et al. (Cooke et al. 2005), and the specific models of the Klein group (Shelley et al. 2001), the Voth group (which is called the Multi-Scale Coarse Grain model) (Izvekov and Voth 2005, 2006), and the Marrink group (called the MARTINI force-field) (Marrink et al. 2007). The above coarse grain models have generally been developed for lipid membranes, however there are also coarse grain force-fields for proteins (as reviewed in Tozzini [2005] and some more recent examples Betancourt and Omovie [2009] and Bereau and Deserno [2009]) and DNA (Khalid et al. 2008; Tepper and Voth 2005).

All-atom representation   United-atom representation

Coarse-grain representation
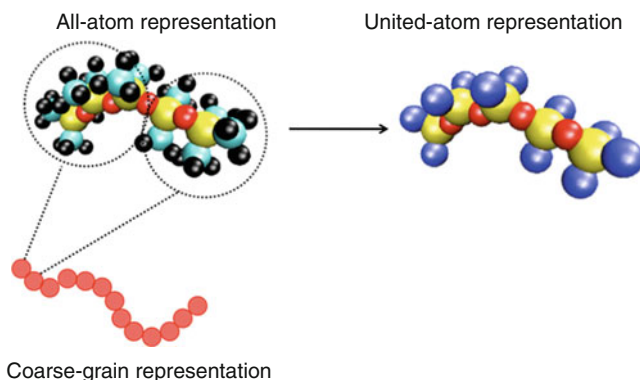
■ Fig. 7-1

Atomistic, united-atom, and coarse grain representations of organic molecules

When developing a coarse grain model for a system, there are two important decisions to be made: (1) how many atoms to combine (coarse grain) into a single interaction site and (2) how to parameterize the coarse grain force-field. In deciding the number of atoms to combine into a single interaction site, one must consider the obvious trade-off of how much detail are you able to sacrifice in order to simulate larger length and/or timescale phenomena and still be able to actually accurately model the phenomena of interest. The least amount of coarse graining that has been used is represented by what is called a "united-atom" representation of a molecule where all "heavy" atoms (generally all non-hydrogen elements in a molecule) are represented and the "light" (i.e., hydrogen) atoms are grouped with the heavy atom to which they are bonded into one interaction site. United atom versions of many of the popular all-atom force-fields listed in section "Classical Force Fields" exist and have been successfully used in several studies. In addition to united-atom models, there are several existing coarse graining methods that will combine different number of atoms together into one interaction site.

In general, coarse grain systems are governed by similar potential terms as are found in atomistic models such as nonbond terms (both pair-wise interactions and electrostatic interactions), bond stretching terms, and then in more sophisticated models even angle and dihedral terms will be included as well. Generally, all specific models are parameterized based on comparison to atomistic simulations and/or detailed experimental data. Effective coarse grain potentials have been extracted from atomistic simulations using inverse Monte Carlo schemes (Elezgaray and Laguerre 2006; Lyubartsev 2005) or force matching approaches (Izvekov and Voth 2005, 2006). Another approach is to develop standard potential functions that are calibrated using thermodynamic data (Marrink et al. 2004). The advantage of the using either the inverse Monte Carlo or force matching schemes is that the resulting force-field will produce a higher level of accuracy and closer resemblance to atomistic simulations. However, these schemes produce force-fields that are useful for a given statepoint and therefore are not transferable. Whereas the advantages of the thermodynamic approach include that it produces a potential that has a broader range of applicability and also the thermodynamic approach does not require atomistic simulations to be done in the first place.

Interaction Potentials/Force Fields

Classical Force Fields

Classical, or empirical, force-fields are generally used to calculate the energy of a system as a function of the nuclear positions of the particles within the system, while ignoring the behavior of the individual electrons. As stated in the section "Born–Oppenheimer Approximation," the Born–Oppenheimer approximation makes it possible to write the energy as a function of the nuclear coordinates. Another approximation that is key to the implementation of classical force-fields is that it is possible to model the relatively complex motion of particles within the system with fairly simple analytical models of inter and intra-molecular interactions. Generally, an empirical force-field consists of terms that model the nonbonded interactions ($E_{nonbond}$), which include both the van der Waals and Coulombic interactions, the bonded interactions (E_{bond}), the angle bending interactions (E_{angle}), and the dihedral (bond rotations) interactions ($E_{dihedral}$):

$$E(\mathbf{R}) = E_{nonbond} + E_{bond} + E_{angle} + E_{dihedral}. \tag{7.35}$$
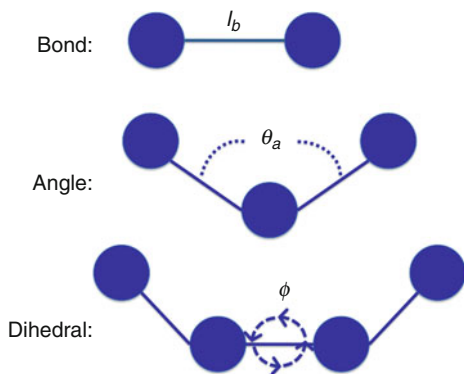
Fig. 7-2
Intramolecular terms of classical force-fields: bond, angle, and dihedral interactions

❯ *Figure 7-2* presents representative cartoons of the bond, angle, and dihedral interactions from a molecular perspective. The form that each of these individual terms takes is dependent on the force-field that you are using. There are several different force-field options available for various systems. The best way to find the most suitable force-field for your specific problem is to conduct a literature and/or internet search in order to find which force-field has the capability to model the molecules you are interested in studying. However, if you are interested in modeling organic/biological molecules, there are several large force-fields that may be a good place to start, including Charmm (MacKerell et al. 1998), OPLS (Jørgensen et al. 1984), Amber (Cornell et al. 1995), and COMPASS (Sun et al. 1998). Likewise, there are several well-known large force-fields that can be used for solids like the BKS potential (van Beest et al. 1990) for oxides and the Embedded Atom Method (EAM) (Daw and Baskes 1983, 1984; Finnis and Sinclair 1984) and Modified Embedded Atom Method (MEAM) (Baskes 1992) force-fields, which are primarily used to model metals. In addition to defining the functional forms used for the various terms in the general potential formulation, a force-field will also define the variables used in the potential which are derived from a combination of quantum simulation results and experimental observations.

In the following sections, each of the terms in ❯ Eq. 7.35 will be discussed further and typical functional forms that are used in the previously mentioned force-fields and others to represent each term will be shown.

We limit the discussion to simple non-polarizable force fields in which the individual atoms carry fixed charges. They capture many-body-effects such as electronic polarization only in an effective way. More sophisticated polarizable force fields have been developed over the past two decades (see for instance Ponder et al. [2010] and references therein) however they are computationally substantially more demanding.

Nonbonded Interactions

There are two general forms of nonbonded interactions that need to be accounted for by a classical force-field: (1) the van der Waals (vdw) interactions and (2) the electrostatic interactions.

van der Waals Interactions

In order to model the van der Waals interactions, we need a simple empirical expression that is not computationally intensive and that models both the dispersion and repulsive interactions that are known to act upon atoms and molecules. The most commonly used functional form of van der Waals energy (E_{vdW}) in classical force-fields is the Lennard-Jones 12-6 function that has the form:

$$E_{vdW}(\mathbf{R}) = \sum_{I>J} 4\epsilon_{IJ}\left[\left(\frac{\sigma_{IJ}}{R_{IJ}}\right)^{12} - \left(\frac{\sigma_{IJ}}{R_{IJ}}\right)^{6}\right], \tag{7.36}$$

where σ_{IJ} is the collision diameter and ϵ_{IJ} is the well depth of the interaction between atoms I and J. Both σ_{IJ} and ϵ_{IJ} are adjustable parameters that will have different values to describe the interactions between different pairs of particles (i.e., the values of σ and ϵ used to describe the interaction between two carbon atoms are different than the values of σ and ϵ used to describe the interaction between a carbon and an oxygen).

❯ Equation 7.36 models both the attractive part (the R^{-6} term) and the repulsive part (the R^{-12} term) of the nonbonded interaction. Other formulations of the Lennard-Jones nonbond potential commonly have the same power law description of the attractive part of the potential, but will have different power law dependence for the repulsive part of the interaction, such as the Lennard-Jones 9-6 function:

$$E_{vdW}(\mathbf{R}) = \sum_{I>J} 4\epsilon_{IJ}\left[\left(\frac{\sigma_{IJ}}{R_{IJ}}\right)^{9} - \left(\frac{\sigma_{IJ}}{R_{IJ}}\right)^{6}\right]. \tag{7.37}$$

When the nonbond interactions of a system that contains multiple particle types and multiple molecules are modeled using a Lennard-Jones type nonbond potential, it is necessary to be able to define the values of σ and ϵ that apply to the interaction between particles of type I and J. The parameters for these cross interactions are generally found using one of the two following mixing rules. One common mixing rule is the Lorentz-Berthelot rule where the value of σ_{IJ} is found from the arithmetic mean of the two pure values and the value of ϵ_{IJ} is the geometric mean of the two pure values:

$$\sigma_{IJ} = \frac{(\sigma_I + \sigma_J)}{2} \tag{7.38}$$

$$\epsilon_{IJ} = \sqrt{\epsilon_I \epsilon_J} \tag{7.39}$$

The other commonly used mixing rule is the one that defines both σ_{IJ} and ϵ_{IJ} as the geometric mean of the values for the pure species:

$$\sigma_{IJ} = \sqrt{\sigma_I \sigma_J} \tag{7.40}$$

$$\epsilon_{IJ} = \sqrt{\epsilon_I \epsilon_J} \tag{7.41}$$

Most force-fields use the Lorentz-Berthelot mixing rule, however the OPLS force-field is one force-field that utilizes the geometric mixing rule.

In other nonbond pairwise potentials, the repulsive portion of the interaction is modeled with an exponential term, which is in better agreement with the functional form of the repulsive term determined from quantum mechanics. One example of such a potential is the Buckingham potential (Buckingham 1938):

$$E_{vdW}(\mathbf{R}) = \sum_{I<J} A_{IJ}\exp(-B_{IJ}R_{IJ}) - \left(\frac{C_{IJ}}{R_{IJ}^6}\right)\right], \tag{7.42}$$

where A_{IJ}, B_{IJ}, and C_{IJ} are adjustable parameters that will have unique values for different types of particles. Another form of the nonbond interaction is the Born–Mayer–Huggins potential (Fumi and Tosi 1964; Tosi and Fumi 1964):

$$E_{vdW}(\mathbf{R}) = \sum_{I<J} A_{IJ} \exp(B_{IJ}(\sigma_{IJ} - R_{IJ})) - \frac{C_{IJ}}{R_{IJ}^6} + \frac{D_{IJ}}{R_{IJ}^8}, \tag{7.43}$$

where A_{IJ}, B_{IJ}, C_{IJ}, D_{IJ} and σ_{IJ} are adjustable parameters that will have unique values for different types of particles. The Born–Mayer–Huggins potential (❯ Eq. 7.43) is identical to the Buckingham potential (❯ Eq. 7.42) when $\sigma = D = 0$.

All of the nonbond potential functional forms that have been presented to this point take into account the effect that one particle has on another particle based solely on the distance between the two particles. However, in some systems like metals and alloys as well as some covalently bonded materials like silicon and carbon, the nonbonded potential is a function of more than just the distance between two particles. In order to model these systems, the embedded-atom method (EAM) (Daw and Baskes 1983, 1984; Finnis and Sinclair 1984) and modified embedded-atom method (MEAM) (Baskes 1992) utilize an embedding energy, F_I, which is a function of the atomic electronic density ρ_I of the embedded atom I and a pair potential interaction ϕ_{IJ} such that

$$E_I(\mathbf{R}) = F_I\left(\sum_{J\neq I} \rho_I(R_{IJ})\right) + \frac{1}{2}\sum_{J\neq I} \phi_{IJ}(R_{IJ}). \tag{7.44}$$

The multi-body nature of the EAM potential is a result of the embedding energy term.

So while the EAM and MEAM potentials have a term to account for multi-body interactions they are still only pair-wise potential, as are all the other nonbond potentials presented to this point. However, there are multi-body potentials that will explicitly account for how the presence of a third, fourth, ...atom affects the nonbond energy felt by any given atom. One example of a three-body potential is the Stillinger-Weber potential (Stillinger and Weber 1985):

$$E(\mathbf{R}) = \sum_I \sum_{J>I} \phi_2(R_{IJ}) + \sum_I \sum_{J\neq I} \sum_{K>J} \phi_3(R_{IJ}, R_{IK}, \theta_{IJK}), \tag{7.45}$$

where there is a two-body term ϕ_2:

$$\phi_2(R_{IJ}) = A_{IJ}\epsilon_{IJ}\left[B_{IJ}\left(\frac{\sigma_{IJ}}{R_{IJ}}\right)^{p_{IJ}} - \left(\frac{\sigma_{IJ}}{R_{IJ}}\right)^{q_{IJ}}\right]\exp\left(\frac{\sigma_{IJ}}{R_{IJ} - a_{IJ}\sigma_{IJ}}\right) \tag{7.46}$$

and a three-body term ϕ_3:

$$\phi_3(R_{IJ}, R_{IK}, \theta_{IJK}) = \lambda_{IJK}\epsilon_{IJK}\left[\cos\theta_{IJK} - \cos\theta_{0,IJK}\right]^2$$
$$\times \exp\left(\frac{\gamma_{IJ}\sigma_{IJ}}{R_{IJ} - a_{IJ}\sigma_{IJ}}\right)$$
$$\times \exp\left(\frac{\gamma_{IK}\sigma_{IK}}{R_{IK} - a_{IK}\sigma_{IK}}\right). \tag{7.47}$$

The Stillinger-Weber potential has generally been used for modeling crystalline silicon; however, more recently it has also been used for organic molecules as well. Another example of a three-body interatomic potential is the Tersoff potential (Tersoff 1988, 1989), which also was created initially in an attempt to accurately model silicon solids.

