

## Chapter 3

### Molecular Dynamics Simulation

Molecular dynamics calculates the biological system in liquid state; this property is the same as the liquid NMR. Besides, the time average property and kinetic property of molecular dynamics simulation is in good accordance with experiments. In order to understand molecular motion, Lipari and Szabo<sup>24; 25</sup> developed the Model-Free approach connecting NMR relaxation and molecular motion. This approach also made NMR relaxation and MD simulation more comparable. Because MD simulates the system usually in ns-ps time scale, the Model-Free approach efficiently calculated magnitude of motion in the same time regime. Thus, this approach increased the connection of MD and NMR. And MD simulation is a good tool to analyze dynamics features in atomic view. That is the reason why we conducted MD simulation to assist us to understand the dynamics in catalytic pathway.

#### 3.1 The Force Field Formalism

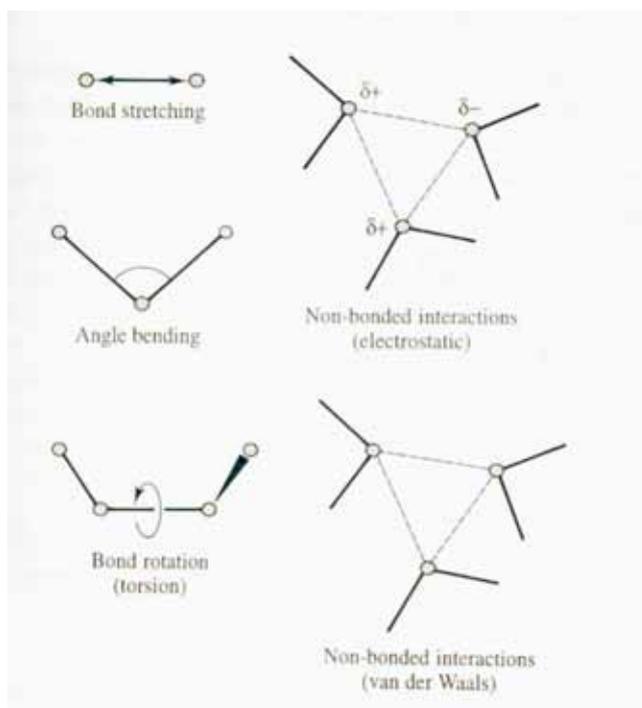
In recent standard molecular dynamics simulation, molecular mechanics is invariably used to perform calculations on systems containing significant numbers of atoms. Molecular mechanics is based upon a rather simple model of the interaction within a system with contributions from processes such as the stretching of bonds, the opening and closing of angles and the rotations about single bonds.

Many of the molecular modeling force field in use today for molecular systems can be interpreted in terms of a relatively simple four-component picture of the intra- and inter- molecular force within the system. Energetic penalties are associated with

the deviation of bonds and angles away from their “reference” or “equilibrium” values, and there is also a function that describes how the energy changes as a bond rotates, and force field also contains terms that described the interaction between non-bonded parts of the system. An attractive feature of this representation is that the various terms can be ascribed to changes in specific internal coordinates such as bond lengths, angles, and the rotation of bonds or movements of atoms relative to each other. This makes it easier to understand how changes in the force field parameters affect its performance, and also helps in the parameterizations process. One functional form for such a force field that can be used to model single molecules or assemblies of atoms or molecule is<sup>36</sup>:

$$\begin{aligned}
 U(\vec{r}^N) = & \sum_{bonds} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 + \sum_{torsions} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \\
 & + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left( 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right) \quad (3.1)
 \end{aligned}$$

$U(\vec{r}^N)$  denotes the potential energy which is a function of the set of the position  $\vec{r}_s$  of N particles (usually atoms). The various contributions are schematically represented in Figure 3.1. The first term in equation [3.1] describes the interaction between pairs of bonded atoms, modeled here by a harmonic potential that gives the increase in energy as the bond length  $l_i$  deviates from the reference value  $l_{i,0}$ . The second term is a summation over all valence angles in the molecule, also modeled by a harmonic potential. The third term is torsional potential that models how the energy changes as a bond rotates. The fourth term contribution is the non-bonded term. This calculation between all pairs of atoms ( $i$  and  $j$ ) that are in different molecules or that are in the same molecule but separated by a least three bonds. In a simple force field the non-bonded term is usually modeled using a Coulomb potential term for electrostatic and a Lennard-Jones potential for van der Waals interactions.



