CHAPTER 3

THEORETICAL BACKGROUND

Computer simulation plays an important role in science today. In the past, physical sciences were catalyzed by the interplay between experiment and theory. In experiment, we measure a system and then obtain some results in numeric form. In theory, we construct a model in a form of a set of mathematical equations. Then we validate the model to describe the system behavior in a few selected cases, simple enough to allow a solution to be computed from the equations. In many cases, this implies a considerable amount of simplification in order to eliminate all the complexities invariably associated with real world problems, and make the problem solvable.

In the past, only some of physical models could be easily tested by some approximations in the calculation. Others, such as the molecules involving a large amount of degree of freedom, the disordered systems and the systems that require accurate treatment of temperature effect, were difficult to test. After the adventure of high-speed computers, the new link between experimental and theory occurred. Theorists provide the model and the machine carries out the calculations by following the suitable algorithm. In this way, we can introduce any complexity and investigate more realistic systems to get a better understanding of real experiments.

The more development of computer simulation the more increase the demand for accuracy of the models. One of the examples is molecular dynamics simulation, which allows evaluating the melting temperature of a material, modeled by means of a certain interaction law. This is the difficult thing to do in the past. Therefore the simulation made models bring to life, disclose some critical areas and provide some suggestions to improve them. The simulation now come closer to the experiment thus some result can sometimes be compared directly with the experimental results. So we can use simulation as a powerful tool to understand and interpret the experiments at microscopic level. Moreover it can be used to study any experiment which are not accessible or any experiment which are expensive. It could be said that computer simulation is a way to do

the experiment, which are just impossible to do in reality. Its outcome increases our understanding of many phenomena greatly.

The basic principles of the simulation methodology designed to study structural, dynamical, and thermodynamic properties of biological macromolecules will be described in this chapter. The molecular mechanics and the statistical mechanics are the two theoretical bases behind the study of molecular dynamics simulation in biomolecules. The former is utilized for describing the numerous pairs of the potential energy arising form the interatomic interactions present in the molecules. The latter is employed for linking the microscopic properties obtained from the former to the observable macroscopic properties.

3.1 Energy minimization

An important method for exploring the potential energy surface is to find configurations that are stable points on the surface [45; 48]. This means finding a point in the configuration space where the net force on each atom vanishes. By simply minimizing the energy, stable conformations can be identified. Two first-order minimization algorithms are frequently used in computer simulation are the method of *steepest descents* and the *conjugate gradient* methods

3.1.1 The steepest descents method

Steepest descents performs the energy minimization through repeated minimizations along the direction of the force. It uses the first derivative of the potential energy with respect to the Cartesian coordinates. The method moves down the steepest slope of the interatomic forces on the potential energy surface. The descent is accomplished by adding an increment to the coordinates in the direction of the negative gradient of the potential energy, or the force.

This method shows reasonable convergence initially so it is commonly used when the gradients are large and the configurations are far from the minimum, which is the case for initial relaxation of poorly refined crystallographic data or graphically built

molecules. Each step in steepest descents requires minimal computing time. However, there is a disadvantage of the method that the progress becomes slow when approaching the minimum. Thus, other minimization methods are carried out after the steepest descents.

3.1.2 Conjugate gradients

The conjugate gradient differs from the steepest descent by using both the current gradient (search direction) and the previous search direction to drive the minimization.

The advantage of this method is that it uses the minimization history to calculate the search direction, so the convergence is faster than the steepest descents. It also contains a scaling factor for determining step size. This makes the step sizes optimal when compared to the steepest descents.

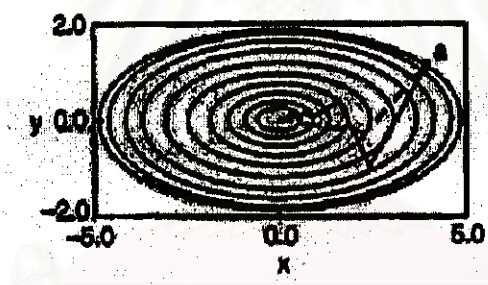

Figure 3.1 Minimization paths for a simple energy surface: —— steepest descents, – – – – conjugate gradients [45].