Electrostatic Interactions

Due to the fact that not all particles in a molecule have the same electronegativity, different particles will have stronger attractions to electrons than others. However, since classical force-fields do not model the flow of electrons, the different particles within a molecule are assigned a partial charge that remains constant during the course of a simulation. Generally these partial charges q_i are assigned to the nuclear centers of the particles. The electrostatic interaction between particles in different molecules or particles that are separated by at least two other atoms in a given molecule is calculated as the sum of the contributions between pairs of these partial charges using Coulomb's law:

$$E_{\text{coul}} = \sum_I \sum_J \frac{q_I q_J}{4\pi\epsilon_0 R_{IJ}} \tag{7.48}$$

where the charges of each particle are q_I and q_J and ϵ_0 is the dielectric constant.

In practice, an Ewald sum (Ewald 1921) is generally used to evaluate the electrostatic interactions within a classical MD simulation. However, this is a very computationally expensive algorithm to implement and it results in a computational cost of $N^{3/2}$, where N is the number of particles in the system. In order to obtain better computational scaling, fast Fourier transforms (FFTs) have been used to calculate the reciprocal space summation required within the Ewald sum. By using the FFT algorithm, one can reduce the cost of the electrostatic algorithm to $N \log N$. The most popular FFT algorithm that has been adopted for use in classical MD simulations is the particle-particle particle-mesh (pppm) approach (Hockney and Eastwood 1981; Luty et al. 1994, 1995).

Bonded Interactions

The bonded interactions are needed to model the energetic penalty that will result from two covalently bonded atoms moving too close or too far away from one another. The most common functional form that is used to model the bond bending interactions is that of a harmonic term:

$$E_{\text{bond}} = \sum_{\text{bonds}} k_b (\ell_b - \ell_b^{(0)})^2 \tag{7.49}$$

where k_b is commonly referred to as the bond constant and is a measure of the bond stiffness and $\ell_b^{(0)}$ is the reference length or often referred to the equilibrium bond length. Each of these parameters will vary depending on the types of particles that the bond is joining.

Angle Bending Interactions

The angle bending interactions are also modeled in order to determine the energetic penalties of angles containing three different particles compressing or overextending such that they distort the geometry of a portion of a molecule away from its desired structure.

Again, the most common functional form to model the angle interactions is a harmonic expression:

$$E_{\text{angle}} = \sum_{\text{angles}} k_a (\theta_a - \theta_a^{(0)})^2 \tag{7.50}$$

where k_a is the angle constant and is a measure of the rigidity of the angle, and $\theta_a^{(0)}$ is the equilibrium or reference angle.

Torsional Interactions

The torsional interactions are generally modeled using some form of a cosine series. The OPLS force-field uses the following expression for its torsional term:

$$E_{\text{dihed}} = \sum_{\text{dihedrals}} \frac{1}{2} K_d^{(1)} [1 + \cos(\phi)] + \frac{1}{2} K_d^{(2)} [1 - \cos(2\phi)] + \frac{1}{2} K_d^{(3)} [1 + \cos(3\phi)]$$

$$+ \frac{1}{2} K_d^{(4)} [1 - \cos(4\phi)] \tag{7.51}$$

where $K_d^{(i)}$ are the force constants for each cosine term and ϕ is the measured dihedral angle. The Charmm force-field uses the following expression:

$$E_{\text{dihed}} = \sum_{\text{dihedrals}} K_d [1 + \cos(n\phi - d_d)], \tag{7.52}$$

where K_d is the force constant, n is the multiplicity of the dihedral angle ϕ, and d_d is the shift of the cosine that allows one to more easily move the minimum of the dihedral energy.

First Principles Electronic Structure Methods

For the electronic ground state, i.e., $k = 0$, Kohn–Sham (KS) density functional theory is commonly used. In this case, the energy is given by

$$E_0 \approx E^{\text{KS}}[\rho] = T_s[\rho] + \int d\mathbf{r} v_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}) + \frac{1}{2} \int d\mathbf{r} v_{\text{H}}(\mathbf{r})\rho(\mathbf{r}) + E_{\text{xc}} \tag{7.53}$$

with the kinetic energy of *noninteracting electrons*, i.e., using a Slater determinant as a wave-function ansatz,

$$\Psi^{\text{KS}} = \frac{1}{\sqrt{n!}} \begin{vmatrix} \psi_1(\mathbf{x}_1) & \psi_2(\mathbf{x}_1) & \cdots & \psi_n(\mathbf{x}_1) \\ \psi_1(\mathbf{x}_2) & \psi_2(\mathbf{x}_2) & \cdots & \psi_n(\mathbf{x}_2) \\ \vdots & \vdots & \vdots & \\ \psi_1(\mathbf{x}_n) & \psi_2(\mathbf{x}_n) & \cdots & \psi_n(\mathbf{x}_n) \end{vmatrix} \tag{7.54}$$

$$T_s[\rho] = -\frac{1}{2} \sum_i^n f_i \int d\mathbf{r}\, \psi_i(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}) \tag{7.55}$$

where f_i is the number of electrons occupying orbital ψ_i, the external potential including nucleus–nucleus repulsion and electron–nucleus attraction,

$$v_{\text{ext}}(\mathbf{r}) = \sum_{I=1}^{N-1} \sum_{J>I}^{N} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} - \sum_{I=1}^{N} \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} \tag{7.56}$$

the Hartree potential (electron–electron interaction)

$$v_{\text{H}}(\mathbf{r}) = \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{7.57}$$

the exchange-correlation energy, E_{xc}, and the electron density

$$\rho(\mathbf{r}) = \sum_i^n f_i |\psi_i(\mathbf{r})|^2 \tag{7.58}$$

The orbitals which minimize the total, many-electron energy (❯ Eq. 7.53) are obtained by solving self-consistently the one-electron Kohn–Sham equations,

$$\left[-\frac{1}{2}\nabla^2 + v_{\text{ext}}(\mathbf{r}) + v_{\text{H}}(\mathbf{r}) + \frac{\delta E_{\text{xc}}[\rho]}{\delta\rho(\mathbf{r})} \right] \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}) \tag{7.59}$$

DFT is exact in principle, provided that $E_{xc}[\rho]$ is known, in which case E^{KS} (see ❯ Eq. 7.53) is an exact representation of the ground state energy E_0 (see ❯ Eq. 7.8). In practice, however, $E_{xc}[\rho]$ is not – and presumably never will be – known exactly; therefore (semiempirical) approximations are used.

The starting point for most density functionals is the *local density approximation* (LDA), which is based on the assumption that one deals with a homogeneous electron gas. E_{xc} is split into an exchange term E_x and a correlation term E_c. Within the LDA, the exchange functional is given exactly by Dirac (1930):

$$E_x^{\text{LDA}}[\rho] = \int \rho(\mathbf{r})\epsilon_x^{\text{LDA}}(\rho(\mathbf{r}))d\mathbf{r} \tag{7.60}$$

where

$$\epsilon_x^{\text{LDA}}(\rho) = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}}\rho(\mathbf{r})^{\frac{1}{3}} \tag{7.61}$$

The LDA correlation functional, on the other hand, can only be approximated. We give here the most commonly used expression by Vosko et al. [1980], derived from Quantum Monte Carlo calculations:

$$E_c^{\text{LDA}}[\rho] = \int \rho(\mathbf{r})\epsilon_c^{\text{LDA}}(\rho(\mathbf{r}))d\mathbf{r} \tag{7.62}$$

where

$$\epsilon_c^{\text{LDA}}(\rho) = A\left\{ \ln\left(\frac{x^2}{X}\right) + \frac{2b}{Q}\tan^{-1}\left(\frac{Q}{2x+b}\right) - \frac{bx_0}{X(x_0)}\left[\ln\left(\frac{(x-x_0)^2}{X}\right) \right.\right.$$
$$\left.\left. + \frac{2(b2x_0)}{Q}\tan^{-1}\left(\frac{Q}{2x+b}\right) \right] \right\} \tag{7.63}$$

with $X = x^2 + bx + c$, $x = \sqrt{r_s}$, $r_s = \sqrt[3]{\frac{3}{4\pi\rho(\mathbf{r})}}$, $Q = \sqrt{4c - b^2}$, $x_0 = -0.104098$, $A = 0.0310907$, $b = 3.72744$, $c = 12.9352$.

This simplest approximation, LDA, is often too inaccurate for chemically relevant problems. A notable improvement is usually offered by so-called *semilocal* or *gradient corrected* functionals (*generalized gradient approximation* (GGA)), in which E_x and E_c are expressed as functionals of ρ and the first variation of the density, $\nabla\rho$:

$$E_x^{\text{GGA}}[\rho, \nabla\rho] = \int \rho(\mathbf{r})\epsilon_x^{\text{GGA}}(\rho(\mathbf{r}), \nabla\rho)d\mathbf{r} \tag{7.64}$$

$$E_c^{\text{GGA}}[\rho, \nabla\rho] = \int \rho(\mathbf{r})\epsilon_c^{\text{GGA}}(\rho(\mathbf{r}), \nabla\rho)d\mathbf{r} \tag{7.65}$$

Popular examples are the BLYP (Becke 1988; Lee et al. 1988), BP (Becke 1988; Polák 1986), and BPW91 (Becke 1988; Perdew et al. 1992) functionals. The expressions for $\epsilon_{x,c}^{\text{GGA}}(\rho(\mathbf{r}), \nabla\rho)$ are complex and shall not be discussed here.

In many cases, accuracy can be further increased by using so-called *hybrid functionals*, which contain an admixture of Hartree–Fock exchange to KS exchange. Probably the most widely used hybrid functional is the three-parameter B3LYP functional (Becke 1993),

$$E_{xc}^{\text{B3LYP}} = aE_x^{\text{LDA}} + (1-a)E_x^{\text{HF}} + b\Delta E_x^{\text{B}} + (1-c)E_c^{\text{LDA}} + cE_c^{\text{LYP}} \tag{7.66}$$

where $a = 0.80$, $b = 0.72$, $c = 0.81$, and E_x^{HF} is the Hartree-Fock exchange energy evaluated using KS orbitals.

New functionals are constantly proposed in search of better approximations to the exact E_{xc}. Often functionals are designed to remedy a particular shortcoming of previous functionals, for instance, for dispersion interactions.

Building the System/Collecting the Ingredients

Setting Up an AIMD Simulation

Building a Molecule

In many cases, the coordinates of a molecular structure are available for download on the web, from crystallographic databases (CCDC 2010; ICSD 2009; PDB 2010; Reciprocal Net 2004; Toth 2009) or journal supplements. For relatively small molecules, an initial guess structure can be built using molecular graphics software packages such as `molden` (2010).

Plane Waves and Pseudopotentials

The most common form of AIMD simulation employs DFT (see section "First Principles Electronic Structure Methods") to calculate atomic forces, in conjunction with periodic boundary conditions and a plane wave basis set. Using a plane wave basis has two major advantages over atom-centered basis functions: (1) there is no *basis set superposition error* (Boys and Bernardi 1970; Marx and Hutter 2000) and (2) the Pulay correction (Pulay 1969, 1987) to the Hellmann–Feynman force, due to basis set incompleteness, vanishes (Marx and Hutter 2000, 2009).

Plane Wave Basis Set
As a consequence of Bloch's theorem, in a periodic lattice, the Kohn–Sham orbitals (see ❯ Eq. 7.59) can be expanded in a set of plane waves (Ashcroft and Mermin 1976; Meyer 2006),

$$\psi_{\mathbf{k},j}(\mathbf{r}) = \sum_{\mathbf{G}} c_{\mathbf{G}}^{\mathbf{k},j} e^{i(\mathbf{k}+\mathbf{G})\mathbf{r}} \tag{7.67}$$

where $\mathbf{k}$ is a wavevector within the Brillouin zone, satisfying Bloch's theorem,

$$\psi(\mathbf{r} + \mathbf{T}) = e^{i\mathbf{kT}}\psi(\mathbf{r}) \tag{7.68}$$

for any lattice vector $\mathbf{T}$,

$$\mathbf{T} = N_1\mathbf{a}_1 + N_2\mathbf{a}_2 + N_3\mathbf{a}_3 \tag{7.69}$$

N_1, N_2, N_3 being integer numbers, and $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ the vectors defining the periodically repeated simulation box.

In ❯ Eq. 7.67, the summation is over all reciprocal lattice vectors **G** which fulfill the condition $\mathbf{G} \cdot \mathbf{T} = 2\pi M$, M being an integer number. In practice, this plane-wave expansion of the Kohn-Sham orbitals is truncated such that the individual terms all yield kinetic energies lower than a specified cutoff value, E_{cut},

$$\frac{\hbar^2}{2m}|\mathbf{k} + \mathbf{G}|^2 \leq E_{\mathrm{cut}} \tag{7.70}$$

The plane-wave basis set thus has the advantage over other basis sets that convergence can be controlled by a single parameter, namely E_{cut}.

