CHAPTER 2

CRYSTALLOGRAPHY & MOLECULAR MODELLING:
A COMPOSITE UNIQUE METHOD OF CHARACTERISING COMPOUNDS
CRYSTALLOGRAPHY

2.1 INTRODUCTION:
A great many solids have a tendency to arrange their constituent atoms in an ordered and periodic pattern, which is called Crystals. Crystallography is concerned with the enumeration and classification of the actual structures of various crystalline substances. The experimental determination of the structure of crystals is done through x-ray diffraction. But x-ray diffracted by a crystal lattice can not be focussed by a lens to give an image as in the case of light. Experimentally, it is impossible to get a molecular image in the image plane. The diffraction of x-rays by crystals was discovered by Max Von Laue in 1912, and the sequence of events that led to the discovery is one of the most exciting chapters in the history of science. X-ray, an electromagnetic (e.m.) wave of short wavelength (~10^-8 cm), need diffraction spacing of comparable dimension. Von Laue showed that, the atomic spacing in the crystal serves as a three dimensional diffraction grating for x-rays.

This conjecture was soon proved experimentally by Sir Lawrence Bragg. Bragg was the first to give a simple explanation of the observed diffraction pattern, using the principle of interference of the waves and he obtained the well known "Bragg condition"\(2d \sin \theta = n \lambda \), which says that the waves interfere constructively whenever the path difference between them, reflected from adjacent planes, is an integral multiple of x-ray wavelength. The picture that one gets on a photographic plate has several intensity peaks of gradually decreasing magnitude. The pattern on
the photographic plate is called a Laue pattern. The pattern, of course, is a two-dimensional picture of reflections arising from planes in a three-dimensional crystal. The relation between the laue spots and the crystal structure in real space is to be determined by stereographic projection. Crystal is an ordered arrangement of atoms, ions, molecular complexes, characterised by three periodicities $a$, $b$, and $c$, called the lattice parameters (in A). The volume defined by the lattice parameters is called the unit cell. If the atomic distribution within the unit cell is known, the whole crystal is reproduced merely by translation of unit cell. The external form and shape of crystals are the proof of the existence of crystal symmetry.

Crystallography, the science of crystals, was born from the observations of minerals. Geometric laws were established in the eighteenth century. Now-a-days crystallography is involved in many fields, in Solid State Physics, Chemical Physics, Material and Biological Sciences. It provides the most general experimental method for the determination of the structure of molecules, large and small. In 1928, Kathelene Lonsdale showed the flat ring structure of benzene molecule from X-ray diffraction analysis. The structure of penicillin and Vitamin B$_{12}$ which baffled chemists for quite sometime, were determined by Dorothy Hodgkin and her collaborators. Similarly, the description of the double helical structure of DNA by Watson and Crick in 1952 from an interpretation of X-ray photograph is a landmark in the history of crystallography.

Crystallography has now become a science present in all fields. Its importance can be seen in perusal of the number of...
Nobel Prizes in the areas of Chemistry, Physics, Medicine and Physiology, that have been awarded throughout the years for the work in the area.

Besides the study of atomic structure of single crystals by X-ray diffraction techniques, neutron and electron diffraction techniques have assumed increasing importance. Neutron diffraction technique provides information about hydrogen and other light atoms. Non-crystalline and partially crystalline materials can also be studied, with the advent of computers and development of synchrotron as an intense X-radiation source and automatic diffractometers, more complex structures such as proteins have been investigated.

In 1985, a pleasant surprise sprung on the scientists, as a mathematician and a physicist were jointly awarded the Nobel Prize in Chemistry. The mathematician is Prof. Herb Hauptman of Medical Foundation of Buffalo and the physicist is Prof. Jerome Karle of U.S. Naval Research Laboratory. They won the Nobel Prize for using statistical methods to radically speed up the techniques by which X-ray Crystallography can map the structure of molecules (Karle, 1986; Hendrickson, 1986).

The recent trend to probe the mechanism of action or function of a molecule is to relate it with its structure i.e. structure function correlation, which needs accurate determination of the architecture of the molecule. And indeed x-ray crystallography is a powerful tool that unveils the mystery of molecular structure. It provides information on the molecular architecture in space and intermolecular interactions, conformations, molecular packing
and hydrogen bonding that collectively explain the molecular properties.

The conventional method of x-ray crystallography is very complex, time consuming and has a limitation. The achievement of these two scientists is to develop and utilise statistical methods and probability in the x-ray crystallography to understand the molecular architecture. It is really a tremendous mathematical model which makes possible the determination of three dimensional crystal structures. This model, now an indispensable tool for analytical chemists, is full of potentialities and throws open new avenues in the field of chemistry.