3.2 Molecular simulation

3.2.1 Molecular mechanics

A simple model is developed in order to study the interactions in molecule in the system, which is the stretching of bonds, the bending of angles and the rotation about the single bonds, etc [45; 48]. The energy of a molecule is a sum of the steric and non-bonded interactions using force fields. Force field is the empirical method to fit the potential energy surface. The force field defines the coordinates used, the mathematical

form of the equations involving the coordinates, and the parameters adjusted in the empirical fit of the potential energy surface. The force fields commonly used for describing molecules employ a combination of internal coordinates (bond distances, bond angles, torsions), to describe the van der Waals and electrostatic interaction between atoms. The function forms range from simple quadratic forms to Morse functions, Fourier expansions, Lennard-Jones potentials, etc. The goal of a force field is to describe entire classes of molecules with reasonable accuracy. In a sense, the force field interpolates and extrapolates from the empirical of the small set of molecules used to parameterize the force field to a larger set of related molecules and structures.

Molecular mechanics force fields use the equations of classical mechanics to describe the potential energy surfaces and physical properties of molecules. The component of a force field is the energy arising from compressing and stretching the bonds, bending the angles, rotating about the single bonds and interacting of non-bond. One function for such a force field that can be used to show the variety of interactions, which are important in protein interactions and seen suitable simple mathematical forms for their representation is

$$\begin{aligned} V(r^N) = & \sum_{bond} K_r (r - r_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 \\ & + \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) \\ & + \sum_{coulombs} \frac{q_i q_j}{\varepsilon r_{ij}} + \sum_{vdW} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right] + \sum_{Hbonds} \left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right] \end{aligned} \tag{3.1}$$

where $V(r^N)$ denotes the potential energy which is a function of the positions r of N particles. Each potential term in Eq.(3.1) will be described as follows.

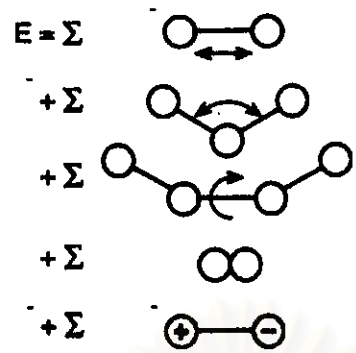

Figure 3.2 Pictorial representation of the terms included in a molecular mechanic force fields [48].

3.2.1.1 Bond stretching

The potential energy curve for a typical bond has the form shown in figure 3.2.

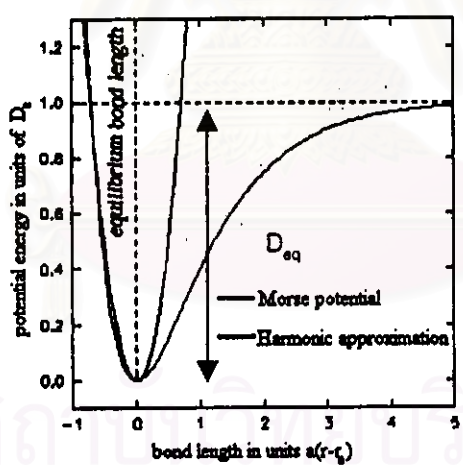

Figure 3.3 Curves showing the variation of bond stretch energy with distance: Morse potential and harmonic potential [48].

The convenient model to represent the potential energy of the diatomic molecules is *Morse Potential*. The function form of the potential is

$$V = D_{eq}\{1 - e^{-a(r-r_0)}\}^2 \qquad (3.2)$$

where r is bond length, r_0 is the equilibrium bond length, D_{eq} is the potential energy for bond formation and *"a"* a parameter controlling the width of the potential well. For large molecules, this is still computationally expensive calculation. Thus, the standard way to approximate the bond stretching potential energy in protein and most other molecules is to employ Hooke's law:

$$V_{bond} = \sum_{bonds} K_r (r - r_0)^2 \qquad (3.3)$$

where K_r is the stretching force constant.

From the mathematical point of view, the two functions (eq. 3.2 and 3.3) have similar form at the potential minimum. This implies that both functions will give the same equilibrium structures. Thus, the force field employed with this harmonic potential is well described for the equilibrium.

3.2.1.2 Angle bending

Bond angles are treated in the same way as bond length and usually described by a harmonic function.

$$V_{angle} = \sum_{angles} K_\theta (\theta - \theta_0)^2 \qquad (3.4)$$

As before K_θ is a bending force constant, θ_0 the equilibrium value for the bond angle.