In this periodic setup, the electron density (see ❯ Eq. 7.58) can be approximated by a sum over a mesh of N_{kpt} **k**-points in the Brillouin zone (Chadi and Cohen 1973; Monkhorst and Pack 1976; Moreno and Soler 1992),

$$\rho(\mathbf{r}) \approx \frac{1}{N_{\mathrm{kpt}}} \sum_{\mathbf{k}} f_{\mathbf{k},j}|\psi_{\mathbf{k},j}(\mathbf{r})|^2 \tag{7.71}$$

Since the volume of the Brillouin zone, $V_{\mathrm{BZ}} = (2\pi)^3/V_{\mathrm{box}}$, decreases with increasing volume of the simulation supercell, V_{box}, only a small number of **k**-points need to be sampled for large supercells. For insulating materials (i.e., large bandgap), a single **k**-point is often sufficient, typically taken to be **k** = 0 (Γ-point approximation).

Pseudopotentials

While plane waves are a good representation of delocalized Kohn–Sham orbitals in metals, a huge number of them would be required in the expansion (❯ Eq. 7.67) to obtain a good approximation of atomic orbitals, in particular near the nucleus where they oscillate rapidly. Therefore, in order to reduce the size of the basis set, only the valence electrons are treated explicitly, while the core electrons (i.e., the inner shells) are taken into account implicitly through pseudopotentials combining their effect on the valence electrons with the nuclear Coulomb potential. This *frozen core approximation* is justified as typically only the valence electrons participate in chemical interactions. To minimize the number of basis functions the pseudopotentials are constructed in such a way as to produce nodeless atomic valence wavefunctions. Beyond a specified cutoff distance from the nucleus, R_{cut} the nodeless pseudo-wavefunctions are required to be identical to the reference all-electron wavefunctions.

Normconserving Pseudopotentials

Normconserving pseudopotentials are generated subject to the condition that the pseudo-wavefunction has the same norm as the all-electron wavefunction and thus gives rise to the same electron density. Although normconserving pseudopotentials have to fulfill a (small) number of mathematical conditions, there remains considerable freedom in how to create them. Hence several different recipes exist (Bachelet et al. 1982; Goedecker et al. 1996; Hamann et al. 1979; Hartwigsen et al. 1998; Kerker 1980; Troullier and Martins 1990, 1991; Vanderbilt 1985).

Since pseudopotentials are generated using *atomic* orbitals as a reference, it is not guaranteed that they are *transferable* to any chemical environment. Generally, transferability is the better the smaller the cutoff radius R_{cut} is chosen. However, the reduction in the number of plane waves required to represent a particular pseudo-wavefunction – i.e., the *softness* of the corresponding pseudopotential – increases as R_{cut} gets larger. So R_{cut} has to be chosen carefully and there is always a trade-off between transferability and softness. An upper limit for

R_{cut} is given by the shortest interatomic distances in the molecule or crystal the pseudopotential will be used for: one needs to make sure that the sum of the two cutoff radii of any two neighboring atoms is smaller than their actual spatial separation.

For each angular momentum l, a separate pseudopotential $V_l^{PS}(r)$ is constructed. The total pseudopotential operator is written as

$$\hat{V}^{PS} = V_{loc}^{PS}(r) + \sum_l V_{nl,l}^{PS}(r)\hat{P}_l \tag{7.72}$$

where the nonlocal part is defined as

$$V_{nl,l}^{PS}(r) = V_l^{PS}(r) - V_{loc}^{PS}(r) \tag{7.73}$$

and the local part $V_{loc}^{PS}(r)$ is taken to be the pseudopotential $V_l^{PS}(r)$ for one specific value of l, typically the highest one for which a pseudopotential was created. The pseudopotential (❯ Eq. 7.72) is called semi-local, since the projector $\hat{P}_l$ only acts on the l-th angular momentum component of the wavefunction, but not on the radius r. (*Note:* a pseudopotential is called nonlocal if it is l-dependent.)

To achieve higher numerical efficiency, it is common practice to transform the semi-local pseudopotential (❯ Eq. 7.72) to a fully nonlocal form,

$$\hat{V}^{PS} = V_{loc}^{PS}(r) + \sum_{ij} |\beta_i > B_{ij} < \beta_j| \tag{7.74}$$

using the Kleinman-Bylander prescription (Kleinman and Bylander 1982).

Vanderbilt Ultrasoft Pseudopotentials

An ultrasoft type of pseudopotential was introduced by Vanderbilt (1990) and Laasonen et al. [1993] to deal with nodeless valence states which are strongly localized in the core region. In this scheme the normconserving condition is lifted and only a small portion of the electron density inside the cutoff radius is recovered by the pseudo-wavefunction, the remainder is added in the form of so-called augmentation charges. Complications arising from this scheme are the nonorthogonality of Kohn–Sham orbitals, the density dependence of the nonlocal pseudopotential, and need to evaluate additional terms in atomic force calculations.

How to Obtain Pseudopotentials?

There are extensive pseudopotential libraries available for download with the simulation packages CPMD (Parrinello et al. 2008), CP2K (Hutter et al. 2009) or online (Vanderbilt Ultra-Soft Pseudopotential Site 2006). However, before applying any pseudopotentials, they should always be tested against all-electron calculations. Pseudopotentials used in conjunction with a particular density functional should have been generated using the same functional.

In many cases, the required pseudopotential will not be available in any accessible library; in this case it may be generated using freely downloadable programs (Vanderbilt Ultra-Soft Pseudopotential Site 2006).

Setting Up a Classical MD Simulation

There are two general stages that make up the preparation to conduct force-field molecular dynamics simulations: (1) gathering preliminary information and (2) building the actual system.

Gathering Preliminary Information

Gathering the preliminary information before conducting the simulation is mostly focussed on making sure that the simulation is possible. First, it is important to identify the type and number of molecules that you wish to model. Then, it is necessary to find the force-field that will allow you to most accurately model the molecules and physical system that you want to simulate. A brief synopsis of some of the larger classical force-field parameter sets is given in section "Classical Force Fields". These force-fields and references may be good starting points in searching for the correct classical force-field to use for a given system, but the best way to find a specific force-field is to just conduct a search for research articles that may have been conducted on the same system. If no force-field parameters exist for the system of interest, then you can use configurations and energies from quantum simulations to parameterize a given force-field for your system. A methodology for how a force-field was parameterized originally is presented in the relevant paper; however, this is a complicated exercise and is probably best left to the experts.

Building the System

After identifying that a force-field exists for the system you wish to model, the next step is to build the initial configuration of the molecules within the system. The initial configuration will consist of initial spatial coordinates of each atom in each given molecule. When building a large system consisting of several molecules of various types, it is easiest to write a computer code that contain the molecular structure and coordinates of each molecule present in the system, and then have the code replicate each molecule how ever many times is necessary in order to build the entire system. Alternatively, most of the molecular dynamics simulation packages previously mentioned have capabilities to build systems from a pdb file; however, these tools are often useful for only certain systems and force-fields. There is unfortunately no one tool which can be used to build any system with any force-field.

These initial configurations can represent a minimum energy structure either from another simulation (i.e., a final structure from a energy minimization in a quantum or classical Monte Carlo simulation can be used as the starting state for classical simulations), from experimental observation (i.e., the pdb database for crystallographic structures of proteins) or building the initial coordinates based upon the equilibrium bond distances and bond angles from the force-field.

The placement of the molecules within the simulation box can be done in a number of different ways as well. The molecules can be placed on the vertices of a regular lattice, or in any other regularly defined geometry that may be useful for conducting your simulation (i.e., in simulating the structural properties of micelles often times the surfactant molecules will initially be placed on the vertices of a buckey ball such that they are in a spherical configuration). Also, molecules can be placed at random positions within the simulation box. The one advantage of placing molecules at regularly spaced positions is that it is easier to insure that there is no overlapping of molecules, whereas with the randomly placed molecules it can be quite difficult to ensure that a placed molecule does not overlap with another molecule in the box (particularly for large or highly branched molecules).

In addition to containing the initial spatial coordinates of all of the molecules in the system, the initial configuration must also contain some additional information about the atoms and

molecules in the systems. Each atom in the configuration must contain a label of what atomic species (i.e., carbon, nitrogen, …) it represents. This label will be different for each simulation code used but all of them will have some type of label as it will inform the simulation code what force-field values to use to represent the interactions of that atom. A list of all of the covalent bonds, the bond angles, and the dihedrals in the system will also need to be included in the initial configuration. The lists of the bonds, angles, and dihedrals contain an identifier for each atom that make up the bond, angle, or dihedral and then an identifier for the type that informs the simulation package which parameters to use in calculating the energy of the bond, angle, or dihedral. The final component of the initial configuration of a classical simulation is a list of all of the various types of atoms, bonds, angles, and dihedrals in the system along with their corresponding force-field parameters (i.e., ϵ and σ for atom types to describe their nonbond interactions, force constants, and equilibrium values for bond, angle, and dihedral types).

Finally, after building the initial configuration, the simulation is about ready to be performed. The last step is to choose the simulation variables and set up the input to the simulation package in order to convey these selections.

These options and the decision process behind choosing from the various options will be presented in the following sections.

Preparing an Input File

Optimization Algorithms

Optimization algorithms are often used to find stationary points on a potential energy surface, i.e., local and global minima and saddle points. The only place where they directly enter MD is in the case of Born–Oppenheimer AIMD, in order to converge the SCF wavefunction for each MD step. It is immediately obvious that the choice of optimization algorithm crucially affects the speed of the simulation.

Steepest Descent

The Steepest Descent method is the simplest optimization algorithm. The initial energy $E[\Psi_0] = E(c_0)$, which depends on the plane wave expansion coefficients $\mathbf{c}$ (see ❯ Eq. 7.67), is lowered by altering $\mathbf{c}$ in the direction of the negative gradient,

$$\mathbf{d}_n = -\frac{\partial E(\mathbf{c}_n)}{\partial \mathbf{c}} \equiv -\mathbf{g}_n \qquad (7.75)$$

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \Delta_n \mathbf{d}_n \qquad (7.76)$$

where $\Delta_n > 0$ is a variable step size chosen such that the energy always decreases, and n is the optimization step index. The steepest descent method is very robust; it is guaranteed to approach the minimum. However, the rate of convergence ever decreases as the energy gets closer to the minimum, making this algorithm rather slow.

Conjugate Gradient Methods

The Conjugate Gradient method generally converges faster than the steepest descent method due to the fact that it avoids moving in a previous search direction. This is achieved by linearly combining the gradient vector and the last search vector,

$$\mathbf{c}_{n+1} = \mathbf{c}_n + \Delta_n \mathbf{d}_n \tag{7.77}$$

where

$$\mathbf{d}_n = -\mathbf{g}_n + \beta_n \mathbf{d}_{n-1} \tag{7.78}$$

Different recipes exist to determine the coefficient β_n (Jensen 2007) among which the Polak–Ribière formula usually performs best for non-quadratic functions,

$$\beta_n = \frac{\mathbf{g}_n(\mathbf{g}_n - \mathbf{g}_{n-1})}{\mathbf{g}_{n-1}\mathbf{g}_{n-1}} \tag{7.79}$$

In the case of a general non-quadratic function, such as the DFT energy, conjugacy is not strictly fulfilled and the optimizer may search in completely inefficient directions after a few steps. It is then recommended to restart the optimizer (setting $\beta = 0$). Convergence can be improved by multiplying $\mathbf{g}_n$ with a preconditioner matrix, e.g., an approximate inverse of the second derivatives matrix (Hessian in the case of geometry optimization) $\tilde{\mathbf{H}}$. The method is then called Preconditioned Conjugate Gradient (PCG). In the CPMD code, the matrix $\tilde{\mathbf{H}}$ is approximated by

$$\tilde{\mathbf{H}} = \left\{ \begin{array}{ll} H_{GG'}^{\mathrm{KS}} & \text{for } G \geq G_{\mathrm{cut}} \\ H_{G_{\mathrm{cut}}G_{\mathrm{cut}}}^{\mathrm{KS}} & \text{for } G < G_{\mathrm{cut}} \end{array} \right\} \tag{7.80}$$

where $H_{GG'}^{\mathrm{KS}}$ is the Kohn–Sham matrix is the plane-wave basis and G_{cut} is a cutoff value for the reciprocal lattice vector $\mathbf{G}$ (set to a default value of 0.5 a.u.).