To comprehend the significance of Hauptman and Karle’s work, a basic knowledge about the determination of crystal structures employing the technology of x-ray crystallography is necessary. It was not developed overnight. Talents like Max Von Laue, W.L.Bragg, P. P. Ewald, K.Lonsdale, K.Banerjee, Perutz, Kendrew, Hodgkin, M. M. Woolfson, H. Lipson, R. W. James have enriched this field with their illustrious contributions and have carried the subject to its present perfection. In brief, the core of the technology is as follows.

The aim of crystal structure analysis is to calculate, using a mathematical treatment, the atomic positions of molecules from its diffraction pattern and intensity data. In this method, one needs primarily to undergo and understand the following affairs in a synergestic manner.

a) X-ray source (x-ray generator, Filter etc.),

b) A good single crystal of the compound concerned with suitable
size,
c) Data Collection (X-ray Cameras, Diffractometer, Area detector, Image plate etc).
d) Decoding of data to structure (data processing, structure solution, refinement of structure etc.).

2.1.1 CRYSTALLIZATION:

The prime thing in X-ray Crystallography is to get a good quality single crystal which can diffract. The essential feature of a crystal is its ordered and three dimensionally periodic internal structure. The crystallization of molecules from solution is a reversible equilibrium phenomenon, and the specific kinetic and thermodynamic parameters will depend on the chemical and physical properties of the solvent and solute involved (McPherson, 1982).

Crystals are formed in a super saturated solution under suitable condition, if the rate of precipitation is controlled. It is a very slow process from fluid state to solid state.

\[ \text{Fluid } \longleftrightarrow \text{ Solid} \]

The nucleation is controlled by some properties of the solute, like dielectric constant, polar character, ability to form hydrogen bond, entropy driven parameter hydrophobicity etc. The crystallization of protein is affected by temperature also. The rate of nucleation decrease with the increase of temperature at high ionic strength.

The nucleation depends on the foreign particles present in the solution. The fluid to solid transition should be started from a point which is lower than its saturation point. The aim is to
grow few crystals with a large size.
The more often used techniques are Vapour diffusion, hanging drop method, dialysis in microbutton, batch crystallization etc.

**Vapour diffusion technique** is widely used for crystallization of protein and organic macro-molecules. To control the rate of nucleation of protein, sometime preceptant like NH₄SO₄ is used.
The compound to be crystallized is taken in the solution of precipitation with a density lower than the optimum density for crystallization. This is taken in a test tube. The tube is placed in a closed container with the precipitant of higher density than optimum. Slow vapourisation occurs from test tube to the container. When optimum density is reached in test tube, the crystal grows.

**Hanging drop technique** is used for the sample which is available in very small quality. In this method, the sample with precipitant is hanged as a droplet from the lid of the container containing precipitant of higher density.
The optimum density for crystallization is in between the density of droplet & the solution in the container. Actually the technique to grow a crystal is more art than science. The optimum condition for crystallization is achieved by trial and error experimentation.

**Selection of Single Crystal:**

As soon as crystals are found, the logical next step is selection of a suitable single crystal.

The crystal is selected under a polarising microscope. The microscope contains an analyser which is situated above the stage and a polarizer below the stage, their axis being perpendicular
to each other. During the complete rotation of the crystal under the microscope, if the crystal shows extinction four times, then it will be a single crystal. Each extinction occurs 90° apart from another.

At a position 45° to their axis, the crystal will show maximum brightness. This uniform behavior is not observed for twinned crystals or crystals other than single.

The dimension of the crystal should not be too large or too small. There is an optimum size of dimension regarding the absorption co-efficient of the crystal. A large crystal absorb large amount of radiation whereas a tiny crystal provide weak reflection. Thus the optimum size should be of the order of 1/μ A where μ is the absorption co-efficient.

