CHAPTER I

Section 1.1 - Introduction to single crystal X-ray diffraction

Section 1.2 – Introduction of hydrogen bonding
Section 1.1
Introduction to single crystal X-ray diffraction

Historical background
The study of the theory of crystallography has caught the attention of many scientists since long time. The first definite contribution was from Kepler in the year 1611. Stensen pointed out the existence of the characteristic interfacial angles in the crystals which later lead to the Miller indexing of the planes and subsequently the classification of crystals under 32 classes independently by Hessel and Godolin. Neumann showed that the laws of symmetry that hold for the external faces hold good even for the properties exhibited by the crystals. The contribution of Huygens and Hauy considered the building blocks of the crystals. Based on the atomic theory of Seeber, Bravais proved the existence of 14 lattices which are currently named after him. Later investigations using group theoretical approach by Fedorov and Schoenflies proved the presence of 230 space groups. Max von Laue worked out the theory of gratings with double periodicity as present in crossed gratings. As the science of optics extended beyond the visible spectrum into the domain of very short waves like X-rays, which were discovered by Rontegen, the fine structure of crystals was demonstrated by Friedrich and Knipping in 1912. This was the starting point of X-ray Crystallography which saw a rapid development of theoretical ideas and techniques to unravel the structures of the wide range of crystalline substances, both natural and synthetic.

Diffraction of X-rays by Crystals
Crystal is a homogeneous solid (Buerger, 1956a; 1956b) and is defined as three dimensional periodic arrangement of atoms or molecules. Within a crystal, the atoms or molecules are arranged in an orderly manner. Such periodically repeating motif in three-dimensions forms a natural grating for diffraction of waves having suitable wavelength. Max von Laue (Laue, 1912) proposed that X-rays are electromagnetic waves and these are diffracted by a crystal. Later, Friedrich and Knipping (Friedrich et al., 1912) showed the diffraction of X-rays from single crystals of coppersulphate pentahydrate.
The periodic repetition of motif or basis is represented by three shortest non-coplanar vectors (James, 1948; Buerger, 1956; 1970) a, b and c are called primitive vectors. The parallelepiped generated by these vectors is called unit cell or primitive cell. All possible linear combinations of these three unit vectors generate an infinite array of discrete points in space referred to as lattice of the crystal. The position of a lattice point is represented by a vector,

\[ \vec{r} = la + mb + nc \]

Where l, m and n are integers. Later, W.L. Bragg and W.H. Bragg treating diffraction as reflection from planes in the lattice, deduced the simple equation,

\[ 2d_{hkl}\sin \theta = n\lambda \]

known as Bragg’s law, where \( \theta \) is the glancing angle for incident X-rays of wavelength \( \lambda \). \( d_{hkl} \) is the interplanar spacing of planes characterized by the Miller indices hkl and n is order of diffraction.

**Reciprocal Lattice**

A reciprocal lattice can be associated with every real crystalline lattice. For every plane in the direct lattice, draw a normal from the origin of the unit cell whose length is restricted to the reciprocal of the interplanar spacing. A point is placed at the end of each limited normal and the collection of such points constitutes a new lattice called the reciprocal lattice. The concept of reciprocal lattice is very helpful in explaining the X-ray diffraction phenomenon.

If k and k’ (|k| = |k’| = \( 2\pi/\lambda \)) are incident and diffracted wave vectors, then diffraction maxima would occur only when the following conditions called Laue equations are satisfied simultaneously.

a. \( \Delta k = 2\pi h \)

b. \( \Delta k = 2\pi k \)

c. \( \Delta k = 2\pi l \)

The vector \( \Delta k = k’ - k \), called as the scattering vector determines the direction of maxima. The solution of the Laue equations turns out to be

\[ \Delta k = 2\pi G_{hkl} \]

where \( G_{hkl} = ha^* + kb^* + lc^* \) is called reciprocal lattice vector, \( a^*, b^*, c^* \) are shortest vectors in the reciprocal space and satisfy the conditions.

\[ aa^* = bb^* = cc^* = 1, \ ab^* = ac^* = ba^* = bc^* = ca^* = cb^* = 0 \]
The solution of the Laue equations implies that the diffraction pattern will occur in reciprocal space or Fourier space.

There is an intimate connection between $G_{hkl}$ which occurs in the solution of the Laue equations and the lattice planes in real space with Miller indices $hkl$.

- $G_{hkl}$ will always be perpendicular to a set of plane of the real lattice with Miller indices $hkl$.
- The magnitude of reciprocal lattice vector $G_{hkl}$ will be equal to reciprocal of the inter planar spacing with Miller indices $hkl$.