Direct Inversion of the Iterative Subspace

Having generated a sequence of optimization steps $\mathbf{c}_i$, the Direct Inversion of the Iterative Subspace (DIIS) method (Császár and Pulay 1984; Hutter et al. 1994; Pulay 1980, 1982) is designed to accelerate convergence by finding the best linear combination of stored $\mathbf{c}_i$ vectors,

$$\mathbf{c}_{n+1} = \sum_{i=1}^{n} a_i \mathbf{c}_i \tag{7.81}$$

Ideally, of course, $\mathbf{c}_{n+1}$ is equal to the optimum vector $\mathbf{c}_{\mathrm{opt}}$. Defining the error vector $\mathbf{e}_i$ for each iteration as

$$\mathbf{e}_i = \mathbf{c}_i - \mathbf{c}_{\mathrm{opt}} \tag{7.82}$$

❯ Eq. 7.81 becomes

$$\sum_{i=1}^{n} a_i \mathbf{c}_{\mathrm{opt}} + \sum_{i=1}^{n} a_i \mathbf{e}_i = \mathbf{c}_{\mathrm{opt}} \tag{7.83}$$

❯ Equation 7.83 is satisfied if

$$\sum_{i=1}^{n} a_i = 1 \tag{7.84}$$

and

$$\sum_{i=1}^{n} a_i \mathbf{e}_i = 0 \tag{7.85}$$

Instead of the ideal case ❯ Eq. 7.85, in practice one minimizes the quantity

$$\langle \sum_{i=1}^{n} a_i \mathbf{e}_i | \sum_{j=1}^{n} a_j \mathbf{e}_j \rangle \tag{7.86}$$

subject to the constraint (❯ Eq. 7.84), which is equivalent to solving the system of linear equations

$$\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} & -1 \\ b_{21} & b_{22} & \cdots & b_{2n} & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} & -1 \\ -1 & -1 & \cdots & -1 & 0 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix} \tag{7.87}$$

where

$$b_{ij} = \langle \mathbf{e}_i | \mathbf{e}_j \rangle \tag{7.88}$$

and the error vectors are approximated by

$$\mathbf{e}_i = -\tilde{\mathbf{H}}^{-1}(\mathbf{c}_{\text{opt}}) \mathbf{g}_i \tag{7.89}$$

using an approximate Hessian matrix $\tilde{\mathbf{H}}$, e.g., ❯ Eq. 7.80.

Controlling Temperature: Thermostats

If understanding the behavior of the system as a function of temperature is the aim of your study, then it is important to be able to control the temperature of your system. The temperature of the system is related to the time average of the kinetic energy, which generally can be calculated by

$$< \mathcal{H} >_{NVT} = \frac{3}{2} N k_B T. \tag{7.90}$$

Below we introduce specific thermostatting techniques for MD simulations at thermodynamic equilibrium, e.g., for calculating equilibrium spatial distribution and time-correlation functions. However, when MD simulations are performed on a system undergoing some *non-equilibrium* process involving exchange of energy between different parts of the system, e.g., when an energetic particle, such as an atom or a molecule, hits a crystal surface, or there is a temperature gradient across the system, one has to resort to specially developed techniques, see for example Kantorovich [2008], Kantorovich and Rompotis [2008]and Toton et al. [2010]. In these methods, based on the so-called Generalized Langevin Equation, the actual system on which MD simulations are performed is considered in contact with one (or more) heat bath(s) kept at constant temperature(s), and the dynamics of the system of interest reflects the fact that there is an interaction and energy transfer between the system and the surrounding heat bath(s).

Rescale Thermostat

One obvious way to control the temperature of a system is to rescale the velocities of the atoms within the system (Woodcock 1971). The rescaling factor λ is determined from $\lambda\sqrt{T_{target}/T_0}$, where T_{target} and T_0 are the target and initial temperatures, respectively. Then, the velocity of each atom is rescaled such that $V_f = \lambda V_i$. In practice, the inputs generally required to use a rescale thermostat include:

- T_0 – Initial temperature
- T_{target} – Target temperature
- τ – Damping constant (i.e., frequency with which to apply the thermostat)
- δT – Maximum allowable temperature difference from T_{target} before thermostat is applied
- $f_{rescale}$ – Fraction of temperature difference between current temperature and T_{target} is corrected during each application of thermostat

If it is desired to have a strict thermostat (i.e., when first starting a simulation that might have particles very near one another), then δT and τ should have values of ~$0.01 T_{target}$ and 1 time step, respectively, and $f_{rescale}$ should be near 1.0. However, if you wish to allow a more lenient thermostat, then the value of δT should be of the same order of magnitude as T_{target}, τ should be ~10^2–10^3 time steps, and $f_{rescale} \sim 0.01$–0.1.

Berendsen Thermostat

Another way to control the temperature is to couple the system to an external heat bath, which is fixed at a desired temperature. This is referred to as a Berendsen thermostat (Berendsen et al. 1984). In this thermostat, the heat bath acts as a reservoir of thermal energy that supplies or removes temperature as necessary. The velocities are rescaled each time step, where the rate of change in temperature is proportional to the difference in the temperature in the system $T(t)$ and the temperature of the external bath T_{bath}:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}(T_{bath} - T(t)) \tag{7.91}$$

which when integrated results in the change in temperature each time step:

$$\Delta T = \frac{\delta t}{\tau}(T_{bath} - T(t)). \tag{7.92}$$

In ❯ Eqs. 7.91 and ❯ 7.92, τ is the damping constant for the thermostat. In practice, the necessary inputs when using the Berendsen thermostat include:

- T_{bath} – temperature of the external heat bath
- τ – damping constant for the thermostat

Obviously the amount of control that the thermostat imposes on the simulation is controlled by the value of τ. If τ is large, then the coupling will be weak and the temperature will fluctuate significantly during the course of the simulation. While if τ is small, then the coupling will be strong and the thermal fluctuations will be small. If $\tau = \delta t$, then the result will be the same as the rescale thermostat, in general.

Nosé–Hoover Thermostat

While the Berendsen thermostat is efficient for achieving a target temperature within your system, the use of a thermostat that represents a canonical ensemble once the system has reached a thermal equilibrium. The extended system method, which was originally introduced by Nosé [1984a, b] and then further developed by Hoover (1985), introduces additional degrees of freedom into the Hamiltonian that describes the system, from which equations of motion can be determined.

The extended system method considers the external heat bath as an integral part of the system by including an additional degree of freedom in the Hamiltonian of the system that is represented by the variable s. As a result, the potential energy of the reservoir is

$$E_{\text{pot}} = (f + 1) k_B T \ln s, \tag{7.93}$$

where f is the number of degrees of freedom in the physical system and T is the target temperature. The kinetic energy of the reservoir is calculated by

$$E_{\text{kin}} = \frac{Q}{2} \left(\frac{ds}{dt} \right)^2, \tag{7.94}$$

where Q is a parameter with dimensions of energy $\times$ (time)^2 and is generally referred to as the "virtual" mass of the extra degree of freedom s. The magnitude of Q determines the coupling between the heat bath and the real system, thus influencing the temperature fluctuations.

Utilizing ❯ Eqs. 7.93 and ❯ 7.94, and substituting the real variables for the corresponding Nosé variables, the equations of motion are found to be as follows:

$$\ddot{\mathbf{R}}_I = \frac{\mathbf{F}_I}{M_I} - \gamma \mathbf{R}_I, \tag{7.95}$$

$$\dot{\gamma} = -\frac{1}{\tau_{\text{NH}}} \left(\frac{f+1}{f} \frac{T_{\text{target}}}{T} - 1 \right), \tag{7.96}$$

where $\gamma = \frac{\dot{s}}{s}$ and $\tau_{\text{NH}} = \frac{Q}{f k_B T_{\text{target}}}$ The variable τ_{NH} is an effective relaxation time, or damping constant.

In practice, the inputs that are necessary when utilizing the Nosé–Hoover thermostat during a molecular dynamics simulation include

- T_{target} – Target temperature
- τ_{NH} – Damping constant
- Q – Fictitious mass of the additional degree of freedom s

The most significant variable in the above list is Q. Large values of Q may cause poor temperature control, with the infinite limit resulting in no energy exchange between the temperature bath and the real system, which is the case of conventional molecular dynamics simulations resulting in the microcanonical ensemble. However, if Q is too small then the energy oscillates and the system will take longer in order to reach a thermal equilibrium.

Controlling Pressure: Barostats

It may be desired to study the behavior of the simulated system while the pressure is held constant (i.e., pressure-induced phase transitions). Many experimental measurements are made

in conditions where the pressure and temperature are held constant and so it is of utmost importance to be able to accurately replicate these conditions in simulations.

One thing of note is that the pressure often fluctuates more than other quantities such as the temperature in an NVT molecular dynamics simulation or the energy in a NVE molecular dynamics simulation. This is due to the fact that the pressure is related to the virial term, which is the product of the positions of the particles in the system and the derivative of the potential energy function. These fluctuations will be observed in the instantaneous values of the system pressure during the course of the simulation, but the average pressure should approach the desired pressure. Since generally the temperature and number of atoms will also be held constant during constant pressure simulations, and the volume of the system will be allowed to change in order to arrive at the desired pressure, therefore, less compressible systems will show larger fluctuations in the pressure than the systems that are more easily compressed.

Berendsen Barostat

Many of the approaches used for controlling the pressure are similar to those that are used for controlling the temperature. One approach is to maintain constant pressure by coupling the system to a constant pressure reservoir as is done in the Berendsen barostat (Berendsen et al. 1984), which is analogous to the way temperature is controlled in the Berendsen thermostat. The pressure change in the system is determined by

$$\frac{dP(t)}{dt} = \frac{1}{\tau_P}(P_0 - P(t)), \tag{7.97}$$

where τ_P is time constant of the barostat, P_0 is the desired pressure and $P(t)$ is the system pressure at any time t. In order to accommodate this change in pressure, the volume of the box is scaled by a factor of μ^3 each time step, therefore the coordinates of each particle in the system are scaled by a factor of μ (i.e., $\mathbf{R}_I(t + \delta t) = \mu^{1/3}\mathbf{R}_I(t)$), where

$$\mu = \left[1 - \frac{\delta t}{\tau_P}(P - P_0)\right]^{\frac{1}{3}}. \tag{7.98}$$

In practice, the inputs for the Berendsen barostat will include:

- P_0 – Desired pressure
- τ_P – Time constant of the barostat

One other input that may be included in the use of the Berendsen barostat is to define which dimensions are coupled during the pressure relaxation. For example, you could define that the pressure is relaxed in a way that the changes in all three dimensions are coupled and therefore all of the dimensions change at the same rate. On the other hand, the pressure relaxation can be handled in an anisotropic manner, such that none of the dimensions are coupled and each dimension will have its own scaling factor that results from the individual pressure components.

Nosé–Hoover Barostat

Similar to the Nosé–Hoover thermostat, the extended system method has been applied to create a barostat (Hoover 1986) that is coupled with a Nosé–Hoover thermostat. In this case, the extra

degree freedom η corresponds to a "piston," and it is added to the Hamiltonian of the system, which results in the following equations of motion:

$$\frac{d\mathbf{R}(t)}{dt} = \mathbf{V}(t) + \eta(t)(\mathbf{R}(t) - \mathbf{R}_{COM}), \tag{7.99}$$

$$\frac{d\mathbf{V}(t)}{dt} = \frac{\mathbf{F}(t)}{M} - [\chi(t) + \eta(t)]\mathbf{V}(t), \tag{7.100}$$

$$\frac{d\chi(t)}{dt} = \frac{1}{\tau_T^2}\left(\frac{T}{T_0} - 1\right), \tag{7.101}$$

$$\frac{d\eta(t)}{dt} = \frac{1}{Nk_BT_0\tau_P^2}V(t)(P - P_0), \tag{7.102}$$

$$\frac{dV(t)}{dt} = 3\eta(t)V(t) \tag{7.103}$$

where $\mathbf{R}_{COM}$ are the coordinates of the center of mass of the system, η is the thermostat extra degree of freedom and can be thought of as a friction coefficient, τ_T is the thermostat time constant, χ is barostat extra degree of freedom and is considered a volume scaling factor and τ_P is the barostat time constant. ❯ Equations 7.102 and ❯ 7.103 explicitly contain the volume of the simulation box, $V(t)$. Generally, this barostat is implemented using the approach described in Melchionna et al. [1993].

In addition to the variables that are a part of the equations of motion, there is a variable Q that represents the "mass" of the "piston." This is analogous to the "mass" variable in the Nosé–Hoover thermostat. In practice, the required input for the Nosé–Hoover barostat will include:

- P_0 – Desired pressure
- T_0 – Desired temperature
- τ_P – Time constant of the barostat
- τ_T – Time constant of the thermostat
- Q – The "mass" of the piston

Like in the case of the Nosé–Hoover thermostat, care must be taken when selecting the value of the variable Q. A small value of Q is representative of a piston with small mass, and thus will have rapid oscillations of the box size and pressure, whereas a large value of Q will have the opposite effect. The infinite limit of Q results in normal molecular dynamics behavior.

Setting the Time Step

Born–Oppenheimer MD

Since BO-MD is classical MD in the sense that the nuclei are classical particles, the same rules concerning the choice of time step apply to both BO-MD and atomistic force-field MD. The largest possible time step, δt, is determined by the fastest oscillation in the system – in many molecules this would be a bond stretching vibration involving hydrogen, e.g., CH, NH, or OH. It is immediately plausible that δt must be smaller than the shortest vibrational period in order to

resolve that motion and for the numerical integrator (see section "Classical Molecular Dynamics") to be stable. Let us assume a particular molecule has an OH vibration at 3,500 cm^{-1}, corresponding to a period of about 10 fs. Then the time step has to be chosen smaller than 10 fs. Using a harmonic approximation it can be shown that the Verlet algorithm is stable for $\omega^2 \delta t^2 < 2$ (Sutmann 2006). In the present example this would dictate a maximum time step of 2 fs. However, although such a choice guarantees numerical stability, it results in deviations from the exact answer. Therefore, in practice smaller time steps – typically around 1 fs – are often used.

