Computational Modeling of Genetic and Biochemical Networks

edited by James M. Bower and Hamid Bolouri

The MIT Press

# 6 Atomic-Level Simulation and Modeling of Biomacromolecules

Nagarajan Vaidehi and William A. Goddard III

6.1 Introduction

In principle, all the problems in biology could be solved by solving the time-dependent Schroedinger equation (quantum mechanics, QM). This would lead to a detailed understanding of the role that molecular-level interactions play in determining the fundamental biochemistry at the heart of biology and pharmacology. The difficulty is the vast range of length and time scales, from a nitrous oxide molecule to an organ (heart, lung), which makes a QM solution both impractical and useless. It is impractical because there are too many degrees of freedom describing the motions of the electrons and atoms, whereas in the functioning of an organ it may be only the rate of transfer across some membrane. The solution to both problems is the hierarchical strategy outlined in figure 6.1. We average over the scale of electrons (from QM) to describe the forces on atoms (the force field, FF), then average over the dynamics of atoms (molecular dynamics, MD) to describe the motions of large biomolecules, then average over the molecular scale to obtain the properties of membranes, then average over the components in a cell, then average over the cells to describe a part of an organ. The strategy is to develop a methodology for going between these various levels so that first principle's theory can be used to predict the properties of new systems.

In this chapter, we describe the first two levels of simulations and the attempts to coarsen the simulations to the mesoscale level in the hierarchy. The complexity of QM limits its applications to systems with only 10 to 200 atoms (depending on the accuracy), leading to distance scales of less than 20 Å and time scales of femtoseconds. The quantum mechanical forces are then translated into a set of parameters describing the bonded and nonbonded forces in a molecule. This set of parameters is known as the *force field*. Given the forces calculated with the FF, we solve Newton's equations to describe the motions of atoms with time, a process referred to as *molecular dynamics*. With MD, one can now consider systems with up to 1 million atoms, allowing practical simulations of systems as large as small viruses (say, 300 Å and 10 ns). The fundamental unit of MD is atoms (not the electrons of QM), allowing us to interpret the chemistry of the systems. A set of parameters that describe the forces between assemblies of atoms (collections of atoms or domains of a protein) in a molecule can be derived from the output of MD simulations. These parameters are further used in mesoscale dynamics.

These atomistic computational methods of chemistry and physics have evolved into sophisticated but practical tools that (using modern supercomputers) now allow many systems to be modeled and simulated. Thus, it is becoming possible to quantitatively predict the three-dimensional structure and dynamics of important biomacromolecules and to ana-
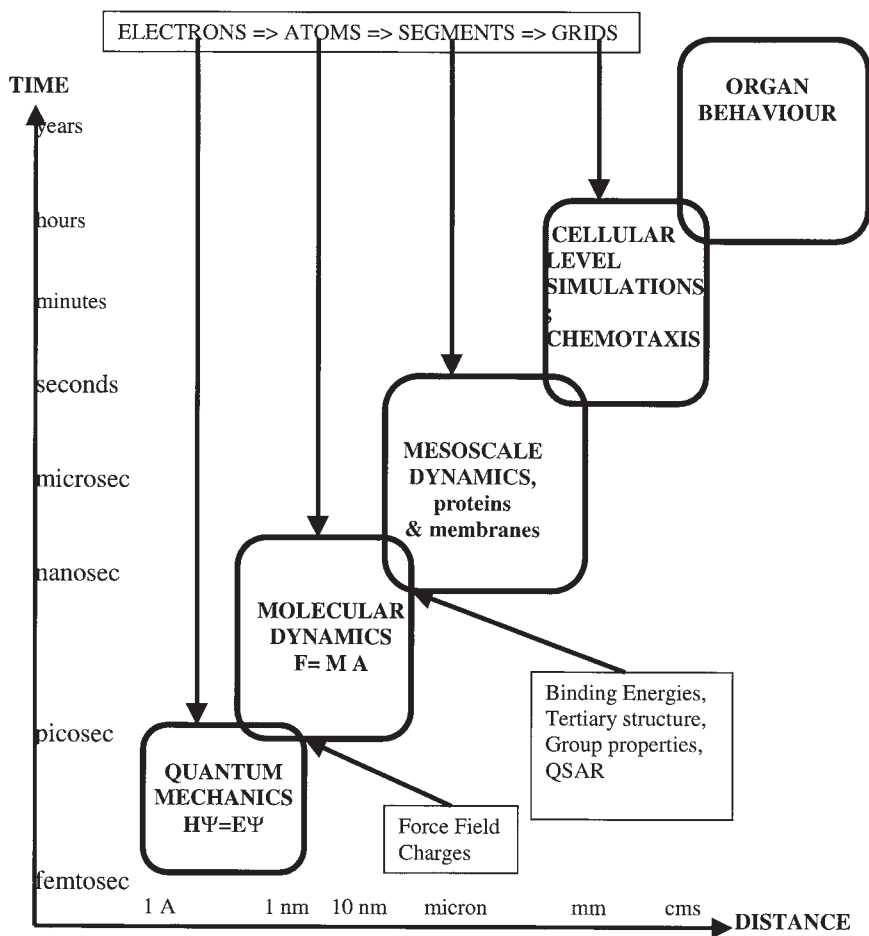
Figure 6.1
The hierarchy of biomolecular simulations. It is necesary to predict reliable properties before starting an experiment. The foundation is quantum mechanics. This allows prediction in advance, but is not practical for time and distance scales of molecular engineering. Thus one must extend from QM to large-scale, cellular-level dynamics by a succession of scales, where at each scale the parameters are determined by averaging over the finer scale.

lyze their interactions with other molecules at appropriate levels of computational resolution (ranging from electronic, to molecular, to the mesoscale cellular level). This makes it practical to begin addressing many complex issues of biological systems in terms of atomistic descriptions that provide quantitative information about the fundamental processes. Such atomistic computer simulations allow one to obtain static and dynamic molecular models for complex biosystems that describe the properties of the macroscale system or process in terms of concepts emphasizing the atomic origins of the phenomena (e.g., how the precise shape and exposed surface of a protein determines its function, which is critical to a drug design). This atomic-level description of the dynamic structure of proteins should be valuable for understanding:

• reaction chemistry at active sites in enzymes

• binding energetics and rates of small molecules or ligands to DNA, enzymes, and receptors (e.g., retinal to the signal receptor to protein rhodopsin)

• binding of antibodies to antigens

• conformational changes in proteins and how these changes modify function (relevant to neurological diseases)

• binding of proteins and other ligands to specific sites on DNA (relevant to expression)

6.2 Molecular Dynamics

Depending on the size of the system to be modeled and the accuracy required, there has evolved a hierarchy of QM methods in which approximations are often used to obtain greater speeds. The most accurate methods use no experimental data (first principles or *ab initio* QM) and are known by such names as Hartree–Fock (HF), density functional theory (DFT), and configuration interaction (CI) (Schaefer 1984, Parr and Yang 1989). However more approximate but much faster semiempirical methods [known by such names as modified intermediate neglect of differential overlap (MINDO), extended Hückel, and AM1) based partly on comparisons with experiments are quite valuable for many problems (Pople and Beveridge 1970). Such QM methods are essential for describing systems in which the nature of the bonds changes; for example, chemical reactions, excited states of molecules, and electron transfer.

In describing the structure and dynamics of large molecules such as proteins and DNA, the nature of the bonds is relatively insensitive to the environment. Instead, the focus of interest usually involves packing and conformation. For such problems, the electrons of QM can be accurately replaced with springs and the dynamics described with Newton's equations rather Schroedinger's. Here the choice of parameters in the FF is critical. They should give a description close to QM (and experiment). This has served well for many

problems; however, it is usually difficult to include polarization, charge transfer, and changes in bond order in the FF. Consequently, direct theoretical descriptions of chemical reactions have been the domain of QM. Recent progress in developing FFs that allow for bond breaking and reactions has been reported (Che et al. 1999, Qi et al. 1999, Demiralp et al. 1998).

6.2.1   The Force Field

The choice of the FF is critical for accurate predictions of the properties of a system. For biomolecules, the FF is described in terms of two types of interaction energies:

$$E_{tot} = E_{nonbond} + E_{valence} , \tag{6.1}$$

where $E_{valence}$ describes interactions involving changes in the covalent bonds and $E_{nonbond}$ describes the nonbonded interactions.

The nonbond energy is separated into electrostatic (Coulomb), van der Waals (VDW), and sometimes explicit hydrogen bond (HB) components. Each atom has associated with it an atomic charge, q_I, leading to an electrostatic energy of the form

$$E_{elec} = \sum_{I,J} \frac{q_I q_J}{\varepsilon R_{IJ}}. \tag{6.2}$$

Here q_I is the charge on atom I, R_{IJ} is the distance between atoms I and J, and ε accounts for the units and dielectric constant of the medium. This raises the issue of how to determine the charges. For modern FF, the charges are determined either directly from QM or from charge equilibration (QEq), a general semiempirical scheme that allows charges to depend on instantaneous structure. Sometimes the dielectric constant is used to replace some effects of the solvent and sometimes the solvent is included explicitly or by a continuous Poisson–Boltzmann approximation.

For biological systems, hydrogen bonding (e.g., between the amide hydrogen and the carbonyl oxygen) is particularly important in determining structure and energetics. Thus, some FFs include specific special HB interaction terms in the VDW part of (6.2) (Levitt 1983, Brooks et al. 1983, Weiner et al. 1986, Mayo et al. 1990, Cornell et al. 1995). However since QM shows that electrostatics dominates hydrogen bond interactions, most modern FFs account for hydrogen bonding through the electrostatics (van Gunsteren et al. 1987, Jorgensen and Tirado-Tives 1988, Hermans et al. 1984).