When crystal is rotated about incident X-ray beam it can be assumed that the reciprocal lattice associated with it also rotated. Hence the Bragg’s law can be rearranged as

$$\sin \theta = \frac{1/d_{hkl}}{2\lambda}$$

P P Ewald gave a geometrical interpretation of Bragg’s law. Consider a crystal at the center of a sphere referred to as Ewald sphere of radius proportional to $1/\lambda$. CP is the direction of diffracted wave vector making an angle $2\theta$ with incident beam in the direction ICO, then

$$\sin \theta = \frac{\Delta k}{2|k|}$$

$$\sin \theta = \frac{G_{hkl}}{2\lambda}$$
Thus diffraction maxima can occur only when $\Delta k$ is a chord of Ewald sphere. It is clear that when the reciprocal lattice point happens to lie on the surface of the Ewald sphere, reflection from the planes corresponding to that reciprocal lattice point will occur.

**Selection of single crystals**

The specimens suitable for X-ray diffraction were selected by examining the crystals under a polarizing microscope, so that crystals free from defects such as twinning, cracks, etc.

**Diffraction amplitude: Structure factor**

X-ray are scattered only by electrons of the atoms in the crystal. To locate the position of an atom it is necessary to calculate the electron density function $\rho(\text{xyz})$ in the unit cell. Thus diffraction amplitude due to the contents of the unit cell would be the Fourier transform of $\rho(\text{xyz})$, obtained as

$$F_{hkl} = \sum_{j=1}^{N} f_{j} e^{2\pi i (hx_{j} + by_{j} + cz_{j})}$$

where, $N$ is total number of atoms in the unit cell, $f_{j}$ is the atomic form factor of jth atom whose fractional coordinates are $x_{j}$, $y_{j}$, $z_{j}$ with respect to origin of the unit cell. The form factor determines the scattering power of a given atom with respect to a single electron. $F_{hkl}$ is the structure factor and it is a complex quantity. The modulus of $F_{hkl}$ gives the amplitude of diffraction maximum.

---

**Data collection flow chart**

![Data collection flow chart](chart.png)
It can also be expressed in terms of electron density function $\rho(xyz)$ in the unit cell as

$$F_{hk} = V \int \int \int \rho(xyz)e^{2\pi i(hx+ly+hz)}dx\,dy\,dz$$

where $V$ is volume of the unit cell.

**Collection of intensity data**

The prime aim of the crystallographers is to know the positions of the atoms within a unit cell. To achieve this, we measure the intensities of the reflections from a given crystal. The measure of the total number of photons which are diffracted in the proper direction by reciprocal lattice point is known as integrated intensity. The intensity data was collected by $\omega$ or $\omega - 2\theta$ scan. From these intensities, the electron density distribution in the crystal cell can be deduced.

While deriving the structure factor $F_{hk}$, certain physical and geometrical factors (Buerger, 1960) affecting the diffracted intensities. The conversion of raw intensity into corrected structure amplitude is known as **data reduction**. Factors affecting the intensity are mainly, polarization factor, Lorentz factor, absorption, extinction and temperature factor.

**The polarization factor**

The incident X-ray beam is unpolarized, which means that the electric vector is in a random direction. It implies that the diffracted beam will be partially or completely polarized. Intensity of diffraction maxima depends on scattering angle. Hence the correction for polarization turns out to be

$$P = \frac{(1+\cos^2\theta)}{2}$$

The correction to be applied to the measured intensity is $1/P$. Since we use graphite crystal monochromator in a diffractometer, the incident monochromatic beam on the crystal itself will be partially polarized. Hence the polarization factor is

$$P = P_t \frac{(\cos^2\theta_m + \cos^2\theta_{hk})}{(1+\cos^2\theta_m)} + (1-P_t) \frac{(\cos\theta_m + \cos^2\theta_{hk})}{(1+\cos\theta_m)}$$

where $\theta_m$ is the Bragg angle for monochromatizing crystal and $P_t$ is the perfection factor for monochromatizing crystal.
Lorentz factor

This geometrical factor which depends on the experimental arrangement for recording intensity and is a measure of relative time opportunity of different reciprocal lattice points to sweep through the Ewald sphere.

When the crystal rotates, the three-dimensional reciprocal lattice also rotates. The reciprocal lattice point near the rotation axis will cut across the Ewald sphere more slowly than one away from the rotation axis. Further opportunity for diffraction of reciprocal lattice points close to upper levels is different from those close to equatorial plane. Thus the time opportunity and geometrical opportunity vary from point to point.

Lorentz factor is in the form of \( \lambda L_{hkl}/\Theta \) which takes account of both time opportunity and the geometrical opportunity for different reciprocal lattice points to diffract. \( 1/\Theta \) allows for time opportunity where as \( L_{hkl} \) considers the geometrical aspect and is shown to be

\[
L_{hkl} = \frac{1}{\sin 2\theta_{hkl}}
\]

Absorption

As the incident and diffracted beams pass through a crystal, their intensities are reduced by absorption. The amount of absorption depends on the path length traversed within the crystal and is therefore different for different diffraction direction. Unless the morphological shape of the crystal is spherical or cylindrical, calculation of the correction factor for the absorption effect is extremely complicated.