Car–Parrinello MD

Although in CP-MD the nuclei are still treated as classical particles, the choice of time step can no longer be based solely on the highest nuclear frequency ω_n^{max}. We also need to consider the fictitious dynamics of the electronic degrees of freedom. In fact, the optimum simulation time step is closely linked to the value of the fictitious electron mass μ as we will see in the following.

The fictitious mass μ has to be chosen small enough to guarantee adiabatic separation of electronic and nuclear motion. This means that the frequency spectrum of the electronic degrees of freedom (Marx and Hutter 2009; Pastore et al. 1991)

$$\omega_{ph} = \sqrt{\frac{2(\epsilon_p - \epsilon_h)}{\mu}} \tag{7.104}$$

must not overlap with the vibrational spectrum of the nuclear system. The lowest electronic frequency according to ❯ Eq. 7.104 is

$$\omega^{min} = \sqrt{\frac{2(\epsilon_{LUMO} - \epsilon_{HOMO})}{\mu}} \tag{7.105}$$

The highest electronic frequency is determined by the plane-wave cutoff energy E_{cut},

$$\omega^{max} \approx \sqrt{\frac{2E_{cut}}{\mu}} \tag{7.106}$$

Thus the maximum simulation time step, which is inversely proportional to ω^{max}, thus obeys the relation

$$\Delta t^{max} \propto \sqrt{\frac{\mu}{E_{cut}}} \tag{7.107}$$

According to ❯ Eq. 7.107 the maximum time step can be increased by simply increasing μ. However, this would also result in a lowering of ω_e^{min} (see ❯ Eq. 7.105) and therefore in a smaller separation $\omega_e^{min} - \omega_n^{max}$ between the nuclear and electronic spectra.

Let us discuss the above using some realistic numbers. In the case of the H_2O molecule, for example, the HOMO-LUMO gap with the BLYP functional is about 5.7 eV. Assuming a typical value of 400 a.u. for μ, the minimum electronic frequency (❯ Eq. 7.105) is ca. 6,900 cm^{-1}. The highest energy molecular vibrational mode in a CP-MD simulation using these parameter values is the asymmetric stretch at about 3,500 cm^{-1}. This means that electronic and nuclear spectra are well separated. A basis set cutoff of E_{cut} = 70 Ry (= 35 a.u.) leads to a maximum electronic frequency (❯ Eq. 7.106) of ≈92,000 cm^{-1} corresponding to a vibrational period of 15 a.u.. Hence the CP-MD time step has to be smaller than this number. For water, a time

step/fictitious mass combination of 4 a.u./400 a.u. has been shown to be a good compromise between efficiency and accuracy (Kuo et al. 2004).

If we were to increase μ to 1,000 a.u., we could afford a larger time step of about 6 a.u. (according to ❷ Eq. 7.107). However, $\omega_e^{\min}$ (❷ Eq. 7.105) would become ca. 4,500 cm^{-1}, dangerously close to $\omega_n^{\max}$. A simple trick that is often used to be able to afford larger time steps is to replace all hydrogen atoms by deuterium atoms thus downshifting $\omega_n^{\max}$. For systems with a small or even vanishing (e.g., metals) bandgap it is increasingly difficult or impossible to achieve adiabatic separation of electronic and nuclear degrees of freedom following the above considerations. A solution to this problem is the use of separate thermostats for the two subsystems (Marx and Hutter 2009; Sprik 1991)

Postprocessing

Data Analysis

Spatial Distribution Functions

For a system of N particles in a volume V at temperature T, the probability of molecule 1 being in the volume element $d\mathbf{R}_1$ around the position $\mathbf{R}_1$, molecule 2 being in $d\mathbf{R}_2$, …, molecule N being in $d\mathbf{R}_N$ is given by McQuarrie [1992]

$$P^{(N)}(\mathbf{R})d\mathbf{R} = P^{(N)}(\mathbf{R}_1,\ldots,\mathbf{R}_N)d\mathbf{R}_1,\ldots,d\mathbf{R}_N = \frac{e^{-E(\mathbf{R})/kT}}{Z_N} \tag{7.108}$$

with the configuration integral

$$Z_N = \int_V e^{-E(\mathbf{R})/kT} d\mathbf{R} \tag{7.109}$$

where $E(\mathbf{R})$ is the potential energy of the system at configuration $\mathbf{R}$ (cf. ❷ Eqs. 7.8 and ❷ 7.10).

For a subset of n molecules, the probability of molecule 1 being in $d\mathbf{R}_1$, …, molecule n being in $d\mathbf{R}_n$ is

$$P^{(n)}(\mathbf{R}_1,\ldots,\mathbf{R}_n) = \frac{\int \cdots \int e^{-E(\mathbf{R})/kT} d\mathbf{R}_{n+1}\ldots d\mathbf{R}_N}{Z_N} \tag{7.110}$$

The probability of *any* molecule being in $d\mathbf{R}_1$, …, *any* molecule n being in $d\mathbf{R}_n$ is

$$\rho^{(n)}(\mathbf{R}_1,\ldots,\mathbf{R}_n) = \frac{N!}{(N-n)!}P^{(n)}(\mathbf{R}_1,\ldots,\mathbf{R}_n) \tag{7.111}$$

In a liquid the probability of finding any one molecule in $d\mathbf{R}_1$, $\rho^{(1)}(\mathbf{R}_1)d\mathbf{R}_1$, is independent of $\mathbf{R}_1$. Therefore

$$\frac{1}{V}\int \rho^{(1)}(\mathbf{R}_1)d\mathbf{R}_1 = \rho^{(1)} = \frac{N}{V} = \rho \tag{7.112}$$

The dependence of the molecules of a liquid on all the other molecules, in other words, their correlation, is captured by the *correlation function* $g^{(n)}(\mathbf{R}_1,\ldots,\mathbf{R}_n)$, which is defined by

$$\rho^{(n)}(\mathbf{R}_1,\ldots,\mathbf{R}_n) = \rho^n g^{(n)}(\mathbf{R}_1,\ldots,\mathbf{R}_n) \tag{7.113}$$

Using ❷ Eq. 7.111 we can thus write

$$g^{(n)}(\mathbf{R}_1,\ldots,\mathbf{R}_n) = \frac{V^n N!}{N^n (N-n)!} \frac{\int \cdots \int e^{-E(\mathbf{R})/kT} d\mathbf{R}_{n+1}\ldots d\mathbf{R}_N}{Z_N} \tag{7.114}$$

The two-body correlation function $g^{(2)}(\mathbf{R}_1, \mathbf{R}_2)$ is of particular interest as it can be determined in X-ray diffraction experiments. In the following we shall only consider the dependence of $g^{(2)}$ on the interparticle distance $R = R_{12} = |\mathbf{R}_1 - \mathbf{R}_2|$, i.e., we have averaged over any angular dependence, and call $g^{(2)}(R_{12}) = g(R)$ the *radial distribution function*. The quantity $\rho g(R) d\mathbf{R}_I$ is proportional to the probability of finding another particle, I, in $d\mathbf{R}_I$ if the reference particle is at the origin. Spherical integration yields

$$\int \rho g(R) 4\pi R^2 dR = N - 1 \approx N \tag{7.115}$$

showing that $\rho g(R) 4\pi R^2 dR$ is the number of particles in the spherical volume element between R and $R + dR$ about the central particle. The radial distribution function $g(R)$ is proportional to the local density $\rho(R) = \rho g(R)$ about a certain molecule. In a fluid, $g(R) \to 1$ as $R \to \infty$, i.e., there is no long-range order and we "see" only the average particle density. At very short range, i.e., $R \to 0$, $g(R) \to 0$, due to the repulsiveness of the molecules. Examples from a CP-MD simulation of liquid water are shown in ❷ *Fig. 7-3*.
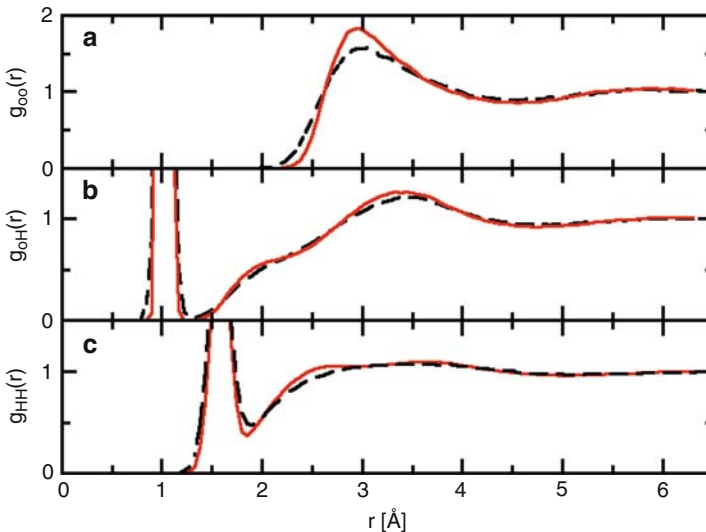
The radial distribution function $g(R)$ provides a useful measure of the quality of a simulation as it can be compared to experimental – X-ray or neutron diffraction – data obtained by Fourier transform of the structure factor

$$h(k) = \rho \int [g(R) - 1] e^{i\mathbf{kR}} d\mathbf{R} \tag{7.116}$$

where $\mathbf{k}$ is the wave vector.

In addition to characterizing the structure of a liquid, the radial distribution function may also be used to calculate thermodynamic properties such as the total energy,

$$E = \frac{3}{2} NkT + 2\pi N\rho \int_0^\infty u(R) g(R) R^2 dR \tag{7.117}$$

❏ Fig. 7-3

Radial distribution function of liquid water from CP-MD simulations at 900 and 1,200 K, respectively

the pressure,

$$p = \rho k T - \frac{2}{3}\pi \rho^2 \int_0^\infty \frac{du(R)}{dR} g(R) R^3 dR \qquad (7.118)$$

and the chemical potential,

$$\mu = k T \ln(\rho \Lambda^3) + 4\pi\rho \int_0^1 d\xi \int_0^\infty u(R) g(R, \xi) R^2 dR \qquad (7.119)$$

where

$$\Lambda = \sqrt{\frac{h^2}{2\pi m k T}} \qquad (7.120)$$

is the thermal de Broglie wavelength. By varying the coupling parameter ξ between 0 and 1, one can effectively take a molecule in and out of the system. It should be stressed that ❯ Eqs. 7.117–7.119 have been derived assuming a pairwise additive intermolecular potential $u(R)$.

We now define the *potential of mean force*, i.e., the interaction between n fixed molecules averaged over the configurations of the remaining molecules $n + 1, \ldots, N$, as

$$w^{(n)}(\mathbf{R}_1, \ldots, \mathbf{R}_n) = -k T \ln g^{(n)}(\mathbf{R}_1, \ldots, \mathbf{R}_n) \qquad (7.121)$$

The mean force acting on molecule J is then obtained from

$$f_J^{(n)} = -\nabla_J w^{(n)} \qquad (7.122)$$

Time Correlation Functions

The classical *time autocorrelation function* of some vectorial function

$$\mathbf{A}(t) = \mathbf{A}(P(t), Q(t)) = \mathbf{A}(P, Q; t) \qquad (7.123)$$

where $Q(t)$ and $P(t)$ are the generalized coordinate and momentum, respectively, is defined as

$$C(t) = < \mathbf{A}(0)\mathbf{A}(t) > = \int \cdots \int dP\, dQ\, \mathbf{A}(P, Q; 0)\mathbf{A}(P, Q; t) f(P, Q) \qquad (7.124)$$

where $f(P, Q)$ is the equilibrium phase space distribution function.

From the velocity autocorrelation function, for example, one can calculate the diffusion coefficient as

$$D = \frac{1}{3} \int_0^\infty < \mathbf{V}_I(0)\mathbf{V}_I(t) > dt \qquad (7.125)$$

where $\mathbf{V}_I$ is the velocity of particle I. Alternatively, one can obtain the diffusion coefficient for long times from the associated *Einstein relation*,

$$6tD = < |\mathbf{R}_I(t) - \mathbf{R}_I(0)|^2 > \qquad (7.126)$$

In practice, D is then determined from a linear fit to the *mean square displacement* (rhs of ❯ Eq. 7.126) as one sixth of the slope. An example is shown in ❯ *Fig. 7-4*.