3.2.1.3 Torsional angles

In molecular mechanics, the dihedral potential function is often implemented as a truncated Fourier series. This periodic function is appropriate for the torsional potential. The standard function form for representing the potential energy for a rotational is

$$V_{torsions} = \sum_{torsions} \frac{V_n}{2}[1+\cos(n\phi-\gamma)] \qquad (3.5)$$

In this representative, V_n is the rotational barrier height, n is the periodicity of rotation, ϕ is the torsion angle, and γ is the phase angle shifts the curve to the left or right. (Fig.3.2).

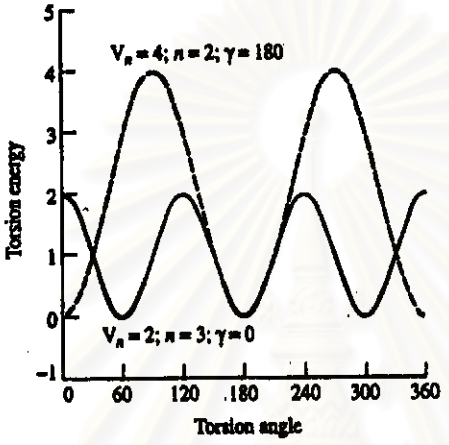

Figure 3.4 Variation of energy with dihedral angle [48].

3.2.1.4 Non-bonded interactions

The non-bonded interactions are distance-dependent calculated as the sum over all atoms with a 1,4 or greater separation. It is usual to consider these interactions as having two components: van der Waals and electrostatic. In the case of *electrostatic interactions*, charges on nuclei and electrons interact according to Coulomb's law. This interaction provides a quantitative measurement of the influence of polarity on the energy and structure.

$$V = \sum_{coulombs} \frac{q_i q_j}{\varepsilon r_{ij}} \qquad (3.6)$$

In this model of electrostatic interactions, two atoms (*i* and *j*) have point charges q_i and q_j. The magnitude of the electrostatic energy varies inversely with the distance between the atoms, r_{ij}. The dielectric constant is given as ε.

van der Waals interaction can be considered as both a size parameter and a representative of electron correlation (resulting from instantaneous dipole interactions). The van der Waals interactions can be determined using the most common function called 6-12 Lennard-Jones potential

$$V_{vdW} = \sum_{i<j}\left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right] \qquad (3.7)$$

where A_{ij} and B_{ij} are calculated from εr_m^{12} and $2\varepsilon r_m^{6}$ respectively, where ε is the well depth (potential energy) at r_m, and r_m is the minimum energy interaction distance. The r^{-12} term accounts short range repulsion whereas the r^{-6} term represents London dispersion-attraction forces.

There is an additional term to be included. That is *hydrogen bonding*. It is the electrostatic interactions between groups, which carry no formal overall electrical charge. The hydrogen bonding is vital for stability and the fundamentally importants to biomolecular structure. This function, 10-12 Lennard-Jones potential, is used to model the interaction between the donor hydrogen atom and the heteroatom acceptor atom. This function is calculated to ensure correct geometries of hydrogen bonding systems. Such force field in certain protein has the form

$$V_{hbond} = \sum\left(\frac{C_{ij}}{r_{ij}^{12}}\right) - \left(\frac{D_{ij}}{r_{ij}^{10}}\right) \qquad (3.8)$$

where *Cij* and *Dij* are hydrogen bond parameters for the interacting pair of atoms. The hydrogen bonding potential given does not contribute significantly to the hydrogen

bonding attraction between two atoms but it is rather implemented to fine-tune the distances between these atoms.

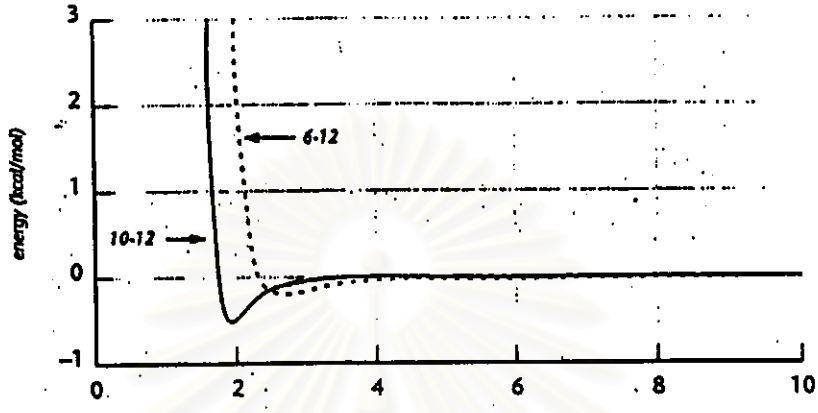

Figure 3.5 The van der Waals (6-12) and hydrogen bond (10-12) potentials [43].

Some non-covalent bonds i.e. hydrogen bonds, electrostatic, and Van derWaals interactions were illustrated in figure 3.5.

3.2.2 Statistical Mechanics

To determine experimentally the value of the property of the system, such properties will depend upon the positions and momenta of the N particles that comprise the system [48]. The instantaneous value of property A can thus be written as $A(p^N(t), r^N(t))$, where $p^N(t)$ and $r^N(t)$ represent the N momenta and positions respectively at time t (i.e. $A(p^N(t), r^N(t)) \equiv A(p_{1x}, p_{1y}, p_{1z}, p_{2x}, \ldots, x_1, y_1, z_1, x_2, \ldots, t)$ where p_{1x} is the momentum of particle 1 in the x direction and x_1 is its coordinate. The fluctuation of the instantaneous value of the property A is a result of interactions between the particles over the times. Such experimental value, which averaged A over the time, is known as *time average* as follows

$$A_{ave} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), r^N(t))dt \qquad (3.9)$$

For any atoms in the system, the force, which acting on each atom due to the interaction with other atoms, is calculated by differentiating the energy function. The acceleration is then determined from force via Newton's second law. The trajectory described how the positions, velocities and accelerations of the system vary with the time are determined using the equation (3.9).

3.2.3 Molecular dynamics

Molecular dynamics (MD) is a computer simulation technique where the time evolution of a set of interacting atoms, macromolecules and the surrounding solvent, is followed by integrating their equation of motion [45; 48]. This method provides most exact and detail information from simulating the dynamics of macromolecules of biological interest.

3.2.3.1 Equation of motion

In molecular dynamics, the motions of a system of particles (atoms) are simulated with respect to forces. The forces, that are acting on each atom and cause the system to change, were collected to give the collective motion of particles over time by integrating Newton's second law.

$$F_i = m_i a_i \qquad (3.11)$$

where F is the force acting on a particle, m its mass, and a its acceleration. This made the following of the evolution of the system over time.

From equation (3.9), one can rearrange the acceleration into the second derivative of displacement(s) with respect to time, $\delta^2 s/\delta t^2$, then (3.9) will be

$$\frac{\delta^2 s}{\delta t^2} = \frac{F_i}{m_i} \tag{3.12}$$

Solving this second order differential equation will give the dynamic behavior for every particles of the system. Integrating (3.12) with respect to time gives

$$\frac{\delta s}{\delta t} = \left(\frac{F_i}{m_i}\right)t + c_1 \tag{3.13}$$

Initial velocity, u_i, is given constant at time $t = 0$. The expression for velocity at any time t is

$$\frac{\delta s_i}{\delta t} = a_i t + u_i \tag{3.14}$$

Integration with respect to time produces

$$s_i = u_i t + \tfrac{1}{2} a_i t^2 + c_2 \tag{3.15}$$

where the constant is the current position. Therefore the displacement can be calculated from an initial velocity, u_i, and the acceleration which can be derived from $a_i = F_i/m_i$.

The displacement can be expressed in truncated Taylor series, which has the form as follows

$$x(t+\Delta t) = x(t) + \left(\frac{\delta x}{\delta t}\right)\Delta t + \left(\frac{\delta^2 x}{\delta t^2}\right)\frac{\Delta t^2}{2} + \ldots \tag{3.16}$$

This assumed that the acceleration remains constant throughout the time step Δt. The integration algorithm was used to overcome to problem of finite time steps and truncation errors.