**Figure 3.1**

Schematic representation of the four key contributions to a molecular mechanics force field: bond stretching, angle bending and torsional terms and non-bonded interactions.

### 3.2 Integration of the Newtonian Equation of Motion

For any arrangement of the atoms in the system, the force acting on each atom due to interactions with other atoms can be calculated by differentiating the energy function. From the force on each atom it is possible to determine its acceleration via Newton's second law. Integration of the equations of motion should then yield a trajectory that describes how the positions, velocities and accelerations of the particles vary with time, and from which the ensemble average values of properties can be determined via the ergodic theorem:

$$\langle A \rangle = \overline{A(t)} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), r^N(t)) dt \quad (3.2)$$

The instantaneous value of the property  $A$  can 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  $t$ .

In molecular dynamics, successive configurations of the system are generated by integrating Newton's laws of motion. The result is a trajectory that specifies how positions and velocities of the particles in the system vary with time. For a system in the micro-canonical ensemble, a trajectory is obtained by solving the differential equations embodied in Newton's second law ( $F=ma$ ):

$$\frac{d^2 x_i}{dt^2} = \frac{F_{x_i}}{m_i} \quad (3.3)$$

This equation describes the motion of a particle of mass  $m_i$  along one coordinate ( $x_i$ ) with  $F_{x_i}$  being the force on the particle in that direction. The essential idea is that the integration is broken down into many small stages, each separated in time by a fixed time  $\delta t$ . The total force on each particle in the configuration at a time  $t$  is calculated as the vector sum of its interactions with other particles. From the force we can determine the accelerations of the particles, which are then combined with the

positions and velocities at a time  $t$  to calculate the positions and velocities at a time  $t + \delta t$ . The force is assumed to be constant during the time step. 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.

It is assumed that the positions and dynamic properties (velocities, accelerations, etc.) can be approximated as Taylor series expansions:

$$r(t + \delta t) = r(t) + \delta t v(t) + \frac{1}{2} \delta t^2 a(t) + \frac{1}{6} \delta t^3 b + \dots$$

where  $v$  is the velocity (the first derivative of the positions with respect to time),  $a$  is the acceleration (the second derivative),  $b$  is the third derivative, and so on. The Verlet algorithm<sup>37</sup> is probably the most widely method for integrating the equations of motion in a molecular dynamics simulation. The Verlet algorithm uses the position and accelerations at time  $t$ , and the position from the previous step,  $r(t - \delta t)$ , to calculate the new positions at  $t + \delta t$ ,  $r(t + \delta t)$ . We can write down the following relationships between these quantities and the velocities at time  $t$ :

$$r(t + \delta t) = r(t) + \delta t v(t) + \frac{1}{2} \delta t^2 a(t) + \dots \quad (3.4)$$

$$r(t - \delta t) = r(t) - \delta t v(t) + \frac{1}{2} \delta t^2 a(t) - \dots \quad (3.5)$$

Adding these two equations gives

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + \delta t^2 a(t) \quad (3.6)$$

The velocities do not explicitly appear in the Verlet integration algorithm. The velocities can be calculated in a variety of ways; a simple approach is to divide the difference in positions at times  $t + \delta t$  and  $t - \delta t$  by  $2\delta t$ :

$$v(t) = [r(t + \delta t) - r(t - \delta t)] / 2\delta t \quad (3.7)$$

Alternatively, the velocities can be estimated at the half-step,  $t + \frac{1}{2}\delta t$  :

$$v(t + \frac{1}{2}\delta t) = [r(t + \delta t) - r(t)] / \delta t \quad (3.8)$$

Implementation of the Verlet algorithm is straightforward and the storage requirements are modest, comprising two sets of positions  $\mathbf{r}(t - \delta t)$  and  $\mathbf{r}(t)$  and the accelerations  $\mathbf{a}(t)$ .