2.1.2 Decoding of X-ray diffraction data:

In the X-ray diffraction experiment using photographic films or diffractometer, we can measure the diffracted intensities I(hkl), where the tripple integer index h,k,l designates uniquely each observation. The scattering electron density, ρ(xyz), being a periodic function, can be represented by a Fourier series,

$$\rho(xyz) = \frac{1}{V} \sum \sum \sum F(hkl) \exp(-2\pi i (hx + ky + lz)) \quad (1)$$

where x,y,z are fractional co-ordinates and F(hkl) is a Fourier coefficient called structure factor. If we consider the Fourier Transform of equation (1), we will get structure factor,

$$F(hkl) = \int \rho(xyz) \exp(2\pi i (hx + ky + lz)) \, dV \quad (2)$$

Another function called Patterson function is defined as,

$$P(uvw) = \frac{2}{V} \sum \sum \sum F(hkl)^2 \cos(2\pi (hu + kv + lw)) \quad (3)$$

where u, v, w are the position vectors.
Thus we have introduced three important functions of diffraction theory, of which two, \( p(\text{xyz}) \) and \( P(\text{uvw}) \) are in real space and the other one \( F(\text{hkl}) \) is in reciprocal space.

One of the major problems in crystal structure determination from the above consideration is that one requires, in addition to the measured intensities, the phase of the diffracted waves. The x-rays will be scattered differently by planes of different atoms at different angles. This leads to slight differences in phase for the diffracted x-rays. As a result of this the intensity of reflected waves is altered and the spots in different parts of the pattern vary in intensity. All that we can measure are the intensities of the spots---intensity being proportional to the square of amplitude—but not the phases. All the information about the phases is lost. This is the well-known "phase problem" in crystallography, a fascinating subject in itself; understanding of the subject requires physical and chemical intuition and mathematical expertise.

Different methodologies are developed to evaluate the phases, since it plays a key role in the crystallography, without which structure of the molecule cannot be solved.

There are basically four procedures to solve the problem:

1) Direct method, 2) Patterson method, 3) Isomorphous method and 4) Anomalous dispersion method.

Among the four methods, the first two methods are widely used for small molecular structure determination and we will discuss these two procedures later.

Once the phase angles are obtained, a scattering density map is computed by summing up the Fourier series in equation (1), and an
approximate set of atomic coordinates is obtained by interpreting this map. Structure factors can be computed using the atomic parameters and approximately chosen thermal vibration parameters.

\[ F(hkl)_{\text{cal}} = \sum f_j \exp(2\pi i (hx_j + ky_j + lz_j)) \exp(-B_j \sin^2 \theta / \lambda^2) \]  

where \( \theta \) is the scattering angle, \( f_j \) and \( B_j \) are the atomic scattering factor and isotropic temperature factor respectively.

The equation (4) is the physical model in which we use the correction factor concerning the vibrational motion of an atom about its equilibrium position. At absolute zero, we may consider the thermal vibration of the atom is ceased and the nuclei is at the rest. But as the temperature increases, the ambient thermal energy causes the atoms to oscillate about their equilibrium position. The atomic vibration can be described in anisotropic form by using six parameters per atom in place of one parameter \( B_j \).

All the parameters along with a scale factor \( K \) can be optimized by the method of Least Squares, where the quantity \( V \), given by,

\[ V = \sum w(hkl) [ |F(hkl)_{\text{obs}}| - |F(hkl)_{\text{cal}}| ]^2 \]  

is minimised. \( w(hkl) \) are weights assigned to the individual observation. A measure of the disagreement between the observed and the calculated structure amplitude is the index \( R \), given by,

\[ R = \sum (|F(hkl)_{\text{obs}}| - |F(hkl)_{\text{cal}}|) / \sum |F(hkl)_{\text{obs}}| \]  

For well refined structures, using experimental data, free of systematic errors, the value of \( R \) is expected to be very small. At the end of the refinement, with reasonably low \( R \) value, the complete set of atomic coordinates is obtained which provides a
complete three dimensional picture of the molecule under investigation.

As mentioned before, usually for small molecular structure solution, the primary choices are Direct method & Patterson method. The essentials of these two important methods are discussed below.

- **Direct Method:**
The 'Direct Methods' in structure analysis are those which attempt to solve the phase problem with no recourse to conventional structural chemical information but radically different approach based on an elaborate statistical correlation of reflection phases-this is the great contribution of Karle and Hauptman (Karle, 1986). These allow highly educated guesses to be made of the phase of reflected x-ray so that a molecular image can be constructed (computed) directly from the diffraction pattern.

In 1953, Hauptman and Karle published a monograph on the "the centrosymmetric crystals" which, together with their 1956 paper in Acta Crystallographica remains the most important landmark in the phase problem (Hauptman et al, 1956). The method is probabilistic in nature and uses rigorous mathematical reasoning. As the electron density no where in the crystal can be negative, the observed amplitude and the corresponding phases are constrained. The positivity requirement was very elegantly expressed by Karle and Hauptman.