The two most common forms for the VDW energy are the Lennard-Jones 6–12

$$E_{LJ} = A R_{IJ}^{-12} - B R_{IJ}^{-6} \tag{6.3}$$

and the Buckingham exponential, −6

$$E_{\exp 6} = Ae^{-CR_{IJ}} - BR_{IJ}^{-6}, \qquad (6.4)$$

where the *A, B,* and *C* parameters are usually defined by a comparison with experimental data or accurate QM calculations on small molecules. Morse functions

$$E_{IJ} = A\{[e^{-C(R-Ri)} - 1]^2 - 1\} \qquad (6.5)$$

are also used (Gerdy 1995, Brameld et al. 1997).

Because the nonbond terms require calculation for all pairs of atoms (scaling as the square of the number of atoms), this is a bottleneck for simulations of very large systems. To reduce these costs, it is common to ignore interactions longer than some cutoff radius (using a spline function to smooth the potential at the cutoff radius). More recently, fast multipole techniques [the cell multipole method (CMM); Ding 1992, Lim et al. 1997, Figueirido et al. 1997] have been used to obtain accurate nonbond energies without cutoffs but scaling linearly with the size of the system. These fast and accurate methods are being used in MD simulations for large-scale biological systems (Vaidehi and Goddard 1997).

The valence energy is usually described as

$$E_{valence} = E_{bond} + E_{angle} + E_{torsion} + E_{inversion}, \qquad (6.6)$$

where bond describes the interaction between two bonded atoms; angle describes the interaction between two bonds sharing a common atom; torsion describes the interaction between a bond *IJ* and a bond *KL* connected through a bond *JK;* and inversion is used to describe nonplanar distortions at atoms with three bonds.

Since the covalent bonds are expected to remain near equilibrium, the bond stretching and angle bending are taken as harmonic:

$$E_{bond} = \frac{1}{2} K_{IJ} (R_{IJ} - R_0)^2 \qquad (6.7)$$

$$E_{angle} = \frac{1}{2} K_{IJK} (\theta - \theta_0)^2 \text{ or } (1/2)C(\cos\theta - \cos\theta_0)^2. \qquad (6.8)$$

Here R_0 is the equilibrium bond distance, K_{IJ} is the bond force constant, θ is the bond angle between bonds *IJ* and *JK,* θ_0 is the equilibrium bond angle, and K_{IJK} (or *C*) is the angle force constant.

The torsion energy is described in terms of the dihedral angle, ϕ, between bonds *IJ* and *KL* along bond *JK*. This is periodic and can be written as

$$E_{torsion} = \frac{1}{2} \sum_{n=1}^{6} K_{\phi,n} [1 - d\cos(n\phi)], \qquad (6.9)$$

where $K_{\phi,n}$ is the torsion energy barrier for periodicity n and $d = \pm 1$ describes whether the torsion angle ($\phi = 0$) is a minimum ($d = +1$) or a maximum. Torsion potentials are essential in describing the dependence of the energies on conformation.

Finally, an inversion term is needed to describe the distortions from planarity of atoms making three bonds (e.g., in aromatic amines or amides). For cases where the equilibrium geometry is planar (e.g., amine N or C), we use

$$E_{\text{inversion}} = K_{\text{inv}}(1 - \cos\omega). \tag{6.10}$$

For cases where the equilibrium geometry is nonplanar (e.g., an amine), we use

$$E_{\text{inversion}} = \frac{1}{2}C_I(\cos\omega - \cos\omega^0)^2 , \tag{6.11}$$

where

$$C_I \frac{K_{\text{inv}}}{(\sin\omega^0)^2}. \tag{6.12}$$

Standard Force Fields   Given the functional forms as described above, the FF is defined by the particular choices for the parameters in the FF (force constants and equilibrium geometries). Three strategies have been used for biological systems.

One is to develop the FFs for a specific class of molecules. The most popular FFs for proteins and DNA are AMBER (Weiner et al. 1984, Cornell et al. 1995) and CHARMM (Brooks 1983, Mackrell et al. 1995), which are parameterized to describe the naturally occurring amino acids and nucleic acids. These parameterizations include the atomic charges required to describe the electrostatics. Such FFs have been quite useful, and the majority of simulations on natural protein and DNA systems use these FF. However, unusual ligands such as drug molecules, cofactors, substrates, or their modifications are difficult to incorporate, as are non-natural amino acids or bases. Useful here are FFs developed for organic systems [OPLS (Pranata et al. 1991) and MM3 (Allinger and Schleyer 1996)] which can describe parameters of non-natural amino acids or bases, along with most molecules that bind to biosystems.

The second strategy is to develop rule-based FFs based on the character and connectivity of the molecules. The simplest such generic FF is Dreiding (Mayo et al. 1990): Equilibrium bond distances are based on atomic radii, and the bond angles, inversion angles, and torsion periodicities are derived from simple rules based on fundamental ideas of bonding. There is only one bond force constant (bond order times 700 kcal/mol Å), one angle force constant 100 kcal/mol rad), and simple rules based on fundamental ideas of bonding. To obtain generic charges, the charge equilibrium method (Rappe and Goddard 1991) was developed in which all charges of all molecules are determined from three parameters per

atom (radius, electronegativity, and hardness). Despite its simplicity, Dreiding gives accurate structures for the main group elements (the B, C, N, O, and F columns) most prevalent in biology.

A more general generic FF that also treats transition metals is the universal force field (UFF; Rappe et al. 1992), which treats all elements of the periodic table (through Lr, element 103). UFF includes simple rules in which the force constants of molecules are derived from atomic parameters. Such generic FFs are most useful for systems with unusual arrangements of atoms or for new molecules for which there are no experimental or QM data. For applications in which it is necessary to have the exact molecular structure, such generic FFs may not be sufficiently accurate. Hence, there is a need for a generic FF that incorporates just enough specificity for accurate simulations of biomolecules while providing the flexibility to model all other organic molecules.

Accurately predicting the vibrational spectra of the molecule, in addition to the geometry and energy, requires the third strategy for FFs in which the parameters are optimized for a specific class. This requires cross-terms coupling different bonds and angles. A general procedure for developing such spectroscopic FFs from QM is the biased Hessian method, which has been used for many systems (Dasgupta et al. 1996). Usually the spectroscopic quality FFs are useful only for limited classes of molecules.

Effect of Solvents The role of solvents (particularly water) is critical in biological simulations, since the secondary and tertiary structures of proteins are determined by the nature of the solvent. Several levels of simulation have been used.

The earliest studies ignored the solvent entirely, usually replacing the effect of solvent polarization by using a dielectric constant larger than one (which is often distance dependent). At this level of approximation, it is important to include counterions to represent the effects of solvent on charged groups. Such simulations were useful in understanding the gross properties of systems.

The most accurate MD treatments include an explicit description of the water using an FF adjusted to describe the bulk properties of water (Jorgensen et al. 1983; Levitt et al. 1997; Rahman and Stillinger 1971; Berendsen et al. 1981, 1987). Although they are accurate, such calculations usually require very long time scales in order to allow the hydrogen-bonding network in the water to equilibrate as the biomolecule undergoes dynamic motion. In addition, for an accurate treatment of solvent effects, the number of solvent atoms may be ten times that of the biomolecule.

An excellent compromise for attaining most of the accuracy of explicit water, while eliminating the atoms and time scale of the solvent, is the dielectric continuum model. Here the electrostatic field of the protein is allowed to polarize the (continuum) solvent, which then acts back on the protein, leading to the Poisson–Boltzmann equation (Sitkoff et al. 1993). Recent developments have led to computationally efficient techniques (Tannor

et al. 1994) that can accurately account for effects of solvation on the forces (geometry) and free energy. For example, solvation energies of neutral molecules are accurate to better than 1 kcal/mol.

6.2.2   Molecular Dynamics Methods

The Fundamental Equations   Given the FF, the dynamics are obtained by solving Newton's equations of motion:

$$-\mathbf{F}_i = m_i \ddot{x}_i, \tag{6.13}$$

where $\mathbf{F}_i$ is the force vector, $\ddot{x}_i$ denotes the acceleration, and m_i is the mass of atom i. Solving equation (6.13) leads to $3N$ coordinates and $3N$ velocities that describe the trajectory of the system as a function of time. Often this dynamic trajectory provides valuable information about a system. Thus, MD methods have been useful for the exploration of structure–activity relationships in biological molecules (McCammon 1987).

However, more often it is the ensemble of conformations near equilibrium that is required to calculate accurate properties. Assuming that the barriers between different relevant structures are sufficiently small that they can be sampled in the time scale of the simulations, we think that the collection of conformations sampled in the dynamics can be used as the ensemble for calculating properties.

The steps in MD simulations are as follows:

1. Start with the structure ($3N$ coordinates), which may be obtained from a crystal structure or from Build software using standard rules of bonding. In addition, it is necessary to have an initial set of velocities, which are chosen statistically to describe a Maxwell–Boltzmann distribution.

2. At each timestep, calculate the potential energy and its derivative to obtain the force on every atom in the molecule. Equation 6.14 is then solved to obtain the $3N$ accelerations at timestep t.

$$\ddot{x}_i = f_i / m_i = -\nabla E_{\text{tot}} / m_i, \tag{6.14}$$

where ∇E_{tot} is the gradient of the potential energy.