3.2.3.2 Leap-frog integration algorithm

The leap-frog algorithm is one of the finite difference methods, which are used for generating molecular dynamics trajectories with continuous potential models. The general idea is that the integration is separated into small time, δt, the total force on each particle in the configuration at time t is calculated as the sum of its interactions with other particles. The acceleration of the particles is combined with the position and velocity at time t, the positions and velocities at time $t+\delta t$ are then calculated. The forces on the particles in their new positions are then determined, leading to new positions and velocities at time $t+2\delta t$, and so on. The positions, velocities, and accelerations can be approximated as Taylor series expansions

$$r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + \frac{1}{6}\delta t^3 b(t) + \frac{1}{24}\delta t^4 c(t) + \ldots \quad (3.17)$$

$$v(t+\delta t) = v(t) + \delta t a(t) + \frac{1}{2}\delta t^2 b(t) + \frac{1}{6}\delta t^3 c(t) + \ldots \quad (3.18)$$

$$a(t+\delta t) = a(t) + \delta t b(t) + \frac{1}{2}\delta t^2 c(t) + \ldots \quad (3.19)$$

$$b(t+\delta t) = b(t) + \delta t c(t) + \ldots \quad (3.20)$$

where v is the velocity (the first derivative of the positions with respect to time), a the acceleration (the second derivative), b the third derivative, and so on.

In leap-frog algorithm, the velocities $v(t+\frac{1}{2}\delta t)$ are first calculated from the velocities at time $t-\frac{1}{2}\delta t$ and the accelerations at time t, where a can be calculated from

$$v(t+\frac{1}{2}\delta t) = v(t-\frac{1}{2}\delta t) + \delta t a(t) \quad (3.21)$$

The new position is then given by

$$r(t+\delta t) = r(t) + \delta t\, v(t+\frac{1}{2}\delta t) \qquad (3.22)$$

3.2.3.3 Periodic Boundary Condition

In the simulation, the imaging box is set up in order to treat the boundary effect, which thus molecule closes to the edge of this box cannot receive all interactions for reliable simulation. Therefore, the periodic boundary condition was set to the system, where the system of interest is surrounded by images of itself. Furthermore, this can help keeping constant number of molecules in the system by entering the image of the leaving molecule.

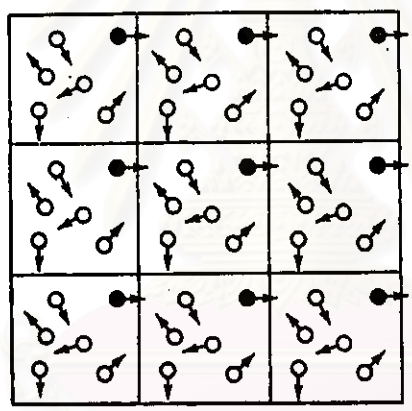

Figure 3.6 Periodic boundary condition in two dimensions [48].

The key point is that now each particle i in the box should be thought as interacting not only with other particles j, but also with their images in the nearby box.

3.2.4 Constant pressure and constant temperature procedure

3.2.4.1 Constant temperature

Coupling to an external bath is the way to maintain the temperature [32]. The bath acts as a thermal energy source to supply or remove heat. At each step, the velocities are scaled such that the rate of change of temperature is proportional to the difference in temperature between the bath and the system.

$$\frac{dT(t)}{dt} = \frac{1}{\tau}\left(T_{bath} - T(t)\right) \tag{3.23}$$

where τ is a coupling parameter whose magnitude determines how tightly the bath and the system are coupled together. The change in temperature between successive time steps is

$$\Delta T = \frac{\delta t}{\tau}\left(T_{bath} - T(t)\right) \tag{3.24}$$

The scaling factor for the velocities is thus:

$$\lambda^2 = 1 + \frac{\delta t}{\tau}\left(\frac{T_{bath}}{T(t)} - 1\right) \tag{3.25}$$

3.2.4.2 Constant pressure

In a macroscopic system, changing its volume is the way to maintain constant pressure. A simulation in isothermal-isobaric ensemble also maintains constant pressure by changing the volume of the simulation cell. So one of the methods used for pressure control is simply the volume scaling. An alternative is to couple the system to a 'pressure bath', analogous to the temperature bath [32]. The rate of change of pressure is given by

$$\frac{dP(t)}{dt} = \frac{1}{\tau_P}\left(P_{bath} - P(t)\right) \tag{3.26}$$

where τ_P is the coupling constant, P_{bath} is the pressure of the bath, and P(t) is the actual pressure at time t. The volume of the simulation box is scaled by the factor λ, which is equivalent to scaling the atomic coordinates by a factor $\lambda^{1/3}$. Thus

$$\lambda = 1 - \mathrm{K}\frac{\delta t}{\tau_P}\left(P - P_{bath}\right) \tag{3.27}$$