These two scientists are not the first to employ the direct method but their work has led to the automation of x-ray crystallography so that most of the structure analysis is now
routine work. And in this method, heavy atom is not needed, so the Chemists are in better position to quickly identify a chemical compound in spatial configuration. And now this method is applicable to molecules containing up to 200 atoms which includes important biological molecules such as steriods, hormones, polypeptides, membrane bound macromolecules and many other drugs.

But the method is not still unambigously powerful enough to apply for biological macromolecules such as proteins. Hauptman is now thinking of extending his approach to the analysis of much larger biological molecules, such as proteins. If the three dimensional analysis of protein structure can be simplified, the basis for biological functioning will be approachable in specific, mechanical terms. Then that would really be a revolutionary breakthrough in the molecular biology. Karle's present idea is to produce molecular assemblies in crystals and different orientations of those assemblies of complex biomolecules in real space.

A systematic account of the determination of the direct methods is beyond the scope of this brief review. Hence certain basic principles together with some working formulas and outline of the procedure of phase determination is stated here. The basis of the whole of present-day direct methods are the two statements, that in a crystal (1) the electron density is positive everywhere (Harker & Kasper 1948 and Karle & Hauptman, 1950) (2) the electron density consists of discrete spherically symmetric atoms.
Direct methods try to evaluate phases directly from the measured diffraction intensities $I_{hkl}$ by using relationship among the phases, whose values are based solely on the intensities.

Roughly one can say that, since the crystal structure can be described by a limited number of parameters (the positions of the atoms) and since many more intensities can be measured, relationships will exist among $F_{hkl}$ (structure factor) and $\phi_{hkl}$ (phases).

Therefore, the first goal of the direct methods is to identify as many of these relationships as possible. In the next step the origin is fixed by specifying the phases of a few suitable reflections numerically. In generalised form we can write down logical steps which will enable one to determine the origin defining reflections for any space group.

The logical steps are:

1) Separate reflections into categories according to the effect on their phases of shifting the origin to all the space group "allowed" position.
2) Plot these categories on a 'reduced reciprocal lattice'.
3) Choose a linearly independent set of three vectors i.e. non-coplanar which together define the primitive unit cell in the reduced reciprocal lattice.
4) Any set of three non-coplanar vectors will therefore give us the three parity groups from which we can choose the origin defining reflections. Then, using the phase relationships new phases are calculated. In general, however, in this way, it will not be possible to phase all strong reflections and hence a few more starting reflections are selected, which act as unknowns
(symbols, ambiguities) in the above described process of phase extension. This process now develops like a snow ball, provided a good choice of origin defining reflections and unknowns have been made. Finally, when most of the strong reflections have got a phase, the numerical values of the unknowns are evaluated and then, using the expression (7) below, an image of the structure is produced.

\[ \rho(x, y, z) = \frac{1}{V} \sum \sum |F_{hkl}| \cos \left[ 2\pi(hx + ky + lz) + \phi_{hkl} \right] \quad -- \ (7) \]

**Structure solution strategy using Direct method:**

In practice these analytical methods of phase determination are carried out on "normalised structure factors" - that is, values of the structure factors, \(|F|\), modified to remove the fall-off in the individual scattering factors, \(f\), with increasing scattering angle, \(\Theta\).

A normalised structure factor, \(E\), represents the ratio of a structure factor, \(F\), to \((\Sigma f_i)^{1/2}\) where the sum is taken over all atoms in the unit cell at the value of \(\sin\Theta/\lambda\) appropriate to the \((hkl)\) values for the reflection includes an overall vibration factor. This use of \(F\) values is approximately equivalent to considering each atom to be a point atom (an extremely sharp peak occupying almost no volume in the electron density map).

From the distribution of \(|E|\) values we get an information on whether the structure is centrosymmetric or noncentrosymmetric. Once a table of \(|E|\) values has been prepared it is usual to rank these \(E\) values in decreasing order of magnitude and work with the strongest 10 percent or so. Then one chooses groups of three reflections that satisfy the conditions.
\[ S(h,k,l) \cdot S(h_1,k_1,l_1) \approx S(h+h_1,k+k_1,l+l_1) \quad (8) \]

where \( S \) means "the sign of" and \((h,k,l), (h_1,k_1,l_1)\) and \((h+h_1, k+k_1, l+l_1)\) are all specific strong reflections. All of these relations are statistical probabilities rather than exact equations as implied by the use of \( \approx \) rather than sign of equality (=). This is called Sayre's equation (Sayre, 1952). The selection of triple products \([E(hkl), E(h_1,k_1,l_1), E(h+h_1, k+k_1, l+l_1)]\) is made by computer. Since each of the three reflections in a triple product has a high \( E \) value, the product of their signs is probably positive. This listing is called the "\( \Sigma_2 \)" listing; \( \Sigma \) is used because, in the probability formula, summations are involved. The "\( \Sigma_1 \)" relations are simpler because they involved only pairs of reflections related by \( E(hkl) \) and \( E(2h,2k,2l) \) and contain the implication that the sign of \( E(2h,2k,2l) \) is probably positive in a centrosymmetric structure.