3. To obtain the velocities and coordinates of each atom as a function of time (the trajectory), we consider a timestep δ and write the acceleration at the nth timestep as

$$\ddot{x}_n = \frac{\dot{x}_{n+\frac{1}{2}} - x_{n-\frac{1}{2}}}{\delta}, \tag{6.15}$$

leading to

$$\dot{x}_{n+\frac{1}{2}} = \dot{x}_{n+\frac{1}{2}} - \frac{\delta}{m_i} \nabla E_n. \tag{6.16}$$

Integrating (6.16) then leads to the coordinates at the next timestep,

$$x_{n+1} = x_n + \delta x_{n+\frac{1}{2}}. \tag{6.17}$$

Equations (6.16) and (6.17) are the fundamental equations for dynamics (the Verlet velocity leapfrog algorithm). The first segment of MD is used to equilibrate the system, removing any bias from the initial conditions.

4. The acceleration and then the velocities are integrated to determine the new atomic positions. This integration is usually performed using the Verlet leapfrog algorithm.

The timestep of integration, δ, must be short enough to provide several points during the period of the fastest vibration. If the hydrogen atoms are described explicitly, the timestep is usually 1 to 2 fs.

The trajectory of the molecular systems may require time scales ranging from picoseconds to hundreds of nanoseconds, depending on the application and the size of the system. Thus, a computationally efficient MD algorithm must allow fast and accurate calculation for atomic forces and use the longest possible timesteps compatible with accurate Verlet integration to simulate molecular motions on the longest time scale.

NPT and NVT Dynamics Newton's equations of motion (6.13) describe a closed system. Thus the total energy (kinetic energy plus potential energy) of the system cannot change, and the system is adiabatic. If the volume is held constant, this simulation generates the microcanonical ensemble of statistical mechanics (denoted NVE for constant number of molecules, volume, and energy). However, most experiments deal with systems in equilibrium with temperature and pressure baths, leading to a Gibbs ensemble (Allen 1987). For the trajectory to generate a Gibbs ensemble, it is necessary to allow the internal temperature of the molecule to fluctuate in the same way it would if it were in contact with a temperature bath, and the volume must fluctuate in the same way it would in contact with a pressure bath.

Several methods (Woodcock 1971, Nose 1984, Hoover 1985, Vaidehi et al. 1996) are used to control the temperature of an MD simulation. The most rigorous method (Nose 1984) introduces into the equations of motion a new dynamic degree of freedom ζ, which is associated with energy transfer to the temperature bath (friction). If the volume is kept fixed, the Nose dynamics generates a Helmholtz canonical ensemble, giving rise to the correct partition function for an NVT system. These partition functions can be used to calculate macroscopic properties of the system. We consider NVT canonical dynamics as adequate for most biochemical problems [such as calculating binding energy and other

molecular properties to be used in deriving quantitative structure–activity relationships (QSAR) that are useful for predicting enzyme activity].

Some interesting biological applications consider the response of proteins to external pressure (Floriano et al. 1998). In this case, we use periodic boundary conditions (described later), placing the molecule and solvent in a periodic cell that can be acted upon by external stresses. Examining the structural deformations of proteins under pressure or under external stress requires that the MD allow the internal pressure of a molecule to fluctuate in the same way as for a system in a constant-pressure environment (Parrinello and Rahman 1981). The modified Newton equations lead to NPT dynamics (constant number of particles, pressure, and temperature).

Constrained Internal Coordinates   The short time scale of 1 fs for MD is required to describe the very rapid oscillations involved in bond stretching and angle bending motions. However, for proteins and nucleic acids, it is the low-frequency motions involved in conformational changes that are of most interest. Several algorithms (Ryckaert et al. 1977, van Gunsteren et al. 1990, Mazur and Abagayan 1989, Jain et al. 1993, Rice and Brunger 1994) have been developed for fixing the bonds and angles in order to focus on the conformational motions. Such constrained dynamics algorithms lead to coupled equations of motion:

$$M(\theta)\ddot{\theta} + C(\theta,\dot{\theta}) = T(\theta) \tag{6.18}$$

for P degrees of freedom (torsions). Here $\ddot{\theta}$ is the angular acceleration; M is the $P \times P$ mass matrix (moment of inertia tensor), which depends on the internal coordinates θ; T is the vector of general forces (tensor) on the atoms; and C is the velocity- dependent Coriolis force. At each timestep we know θ, M, T, and C, and we must solve the matrix equation (6.18) to obtain the acceleration,

$$\ddot{\theta} = M^{-1}(T - C). \tag{6.19}$$

Integration of $\ddot{\theta}$ gives the velocity $\dot{\theta}$ and further integration leads to the torsion or the dihedral angles θ from which the coordinates can be obtained. The problem is that solving equation (6.19) requires inverting the $P \times P$ dense mass matrix M at every timestep. For a system with, say, 10,000 atoms, there might be 3000 torsional degrees of freedom, making the solution of (6.19) for every timestep impractical (the cost of inverting M scales as P^3). Recently we developed the Newton–Euler inverse mass operator (NEIMO) method (Jain et al. 1993, Mathiowetz et al. 1994, Vaidehi and Goddard 1996), which solves (6.19) at a computational cost proportional to P. The NEIMO method considers the molecule to be a collection of rigid "clusters" connected by flexible "hinges." A rigid cluster can be a single atom, a group of atoms (a peptide bond), or a secondary structure (a helix or even an entire domain of a protein). Such constrained models allow much larger timesteps.

Table 6.1
Time steps obtained in the hierarchical modeling of two proteins

Protein	MD method	Degrees of freedom	Time step (ps)
Protein A[1]	Newtonian	1062	0.001
	NEIMO (all torsions)	219	0.010
	H-NEIMO	92	0.020
PGK	Newtonian (all atom)	12525	0.001
	NEIMO (all torsions)	2210	0.005
	H-NEIMO	80	0.010

[1]Protein A is a helix-coil-helix segment from staphylococcus aureus

With fine-grain all-torsion NEIMO dynamics, we normally treat double bonds, terminal single bonds, and rings (benzene) as rigid bodies. In hierarchical or H-NEIMO dynamics (Vaidehi et al. 2000), we allow higher levels of coarseness, keeping various segments or parts of a domain of a protein rigid during the dynamics. This allows larger timesteps, as illustrated in table 6.1.

Using hierarchical NEIMO simulations on the glycolytic enzyme phosphoglycerate kinase (PGK), we are able to follow the long time-scale domain motions in PGK responsible for its function. PGK catalyzes an essential phosphorylation step in the glycolytic pathway. Under physiological conditions, PGK facilitates the phosphoryl transfer from adenosine diphosphate (ADP) to adenosine triphosphate (ATP). PGK consists of two major domains (1400 atoms in each domain) denoted as the C-domain and the N-domain. In most crystal structures (McPhillips et al. 1996, Bernstein et al. 1997), the substrates are found bound to the opposite domains at a distance of ~13 Å. Thus, Blake (1997) proposed a hinge-bending mechanism by which the protein brings the substrates together to react. Using H-NEIMO (figure 6.2; see also plate 19), we found low-frequency domain motions that take the open structure examined by yeast PGK (McPhillips 1996) to the closed structure examined by *Trypanosoma brucei* PGK (Bernstein et al. 1997). The NEIMO dynamics suggest that PGK undergoes long time-scale motions that put the substrate binding sites together, then take them apart, then put them together again. The extent and rate of the domain motions depend on the nature of the substrates bound.

MPSim   In order to allow long time simulations on very large systems (up to a million atoms), we developed the massively parallel simulation (MPSim) program (Lim et al. 1997) to operate efficiently with massively parallel computers. MPSim includes important algorithm developments such as CMM (for calculating long-range nonbond interactions) (Ding et al. 1992), NEIMO (Jain et al. 1993, Mathiowetz et al. 1994, Vaidehi et al. 1996), and the Poisson–Boltzmann solution (Tannor et al. 1994). It is compatible with massively parallel high-performance computers (SGI-Origin-2000, IBM-SP2, Cray T3D/T3E, HP/Convex-Exemplar, and Intel Paragon). MPSim has been used to understand the action of drugs on
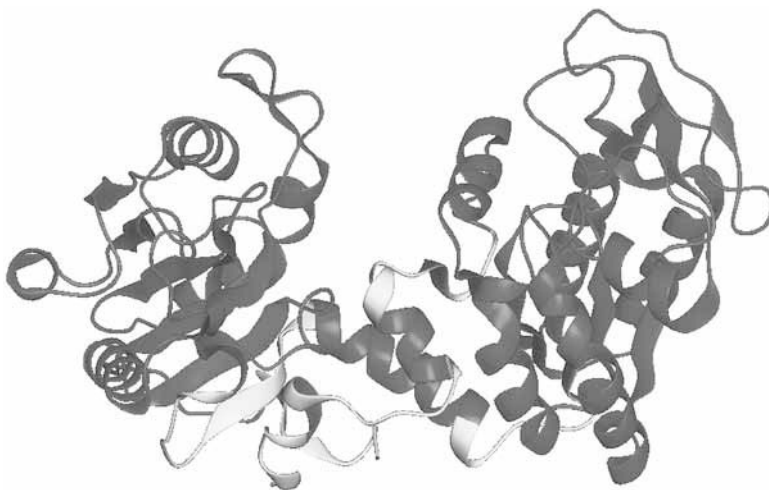
Figure 6.2
The H-NEIMO model for phosphoglycerate kinase. Here the parts of two domains that bind the substrates (shown in yellow) are kept rigid while the loops (shown in green) are allowed to have full torsional freedom. H-NEIMO MD simulations exhibit large-scale domain motions that bring the substrates close together and then apart. (See plate 19.)

the human rhinovirus (Vaidehi et al. 1996) as summarized in section 6.3.2. It is also being used for materials science problems (Miklis et al. 1997, Demiralp et al. 1999).