Another common application of correlation functions is the calculation of IR absorption spectra. The lineshape function, $I(\omega)$, is given by the Fourier transform of the autocorrelation function of the electric dipole moment $\mathbf{M}$,

$$I(\omega) = \frac{1}{2\pi} \int_{-\infty}^\infty < \mathbf{M}(0)\mathbf{M}(t) > e^{-i\omega t} dt \qquad (7.127)$$
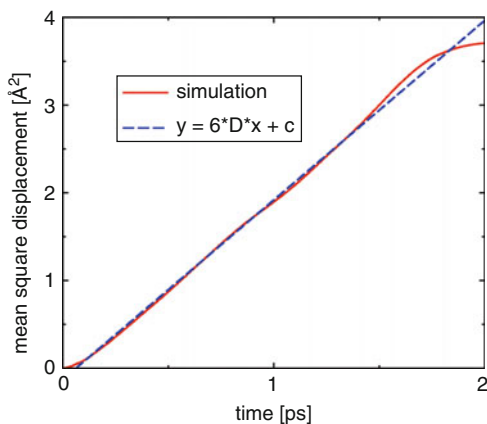
◾ **Fig. 7-4**
Mean square displacement of liquid water from CP-MD simulations at 900 K and linear fit to determine the diffusion constant *D* using ❯ Eq. 7.126
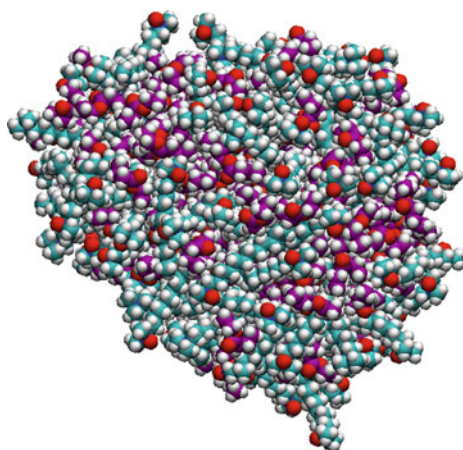
◾ **Fig. 7-5**
A snapshot of a micelle formed from DDAO molecules and oil molecules formed using the VMD software package (VMD **2009)**

Visualization

Due to the nature of MD simulations, one of the most productive forms of analysis of a simulation is to be able to visualize the trajectory of the molecules of interest. This is particularly useful since experimental techniques are not able to produce visual pictures of atomistic interactions and therefore it is something that only simulations (at this point) are able to provide. In order to visualize a simulation trajectory there are several different very powerful computer packages that are commonly used. These software packages include VMD (2009), PyMol (2010), RasMol (2008), and several others (Free Molecular Visualization Software 2008). ❯ *Figure 7-5* shows an example of the type of pictures that can be made using the visualization software.

Each of these codes will generally accept the trajectory in any number of standard inputs (i.e., pdb, xyz,…) and then will generate snapshots which can be rendered individually or as a movie. In addition to providing the visualization, these codes have become progressively powerful analysis codes in their own right. They now have the ability to measure bond lengths, angles, and dihedrals as a function of time, determine the solvent accessible surface area, hydrogen bond network, and many other useful structural related properties of the system.

References

ABINIT. (2010). http://www.abinit.org. Accessed 02 July 2011.

Allen, M. P., & Tildesley, D. J. (1987). *Computer simulation of liquids*. Oxford: Clarendon Press.

Amara, P., Field, M. J., Alhambra, C., & Gao, J. (2000). The generalized hybrid orbital method for combined quantum mechanical/molecular mechanical calculations: Formulation and tests of the analytical derivatives. *Theoretical Chemistry Accounts, 104*, 336.

Anderson, J. A., Lorenz, C. D., & Travesset, A. (2008). General purpose molecular dynamics simulations fully implemented on graphics processing units. *Journal of Computational Physics, 227*, 5342.

Aqvist, J., & Warshel, A. (1993). Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chemical Reviews, 93*, 2523.

Artacho, E., Anglada, E., Dieguez, O., Gale, J. D., Garcia, A., Junquera, J., Martin, R. M., Ordejón, P., Pruneda, J. M., Sánchez-Portal, D., & Soler, J. M. (2008). The SIESTA method; developments and applicability. *Journal of Physics: Condensed Matter, 20*, 064208.

Ashcroft, N. W., & Mermin, N. D. (1976). *Solid state physics*. Philadelphia: Saunders College Publishing.

Assfeld, X., & Rivail, J. L. (1996). Quantum chemical computations on parts of large molecules: The ab initio local self consistent field method. *Chemical Physics Letters, 263*, 100.

Assfeld, X., Ferré, N., & Rivail, J. L. (1998). In J. Gao & M. A. Thompson (Eds.), *Combined quantum mechanical and molecular mechanical methods, ACS Symp. Ser.* (Vol. 712, p. 234). Washington: American Chemical Society.

Aulbur, W. G., Jonsson, L., & Wilkins, J. W. (2000). Quasiparticle calculations in solids. *Solid State Physics, 54*, 1.

Bachelet, G. B., Hamann, D. R., Schlüter, M. (1982). Pseudopotentials that work: From H to Pu. *Physical Review B, 26*, 4199.

Balint-Kurti, G. G. (2008). Time-dependent and time-independent wavepacket approaches to reactive scattering and photodissociation dynamics. *International Reviews in Physical Chemistry, 27*, 507.

Baskes, M. I. (1992). Modified embedded-atom potentials for cubic materials and impurities. *Physical Review B, 46*, 2727.

Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A, 38*, 3098.

Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *Journal of Chemical Physics, 98*, 5648.

Bereau, T., & Deserno, M. (2009). Generic coarse-grained model for protein folding and aggregation. *Journal of Chemical Physics, 130*, 235106.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Nola, A. D., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics, 81*, 3684.

Betancourt, M. R., & Omovie, S. J. (2009). Pairwise energies for polypeptide coarse-grained models derived from atomic force fields. *Journal of Chemical Physics, 130*, 195103.

Bhandarkar, M., Bhatele, A., Bohm, E., Brunner, R., Buelens, F., Chipot, C., Dalke, A., Dixit, S., Fiorin, G., Freddolino, P., Grayson, P., Gullingsrud, J., Gursoy, A., Hardy, D., Harrison, C., Hénin, J., Humphrey, W., Hurwitz, D., Krawetz, N., Kumar, S., Kunzman, D., Lee, C., Mei, C., Nelson, M., Phillips, J., Sarood, O., Shinozaki, A., Zheng, G., & Zhu, F. (2009). NAMD User's Guide. Theoretical Biophysics Group, University of Illinois and Beckman Institute. http://www.ks.uiuc.edu/Research/namd/. Accessed 02 July 2011.

Blochl, P. E. (1994). Projector augmented-wave method. *Physical Review B, 50*, 17953.

Blochl, P. E., Forst, C. J., & Schimpl, J. (2003). Projector augmented wave method: Ab initio molecular dynamics with full wave functions. *Bulletin of Material Science, 26*, 33.

Bockstedte, M., Kley, A., Neugebauer, J., & Scheffler, M. (1997). Density-functional theory calculations for poly-atomic systems: Electronic structure, static and elastic properties and ab initio molecular dynamics. *Computer Physics Communications, 107*, 187.

Boeck, S. (2009). *Development and application of the S/PHI/nX library*. Saarbrücken: Südwestdeutscher Verlag für Hochschulschriften.

Bowler, D. R., Choudhury, R., Gillan, M. J., & Miyazaki, T. (2006). Recent progress with large-scale ab initio calculations: The CONQUEST code. *Physica Status Solidi B, 243*, 989.

Boys, S. F., & Bernardi, F. (1970). The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Molecular Physics, 19*, 553.

Bredow, T., & Jug, K. (2005). Theory and range of modern semiempirical molecular orbital methods. *Theoretical Chemistry Accounts, 113*, 1.

Brooks, B. R., Brooks, C. L., III, Mackerell, A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., & Karplus, M. (2009). CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry, 30*, 1545.

Buckingham, R. A. (1938). The classical equation of state of gaseous helium, neon and argon. *Proceedings of the Royal Society of London A, 168*, 264.

Bulo, R. E., Ensing, B., Sikkema, J., & Visscher, L. (2009). Toward a practical method for adaptive QM/MM simulations. *Journal of Chemical Theory and Computation, 5*, 2212.

Car, R., & Parrinello, M. (1985). Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters, 55*, 2471.

Case, D. A., Cheatham, T. E., III, Darden, T., Gohlke, H., Luo, R., Jr., K. M. M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. (2005). The Amber biomolecular simulation programs. *Journal of Computational Chemistry, 26*, 1668.

Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Crowley, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Kolossváry, I., Wong, F. K., Paesani, F., Vanicek, J., Wu, X., Brozell, S. R., Steinbrecher, T., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Mathews, D. H., Seetin, M. G., Sagui, C., Babin, V., & Kollman, P. (2008). AMBER 10.

San Francisco: University of California. http://ambermd.org/. Accessed 02 July 2011.

CASTEP. (2009). http://www.tcm.phy.cam.ac.uk/castep/. Accessed 02 July 2011.

CCDC. (2010). Cambridge crystallographic data centre. http://www.ccdc.cam.ac.uk. Accessed 02 July 2011.

Chadi, D. J., & Cohen, M. L. (1973). Special points in the brillouin zone. *Physical Review B, 8*, 5747.

CHARMM. (2009). http://www.charmm.org/. Accessed 02 July 2011.

Clark, S. J., Segall, M. D., Pickard, C. J., Hasnip, P. J., Probert, M. J., Refson, K., & Payne, M. C. (2005). First principles methods using CASTEP. *Zeitschrift für Kristallographie, 220*, 567.

CONQUEST. (2009). http://hamlin.phys.ucl.ac.uk/NewCQWeb/bin/view. Accessed 02 July 2011.

Cooke, I. R., Kremer, K., & Deserno, M. (2005). Tunable generic model for fluid bilayer membranes. *Physical Review E, 72*, 011506.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of American Chemical Society, 117*, 5179.

CP-PAW. (2006). http://orion.pt.tu-clausthal.de/paw/. Accessed 02 July 2011.

Császár, P., & Pulay, P. (1984). Geometry optimization by direct inversion in the iterative subspace. *Journal of Molecular Structure, 114*, 31.

DACAPO. (2006). https://wiki.fysik.dtu.dk/dacapo. Accessed 02 July 2011.

Daw, M. S., & Baskes, M. I. (1983). Semiempirical, quantum mechanical calculation of hydrogen embrittlement in metals. *Physical Review Letters, 50*, 1285.

Daw, M. S., & Baskes, M. I. (1984). Embedded-atom method: Derivation and application to impurities, surfaces, and other defects in metals. *Physical Review B, 29*, 6443.

Day, P. N., Jensen, J. H., Gordon, M. S., Webb, S. P., Stevens, W. J., Krauss, M., Garmer, D., Basch, H., & Cohen, D. (1996). An effective fragment method for modeling solvent effects in quantum mechanical calculations. *Journal of Chemical Physics, 105*, 1968.

Dirac, P. A. M. (1930). Note on exchange phenomena in the Thomas atom. *Proceedings of the Cambridge Philosophical Society, 26*, 376.

Doltsinis, N. L. (2006). In J. Grotendorst, S. Blügel, & D. Marx (Eds.), *Computational nanoscience: Do it yourself!* Jülich: NIC. http://www2.fz-juelich.de/nic-series/volume31/doltsinis3.pdf. Accessed 02 July 2011.

Doltsinis, N. L., & Marx, D. (2002a). Nonadiabatic car-parrinello molecular dynamics. *Physical Review Letters, 88*, 166402.

Doltsinis, N. L., & Marx, D. (2002b). First-principles molecular dynamics involving excited states and nonadiabatic transitions. *Journal of Theoretical and Computational Chemistry, 1*, 319–349.

Dreizler, R. M., & Gross, E. K. U. (1990). *Density–functional theory*. Berlin: Springer.

Elezgaray, J., & Laguerre, M. (2006). A systematic method to derive force fields for coarse-grained simulations of phospholipids. *Computer Physics Communications, 175*, 264.

Elstner, M., Porezag, D., Jungnickel, G., Elsner, J., Haugk, M., Frauenheim, T., Suhai, S., & Seifert, G. (1998). Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Physical Review B, 58*, 7260.

Ewald, P. P. (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annals of Physics, 369*, 253.

Ferré, N., Assfeld, X., & Rivail, J. L. (2002). Specific force field parameters determination for the hybrid ab initio QM/MM LSCF method. *Journal of Computational Chemistry, 23*, 610.

FHI98md. (2002). http://www.fhi-berlin.mpg.de/th/fhimd/. Accessed 02 July 2011.

Finnis, M. W., & Sinclair, J. E. (1984). A simple empirical n-body potential for transition metals. *Philosophical Magazine A, 29*, 6443.

Fornili, A., Sironi, M., & Raimondi, M. (2003). Determination of extremely localized molecular orbitals and their application to quantum mechanics/molecular mechanics methods and to the study of intramolecular hydrogen bonding. *Journal of Molecular Structure (THEOCHEM), 632*, 157.

Fornili, A., Loos, P.-F., Sironi, M., & Assfeld, X. (2006a). Frozen core orbitals as an alternative to specific frontier bond potential in hybrid Quantum Mechanics/Molecular Mechanics methods. *Chemical Physics Letters, 427*, 236.

Fornili, A., Moreau, Y., Sironi, M., & Assfeld, X. (2006b). On the suitability of strictly localized orbitals for hybrid QM/MM calculations. *Journal of Computational Chemistry, 27*, 515.

Free Molecular Visualization Software. (2008). http://www.umass.edu/microbio/rasmol/othersof.htm. Accessed 02 July 2011.

Fumi, F. G., & Tosi, M. P. (1964). Ionic sizes and born repulsive parameters in the NaCl-type alkali halides I : The Huggins-Mayer and Pauling forms. *Journal of Physics and Chemistry of Solids, 25*, 31.

Gao, J., Amara, P., Alhambra, C., & Field, M. J. (1998). A Generalized Hybrid Orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *Journal of Physics Chemistry A, 102*, 4714.

Garcia-Viloca, M., & Gao, J. (2004). Generalized hybrid orbital for the treatment of boundary atoms in combined quantum mechanical and molecular mechanical calculations using the semiempirical parameterized model 3 method. *Theoretical Chemistry Accounts, 111*, 280.

Giannozzi, P., Baroni, S., Bonini, N., Calandra, M., Car, R., Cavazzoni, C., Ceresoli, D., Chiarotti, G. L., Cococcioni, M., Dabo, I., Corso, A. D., Fabris, S., Fratesi, G., de Gironcoli, S., Gebauer, R., Gerstmann, U., Gougoussis, C., Kokalj, A., Lazzeri, M., Martin-Samos, L., Marzari, N., Mauri, F., Mazzarello, R., Paolini, S., Pasquarello, A., Paulatto, L., Sbraccia, C., Scandolo, S., Sclauzero, G., Seitsonen, A. P., Smogunov, A., Umari, P., & Wentzcovitch, R. M. (2009). QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter, 21*, 395502.

Goedecker, S., Teter, M., & Hutter, J. (1996). Separable dual-space Gaussian pseudopotentials. *Physical Review B, 54*, 1703.

Goetz, R., Compper, G., & Lipowsky, R. (1999). Mobility and elasticity of self-assembled membranes. *Physical Review Letters, 82*, 221.

Gordon, M. S., Freitag, M. A., Bandyopadhyay, P., Jensen, J. H., Kairys, V., & Stevens, W. J. (2001). The effective fragment potential method: A QM-based MM approach to modeling environmental effects in chemistry. *The Journal of Physical Chemistry A, 105*, 105.

Grigorenko, B. L., Nemukhin, A. V., Topol, I. A., & Burt, S. K. (2002). Modeling of biomolecular systems with the quantum mechanical and molecular mechanical method based on the effective fragment potential technique: Proposal of flexible fragments. *The Journal of Physical Chemistry A, 106*, 10663.

GROMOS. (2007). BIOMOS b.v, laboratory of physical chemistry ETH Hnggerberg, HCI. http://www.gromos.net/. Accessed 02 July 2011.

Hamann, D. R., Schlüter, M., Chiang, C. (1979). Norm-conserving pseudopotentials. *Physical Review Letters, 43*, 1494.

Hartwigsen, C., Goedecker, S., & Hutter, J. (1998). Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Physical Review B, 58*, 3641.

Heyden, A., Lin, H., & Truhlar, D. G. (2007). Adaptive partitioning in combined quantum mechanical

and molecular mechanical calculations of potential energy functions for multiscale simulations. *The Journal of Physical Chemistry B, 111*, 2231.

Hockney, R. W. (1970). The potential calculation and some applications. *Methods in Computational Physics, 9*, 135.

Hockney, R., & Eastwood, J. (1981). *Computer simulation using particles*. New York: McGraw-Hill.

Hofer, T. S., Pribil, A. B., Randolf, B. R., & Rode, B. M. (2005). Structure and dynamics of solvated Sn(II) in aqueous solution: An ab initio QM/MM MD approach. *Journal of American Chemical Society, 127*, 14231.

HOOMD. (2009). http://codeblue.umich.edu/hoomd-blue. Accessed 02 July 2011.

Hoover, W. G. (1985). Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A, 31*, 1695.

Hoover, W. G. (1986). Constant-pressure equations of motion. *Physical Review A, 34*, 2499.

Hutter, J., Lüthi, H. P., & Parrinello, M. (1994). Electronic structure optimization in plane-wave-based density functional calculations by direct inversion in the iterative subspace. *Computational Materials Science, 2*, 244.

Hutter, J., Kohlmeyer, A., Mundy, C. J., Mohamed, F., Schiffmann, F., Tabacchi, G., Forbert, H., Bethune, I., Kuo, W., Krack, M., Iannuzzi, M., Guidon, M., McGrath, M., Kuehne, T. D., Laino, T., Borstnik, U., VandeVondele, J., & Weber, V. (2009). CP2K – a general program to perform molecular dynamics simulations. http://cp2k.berlios.de. Accessed 02 July 2011.

ICSD. (2009). Inorganic crystal structure database. http://www.fiz-karlsruhe.de/icsd.html. Accessed 02 July 2011.

Izvekov, S., & Voth, G. A. (2005). A multiscale coarse-graining method for biomolecular systems. *The Journal of Physical Chemistry B, 109*, 2469.

Izvekov, S., & Voth, G. A. (2006). Multiscale coarse-graining of mixed phospholipid/cholesterol bilayers. *Journal of Chemical Theory and Computation, 2*, 637.

Jensen, F. (2007). *Introduction to computational chemistry*. Chichester: Wiley.

Jensen, J. H., Day, P. N., Gordon, M. S., Basch, H., Cohen, D., Garmer, D. R., Krauss, M., & Stevens, W. J. (1994). In D. A. Smith (Ed.), *Modeling the hydrogen bond*, ACS Symp. Ser. (Vol. 569, p. 139). Washington, DC: American Chemical Society.

Jørgensen, W. L., Madura, J. D., & Swenson, C. J. (1984). Optimized intermolecular potential functions for liquid hydrocarbons. *Journal of American Chemical Society, 106*, 6638.

Jung, J., Choi, C. H., Sugita, Y., & Ten-no, S. (2007). New implementation of a combined quantum mechanical and molecular mechanical method using modified generalized hybrid orbitals. *Journal of Chemical Physics, 127*, 204102.

Kairys, V., & Jensen, J. H. (2000). QM/MM boundaries across covalent bonds: A frozen localized molecular orbital-based approach for the effective fragment potential method. *The Journal of Physical Chemistry A, 104*, 6656.

Kantorovich, L. (2008). Generalized Langevin equation for solids. I. Rigorous derivation and main properties. *Physical Review B, 78*, 094304.

Kantorovich, L., & Rompotis, N. (2008). Generalized Langevin equation for solids. II. Stochastic boundary conditions for nonequilibrium molecular dynamics simulations. *Physical Review B, 78*, 094305.

Kendall, R. A., Apra, E., Bernholdt, D. E., Bylaska, E. J., Dupuis, M., Fann, G. I., Harrison, R. J., Ju, J., Nichols, J. A., Nieplocha, J., Straatsma, T. P., Windus, T. L., & Wong, A. T. (2000). High performance computational chemistry: An overview of NWChem a distributed parallel application. *Computer Physics Communications, 128*, 260.

Kerdcharoen, T., Liedl, K. R., & Rode, B. M. (1996). A QM/MM simulation method applied to the solution of Li$^+$ in liquid ammonia. *Chemical Physics, 211*, 313.

Kerdcharoen, T., & Morokuma, K. (2002). ONIOM-XS: An extension of the ONIOM method for molecular simulation in condensed phase. *Chemical Physics Letters, 355*, 257.

Kerker, G. (1980). Non-singular atomic pseudopotentials for solid state applications. *The Journal of Physical Chemistry C, 13*, L189.

Khalid, S., Bond, P. J., Holyoake, J., Hawtin, R. W., & Sansom, M. S. P. (2008). DNA and lipid bilayers: Self-assembly and insertion. *Journal of the Royal Society Interface, 5*, S241.

Kleinman, L., & Bylander, D. M. (1982). *Physical Review Letters, 48*, 1425.

Kołos, W. (1970). Adiabatic approximation and its accuracy. *Advances in Quantum Chemistry, 5*, 99.

Kremer, K., & Grest, G. (1990). Dynamics of entangled linear polymer melts: A molecular-dynamics simulation. *Journal of Chemical Physics, 92*, 5057.

Kresse, G., & Furthmüller, J. (1996). Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B, 54*, 11169.

Kuo, I.-F. W., Mundy, C. J., McGrath, M. J., Siepmann, J. I., VandeVondele, J., Sprik, M., Hutter, J., Chen, B., Klein, M. L., Mohamed, F., Krack, M., & Parrinello, M. (2004). Liquid water from first principles: Investigation of different sampling approaches. *The Journal of Physical Chemistry B, 108*, 12990.

Kutzelnigg, W. (1997). The adiabatic approximation I. The physical background of the Born-Handy ansatz. *Molecular Physics, 90*, 909.

Laasonen, K., Pasquarello, A., Car, R., Lee, C., & Vanderbilt, D. (1993). Car-Parrinello molecular dynamics with Vanderbilt ultrasoft pseudopotentials. *Physical Review B, 47*, 10142.

Laio, A., VandeVondele, J., & Rothlisberger, U. (2002). A Hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations. *Journal of Chemical Physics, 116*, 6941.

LAMMPS. (2010). http://lammps.sandia.gov/. Accessed 02 July 2011.

Lee, C., Yang, W., & Parr, R. C. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B, 37*, 785.

Lindahl, E., Hess, B., & van der Spoel, D. (2001). GROMACS 3.0: A package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling, 7*, 306.

Loos, P.-F., & Assfeld, X. (2007). Self-consistent strictly localized orbitals. *Journal of Chemical Theory and Computation, 3*, 1047.

Luty, B. A., Davis, M. E., Tironi, I. G., & van Gunsteren, W. F. (1994). A comparison of particle-particle, particle-mesh and Ewald methods for calculating electrostatic interactions in periodic molecular systems. *Molecular Simulation, 14*, 11.

Luty, B. A., Tironi, I. G., & van Gunsteren, W. F. (1995). Lattice-sum methods for calculating electrostatic interactions in molecular simulations. *Journal of Chemical Physics, 103*, 3014.

Lyubartsev, A. P. (2005). Multiscale modeling of lipids and lipid bilayers. *European Biophysics Journal, 35*, 53.

MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodham, B., Reihar, W. E., III, Raux, B., Schlenjrich, M., Smith, J. C., Store, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., & Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B, 102*, 3586.

Marrink, S. J., de Vries, A. H., & Mark, A. E. (2004). Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B, 108*, 750.

Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., & de Vries, A. H. (2007). The MARTINI force field: Coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B, 111*, 7812.

Marx, D., & Hutter, J. (2000). Ab initio molecular dynamics: Theory and implementation. In J. Grotendorst (Ed.), *Modern methods and algorithms of quantum chemistry*. Jülich: NIC. http://www.theochem.ruhr-uni-bochum.de/research/marx/marx.pdf. Accessed 02 July 2011.

Marx, D., & Hutter, J. (2009). *Ab initio molecular dynamics: Basic theory and advanced methods*. Cambridge: Cambridge University Press.

McQuarrie, D. A. (1992). *Statistical mechanics*. London: Academic.

Melchionna, S., Ciccotti, G., & Holian, B. L. (1993). Hoover NPT dynamics for systems varying in shape and size. *Molecular Physics, 78*, 533.

Mervis, J. (2001). NSF launches TeraGrid for academic research. *Science, 293*, 1235.

Meyer, B. (2006). The pseudopotential plane wave approach. In J. Grotendorst, S. Blügel, & D. Marx (Eds.), *Computational nanoscience: Do it yourself!* Jülich: NIC. http://www2.fz-juelich.de/nic-series/volume31/meyer1.pdf. Accessed 02 July 2011.

Molden. (2010). *A pre- and post processing program of molecular and electronic structure*. http://www.cmbi.ru.nl/molden/. Accessed 02 July 2011.

Monard, G., Loos, M., Théry, V., Baka, K., & Rivail, J. L. (1996). Hybrid classical quantum force field for modeling very large molecules. *International Journal of Quantum Chemistry, 58*, 153.

Monkhorst, H. J., & Pack, J. D. (1976). Special points for Brillouin-zone integrations. *Physical Review B, 13*, 5188.

Moreno, J., & Soler, J. M. (1992). Optimal meshes for integrals in real- and reciprocal-space unit cells. *Physical Review B, 45*, 13891.

Murphy, R. B., Philipp, D. M., & Friesner, R. A. (2000). Frozen orbital QM/MM methods for density functional theory. *Chemical Physics Letters, 321*, 113.

Nemukhin, A. V., Grigorenko, B. L., Bochenkova, A. V., Topol, I. A., & Burt, S. K. (2002). A QM/MM approach with effective fragment potentials applied to the dipeptideâŁ"water structures. *Journal of Molecular Structure (THEOCHEM), 581*, 167.

Nemukhin, A. V., Grigorenko, B. L., Topol, I. A., & Burt, S. K. (2003). *Journal of Computational Chemistry, 24*, 1410.

Nosé, S. (1984a). A unified formulation of the constant temperature molecular dynamics methods. *Journal of Chemical Physics, 81*, 511.

Nosé, S. (1984b). A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics, 52*, 255.

NWChem. (2008). http://www.nwchem-sw.org/index.php/Main_Page. Accessed 02 July 2011.

ONETEP. (2005). http://www2.tcm.phy.cam.ac.uk/onetep/. Accessed 02 July 2011.

Parr, R. G., & Yang, W. (1989). *Density functional theory of atoms and molecules*. Oxford: Oxford University Press.

Parrinello, M., Hutter, J., Marx, D., Focher, P., Tuckerman, M., Andreoni, W., Curioni, A., Fois, E., Roetlisberger, U., Giannozzi, P., Deutsch, T., Alavi, A., Sebastiani, D., Laio, A., Vande-Vondele, J., Seitsonen, A., & Billeter, S. (2008). Car–Parrinello molecular dynamics: An ab initio electronic structure and molecular dynamics program. http://www.cpmd.org. Accessed 02 July 2011.

Pastore, G., Smargiassi, E., & Buda, F. (1991). Theory of ab initio molecular-dynamics calculations. *Physical Review A44*, 6334.

PDB. (2010). RCSB protein data bank. http://www.rcsb.org/pdb. Accessed 02 July 2011.

Perdew, J. P., Chevary, J. A., Vosko, S. H., Jackson, K. A., Pederson, M. R., Singh, D. J., & Fiolhais, C. (1992). Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B, 46*, 6671.

Philipp, D. M., & Friesner, R. A. (1999). Mixed ab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide. *Journal of Computational Chemistry, 20*, 1468.

Phillips, J. C., Braun, R., Wang, W., Cumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., & Schulten, K. (2005a). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry, 26*, 1781.

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., & Schulten, K. (2005b). *Journal of Computational Chemistry, 26*, 1781.

PINY. (2005). http://homepages.nyu.edu/~mt33/PINY_MD/PINY.html. Accessed 02 July 2011.

Plimpton, S. J. (1995). Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics, 117*, 1.

Polák, R. (1986). An investigation of the importance of many-centre effects in the diatomics-in-molecules approach. *Chemical Physics, 103*, 277.

Ponder, J. W., Wu, C., Ren, P., Pande, V. S., Chodera, J. D., Schnieders, M. J., Haque, I., Mobley, D. L., Lambrecht, D. S., DiStasio, R. A., Jr., Head-Gordon, M., Clark, G. N. I., Johnson, M. E., & Head-Gordon, T. (2010). Current status of the AMOEBA polarizable force field. *The Journal of Physical Chemistry B, 114*, 2549.

Pu, J., Gao, J., & Truhlar, D. G. (2004a). Combining Self-Consistent-Charge Density-Functional Tight-Binding (SCC-DFTB) with molecular mechanics by the Generalized Hybrid Orbital (GHO) method. *The Journal of Physical Chemistry A, 108*, 5454.

Pu, J., Gao, J., & Truhlar, D. G. (2004b). Generalized Hybrid Orbital (GHO) method for combining ab initio Hartree-Fock wave functions with molecular mechanics. *The Journal of Physical Chemistry A, 108*, 632.

Pu, J., Gao, J., & Truhlar, D. G. (2005). Generalized Hybrid-Orbital method for combining density functional theory with molecular mechanicals. *ChemPhysChem, 6*, 1853.

Pulay, P. (1969). Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules I. Theory. *Molecular Physics, 17*, 197.

Pulay, P. (1980). Convergence acceleration of iterative sequences. The case of SCF iteration. *Chemical Physics Letters, 73*, 393.

Pulay, P. (1982). Improved SCF convergence acceleration. *Journal of Computational Chemistry, 3*, 556.

Pulay, P. (1987). Analytical derivative methods in quantum chemistry. *Advances in Chemical Physics, 69*, 241.

PWscf. (2009). http://www.pwscf.org/home.htm. Accessed 02 July 2011.

PyMOL. (2010). http://www.pymol.org. Accessed 02 July 2011.

QuantumEspresso. (2009). http://www.quantum-espresso.org. Accessed 02 July 2011.

RasMol. (2008). http://rasmol.org/. Accessed 02 July 2011.

Reciprocal Net. (2004). A distributed crystallography network for researchers, students and the general public. http://www.reciprocalnet.org. Accessed 02 July 2011.

Reed, D. A. (2003). Grids, the TeraGrid, and beyond. *Computer, 36*, 62.

Refson, K. (2000). Moldy: A portable molecular dynamics simulation program for serial and parallel computers. *Computer Physics Communications, 126*, 310.

Refson, K. (2001). Moldy user's manual. http://www.ccp5.ac.uk/moldy/moldy.html/. Accessed 02 July 2011.

Scott, W. R. P., Hünenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Krüger, P., & van Gunsteren, W. F. (1999). The GROMOS biomolecular simulation program package. *The Journal of Physical Chemistry A, 103*, 3596.

Segall, M. D., Lindan, P. L. D., Probert, M. J., Pickard, C. J., Hasnip, P. J., Clark, S. J., & Payne, M. C. (2002). First-principles simulation: Ideas, illustrations and the CASTEP code. *Journal of Physics: Condensed Matter, 14*, 2717.

Senn, H. M., & Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition, 48*, 1198.

Shelley, J. C., Shelley, M. Y., Reeder, R. C., Bandyopadhyay, S., Moore, P. B., & Klein, M. L. (2001). Simulations of phospholipids using a coarse grain model. *The Journal of Physical Chemistry B, 105*, 9785.

Shillcock, J. C., & Lipowsky, R. (2002). Equilibrium structure and lateral stress distribution of amphiphilic bilayers from dissipative particle dynamics simulations. *Journal of Chemical Physics, 117*, 5048.

Shirts, M., & Pande, V. S. (2000). Screen savers of the world unite! *Science, 290*, 1903.

Shurki, A., & Warshel, A. (2003). Structure/function correlations of enzymes using MM, QM/MM and related approaches; methods, concepts, Pitfalls and current progress. In V. Dagett (Ed.), *Protein simulations*, Adv. Protein Chem. (Vol. 66, p. 249). San Diego: Academic.

SIESTA. (2010). http://www.icmab.es/siesta/. Accessed 02 July 2011.

Sironi, M., Genoni, A., Civera, M., Pieraccini, S., & Ghitti, M. (2007). Extremely localized molecular orbitals: Theory and applications. *Theoretical Chemistry Accounts, 117*, 685.

Skylaris, C.-K., Haynes, P. D., Mostofi, A. A., & Payne, M. C. (2005). Introducing ONETEP: Linear-scaling density functional simulations on parallel computers. *Journal of Chemical Physics, 122*, 084119.

Smith, W., Yong, C. W., & Rodger, P. M. (2002). DL_POLY: Application to molecular simulation. *Molecular Simulation, 28*, 385.

Soler, J. M., Artacho, E., Gale, J. D., Garcia, A., Junquera, J., Ordejón, P., & Sánchez-Portal, D. (2002). The SIESTA method for ab initio order-N materials simulation. *Journal of Physics: Condensed Matter, 14*, 2745.

S/PHI/nX. (2009). http://www.mpie.de/index.php?id=sxlib. Accessed 02 July 2011.

Sprik, M. (1991). Computer simulation of the dynamics of induced polarization fluctuations in water. *The Journal of Physical Chemistry, 95*, 2283.

Stillinger, F., & Weber, T. A. (1985). Computer simulation of local order in condensed phases of silicon. *Physical Review B, 31*, 5262.

Sun, H., Ren, P., & Fried, J. R. (1998). The COMPASS force field: Parameterization and validation for phosphazenes. *Computational and Theoretical Polymer Science, 8*, 229.

Sutmann, G. (2002). Classical molecular dynamics. In J. Grotendorst, D. Marx, & A. Muramatsu (Eds.), *Quantum simulations of complex many-body systems: From theory to algorithms*. Jülich: NIC. http://www2.fz-juelich.de/nic-series/volume10/sutmann.pdf. Accessed 02 July 2011.

Sutmann, G. (2006). Molecular dynamics – vision and reality. In J. Grotendorst, S., Blügel, & D. Marx (Eds.), *Computational nanoscience: Do it yourself!* Jülich: NIC. http://www2.fz-juelich.de/nic-series/volume31/sutmann.pdf. Accessed 02 July 2011.

Swope, W. C., Anderson, H. C., Berens, P. H., & Wilson, K. R. (1982). *Journal of Chemical Physics, 76*, 637.

Tepper, H. L., & Voth, G. A. (2005). A coarse-grained model for double-helix molecules in solution: Spontaneous helix formation and equilibrium properties. *Journal of Chemical Physics, 122*, 124906.

Tersoff, J. (1988). New empirical approach for the structure and energy of covalent systems. *Physical Review B, 37*, 6991.

Tersoff, J. (1989). Modeling solid-state chemistry: Interatomic potentials for multicomponent systems. *Physical Review B, 39*, 5566.

Théry, V., Rinaldi, D., Rivail, J. L., Maigret, B., & Ferenczy, G. G. (1994). Quantum mechanical computations on very large molecular systems: The local self-consistent field method. *Journal of Computational Chemistry, 15*, 269.

Thiel, W. (2009). QM/MM methodology: Fundamentals, scope, and limitations. In J. Grotendorst, N. Attig, S. Blgel, & D. Marx (Eds.), *Multiscale simulation methods in molecular sciences*. Jülich: NIC. http://www2.fz-juelich.de/nic-series/volume42/thiel.pdf. Accessed 02 July 2011.

Todorov, I. T., & Smith, W. (2009). *The DL_POLY_3 user manual*. http://www.cse.scitech.ac.uk/ccg/software/DL_POLY/. Accessed 02 July 2011.

Tosi, M. P., & Fumi, F. G. (1964). Ionic sizes and born repulsive parameters in the NaCl-type alkali halides II : The generalized Huggins-Mayer form. *Journal of Physics and Chemitry of Solids, 25*, 45.

Toth (2009). Information systems. http://www.tothcanada.com. Accessed 02 July 2011.

Toton, D., Lorenz, C. D., Rompotis, N., Martsinovich, N., & Kantorovich, L. (2010). Temperature control in molecular dynamic simulations of non-equilibrium processes. *Journal of Physics Condensed Matter, 22*, 074205.

Tozzini, V. (2005). Coarse-grained models for proteins.*Current Opinion in Structural Biology, 15*, 144.

Troullier, N., & Martins, J. L. (1990). A straightforward method for generating soft transferable

pseudopotentials. *Solid State Communications, 74*, 613.

Troullier, N., & Martins, J. L. (1991). Efficient pseudopotentials for plane-wave calculations. *Physical Review B, 43*, 1993.

Tuckerman, M. E. (2002). Path integration via molecular dynamics. In J. Grotendorst, D. Marx, & A. Muramatsu (Eds.), *Quantum simulations of complex many-body systems: From theory to algorithms*, Jülich: NIC. http://www.fz-juelich.de/nic-series/volume10/tuckerman1.pdf. Accessed 02 July 2011.

Tuckerman, M. E., & Hughes, A. (1998). Path integral molecular dynamics. In B. J. Berne, G. Ciccotti, & D. F. Coker (Eds.), *Classical and quantum dynamics in condensed phase simulations* (p. 311). Singapore: World Scientific.

Tuckerman, M. E., Berne, B. J., Martyna, G. J., & Klein, M. L. (1993). Efficient molecular dynamics and hybrid Monte Carlo algorithms for path integrals. *Journal of Chemical Physics, 99*, 2796.

van Beest, B. W. H., Kramer, G. J., & van Santen, R. A. (1990). Force fields for silicas and aluminophosphates based on ab initio calculations. *Physical Review Letters, 64*, 1955.

van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005a). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry, 26*, 1701.

van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A. R., Apol, E., Meulenhoff, P. J., Tieleman, D. P., Sijbers, A. L. T. M., Feenstra, K. A., van Drunen, R., & Berendsen, H. J. C. (2005b). Gromacs user manual version 4.0. http://www.gromacs.org/. Accessed 02 July 2011.

Vanderbilt, D. (1985). Optimally smooth norm-conserving pseudopotentials. *Physical Review B, 32*, 8412.

Vanderbilt, D. (1990). Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Physical Review B, 41*, 7892.

Vanderbilt Ultra-Soft Pseudopotential Site. (2006). http://www.physics.rutgers.edu/~dhv/uspp/. Accessed 02 July 2011.

VandeVondele, J., Krack, M., Mohamed, F., Parrinello, M., Chassaing, T., & Hutter, J. (2005). Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Computer Physics Communications, 167*, 103.

VandeVondele, J., Iannuzzi, M., & Hutter, J. (2006). Large scale condensed matter calculations using the Gaussian and augmented plane waves method. In *Computer simulations in condensed matter systems: From materials to chemical biology, Volume 1. Lecture notes in physics* (Vol. 703, p. 287). Berlin/Heidelberg: Springer.

VASP. (2009). http://cms.mpi.univie.ac.at/vasp/. Accessed 02 July 2011.

Verlet, L. (1967). Computer "Experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical Review, 159*, 98.

VMD. (2009). http://www.ks.uiuc.edu/Research/vmd/. Accessed 02 July 2011.

Vosko, S. H., Wilk, L., & Nusair, M. (1980). Accurate spin-dependent electron liquid correlation energies for local spin-density calculations – A critical analysis. *Canadian Journal of Physics, 58*, 1200.

Wang, J. M., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Toward direct determination of conformations of protein building units from multidimensional NMR experiments. V. NMR chemical shielding analysis of N-formyl-serinamide, a model for polar side-chain containing peptides. *Journal of Computational Chemistry, 24*, 1157.

Warshel, A. (1991). *Computer modeling of chemical reactions in enzymes and solutions*. New York: Wiley.

Warshel, A. (2003). Computer simulations of enzyme catalysis: Methods, progress, and insights. *Annual Review of Biophysics and Biomolecular Structure, 32*, 425.

Warshel, A., & Levitt, M. (1976). Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology, 103*, 227.

Woodcock, L. V. (1971). Isothermal molecular dynamics calculations for liquid salts. *Chemical Physics Letters, 10*, 257.

Worth, G. A., Meyer, H. D., Koeppel, H., Cederbaum, L. S., & Burghardt, I. (2008). Using the MCTDH wavepacket propagation method to describe multimode non-adiabatic dynamics. *International Reviews in Physical Chemistry, 27*, 569.

Zhang, Y. (2005). A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *Journal of Chemical Physics, 122*, 024114.

Zhang, Y. (2006). Pseudobond ab initio QM/MM approach and its applications to enzyme reactions. *Theoretical Chemistry Accounts, 116*, 43.

Zhang, Y., Lee, T.-S., & Yang, W. (1999). A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *Journal of Chemical Physics, 110*, 46.