The probability aspects in these sign relationships are very important. The probability that a triple product is positive & can be written as

\[ P_+ = \frac{1}{2} + \frac{1}{2} \tanh \left( \frac{|E_{hkl} E_{h_1 k_1 l_1} E_{h-h_1, k-k_1, l-l_1}|}{N} \right) \quad (9) \]

\[ = \frac{1}{2} + \frac{1}{2} \tanh \left( \frac{|E_H E_K E_{H-K}|}{N^2} \right) \quad (9) \]

where \( N \) is the no. of atoms in the unit cell, \( H \equiv h,k,l \), and \( K \equiv h_1, k_1, l_1 \).

In the final stage, \( E \) maps are calculated using the sets of phases as indicated by the figures of merit. The complete interpretation of the map is done in three stages.

(a) peak search (b) separation of peaks into potentially bonded clusters, (c) application of simple stereochemical criteria to identify possible molecular fragments.
The stages (a) and (b) are very similar to those described by Declercq, Germain, Main & Woolfson (1973) and produces a list of peak coordinates. Stage (c) is described by Main & Hull (1978). Here, stereochemical criteria of maximum and minimum bond lengths & angles are applied to the E-map peaks to identify chemically sensible molecular fragments. Peaks which do not fulfil the criteria are eliminated in such a way as to maximise the number of high peaks which are accepted.

Sometimes only part of the structure is revealed in an interpretable way and the rest may be found from successive difference fourier maps.

**Patterson Method:**

A significant attempt to obtain structural information from the measured intensities was made by A. L. Patterson et al, in 1934 & 1935. He developed a Fourier series which has at its coefficients the magnitude of the square of the structure factors rather than the structure factor themselves. The phases may then be eliminated from this equation, $F_h = \int p(r) \exp(2\pi n h \cdot r) \, dV$ by multiplying with its complex conjugate to obtain,

$$|F_h|^2 = F_h F^*_h \quad \ldots \ldots \quad (10)$$

The fourier transform of this equation (10) is known as the Patterson function.

$$P(u) = \int F_h F^*_h \exp(-2\pi n h \cdot u) \, dV^* \quad \ldots \ldots \quad (11)$$

Note that the argument of $P$ is usually denoted by a 'u' and not 'r' although it is a vector of physical space as is the argument 'r' of $p(r)$, but there is a geometric difference between these two quantities. So Patterson function is the convolution of the square of the electron density $p(r)$ and expressed as,
\[ P(u) = \int p(r) p(u-r) dV \] .......................... (12)

All the procedures, called 'heavy atom' or 'Patterson' methods are based on the equation (12).

The highlight features of Patterson methods are:

a) \(|F_h|\) is a phaseless quantity and Patterson function \(P(u)\) is always centrosymmetric even when \(p(r)\) is not.

b) For 'n' point atoms with atomic weights equal to atomic numbers, the Patterson function consists of peaks corresponding to all possible interatomic vectors within the unit cell. The strength of each Patterson peak is proportional to the product of atomic numbers of the atoms connected by the vector \(u\) multiplied by \(M\), where \(M\) is the multiplicity number.

c) The number of peaks in Patterson map is much than that in \(p(r)\). This makes the study of Patterson map difficult. Patterson map has \(N(N-1)\) peaks (other than that at the origin) instead of \(N\) peaks in \(p(r)\) map. Each Patterson peak is roughly twice as wide as an electron density peak. In real structures, the atoms are not points, consequently there will be overlapping of peaks. To reduce the degree of overlap, \(|E_h|\) or \(|E_hF_h|\) are used as co-efficient in the \(P(u)\). The large origin peak which may hide some short of vectors, is eliminated by using the co-efficient \(|F_h'|^2 = |F_h|^2 - \Sigma f_j^2\) where \(|F_h|^2\) must be on absolute scale.

d) The map will have prominent maximum when, (i) the structure contains a limited number of heavy atoms \([Z_p > Z_1\) (atomic number of lighter atoms)] giving rise to 3 types of peaks, (ii) If the molecular geometry of the structure is such that it will give rise to several interatomic vectors with almost same magnitude.
and direction (e.g. systems with condensed aromatic rings); these correspond to appreciable value for the multiplicity factors (M).