### 3.3 Temperature and Pressure Coupling

Molecular dynamics is traditionally performed in the constant  $NVE$  (or  $NVEP$ ) ensemble. Although thermodynamic results can be transformed between ensembles, this is strictly only possible in the limit of infinite system size ('the thermodynamic limit'). It may therefore be desired to perform the simulation in a different ensemble. The two most common alternative ensembles are the constant  $NVT$  and the constant  $NPT$  ensembles. In our system, we used the  $NPT$  ensembles.

In order to control the temperature of molecular system, we coupled the system to an external heat bath that is fixed at the desired temperature<sup>38</sup>. The bath acted as a source of thermal energy, supplying or removing heat from the system as appropriate. The velocities are scaled at each step, such that the rate of change of temperature is proportional to the difference in temperature between the bath and the system:

$$\frac{dT(t)}{dt} = \frac{1}{\tau} (T_{bath} - T(t)) \quad (3.9)$$

where  $\tau$  is a coupling parameter whose magnitude determines how tightly the bath and the system are coupled together. This method gives an exponential decay of the system towards the desired temperature. The change in temperature between successive steps is:

$$\Delta T = \frac{\delta t}{\tau} (T_{bath} - T(t)) \quad (3.10)$$

If  $\tau$  is large, then the coupling will weak; and if  $\tau$  is small, the coupling will be strong. This approach permits the system to fluctuate about the desired temperature.

A simulation in the isothermal-isobaric ensemble maintains constant pressure by changing the volume of the simulation cell. The amount of volume fluctuation is related to the isothermal compressibility,  $\kappa$ :

$$\kappa = -\frac{1}{V} \left( \frac{\partial V}{\partial P} \right)_T \quad (3.9)$$

An easily compressible substance has a larger value of  $\kappa$ , so large volume fluctuations occur at a given pressure than in a more incompressible substance. Many of the methods used for pressure control are analogous to those used for temperature control. Thus, the pressure can be maintained at a constant value by simply scaling the volume. An alternative is to couple the system to a ‘pressure bath’, analogous to a temperature bath<sup>38</sup>. The rate of change of pressure is given by:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_p} (P_{bath} - P(t)) \quad (3.10)$$

$\tau_p$  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 a factor  $\lambda$ , which is equivalent to scaling the atomic coordinates by a factor  $\lambda^{1/3}$ . Thus:

$$\lambda = 1 - \kappa \frac{\delta t}{\tau_p} (P - P_{bath}) \quad (3.11)$$

The new positions are given by:

$$r'_i = \lambda^{1/3} r_i \quad (3.12)$$

The constant  $\kappa$  can be combined with the relaxation constant  $\tau_p$  as a single constant. This expression can be applied isotropically or anisotropically. In general, it is best to use the anisotropic approach as this enables the box dimensions to change independently.

In the extended pressure-coupling system methods, first introduced by Anderson<sup>39</sup> an extra degree of freedom, corresponding to the volume of the box, is added to the system. The kinetic energy associated with this degree of freedom is  $\frac{1}{2} Q (dV/dt)^2$ , where  $Q$  is the ‘mass’ of the piston. The piston also has potential energy  $PV$ , where  $P$

is the desired pressure and  $V$  is the volume of the system. The volume can vary during the simulation, with the average volume being determined by the balance between the internal pressure of the system and the desired external pressure. The extended-system temperature-scaling method uses a scaled time; in the extend pressure method the coordinates of the extended system are related to the ‘real’ coordinates by

$$r'_i = V^{1/3} r_i \quad (3.13)$$

### 3.4 Constrained Dynamics

When the simulation involves conformational flexible molecules then the motion is inevitably desired in terms of the atomic Cartesian coordinates. The high frequency motions (e.g. the vibration of bonds between hydrogen atoms and heavy atoms) are usually of less interest than the lower frequency modes, which often correspond to major conformational changes. Unfortunately, the time step of a molecular dynamics simulation is dictated by the highest frequency motion present in the system. In order to increase the time step without affecting the other internal degrees of freedom, the constrained dynamics is applied in many biological systems.