Periodic Boundary Conditions   In order to control the pressure on a system and to include explicit solvents without introducing complications of free surfaces, it is convenient to place the molecule in a large box (much larger than the molecule) and then reproduce it periodically to fill space. In addition, it is particularly useful to describe some systems, such as DNA, as periodic and repeating in one direction. The computational box containing the molecular system is surrounded by an infinite number of copies of itself (see figure 6.3).

Because periodic systems involve an infinite number of atoms, some care must be taken in calculating the long-range forces. Otherwise, singularities or wild oscillations can occur. The most common accurate method is Ewald summation (Ewald 1921, de Leeuw et al. 1980, Heyes 1981, Allen 1987, Karasawa and Goddard 1989, Chen et al. 1997), which

Figure 6.3
Molecules in a periodic system. (a) Two-dimensional view of a tripeptide in a periodic box. (b) The simulation unit cell is the box outlined in the center. It is surrounded by an infinite array of equivalent boxes to that there is no free surface. Molecules are allowed to move from box to box, but the number in the unit cell is always constant.
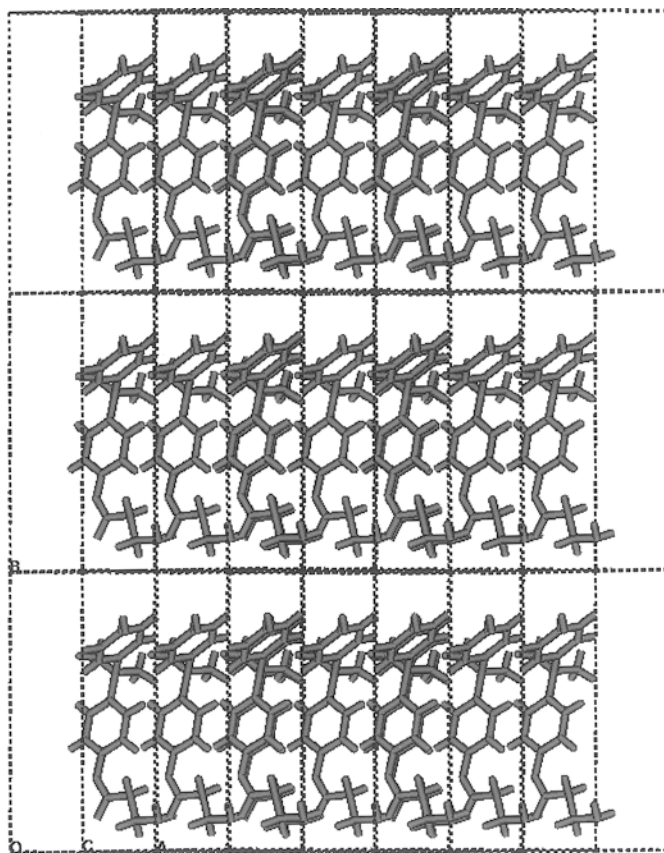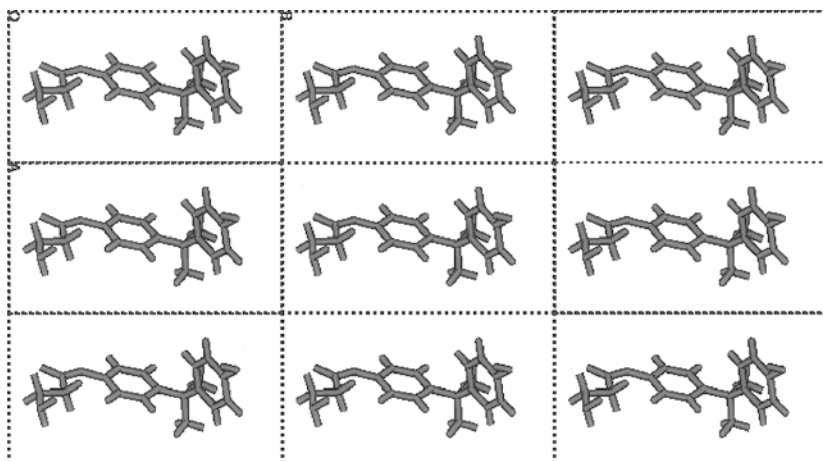
considers point charges smeared over a region of finite size $1/\eta$, chosen to converge rapidly (in real space), and then Fourier transforms the difference between the smeared charges and the point charges to obtain rapid convergence (in the reciprocal space sums). Karasawa and Goddard (1989) showed how to choose the optimum η to minimize computational cost for a given level of accuracy. This leads to costs scaling as $N^{3/2}$ for systems with N atoms per unit cell. Ding et al. (1992b) showed how to use the reduced CMM method to achieve linear scaling.

Monte Carlo Methods   For many systems, the barriers between low-lying structures may be too large for MD to sample all the structures. In such cases, we often use Monte Carlo or statistical sampling techniques, using a random search algorithm such as Monte Carlo metropolis. With a sufficiently large number of samples, the occurrence of each conformation is proportional to the Boltzmann factor, leading to a canonical distribution. The steps involved in the Monte Carlo simulation procedure are:

1. Starting from a given molecular conformation, a new conformation is generated by random displacement of one or more atoms. The random displacements should be such that in the limit of a large number of successive displacements, the available conformation space is uniformly sampled.

2. The newly generated conformation is accepted or rejected based on the change in the potential energy of the current step compared with the previous step. The new conformation is accepted if the change in potential energy $\Delta V = V$ (present step) $- V$ (previous step) ≤ 0, or if $\Delta V > 0$ when the Boltzmann factor is greater than a random number ***R.***

   Upon acceptance, the new conformation is counted and used as a starting point for the generation of the next random displacement. If the criteria are not met, then the new conformation is rejected and the previous conformation is counted again as a starting point for another random displacement. This method thus generates a Boltzmann ensemble of conformations. Many Monte Carlo methods are available (Allen 1987) and have been used as a fast conformational search tool in protein folding (Sternberg 1996). A recent advance, continuous conformation Boltzmann biased direct Monte Carlo (Sadanobu and Goddard 1997) has been used to determine the complete set of folding topologies for proteins with up to 100 residues (Debe et al. 1999). Such methods show considerable promise for solving the protein-folding problem (predicting the tertiary structure from a primary sequence).

6.3   Application to Biological Problems

In this section, we summarize some recent applications of quantum chemistry and molecular dynamics to problems in structural biology.
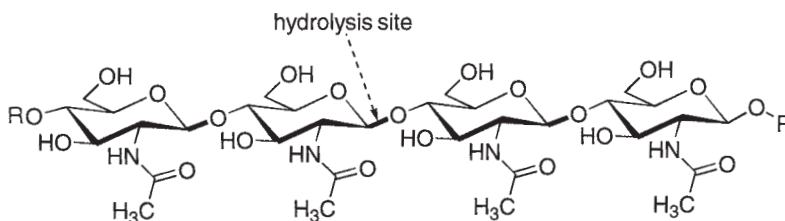
Figure 6.4
An arrow marks the hydrolysis site of chitin, the Symbol"b (1,4)-*N*-acetylglucosamine (GlcNAc) polysaccharide substrate of chitinases.

6.3.1   Study of Enzyme Reaction Mechanisms

The reaction mechanisms for many enzymes have been studied using a combination of QM and MD methods (McCammon 1987, Cunningham and Bach 1997). Examples include the hydrolysis of a peptide bond by serine proteases, and hydrolysis by the metalloenzyme, staphylococcal nuclease of both DNA and RNA, etc. (Warshel 1991). We describe here a recent QM and MD study for the elucidation of the mechanism of family 18 (Brameld and Goddard 1998a) and family 19 chitinases (Brameld and Goddard 1998b). This application relies heavily on a combination of QM and MD methods, demonstrating the feasibility of solving difficult problems using modern computational methods.

Chitin (see figure 6.4) is a β (1,4)-linked *N*-acetylglucosamine (GlcNAc) polysaccharide that is a major structural component of fungal cell walls and the exoskeletons of invertebrates (including insects and crustaceans). This linear polymer may be degraded through the enzymatic hydrolysis action of chitinases. Chitinases have been found in a wide range of organisms, including bacteria (Watanabe et al. 1990; plants (Collinge et al. 1993), fungi (Bartinicki-Garcia 1968), insects (Kramer et al. 1985), and crustaceans (Koga et al. 1987). For organisms that utilize the structural properties of chitin, chitinases are critical for the normal life-cycle functions of molting and cell division (Fukamizo and Kramer 1985, Kuranda and Robbins 1991). Because chitin is not found in vertebrates, inhibition of chitinases is a promising strategy for treatment of fungal infections and human parasitosis (Robertus 1995).