In the Patterson map, the symmetry of the crystal is reflected into the vector map. All the vectors are translated to a common origin. The point group symmetry remain, but translational symmetry disappear, -- the screw axes become rotation axes and glide planes become mirror planes. As a result of this, it has been shown that space groups drop from 230 to only 24 space groups (centering remains unaltered). However the symmetry operators present in the crystal leave a trace in the Patterson map. It was pointed out by Harker (Harker et al, 1948 ) that the symmetry of the Patterson map often arises in a highly systematic manner which of itself can give useful information. Specially, he showed that certain symmetry elements within the real crystal are associated with planes or lines within the Patterson map where there will be local concentrations of peaks. These particular clustering of vector maximum on specific lines or planes of the map called Harker lines or Harker planes. And identification (presence or absence of Harker lines or planes ) of these is often very useful in giving us further symmetry information in addition to that which we derive from the symmetry of the diffraction pattern and the recognition of systematic absences.

These vectors contain one or two constant components and these vectors are between the equivalent atoms related by symmetry element other than the centre of inversion.

e.g for \( P2_1 \) space group \( u = 2x, v = 1/2, w = 2z \)

\( P2 \quad \) 

\( Pm \quad \) 

\( Pc \quad \) 

\( u = 0, \quad v = 2y, \quad w = 0 \)

\( u = 0, \quad v = 2y, \quad w = 1/2 \)
It has been used as a guideline in the selection of a heavy atom as.

\[ \frac{\sum z^2_{\text{heavy}}}{\sum z^2_{\text{light}}} \sim 1 \]

This relation is merely a guide, however, and large deviation from it can often be tolerated (Sim, 1957 and Parthasarathy, 1965). When the ratio \( < 1 \), the interpretation of \( P(u) \) and process of completing the structure is more difficult but the accuracy of the position of the light atoms increase. But if the ratio \( > 1 \) e.g. 2.07, the structure is easily solved but the accuracy is low.

Once the heavy atoms and hopefully, a few light atoms are located, the complete model of the structure may be built up by successive Fourier synthesis. Hence it is possible to identify the many peaks in the Patterson map as a set of interatomic vectors, and so a model structure may be proposed. But as the size of the molecule under study increases, the Patterson map becomes much more complicated and indistinct, and certainly as regards biopolymers, it is quite impossible to deduce detailed structural information from a Patterson map alone. Nevertheless, this is vital for the success of all methods of solving the phase problem.

For solving crystal structure by Direct method and Patterson method, we mainly use Multan78 (Main, 1978) and Shelx86 (Sheldrick, 1986) in MicroVaxII and PC/AT 386 in our department.

2.1.3 Improvement of the model:

The next task is to improve the computer derived structural model's atomic positions (if chemically meaningful) against actual observed data. In the small molecular structures, their improvement is achieved by Fourier analysis and least squares
techniques of refinement followed by reliable index check,

\[ R = \frac{\sum |F_{ob} - |F_{cal}|)}{\sum |F_{ob}|} \]

where \( F_{ob} \) is the observed structure factor and \( F_{cal} \) is the calculated structure factor. The calculations are done by SHELX76 program (Sheldrick, 1976) in MICROVAX II computer. Once the refined structure of reasonable low R value is at hand, we can envision the details of the molecule, bond lengths, bond angles, torsion angles and inter and intramolecular interactions stabilising the molecular structure in space. For this calculations, we use PARST88 (Nardelli, 1983) in MICROVAX II.

**MOLECULAR MODELLING:**

2.2 INTRODUCTION:

Molecular modelling has emerged as a basic tool in characterising molecular structures in detail. The effective programmes that are associated with molecular modelling are surprising. They can build a molecule, minimise energies, design and develop new molecules of biological importance and last but not the least, this can open up a vision in molecular assembly, structure and its functionality. A brief account of different aspects & strategies of molecular modelling is given since we have used molecular modelling at various stages of our structure-function interpretation.