The most commonly used method for applying constraints, particularly in molecular dynamics, is the SHAKE procedure of Ryckaert, Ciccotti and Berendsen<sup>40</sup>. In constraint dynamics the equations of motion are solved while simultaneously satisfying the imposed constraints. Constrained systems have been much studied in classical mechanics. Constraints are often categorized as holonomic or non-holonomic. Holonomic constraints can be expressed in the form

$$f(q_1, q_2, q_3, \dots, t) = 0 \quad (3.14)$$

$q_1, q_2$ , etc., are the coordinates of the particles. Non-holonomic constraints can not be expressed in this way. SHAKE uses holonomic constraints. In a constrained system the coordinates of the particles are not independent and the equations of motion in each of the coordinates of the particles are not independent and the equations of motion in each of the coordinate directions are connected. As we know, the motion of a system of  $N$  particles can be described in terms of  $3N$  independent coordinates or degree of freedom. If there are  $k$  holonomic constraints then the number of degrees of freedom is reduced to  $3N-k$ .

In the general case, the equations of motion for a constrained system involve two type of force: the ‘normal’ forces arising from the intra- and intermolecular interactions, and the forces due to the constraints. The constraint  $\sigma_k$  requires the bond between atoms  $i$  and  $j$  to remain fixed. The constraints influences the Cartesian coordinates of atom  $i$  and  $j$ . The force due to this constraint can be written as follows:

$$F_{ckx} = \lambda_k \frac{\partial \sigma_k}{\partial x} \quad (3.15)$$

where  $\lambda_k$  is the Lagrange multiplier and  $x$  represents one of the Cartesian coordinates of the two atoms. If an atom is involved in a number of constraints then the local constraint force equals the sum of all such terms. The nature of the constraint for a bond between atoms  $i$  and  $j$  is:

$$\sigma_{ij} = (r_i - r_j)^2 - d_{ij}^2 = 0 \quad (3.16)$$

The constraint force lies along the bond at all times. For each constraint bond, there is an equal and opposite force on the two atoms that comprise the bond. The overall effect is that the constraint forces do no work. Suppose the constraint  $k$  corresponds to the bond length with respect to the coordinates of atoms  $i$  and  $j$  and multiplying by an undetermined multiplier,

$$\begin{aligned} \frac{\partial \sigma_k}{\partial r_i} &= 2(r_i - r_j) \quad \text{so} \quad F_{ci} = \lambda(r_i - r_j) \\ \frac{\partial \sigma_k}{\partial r_j} &= -2(r_i - r_j) \quad \text{and} \quad F_{cj} = -\lambda(r_i - r_j) \end{aligned} \quad (3.17)$$

The above expression for the forces can be incorporated into the Verlet algorithm as follows:

$$r_i(t + \delta t) = 2r_i(t) - r_i(t - \delta t) + \frac{\delta t^2}{m_i} F_i(t) + \sum_k \frac{\lambda_k \delta t^2}{m_i} r_{ij}(t) \quad (3.18)$$

The SHAKE method uses an alternative approach in which constraint is

considered in turn and solved. Satisfying one constraint may cause another constraint to be violated, and so it is necessary to iterate around the constraints until they are all satisfied to within some tolerance. The tolerance should be tight enough to ensure that the fluctuations in the simulation due to the SHAKE algorithm are much smaller than the fluctuations due to other sources, such as the use of cutoffs. Another important requirement is that the constrained degree of freedom should be only weakly coupled to the remaining degree of freedom, so that the motion of the molecule is not affected by the application of the constraints. The sampling of unconstrained degree of freedom should also be unaffected. Therefore, angle constraints can be easily accommodated in the SHAKE scheme by recognizing that an angle constraint simply corresponds to an additional distance constraint. The SHAKE method has been extended by Tobias and Brooks <sup>41</sup> to enable constraints to be applied to arbitrary internal coordinates. This enables the torsion angle of a rotatable bond to be constrained to a particular value during a molecular dynamics simulation.