Based on amino acid sequence, the glycosyl hydrolases have been classified into 45 families. Using this classification method, the chitinases form families 18 and 19, which are unrelated, differing both in structure and in mechanism. Sequence analysis shows little homology between these classes of chitinases. Family 19 chitinases (found in plants) share the bilobal α + β folding motif of lysozyme, which forms a well-defined substrate binding cleft between the lobes. In contrast, family 18 chitinases share two short sequence motifs, which form the catalytic (βα) 8-barrel active site. Family 18 chitinases with diverse
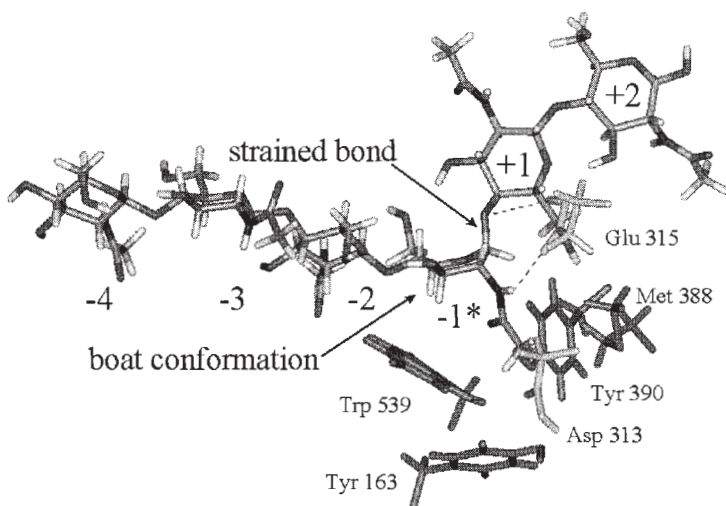
Figure 6.5
The minimum energy structure for the −1-–boat hexaNAG conformation. A boat geometry for GlcNAc residue −1 and the twist between residues −1 and +1 strains the linking glycosidic bond. This distortion observed in the simulations should be included when designing new inhibitors.

sequences have been isolated from a wide range of eukaryotes and prokaryotes. The hydrolysis site of chitin is shown in figure 6.4.

Brameld investigated the hydrolysis mechanisms of the chitinases by examining the reactivity of the chitin substrate alone and in the presence of the enzyme. This was done using *ab initio* quantum mechanical calculations on three possible reaction intermediates for the enzymatic hydrolysis of chitin. He found that anchimeric assistance from the neighboring *N*-acetyl group of the chitin is critical in stabilizing the resulting oxazoline ion intermediate.

MD simulations of the complete enzyme with bound substrate led to further insights into the mechanisms of family 18 and 19 chitinases, which differ substantially. All MD simulations were carried out using the MSC-PolyGraf program using the Dreiding FF (Mayo et al. 1990). QEq charges (Rappe and Goddard 1991) were used for all GlcNAc residues. All nonbond interactions were considered explicitly (using a distance-independent dielectric constant), with a cutoff of 9.5 Å for MD simulations and 13.5 Å for energies. Solvation energies were estimated using the continuum solvent model in the Delphi program (Tannor et al. 1994). Two possible intermediates, namely, the oxazoline ion and the oxocarbenium ion, were proposed. The formation of the oxazoline ion intermediate results from substrate distortion (see figure 6.5) induced within the active site of family 18 chitinases. This substrate distortion was observed in the MD simulations and is necessary

to design inhibitors for the family 18 chitinase. Yet surprisingly, the family 19 chitinases do not utilize an oxazoline ion intermediate and undergo a considerable change in enzyme conformation to stabilize the resulting oxocarbenium ion intermediate.

The oxazoline transition state serves as a target for the rational design of more potent glycosidase inhibitors specific to family 18 chitinases. Simple analogs of allosamidin that incorporate the key features of a delocalized positive charge while maintaining a chairlike sugar conformation may prove to be synthetically more accessible than allosamidin. Such analogs could lead to a new generation of chitinase transition-state inhibitors.

6.3.2   A Model for Drug Action on Rhinovirus-1A and Rhinovirus-14

Human rhinovirus (HRV) belongs to the picornavirus family. It has over 100 serotypes, providing a challenge to drug design. The serotypes of HRV are classified into two groups. The major receptor group (including HRV-14 and HRV-16) binds to the intercellular adhesion molecule 1 (ICAM-1) receptors. The minor receptor group (including HRV-1A) binds to low-density lipoprotein-type receptor molecules. The protein capsid of HRV consists of an icosahedral shell with 60 copies of the four viral proteins (VP1, VP2, VP3, and VP4) totaling 480,000 atoms (300 Å in diameter). Figure 6.6 (see also plate 20) shows the icosahedral shell, its elements such as the pentamers that make up the icosahedron, the asymmetric unit that makes up the pentamer, and the basic four viral proteins of the asymmetric unit. A single-stranded RNA is enclosed in the protein shell. The sequence of events involved in the endocytosis is not clear yet, but circumstantial evidence suggests that the RNA is released through the pentamer in the virus coat (Rueckert 1991).

There are several known isoxazole-derived drugs for HRV-1A and HRV-14 (Couch 1990). It is known that binding of these drugs to HRV-14 prevents binding of the virus to the ICAM-1 receptors. However for HRV-1A, binding of thee drugs does not block receptor attachment; rather, it prevents uncoating of the virus. One speculation is that this binding leads to stiffening of the viral capsid. Based on this speculation and the fact that the RNA is released through the pentamer channel, Vaidehi and Goddard (1997) proposed the pentamer channel stiffening Model (PCSM). Drug action on HRV-1A constricts or stiffens the pentamer channel sufficiently that the RNA cannot exit, thus preventing uncoating.

Using MPSim (see the section on this program) on the KSR-64 processor parallel computer, we showed a strong correlation (see figure 6.7) of drug effectiveness (minimum inhibitory concentration, MIC) with the strain energy increase calculated for various drugs. Here we defined the strain energy as the energy required to expand the pentamer channel to 25 Å. The strain energy was calculated using MPSim with the Amber FF for the viral proteins and the Dreiding FF for the drugs.

Figure 6.7 shows that all effective drugs cause an increase in the strain energy required to open the pentamer channel. The best drug, WIN56291 (MIC = 0.1 μM), shows the sharpest increase in the strain energy compared with the native HRV-1A or an ineffective
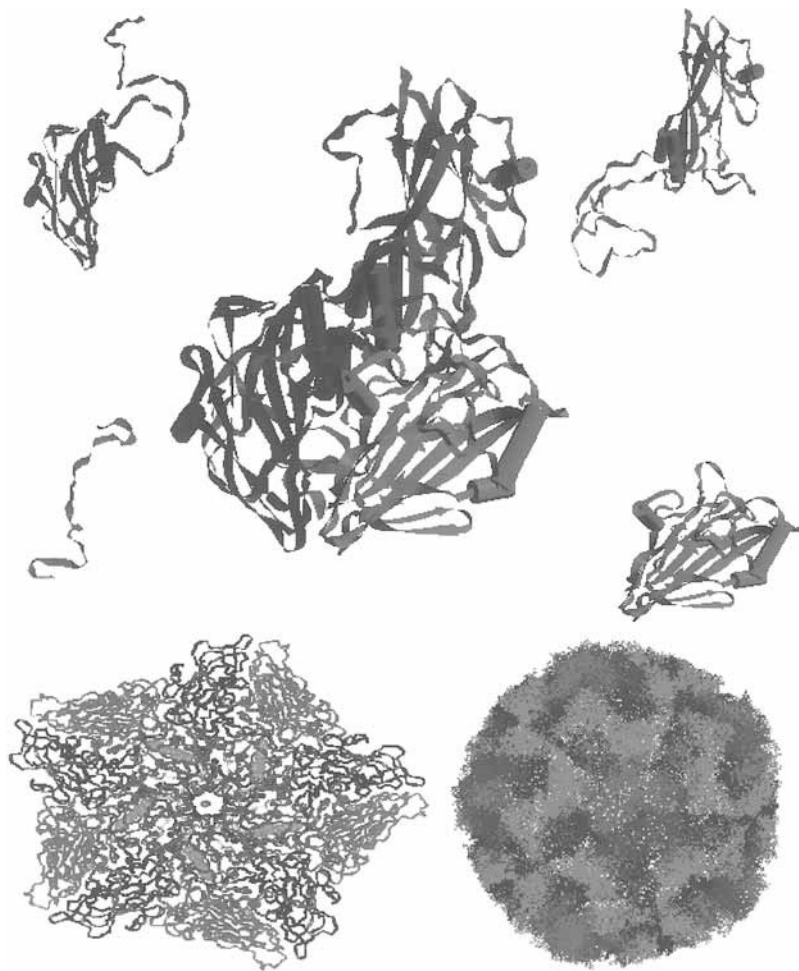
Figure 6.6
Structure of a rhinovirus. The complete icosahedral viral capsid (300 Å diameter) is shown at the bottom right.
The four viral proteins making up the asymmetric unit of the virus are shown in the top half. The five asymmetric
units forming the pentamer are shown in the bottom left. Twelve such pentamers form the complete viral capsid.
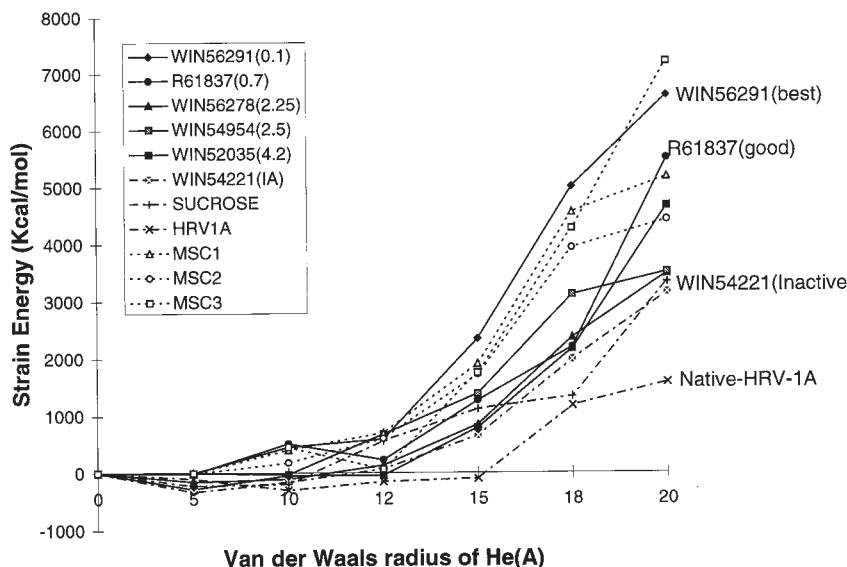(See plate 20.)