Molecular mechanics methods are based on pragmatic view of the molecular structure that is considered as a set of balls and springs with series of potential energy functions expressing the
molecular force field. Molecular mechanics of energy minimisation involves successive iterative computations, where an initial conformation is submitted to full geometry optimisation. The goal is to reach a local minimum on the potential surface within the minimum amount of time. The more sophisticated methods use the first and occasionally the second derivatives of the energy function for guiding the minimisation. No method can guarantee of finding the absolute lowest energy structure - the global minimum. Energy minimisation encountered, without realizing that much deeper, more stable minima may be accessible. Molecular dynamics method is able to climb small barriers and is therefore much more efficient at locating deep local minima than simple minimization.

Recent developments in computer hardware and software have greatly facilitated molecular modelling on Computer Work-stations and the technique is now used routinely in the medicinal chemistry environment. Molecular modelling systems provide powerful tools for building, visualizing, analysing and storing models of complex molecular systems that can help to interpret structure-activity relationships. Certainly the goal of molecular modelling should not be limited only to providing insight, but it should also help to suggest new experiments i.e., new structures tailored to have the desired biological activity.

A significant branch of chemical research has for many years been devoted to the calculation, a priori, the molecular energies. Methods have been developed which allow preparation of a reasonably accurate estimate of the energy of a molecule, once its precise structure is known.
With the advancement of computer, the task is made easier and the result is that the molecular modelling has taken an essential part in Medicinal Chemistry, complementary to x-ray crystallography. But Molecular modelling can not yet produce quantitative predictions of activity except in very special cases, but it can provide valuable qualitative guidelines that help to design new structure.

2.2.1 Potential Energy Functions:

The first step in Molecular mechanics calculation is to construct a potential energy surface, the energy of the system as a function of atomic co-ordinates. The potential energy of a system can be used to determine the forces acting on the atoms of the system and the first derivative of the potential energy w.r.t atomic co-ordinates gives the forces acting on each atom of a macromolecule. These forces can be used to determine dynamical properties of the system by solving Newton's equations of motion to describe how the atomic positions change with time. The common features of most of the potential functions are a harmonic restoring force between bonded nearest neighbours, a penalty for deforming the angle between three neighbouring atoms, a dihedral torsional and improper torsional potential to allow for the hindered rotation of groups about a bond, and non-bonded interactions between separated atoms. Non bonded interactions are common to both intermolecular and intramolecular potentials, whereas the other terms are strictly intramolecular (Brooks et al, 1983).

Thus empirical energy function is made up of a sum of many terms.
The Total energy is expressed in the form:

$$E_T = (E_B + E_\Theta + E_\Phi + E_w) + (E_{vdw} + E_{el} + E_{hb}) + (E_{cr} + E_{c\Phi})$$

- **Internal energy terms:**
  - Bond potential: $E_b = \sum k_b (r - r_0)^2$
  - Bond angle potential: $E_\Theta = \sum k_\Theta (\Theta - \Theta_0)^2$
  - Dihedral angle potential: $E_\Phi = \sum k_\Phi - K_\Phi \cos(n\Phi), n=1,2,3...$
  - Improper torsion potential: $E_w = \sum k_w (W - W_0)^2$

The force constants ($k_b, k_\Theta, k_\Phi, k_w$) and geometric constants ($r_0, \Theta_0, n, W_0$) are selected from parameter file supplied by the program.

- **Non-bonded energy terms:**

  The important non-bonded interactions between atoms consists of the Van der waals term and electrostatic term.

  Van der waal energy is the combination of two types of interactions. First one is the repulsive forces arise from a combination of internuclear repulsions and the Pauli exclusion principle and is expressed by an inverse 12th power

  $$E_{rep} = \frac{A_{ij}}{r_{ij}^{12}}, \text{ where } A_{ij} \text{ is an adjustable parameter.}$$

  The second one is the attractive dispersion force called London force (London, 1937) arises from the small fluctuation of charge distribution of an atom in the presence of another atom. These fluctuations give rise to an attractive dipole-dipole interaction that London first showed decreases with the inverse sixth power of the separation distance $r_{ij}$.

  $$E_{att} = \frac{-B_{ij}}{r_{ij}^6}, \text{ the parameter } B_{ij} \text{ is a function of } r_{ij}^6,$$
atomic polarizabilities and ionization potentials of the interacting atoms (Pitzer, 1959).

Combining the two potential functions, we arrive at most familiar expressions known as Lennard-Jones 6-12 potential

$$E_{vdw} = \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right)$$

The other contribution of non-bonded interactions arises from the effective partial charges that reside on the atoms. When atoms are bonded to each other, the electronic charges flow from less electronegative atom to the more electronegative atom until a balance is achieved. The arrangement of charge distribution along the chemical bond is assigned by calculating partial charges of the atoms in the molecule.

