20.2. Molecular-dynamics simulations of biological macromolecules

BY C. B. POST AND V. M. DADARLAT

20.2.1. Introduction

Molecular dynamics (MD) is the simulation of motion for a system of particles. Advances in the theory of atomic interactions and the increasing availability of high-power computers have led to rapid development of this field and greater understanding of macromolecular motions. In the earliest molecular-dynamics simulations of protein molecules (McCammon *et al.*, 1977; McCammon & Harvey, 1987), the systems were greatly simplified in order to fit within the computing capabilities of that time. Simplifications included the exclusion of water molecules and even of explicit hydrogen atoms; the effect of hydrogen atoms was built into the heavy-atom properties using so-called extended-atom parameters. Simulation time periods were limited to tens of picoseconds for systems of less than 10^3 atoms. Modern simulations, by contrast, are based on improved force fields (MacKerell *et al.*, 1998) and benefit from considerable development in algorithms. In addition, the possible size and time period of simulations have increased by orders of magnitude; large systems of the order of 10^4 atoms (including explicit solvent molecules) and nanosecond time periods are accessible. With dedicated computer time, the microsecond regime is possible (Duan & Kollman, 1998). Interestingly, the first 100 ps simulation of an enzyme complex was of hen egg-white lysozyme (Post *et al.*, 1986), the first enzyme whose structure was solved by X-ray crystallography. Then the simulation required several months of dedicated time on a Cray supercomputer, but now it can be accomplished in less than a week on a common workstation.

A consequence of this enormous growth in computing power has been the particularly successful application of molecular dynamics of biological molecules to three-dimensional structure determination and refinement. It is now practical to use molecular dynamics, in combination with crystallographic and NMR data, to search the large conformational space of proteins and nucleic acids to find structures consistent with the data and to improve the agreement with the data. The advantages of molecular dynamics over manual rebuilding and least-squares refinement are the abilities to overcome the local minimum problem in an automated fashion and to search the complex conformational space of a macromolecule more extensively (Brünger *et al.*, 1987).

20.2.2. The simulation method

Molecular mechanics, whereby the energy of the system is expressed in classical terms as a function of atomic coordinates, is well established as a useful approach for describing atomic interactions (Brooks *et al.*, 1988; Goodfellow & Levy, 1998). Owing to the size of proteins and nucleic acids, the potential-energy function for large biomolecules is empirically based rather than derived from quantum-mechanical calculations. The total force on each atom, $\mathbf{F}_i$, is calculated from the gradient, or the first derivative of this potential energy, with respect to the atomic coordinates. The motion of the atom resulting from the net force is described by Newton's equation of motion,

$$\mathbf{F}_i = m_i\mathbf{a}_i, \qquad (20.2.2.1)$$

where m_i is the mass of atom i and $\mathbf{a}_i$ is the acceleration. Integration of equation (20.2.2.1) gives

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \mathbf{F}_i(t)(\Delta t)^2/(2m_i), \qquad (20.2.2.2)$$

where Δt is the time step in the integration, $\mathbf{r}_i(t+\Delta t)$ is the atomic position at time $(t+\Delta t)$, given the position $\mathbf{r}_i(t)$ at time t, and $\mathbf{v}_i(t)$ is the velocity. The forces on the particles change continuously so that a numerical solution of the equation is required. The Verlet algorithm (Verlet, 1967) or a variation, Leapfrog, is commonly used.

20.2.3. Potential-energy function

20.2.3.1. *Empirical energy*

The central element of simulations is the interaction potential between atoms as a function of atomic position, **r**. The success of simulations in describing the average structure of proteins and other biological features suggests that such relatively simple potential functions adequately represent proteins and nucleic acids. The empirically based components of the energy function, $E_{\rm empir}$, include geometric terms for bond lengths, bond angles and torsion angles, and non-bonding terms for steric van der Waals interactions and electrostatic interactions. A commonly used energy function is

$$E_{\rm empir} = E_{\rm geom} + E_{\rm nonb}, \qquad (20.2.3.1)$$

$$E_{\rm geom} = \sum_{\rm bonds} (1/2)k_b(b-b_{\rm eq})^2 + \sum_{\rm angles} (1/2)k_\theta(\theta-\theta_{\rm eq})^2 + \sum_{\rm torsions} (1/2)k_\varphi[1+\cos(n\varphi-\delta)], \qquad (20.2.3.2)$$

$$E_{\rm nonb} = \sum_{\rm nbpairs} \left(q_iq_j/Dr_{ij}\right) + 4\varepsilon_{ij}\left[\left(\sigma_{ij}/r_{ij}\right)^{12} - \left(\sigma_{ij}/r_{ij}\right)^{6}\right], \qquad (20.2.3.3)$$

where k_b, k_θ and k_φ are force constants, $b_{\rm eq}$ and $\theta_{\rm eq}$ are equilibrium values for bond lengths, b, and angles, θ, respectively, and φ is the torsion angle of periodicity n and phase δ. The non-bonded terms depend on the interatomic distance r_{ij}, the dielectric constant D, the partial atomic charge q_i, and the van der Waals parameters ε_{ij} and σ_{ij}. The bond-stretching and angle-bending contributions are represented by harmonic potentials, while the energy associated with rotation about a bond, the torsional potential, is modelled by a cosine function [equation (20.2.3.2)]. The electrostatic component of the non-bonded interactions [first term of equation (20.2.3.3)] follows Coulomb's Law, and a Lennard–Jones 6–12 potential function [second term of equation (20.2.3.3)] is used to model steric repulsion and attractive dispersion interactions. $E_{\rm nonb}$, as a sum over pairs of atoms not involved in either a bond or bond angle, requires the use of a pairwise list between atoms. The small contribution from pairs separated by a large distance allows the use of cutoff limits for this list, but at some cost in accuracy.

Initial values for the atomic coordinates and velocities are required to begin the molecular-dynamics simulation. While initial coordinates are obtained from the model built into the electron-density map, it is necessary to generate initial velocities computationally. The most common approach is to assign random values for each atom, i, consistent with the temperature chosen for the system: $3kT/2 = (1/2)\sum m_i v_i^2$.

Integration of the equation of motion [equation (20.2.2.1)] also requires specification of the time step Δt. In the case of structure determination, this choice is limited only by the numerical stability of the calculation. Too large a value for Δt results in errors in the integration, manifested by a rapid and unacceptable increase in energy. Whereas a value for Δt of 1 to 2 fs is required for accurate trajectories and strict conservation of the energy, structure-determination protocols can employ larger values and are limited only by the need for numerical stability.

Both the temperature and Δt influence the sampling rate of conformational space. Enhanced sampling increases the rate of

references