Figure 6.7
Pentamer stifness for HRV-1A with various Winthrop drugs. The MIC values for the drugs are given in parenthesis. MSC1, MSC2, and MSC3 denote new candidates for drug molecules.

drug, WIN54954 (MIC = 2.5 μM). This suggests that the PCSM can be used to predict the efficacy of a drug before synthesis and testing. Three drugs—MSC1, MSC2, and MSC3—were tested in this way.

6.3.3 Calculation of Binding Energy Using Free Energy Perturbation Theory

A major application of MD simulations is for drug design, where high binding energy is expected to be a necessary condition for a good drug. MD provides microscopic-level information (atomic and molecular positions, velocities, etc.) about the dynamics of macromolecules; this information can be averaged over time (using appropriate formulas from statistical mechanics) to obtain the macroscopic thermodynamic properties of a system, such as the free energy, temperature, and heat capacity. Thus, one assumes that the macroscopic property G_{obs} can be expressed as an average over the ensemble of conformation from the MD, $G_{obs} = <G>_{ens}$, and that this can be expressed as the time average of $G(r_i)$ over the trajectory over a sufficiently long time interval. This is written as

$$G_{obs} = <G(r_i)>_{time} = \lim_{t->\infty} \frac{1}{t} \int_0^t G(r)dt, \qquad (6.20)$$

where $G(r_i)$, r_i is a function of time and is generated as the trajectory of the MD simulations.

The problem is that the convergence of (6.20) is sufficiently slow that the error may be large compared with the difference in bond energies for the drugs. This problem is solved by transforming the system slowly from drug A ($\lambda = 0$) to drug B ($\lambda = 1$) and integrating the difference in $G(\lambda)$ along the trajectory. This is called *free energy perturbation* (FEP) theory. It has been demonstrated to give a quantitative estimate of the relative free energy of binding of various drug molecules or inhibitors to its receptors. The integration in (6.20) is replaced by finite sum over timesteps. The free energy of molecular systems can be calculated (van Gunsteren and Weiner 1989) using equation (6.20). The free energy difference given by

$$G(\lambda) - G(\lambda_i) = -RT \ln < \exp\left[-\left(\frac{V_\lambda - V_{\lambda i}}{RT}\right)\right] > \lambda_i \qquad (6.21)$$

provides a means of calculating the free energy difference any two states $A(\lambda)$ and $A(\lambda_i)$. However, unless the states share a significant fraction of conformation space, convergence is very slow. The convergence time can be reduced by calculating the relative free energy difference between closely related states, using the thermodynamic cycle. To visualize the method (see figure 6.8), we consider the relative binding of two ligands, L_1 and L_2, to a receptor that is a protein or DNA. The appropriate thermodynamic perturbation cycle (Zwanzig 1956) for obtaining the relative binding constant is given in figure 6.8.

The ratio of the binding constants for L_1 and L_2 can be calculated from the equation

$$\frac{K_{L2}}{K_{L1}} = \exp[-(\Delta G_{L2} - \Delta G_{L1})/RT \qquad (6.22)$$

where R denotes the gas constant and K_{L1} and K_{L2} are the binding constants for L1 and L2. Simulation of steps 1 and 3 involves replacement of solvent molecules in the binding pocket of the protein (receptor) by the ligand and removing the ligand from the solvent and binding it to the protein. These are very slow processes for the time scale of MD simulations. However, it is possible to simulate steps 2 and 4 using MD if the chemical composition and the structure of the ligands are very closely related, as described in the example below. Hence the free energy difference for steps 1 and 2 is related to the free energy difference between steps 2 and 4 by (figure 6.8):

$$\Delta G_3 - \Delta G_1 = \Delta G_2 - \Delta G_4$$

The crucial factor in the simulations of steps 2 and 4 is in the adequacy of the sampling in the configuration space of the system. Longer MD simulations are required for accurate free energies.
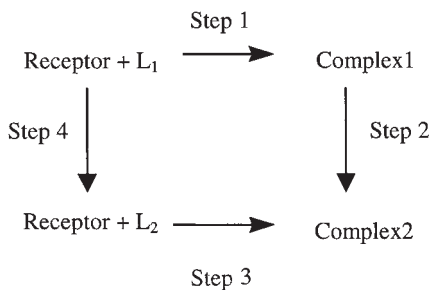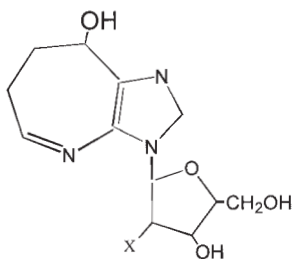
Figure 6.8
Thermodynamic cycle for the calculation of the relative free energies of binding.

This method has been used extensively to study the relative binding energies of drugs and inhibitors for enzymes (van Gunsteren and Weiner 1989; Marrone et al. 1997; Plaxco and Goddard 1994) and inhibitors. Here we describe briefly one of the recent applications (Marrone et al. 1996) of the free energy perturbation simulations to the study of inhibition of the enzyme adenosine deaminase by 8R-coformycin and (8R)-deoxycoformycin. The inhibition of the enzyme adenosine deaminase, which deaminates the base adenosine, could provide an effective treatment of some immunological disorders. Thus MD simulations can play a critical role in designing inhibitors for this enzyme since it gives a good model for the binding site of the inhibitors in the enzyme and also provides an estimate of their binding energies. The coformycin and deoxycoformycin molecules differ in the sugar moiety attached to them, which is ribose in the case of coformycin and deoxyribose for deoxycoformycin. The molecular structure of these two substrates is shown in figure 6.9.

It is clear that these two inhibitors differ only by only a small functional group and hence provide a case well suited for free energy perturbation calculations. Molecular dynamics and free energy simulations of coformycin and deoxycoformycin and their complexes with adenosine deaminase show a difference of -1.4 kcal/mol in binding energy between deoxycoformycin and coformycin. Deoxycoformycin and coformycin differ by a hydroxyl group, but the relative binding energy is small enough, showing that this hydroxyl group is buried near a flexible hydrophilic region of the enzyme conformation rather than being sequestered in a hydrophobic pocket. Thus, detailed structural aspects of the transition state analog have been derived in this study (Marrone et al. 1996).

6.3.4 Quantitative Structure–Activity Relationships

Following the calculation of relative binding constants, one of the most important steps in the process of drug discovery is the study of QSAR (Franke 1984; Fanelli et al. 1998). The aim is to explain the experimental data obtained for binding of various ligands to a recep-
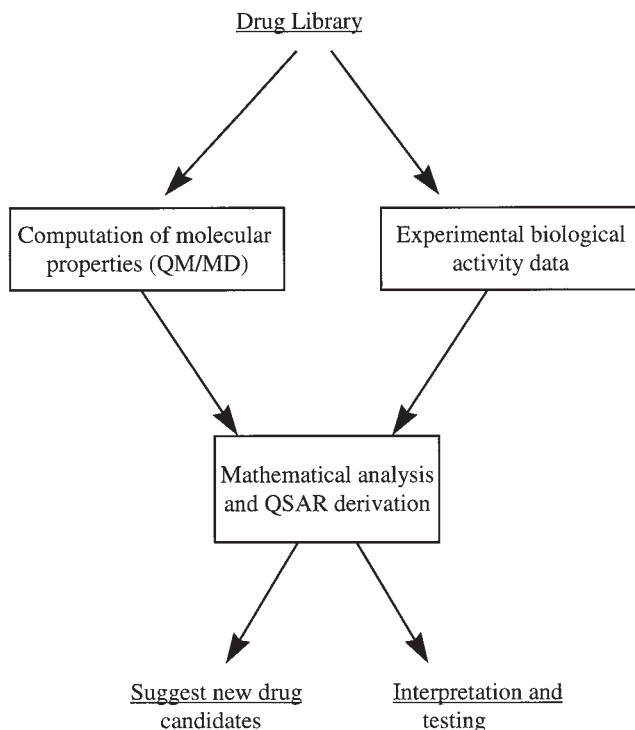
Figure 6.9
Molecular structure of adenosine deaminase inhibitors X = H for deoxycoformycin and X = OH for coformycin.

tor, at the molecular level, in terms of physiochemical properties of the ligands, and to pre-
dict or estimate similar biodata for new analogs. QSARs are equations that relate an
observed experimental quantity for ligand binding, for example, the minimum inhibitory
concentration measured routinely in drug industries, to the calculated molecular properties
of various regions of a drug molecule. QSARs are widely used today in the pharmaceutical
industry to design new drugs or inhibitors. Figure 6.10 shows the major steps involved in a
QSAR analysis.