Thus the interaction energy of two atoms i & j due to their partial charges $q_i$ & $q_j$ in dielectric medium of $\varepsilon$ is expressed by coulomb expression;

$$E_{el} = \frac{q_i q_j}{\epsilon r_{ij}} \text{ where } r_{ij} \text{ is the distance between two atoms.}$$

**Hydrogen bond energy**: Hydrogen bond energy is the energy of formation of a bond between two electronegative atoms D and A of the type D-H...A-A where AA, A,H and D is called acceptor antecedent, acceptor, hydrogen and donor.

$$E_{hb} = \left( \frac{A'}{r_{12}^{12}} - \frac{B'}{r_{10}^{10}} \right) f(\theta_1) f(\theta_2)$$

where $r$ is the distance between acceptor (A) and donor (D), $\theta_1$ is the angle between acceptor (A), hydrogen (H) & donor (D) and $\theta_2$ is the angle between acceptor antecedent(AA), acceptor(A) and donor (D). A' and B' are constants. When $\theta_1 \leq 90^\circ$, $f(\theta_1)=0$ & hence the total term is zero. Previously only 10-12 exponent of distance function was used (McGuire et al, 1972). But later, abinitio calculations indicated the dependence of H-bond energy
on the angle parameter and Brooks et.al (Brooks et al, 1983) gave the generalised expression for H-bond potential energy which is now widely used for molecular mechanics & molecular dynamics calculation of macromolecules.

Constraint energies are due to atom harmonics and dihedral constraints.

2.2.2 Energy minimization:

For a potential energy function, it is often desirable to find minimum energy configurations of a system. Minimization is performed to relieve strain (Brooks et al, 1988). For macromolecular systems, the number of local minima and the cost of the computations prevent exhaustive search of the energy surface, so it is frequently impossible to determine the global energy minimum, generally, a local minimum in neighborhood of the X-ray structures, if available is examined. There are different iterative minimization algorithms. We have applied two most commonly used first derivative approaches—(1) Steepest descent (SD) (Levitt et al, 1969) and (2) Conjugate Gradient (CG) (Fletcher et al, 1964).

In the Steepest Descent (SD) method, a displacement opposite to the potential energy gradient (i.e. in the direction of the force) is added to the co-ordinates at each step. This can be written,

$$\delta n = -K_n(\nabla n E_T),$$

and in this iterative approach, the positional vector (expressed in cartesion co-ordinates) may be symbolised by

$$r_{n+1} = r_n + \delta n$$

where the subscripts refer to the number of iterate and \(\delta\) is the
displacement along the direction of force. Kn is a parameter that adjusts step size to take account of the fact that energy may increase, as well as decrease, after a step is taken. If the energy decreases, Kn is increased for the next step, while if the energy increases, presumably because the step size was too large, Kn is decreased. The disadvantages of this method is that it suffers poor convergence, but this method is very useful to remove bad contacts by small change in positional co-ordinates. It is not an efficient method to locate minima for a complex potential energy surface that characterize most macromolecules.

A second method is the Conjugate Gradient (CG) technique which has better convergence characteristics based on first derivative information (Fletcher, 1964). The method makes use of the previous history of minimization steps as well as the current gradient to the next step. In addition, the step size $\delta_n$ is modulated by a parameter $\sigma$, which is chosen to give the optimal step. The parameter $\sigma$ is determined by simple line search, which requires a few extra energy evaluations per step. Thus conjugate Gradient algorithm can be written in terms of the parameters $\delta_n$ and $\sigma$ in the form,

$$\delta_n = -g_n + \delta_{n-1} \cdot g_{n-1}^2/g_{n-1}^2 & r_{n+1} = r_n + \sigma \delta_n$$

where $g_n = \nabla_n E_T$ denotes the gradient vector for nth co-ordinate set. The above equation shows that CG technique makes a given step of a linear combination of the current gradient and the previous step. For a N-dimensional quadratic surface, the CG method reaches the minimum in the order of N-steps. When the algorithm is found to be making little progress on a non quadratic surface, it may be reinitialised by setting the
contribution from previous steps $g_n$ to zero and continuing. The algorithm usually requires more energy evaluations per step than SD algorithm and it converges more rapidly and often produces a substantially lower energy when it converged.

For the calculation of potential energy functions and subsequent energy minimization, we have used Charmm version 2.1 (Brooks et al, 1983).

REFERENCES:


