Chapter – 3

Computer simulation technique using ESFF

(Extensive Systematic Force Field)

The usual approach to complex formation of substances is the experimental method. Before making an approach to experimental studies, a theoretical approach to the expected structure was carried out using computer simulation technique. The experimental results were then correlated to the theoretical model to see whether they agree.

The past few years have seen impressive advances in the power and scope of computer simulation techniques for biomolecules. Molecular simulations have proved to be an excellent complementary tool to experimental techniques for the determination of the structure as well as the related properties. \(^{58,59}\)

Molecular mechanics (MM) has evolved into an important and widely used theoretical method that allowed researchers in chemistry, physics and biology to model the atomistic level details of many different types of systems including gases, liquids, solids, surfaces and clusters. It is
well known that both sequence and composition of DNA influences the binding of various metal complexes. The energetic of binding of the transition metal ions is essential for understanding the mechanism involved in their reaction.

**Basic principles of molecular mechanics**

Molecular mechanics (MM) is used to describe (a) the structure and stability of a molecular system, (b) the free energy of different states of a molecular system and (c) the reaction processes with molecular systems in terms of interactions at the atomic level based on the force field. To accomplish this, MM codes must generate an equation for each molecule that relates its potential energy to its nuclear coordinates. This equation defines a potential hypersurface of $3N+1$ dimensions, where $3N$ dimensions are used to specify the positions of each of the $N$ atoms in the molecule and the extra dimension defines the potential energy. Each point on the potential surface corresponds to the geometry of the molecule. Minima on the potential surface correspond to the stable conformation of the molecule. MM codes are able to locate these minima and identify conformations.
potential energy associated with each stable conformation is used to assign their relative stability.\textsuperscript{61}

In MM models, the equation that relates a molecule's potential energy to its conformation is composed of a sum of terms that represent different types of energy contributions. The types of term that are included in a potential energy equation often differ from one force field parameter to the other. In the simplest model, the total potential energy is broken down into four components.

\[ U_{\text{total}} = U_r + U_\theta + U_\phi + U_{\text{vdw}} \]

The sum \( U_{\text{total}} \) represents the contributions of the potential energy, due to bond stretching compression terms \( U_r \), valence angle bending terms, \( U_\theta \), internal rotational or torsional terms \( U_\phi \) and van der walls interactions \( U_{\text{vdw}} \).\textsuperscript{62}

**Force field parameters**

The terms in the potential energy equation are described by simple analytical expressions with adjustable parameters. These are called potential functions. A set of potential functions and its corresponding set of parameters together are called as force fields. Force fields are categorized by the type of potential functions they contain.
The force fields commonly used for describing molecules employ a combination of internal coordinates and terms (bond distances, bond angles, torsions, etc) to describe part of the potential energy surface due to interactions between bonded atoms and nonbonded terms to describe the van der waals and electrostatic interactions between atoms. The actual coordinates of a model combined with the force field data lead to the energy expression for the model. This energy expression is the equation that describes the potential energy surface of a particular model as a function of its atomic coordinates.

A typical molecular force field or effective potential for a system of N atoms with masses \( m_i \) \((i=1,2,\ldots,N)\) and cartesian vector \( r_i \) has the following form:

\[
V(r_1, r_2, \ldots, r_N) = \sum_{\text{bonds}} \frac{1}{2} K_b (b-b_0)^2 + \sum_{\text{angles}} \frac{1}{2} K_0 (\theta - \theta_0)^2 \\
+ \sum_{\text{improper dihedrals}} \frac{1}{2} K_\xi (\xi - \xi_0)^2 + \sum_{\text{dihedrals}} K_\phi [1+\cos(n\phi-\delta)] \\
+ \sum_{\text{pairs}(i,j)} [C_{12}(i,j)/r_{ij}^{12}] - c_6(i,j)/r_{ij}^6 + q_i q_j /4\pi\varepsilon_0 \varepsilon_r r_{ij}
\]

The first term represents the covalent bond stretching interaction along the bond. It is a harmonic potential in which the minimum energy bond length \( b_0 \) and the force constant \( K_b \) vary with the particular type of bond. The
second term describes the bond angle bending interaction in the similar form. Two forms are used for the dihedral angle interactions; a harmonic term for dihedral angles $\xi$, that are not allowed to make transitions, (dihedral angles within aromatic rings) and sinusoidal term for the other dihedral angle $\phi$, which may make 360 degree turn. The last term is a sum of all pairs of atoms and represents the effective non bonded interaction, composed of the Van der waals and the coloumb interaction between atoms $i$ and $j$ with charges $q_i$ and $q_j$ at a distance $r_{ij}$.

One should fit the force field parameters to properties of small molecules, which may be considered as building blocks of larger molecules such as proteins and DNA, and subsequently apply them to these larger molecules without any further adaptations. The choice of the particular force field should depend on the type of the system for which it has been designed.

The AMBER and CHARMM force fields are aimed at the description of isolated polypeptides and poly nucleotides, in which the absence of a solvent environment is compensated by the use of the distance dependent dielectric
constant $\varepsilon_r$. The Consistent valence force field (CVFF) is a classic force field, which is the traditional default force field in the program having some anharmonic and cross term enhancement. It has been used extensively for peptides and proteins. The CFF family (CFF91, PCFF, CFF, COMPASS) is closely related to second generation force fields. The CFF family of force fields was parameterized against wide range of experiments, which were observable for organic compounds containing halogen atoms and ions, alkali metal cations and several important divalent metal cations. PCFF is based on CFF91, extended to organic polymers, metal and zeolites. COMPASS is a new version of PCFF$^{66}$.

The universal force field is an excellent general purpose force field$^{67}$. It was parameterized for the full periodic table and has been carefully validated for main-group compounds, organic molecules, and metal complexes.$^{68}$ The extensible and systematic force field (ESFF) supports 879 atom types covering all the elements of the periodic table upto Rn. This force field has been used for the modeling of transition metal complexes$^{69}$. 

28
Minimization

Minimization is an important method for exploring the potential energy surface to find the configurations, which are stable points on the surface. It is performed in two steps. First, the energy expression must be defined and evaluated for a given conformation. Next the conformation is adjusted to lower the value of the energy expression.

A minimum may be found after one adjustment or may require thousands of iteration depending on the nature of the algorithm, the form of the energy expression and the size of the model. The minimisers used are discover and discover_3 module in Insight II package. The popular and widely used minimization algorithms are steepest descent and conjugate gradient method. In the steepest descent methods, the line search direction is defined along the direction of the local downhill gradient. The exclusive reliance of the steepest descent method on the gradient is its weakness as well as strength. Convergence is slow near the minimum because the gradient approaches zero, but the method is extremely robust, even for the systems that are far from harmonic. Hence, this method is often used when
gradients are large and the configurations are far from the minimum. This is commonly the case with initial relaxation of poorly refined models. The more advanced algorithms like conjugate gradient are often designed to begin with steepest descent as the first step. In this algorithm, the time per iteration may be longer than for the steepest, but more efficient convergence to the minimum is achieved by conjugate gradients.\textsuperscript{71}

**Computational details**

Molecular modeling was performed on a Silicon Graphics 02 work station using the Biosym Modeling Package from Molecular simulation. Inc. (San Diego, CA). The details of L-Cystine structure building and their interactions with metal complexes of cadmium and copper are explained in the following section.

Schneider and Behrens have extensively used molecular mechanics calculations using ESFF force field to study the organometallic complexes in crystalline silica matrices\textsuperscript{72-74}.
Eventhough the interaction of several transition metal ions with DNA has been the subject of many investigations, binding of copper and cadmium complexes with amino acid like L-Cystine has not attracted much attention by molecular mechanical study. In this report an examination of the interaction of copper and cadmium sulphates with L-Cystine by means of molecular mechanics calculation had been made. Hence, an attempt has been made to rationalize sequence and structural dependent interaction of copper and cadmium sulphates with L-Cystine.

Formation of as many hydrogen bonds as possible between L – Cystine and its complexes without introducing severe distortion in the complex structure was attempted. Both the amino acid complexes understudy were subjected to minimization using ESFF force field with a sigmoidal distance dependent dielectric function $\varepsilon = 4r_{ij}$, which has been demonstrated to be an appropriate implicit treatment of the dielectric function in computing the electrostatic potential of L-Cystine. The energy minimization employed steepest descent followed by conjugate gradient algorithms until the convergence criterian was reached$^{75}$. 
Interaction energies of the L-Cystine copper sulphate and L-Cystine cadmium sulphate can be estimated by calculating the differences between the total energies and the sum of lowest energies found for the optimized structures of the free L-Cystine and copper sulphate and L-Cystine and cadmium sulphate complexes. The negative of the interaction energy is the binding energy\(^76\).

\[
I.E = T.E - (\text{Sum of the individual energies})
\]

\[
B.E = -I.E
\]

Where I.E. = Interaction energy

\[
T.E = \text{Total energy of L - Cystine / complexes}
\]

and B.E = Binding energy

**Molecular simulations**

Molecular simulations can be performed on well characterized systems leading to a better fundamental understanding of atomic level interactions. Accurate force fields can be obtained by defining parameters obtained from \(\text{ab initio}\) initiated calculations and spectroscopic data. This information can then be used to predict useful physical properties of systems that are less well characterised\(^77\).
Potential Energy Surface

In the molecular simulations, the essential part is to know how the potential energy of the molecular system varies with the positions of its constituent atoms. In principle, the potential could be obtained by solving quantum mechanical equations to determine the ground state energy of the electrons and nuclei in the system at each possible set of nuclear positions. The resulting energies form a continuous "Born – Oppenheimer Surface" as a function of nuclear positions and the surface describes the potential energy cost for the kinds of atomic motion considered.

Energy minimization methods adopted

Optimisation of energy can be classified into the following methods. (a). Zeroth order method (b) First order method (c) Second order method.

(a) Zeroth order method – Grid search

The simplest way of locating the minimum of $V(x)$ is to scan the space defined by $X$ in regular increments of $X_1$, regular increments of $X_2$ and so on. For the two dimensional problems, this is equivalent to laying a grid over the surface and picking that grid point at which $V(x)$ is minimum. The
ability of this grid searching procedure to locate the global minimum which depends on the fineness of the grid and the roughness of the surface. An iterative procedure can be adopted by beginning with a relatively coarse grid and using successively finer grids to examine the neighbourhood of the minimum located by the previous cycle. For surfaces with multiple local minima, these methods may not give correct results.\textsuperscript{81}

(b) Steepest descent method

Steepest descent is a conservative algorithm in the sense that it tries to head downhill without crossing local barrier to search for deeper nearby minima. Physically, it makes very limited perturbations to the starting molecular structure\textsuperscript{82}. Since the largest interatomic forces determine the direction of the gradient, steepest descent can eliminate the worst steric conflicts, bring bond length and bond angles to values near their canonical value. However, it will not produce the collective motions that are necessary to generate optimum overall stereochemical structures\textsuperscript{83}. It is a non convergent method and it is particularly inefficient for multidimensional problems that involve irregular potential
surfaces with many local minima, such as those characteristic of macromolecular energy calculations.

**Conjugate gradient**

A generally more efficient descent technique is the conjugate gradient algorithm, whereas steepest descent chooses the descent direction based entirely on the gradient at the current step, conjugate gradient combines information on the current gradient with that based on the gradient at previous steps.

(c) Second order method: Newton Raphson method

The Newton – Raphson method is a second derivative method that is based on the assumption that, in the region of the minimum, the energy depends approximately quadratically on the independent variables. For the one dimensional case, $V(x)$ in the neighbourhood of the minimum is assumed to be of the form.

$$V(x) = a + bx + cx^2$$

Where $a, b$ and $c$ are constants. The first two derivatives are

$$V'(x) = b + 2cx \text{ and } V''(x) = 2c$$
At the minimum \( V'(x^*) = 0 \) and \( x^* \) can be calculated from the first derivative of the potential \( x^* = -b/2c \). Substituting the values of \( b \) and \( c \), the value of \( X^* \) is.

\[
X^* = x - \frac{V'(x)}{V''(x)}
\]

If the system is at the point \( X \) and if the first two derivatives are known at this point, the value \( X^* \) predicts the location of the minimum. The two reasons for not using this method frequently in problems of energy minimization or complex molecules is that, (i) the highly non quadratic character of the energy surface and the presence of multiple local minima render the surface unsuitable for examination by Newton – Raphson algorithm (ii) The potential energy function is expressed in-terms of a mixture of internal coordinates.

Conclusions from simulation analysis

Using the first order method of steepest descent and conjugate gradience, the interaction energies for O – Cu, S–Cd, S–Cu and the Tetramer in kcal / mol were calculated and tabulated as shown in Table (I).
Using the ESFF force field, all these interaction energies were minimized and the internal energy, bond energy, angle energy, torsion energy etc; were also calculated and these are tabulated in Table : (II)
Table : II

Minimised energies using ESFF

(All energies are in k.cal / mol)

<table>
<thead>
<tr>
<th>Type of Energy</th>
<th>N-Cd</th>
<th>N-Cu</th>
<th>O-Cd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Potential Energy</td>
<td>-272.857</td>
<td>-297.735</td>
<td>-252.762</td>
</tr>
<tr>
<td>Internal Energy</td>
<td>15.592</td>
<td>19.040</td>
<td>13.305</td>
</tr>
<tr>
<td>Bond Energy</td>
<td>3.994</td>
<td>4.314</td>
<td>3.391</td>
</tr>
<tr>
<td>Angle Energy</td>
<td>10.422</td>
<td>13.432</td>
<td>9.086</td>
</tr>
<tr>
<td>Torsion Energy</td>
<td>1.176</td>
<td>1.293</td>
<td>0.827</td>
</tr>
<tr>
<td>O/p Energy</td>
<td>0.003</td>
<td>0.000</td>
<td>0.002</td>
</tr>
<tr>
<td>Nonbonded Energy</td>
<td>-288.449</td>
<td>-316.775</td>
<td>-266.068</td>
</tr>
<tr>
<td>vdw Energy</td>
<td>3.118</td>
<td>2.763</td>
<td>0.920</td>
</tr>
<tr>
<td>vdw-dispersion Energy</td>
<td>-18.693</td>
<td>-17.621</td>
<td>-12.735</td>
</tr>
<tr>
<td>Electrostatic Energy</td>
<td>-291.568</td>
<td>-319.538</td>
<td>-266.988</td>
</tr>
</tbody>
</table>

The computer simulation models of L-cystine CuSO₄ and L-cystine CdSO₄ complexes are shown in Figures (1) and (3) respectively. The arrangement of other atoms and molecules with respect to copper and cadmium atoms are shown in figures (2) and (4) respectively.
Fig. 1: Computer simulation model of L-cystine copper sulphate complex.
Fig. 2: Bond names and their arrangement with respect to copper atoms in the L-cystine copper sulphate complex.
Fig. 3: Computer simulation model of L-cystine cadmium sulphate complex
Fig. 4: Bond names and their arrangement with respect to cadmium atoms in the L-cystine cadmium sulphate complex