   Today computational chemistry using QM–MD simulations allows us to define and
compute ad hoc shape and size descriptors for the different conformations assumed by
drugs in biotest solutions. Together with the statistically sound experimental data mea-
sured on well-identified target receptors, these descriptors are essential elements for ob-
taining simple, consistent, comparable, and easily interpretable theoretical QSAR models
based on ligand similarity–target receptor complementarity paradigms. The molecular
properties calculated can be of varied natures. For example, based on the structural for-
mula of the drugs, they can be broken into fragments, and the properties of the subsystem
or fragments that contribute to the drug activity can be calculated. Intermolecular interac-
tion properties such as polarizability or the hydrophilic or hydrophobic nature of groups
can be computed. Thus, QM–MD simulations are used extensively in pharmaceutical com-
panies to derive QSARs (Fanelli et al. 1998).

6.4   Summary

Atomistic simulations constitute a powerful tool for elucidating the dynamic behavior of
biological systems. They can be used in a variety of problems in biological research and
drug design, as suggested in this chapter. The major thrust in this area is the simulation of
large systems (hundreds of angstroms) for long time scales (microseconds).

Figure 6.10
Principal steps involved in QSAR analysis.

The next challenge is to find accurate ways to describe the dynamics of biological systems at a coarse grain level while retaining the accuracy of atomic-level MD in order to examine dynamic behavior for very long times and large distance scales. The dream here would be eventually to simulate the processes of an entire cell, where the coarse grain description would describe the elastic properties of the cell membrane plus the chemical nature of the receptors, but without explicit atoms. In this approach, the grid points of the coarse grain would embody chemical and physical properties so that expanding the grid would lead automatically to a description of binding some new molecule or the active transport of a molecule down a channel. Thus, one can focus down to the atomic level to determine the molecular properties and then back to the coarse grain for the large-scale motions.

Probably the most important challenge in biology is the prediction of accurate three-dimensional structures and the functions of proteins (tertiary structure) entirely from the

gene sequence (primary structure). Enormous progress is being made here (Debe et al. 1999), but there is not yet success in *ab initio* predictions.

Acknowledgments

References

Allen, M. P., and Tildesley, D. J. (1987). *Computer Simulations of Liquids.* Clarendon, Oxford.

Allinger, N. L., and Schleyer, P. V. (1996). Special issue on molecular mechanics. *J. Comp. Chem.* 17: 385–385.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., and Hermans, J. (1981). Interaction models for water in relation to protein hydration. *Jerusalem Symp. Quantum Chem. Biochem.* 14: 331–342.

Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. (1987). The missing term in effective pair potentials. *J. Phys. Chem.* 91: 6269–6271.

Bernstein, B. E., Michels, P. A. M., and Hol, W. G. J. (1997). Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation. *Nature* 385: 275–278.

Blake, C. F. (1997). Glycolysis-phosphotransfer hinges in PGK. *Nature* 385: 204–205.

Brameld, K. A., Dasgupta, S., and Goddard III, W. A. (1997). *Ab initio* derived spectroscopic quality force fields for molecular modeling and dynamics. *J. Phys. Chem.* B 101: 4851–4859.

Brameld, K. A., and Goddard III, W. A. (1998a). Substrate distortion to a boat conformation at subsite-1 is critical in the mechanism of family 18 chitinases. *J. Am. Chem. Soc.* 120: 3571–3580.

Brameld, K. A., and Goddard III, W. A. (1998b). The role of enzyme distortion in the single displacement mechanism of family 19 chitinases. *Proc. Natl. Acad. Sci.* 95: 4276–4281.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, and Karplus, M. (1983). CHARMM—a program for macromolecular energy, minimization, and dynamics calculations, *J. Comput. Chem.* 4: 187–217.

Che, J., Cagin, T., and Goddard III, W. A. (1999). Studies of fullerenes and carbon nanotubes by an extended bond order potential. *Nanotechnology* 10: 263–268.

Chen, Z. M., Cagin, T., and Goddard III, W. A. (1997). Fast Ewald sums for general van der Waals potentials. *J. Comp. Chem.* 18: 1365–1370.

Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995). A second generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* 117: 5179–5197.

Collinge, D. B., Kragh, K. M., Mikkelsen, J. D., Nielsen, K. K., Rasmussen, U., and Vad, K. (1993). Plant kinases. *Plant J.* 3: 31–40.

Couch, R. B. (1990). Rhinoviruses. In *Virology,* B. N. Fields, and D. M. Knipe, eds., pp. 607–629. Raven, New York.

Cunningham, M. A., Ho, L. L., Nguyen, D. T., Gillilan, R. E., and Bash, P. A. (1997). Simulation of the enzyme reaction mechanism of malate dehydrogenase. *Biochemistry* 36: 4800–4816.

Dasgupta, S., Yamasaki, T., and Goddard III, W. A. (1996). The Hessian biased singular value decomposition method for optimization and analysis of force fields. *J. Chem. Phys.* 104: 2898–2920.

Debe, D., Carlson, M. J., and Goddard III, W. A. (1999). The topomer-sampling model of protein folding. *Proc. Natl. Acad. Sci. USA* 96: 2596–2601.

Demiralp, E., Cagin, T., Huff, N. T., and Goddard III, W. A. (1998). New interatomic potentials for silica. *XVIII Intl. Congress on Glass Proc.,* M. K. Choudhary, N. T. Huff, and C. H. Drummond III, ed., pp. 11–15.

Demiralp, E., Cagin, T., and Goddard III, W. A. (1999). Morse stretch potential charge equilibrium force field for ceramics: application to the quartz-stishovite phase transition and to silica glass. *Phys. Rev. Lett.* 82: 1708–1711.

Ding, H. Q., Karasawa, N., and Goddard III, W. A. (1992). Atomic level simulations on a million particles—the cell multipole method for Coulomb and London nonbond interactions. *J. Chem. Phys.* 97: 4309–4315.

Ding H. Q., Karasawa N., and Goddard III, W. A. (1992b). The reduced cell multipole method for coulomb interactions in periodic-systems with million-atom unit cells. *Chem. Phys. Lett.* 196: 6–10.

De Leeuw, S. W., Perram, J. W., and Smith, E. R. (1980). Simulation of electrostatic systems in periodic boundary conditions. *Proc. Roy. Soc. London* A373: 27–56.

Ewald, P. (1921). Die Berechnung optischer und elektrostatischer gitterpotentiale. *Ann. Phys.* (Leipzig) 64: 253–287.

Fanelli, F., Menziani, C., Scheer, A., Cotecchia, S., and De Benedetti, P. D. (1998). Ab initio modeling and molecular dynamics simulation of the alpha(lb)-adrenergic receptor activation). *Methods* 14: 302–317.

Figueirido, F., Levy, R. M., Zhou, R. H., and Berne, B. J. (1997). Large scale simulation of macromolecules in solution: Combining the periodic fast multipole method with multiple time step integrators. *J. Chem. Phys.* 107: 7002–700.

Floriano, W. B., Nascimento, M. A. C., Domont, G. B., and Goddard III, W. A. (1998). Effects of pressure on the structure of metmyoglobin: molecular dynamics predictions for pressure unfolding through a molten globule intermediate. *Protein Science* 7: 2301–2313.

Franke, A. R., (1984). *Theoretical Drug Design Methods.* Elsevier, Amsterdam.

Fujita, T. (1990). The extrathermodynamic approach to drug design. In *Comprehensive Medicinal Chemistry,* Hansch, C., ed., vol. 4, pp. 497–560. Pergamon, Oxford.

Fukamizo, T., and Kramer, K. J. (1985). mechanism of chitin hydrolysis by the binary chitinase system in insect molting fluid. *Insect. Biochem.* 15: 141–145.

Gerdy, J. J. (1995). Accurate Interatomic Potentials for Simulations. Ph.D thesis, California Institute of Technology.

Henrissat, B., and Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino-acid-sequence similarities. *Biochem. J.* 293: 781–788.

Hermans, J., Berendsen, H. J. C., van Gunsteren, W. F., and Postma, J. P. M. (1984). A consistent empirical potential for water-protein interactions. *Biopolymers* 23: 1513–1518.

Heyes, D. M. (1981). Electrostatic potentials and fields in infinite point charge lattices. *J. Chem. Phys.* 74: 1924–1929.

Hoover W. G. (1985). Canonical dynamics—equilibrium phase-space distributions. *Phys. Rev. A* 31: 1695–1697.

Jain, A., Vaidehi, N., and Rodriguez, G. (1993). A fast recursive algorithm for molecular dynamics simulation. *J. Comp. Phys.* 106: 258–268.

Jorgensen, W. L., Chandrasekar, J., Madura, J. D., Impey, R. W., and Klein. M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79: 962.

Jorgensen, W. L., and Tirado-Tives, J. (1988). The OPLS potential functions for proteins—energy minimizations for crystals of cyclic-peptides and crambin. *J. Am. Chem. Soc.* 110: 1666–1671.

Karasawa, N., and Goddard III, W. A. (1989). Acceleration of convergence for lattice sums. *J. Phys. Chem.* 93: 7320–7327.

Kier, L. B., and Testa, B. (1995). Complexity and emergence in drug research In *Advances in Drug Research,* B. Testa, and U. A. Meyer eds., vol. 26, pp. 1–43. Academic Press, London.

Koga, D., Isogai, A., Sakuda, S., Matsumoto, S., Suzuki, A., Kimura, S., and Ide, A. (1987). Specific-inhibition of bombyx-mori chitinase by allosamidin. *Agri. Biol. Chem.* 51: 471–476.

Kramer, K. J., Dziadik-Turner, C., and Koga, D. (1985). In *Comprehensive Insect Physiology, Biochemistry and Pharmacology: Integument, Respiration and Circulation.* Pergamon Press, Oxford.

Kuranda, M. J., and Robbins, P. W. (1991). Chitinase is required for cell-separation during growth of saccharomyces-cerevisiae. *J. Biol. Chem.* 266: 19758–19767.

Levitt, M. (1983). Molecular dynamics of native protein. 1. Computer simulation of trajectories. *J. Mol. Biol.* 168: 595–620.

Levitt, M., Hirschberg, M., Sharon-R., Laidig, K. E., and Dagett, V. (1997). Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *J. Phys. Chem.* B101: 5051–5061.

Lim, K. T., Brunett, S., Iotov, M., McClurg, B., Vaidehi, N., Dasgupta, S. Taylor, S., and Goddard III, W. A. (1997). Molecular dynamics for very large systems on massively parallel computers: the MPSim program. *J. Comp. Chem.* 18: 501–521.

Mackerell, A. D., Wiorkiewicz-Kuczera, J., and Karplus, M., (1995). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Am. Chem. Soc.* 117: 11946–11975.

Marrone, T. J., Straatsma, T. P., Briggs, J. M., Wilson D. K., Quiocho, F. A., and McCammon, J. A. (1996). Theoretical study of inhibition of adenosine deaminase by (8R)-coformycin and (8R)-deoxycoformycin, *J. Med. Chem.* 39: 277–284.

Marrone, T. J., Briggs, J. M., and McCammon, J. A. (1997). Structure-based drug design: computational advances. *Ann. Rev. Pharmacol. Toxicol.* 37: 71–90.

Mathiowetz, A., Jain, A., Karasawa, N., and Goddard III W. A. (1994) Protein simulations using techniques suitable for very large systems: the cell multipole method for nonbond interactions and the newton-euler inverse mass operator method for internal coordinate dynamics. *Proteins* 20: 227–247.

Mayo, S. L., Olafson, B. D., and Goddard, W. A. (1990). DREIDING—a generic force field for molecular simulations. *J. Phys. Chem.* 94: 8897–8909.

Mazur, A., and Abagayan, R. J. (1989). New methodology for computer-aided modeling of biomolecular structure and dynamics. 1. Non-cyclic structures. *Biomol. Struct. Dyn.* 6: 815–832.

McCammon, J. A. (1987). In *Dynamics of Proteins and Nucleic Acids.* Cambridge University Press, Cambridge.

McPhillips, T. M., Hsu, B. T., Sherman, M. A., Mas, M. T., and Rees, D. C., (1996). Structure of the R65Q mutant of yeast 3-phosphoglycerate kinase complexed with MG-AMP-PNP and 3-phospho-D-glyceratefe. *Biochem.* 35: 4118–4127.

Miklis, P., Cagin, T., and Goddard III, W. A. (1997). Dynamics of bengal rose encapsulated in the meijer dendrimer box. *J. Am. Chem. Soc.* 119: 7458–7462.

Montorsi, M., Menziani, M. C., Cocchi, M., Fanelli, F., and De Benedetti, P. G. (1998). Computer modeling of size and shape descriptors of alpha(1)-adrenergic receptor antagonists and quantitative structure-affinity/selectivity relationships. *Methods* 145: 239–254.

Nose, S. (1984). A unified formulation of the constant temperature molecular-dynamics methods. *Mol. Phys.* 52: 255–268.

Parr, R. G., and Yang, W. (1989). in *Density Functional Theory of Atoms and Molecules.* Clarendon Press, Oxford.

Parinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52: 7182–7190.

Plaxco, K. W., and Goddard III, W. A. (1994). Contributions of the thymine methyl-group to the specific recognition of polynucleotides and mononucleotides—an analysis of the relative free-energies of solvation of thymine and uracil. *Biochem* 33: 3050–3054.

Pople, J. A., and Beveridge, D. L. (1970). In *Approximate Molecular Orbital Theory.* McGraw-Hill, New York.

Pranata, J., Wierschke, S. G., and Jorgensen, W. L. (1991). OPLS potential functions for nucleotide bases—Relative association constants of hydrogen-bonded base-pairs in chloroform. *J. Am. Chem. Soc.* 113: 2810–2819.

Qi, Y., Cagin, T., Kimura, Y., and Goddard III, W. A. (1999). Molecular-dynamics simulations of glass formation and crystallization in binary liquid metals: Cu-Ag and Cu-Ni. *Phys. Rev. B* 59: 3527–3533.

Rahman, A., and Stillinger, F. H. (1971). Molecular dynamics study of liquid water. *J. Chem. Phys.* 55: 3336–3359.

Rappé, A. K., and Goddard III, W. A. (1991). *J. Chem. Phys.* 95: 3358–3363.

Rappé, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A., III, and Skiff, W. M. (1992). UFF, A full periodic-table force-field for molecular mechanics and molecular-dynamics simulations. *J. Am. Chem. Soc.* 114: 10024–10035.

Rice, L. M., and Brunger, A. T. (1994). Torsion angle dynamics—Reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins* 19: 277–290.

Ringnalda, M. N., Langlois, J.-M., Greeley, B. H., Murphy, R. B., Russo, T. V., Cortis, C., Muller, R. P., Marten, B., Donnelly, R. E., Mainz, D. T., Wright, J. R., Pollard, W. T., Cao, Y., Won, Y., Miller, G. H., Goddard, W. A., III, and Friesner, R. A. Jaguar 3.0 from Schrödinger, Inc. Portland, OR.

Robertus J. D., Hart P. J., Monzingo A. F., Marcotte E., and Hollis, T. (1995). The structure of chitinases and prospects for structure-based drug design. *Can. J. Bot.* 73: S1142–S1146, Suppl. 1 E–H.

Rueckert, R. R. (1991). Picornaviridae and their replication. In *Fundamental Virology,* B. N. Fields, and D. M. Knipe, eds., pp. 409–450. Raven, New York.

Ryckaert, J. P., Cicotti, G., and Berendsen, H. J. C. (1977). Numerical Integration of Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comp. Phys.* 23: 327–341.

Sadanobu, J., and Goddard III, W. A. (1997). The continuous configurational Boltzmann biased direct Monte Carlo method for free energy properties of polymer chains. *J. Chem. Phys.* 106: 6722–6729.

Schaefer, H. F. (1984). In *Quantum Chemistry: The Development of Ab Initio Methods in Molecular Electronic Structure Theory.* Clarendon Press, Oxford.

Sitkoff D., Sharp K. A., and Honig B. (1993), A rapid and accurate route to calculating solvation energy using a continuum model which includes solute polarizability. *Biophys. J.* 64: A65–A65.

Sternberg, M. J. E. (1996). *Protein Structure Prediction.* Oxford University Press, Oxford.

Tannor, D. J., Marten, B., Murphy, R., Friesner, R. A., Sitkoff, D., Nicholls, A., Ringnalda, M., Goddard III, W. A., and Honig, B. (1994). Accurate first principles calculation of molecular charge-distributions and solvation energies from ab-initio quantum-mechanics and continuum dielectric theory. *J. Am. Chem. Soc.* 116: 11875–11882.

Vaidehi, N., Jain, A., and Goddard III, W. A. (1996). Constant temperature constrained molecular dynamics: the Newton-Euler inverse mass operator method. *J. Phys. Chem.* 100: 10508–10517.

Vaidehi, N., and Goddard III, W. A. (1997). The pentamer channel stiffening model for drug action on human rhinovirus HRV-1A. *Proc. Natl Acad. Sci. USA* 94: 2466–2471.

Vaidehi, N., and Goddard III, W. A. (2000) Domain motions in phospho glycerate kinase using hierarchical NEIMO simulations. *J. Phys. Chem.* A 104: 2375–2383.

van Gunsteren, W. F., and Berendsen, H. J. C. 1987. *Groningen Molecular Simulation (GROMOS) Library Manual.* Biomos, Groningen.

van Gunsteren, W. F., and Weiner, P. K. (1989). In *Computer Simulation of Biomolecular Systems.* Leiden.

van Gunsteren, W. F., and Berendsen H. J. C. (1990). Computer-simulation of molecular-dynamics—methodology, applications, and perspectives in chemistry. *Ang. Chem. Int. Ed. Engl.* 29: 992–1023.

Warshel, A. (1991). In *Computer Modeling of Chemical Reactions in Enzymes and Solutions.* Wiley, New York.

Watanabe, T., Suzuki, K., Oyanagi, W., Ohnishi, K., and Tanaka, H. (1990). Gene cloning of chitinase-A1 from bacillus-circulans W1-12 revealed its evolutionary relationship to serratia chitinase and to the type-iii homology units of fibronectin. *J. Biol. Chem.* 265: 15659–15665.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. J. (1984). A new force-field for molecular mechanical simulation of nucleic-acids and proteins. *J. Am. Chem. Soc.* 106: 765–784.

Weiner, S. J., Kollman, P. A., Nguyen, D. T., and Case, D. A. (1986). An all atom force-field for simulations of proteins and nucleic-acids. *J. Comput. Chem.* 7: 230–252.

Woodcock, L. V. (1971). Isothermal molecular dynamics calculations for liquid salts. *Chem. Phys. Lett.* 10: 257–261.