If \( I_0 \) is the intensity of the incident beam, after traversing the thickness \( t \) of the crystal, the intensity becomes

\[
I = I_0 e^{-\mu t}
\]

where \( \mu \) is the linear absorption coefficient.

If the compound has very large absorption coefficient for the radiation used, the diffraction intensities need absorption correction. Among the methods used for absorption correction the numerical method (Coppens et al., 1965) and \( \psi \)-scan method are very promising.
**Extinction**

Attenuation of diffracted beams may also occur due to extinction effects: primary and secondary.

The loss of intensity of the diffracted beam due to destructive interference of rays arising out of multiple reflections from planes satisfying the reflection condition is known as primary extinction. This effect is severe in the case of an ideally perfect crystal.

As the incident X-ray penetrates inside the crystal, it goes on losing considerable fraction of its energy by reflection in the previous planes it encounters. Thus the planes lying deeper receive less and less incident intensity and consequently the diffracted intensity is attenuated. This effect is known as secondary extinction and is due to mosaic nature of crystal. Secondary extinction is more pronounced at low \( \sin \theta / \lambda \) and correction has been suggested by Darwin. The absorption coefficient is modified by adding a term \( g \) to account for secondary extinction. Correction for the secondary extinction is generally considered at the end of the refinement and can be given by approximate equation.

\[
|F_c| = k |F_{rel}| (1 + gI_c)
\]

where \( k \) is the scale constant and \( g \) is secondary extinction coefficient and is characteristic of the crystal for a given radiation.

**Anomalous scattering**

When the energy of incident X-ray beam is close to that which will give rise to a change in the quantum state of the electron, the expression for Thomson scattering will not be valid. Due to perturbation of electron distribution, spherical symmetry and centrosymmetry are destroyed. Then the atom is said to be an anomalous scatterer. The atomic scattering factor \( f_j \) is no longer real and takes the form,

\[
f_j = (f_j)_c + \Delta f' + i\Delta f''
\]

where \( (f_j)_c \) is the form factor calculated on the assumption of spherical symmetry and is corrected by a real part \( \Delta f' \) and an imaginary part \( \Delta f'' \) both of which may be calculated theoretically. When \( f_j \) is real,

\[
|F_{hlu}| = |F_{hal}|
\]
which is the well known Friedel’s law. Since we observe intensities, Friedel’s law implies

\[ I_{hkl} = I_{\overline{h}\overline{k}\overline{l}} \]

When an anomalously scattering atom is present, Friedel’s law breaks down and

\[ I_{hkl} \neq I_{\overline{h}\overline{k}\overline{l}} \]

Thus, Friedel’s law helps to identify existence of inversion center and in the solution of phase problem.

**Temperature factor**

So far we have considered the atoms in the crystal are at rest. But in reality, the atoms vibrate about a mean position, the extent of vibration depending on the temperature. Because of the finite size of the electron cloud about the nucleus, the density scattered by an atom falls off with increasing \( \sin \theta / \lambda \). Thermal vibration of the atom has the effect of smearing out the electron cloud in a larger volume, consequently enhancing the fall-off rate. If \( f_0 \) be the scattering factor of a stationary atom, the scattering factor for a vibrating one is given by Waller, as

\[ F = f_0 e^{-B \sin^2 \theta / \lambda^2} \]

where \( B \) is known as the isotropic thermal parameter and is given by \( B = 8\pi^2 \overline{U}^2 \), \( \overline{U}^2 \) being the mean square displacement of the atom. \( B \sin^2 \theta / \lambda^2 \) is the Debye-Waller factor.

In an anisotropic vibration of an atom the electron cloud is smeared out in the form of an ellipsoid. \( B \) is expressed in terms of six anisotropic parameters expressed as \( U_{ij} \) and the temperature factor takes the form

\[ \exp -2\pi^2 (h^2 a^*2 U_{11} + k^2 b^*2 U_{22} + l^2 c^*2 U_{33} + 2hk a*b*U_{12} + 2kl b*c*U_{23} + 2hl a*c*U_{13}) \]

where \( U_{ij} \)’s having dimension Å\(^2\) and \( b_i \)’s being reciprocal lattice unit vectors.

At the beginning, when practically nothing is known about the structure, the scale factor \( K \) has also to be found out to bring the observed \( |F_0| \) to the absolute level \( |F| \), such that

\[ \frac{|F_0|^2}{K} = |F|^2 \]
The method of estimating the scale factor $K$ as well as overall temperature factor $B$ was suggested (Wilson, 1942). Wilson showed that for a random distribution of $N$ atoms in the unit cell of a crystal,

$$<|F|^2> = \sum_{j=1}^{N} f_j^2$$

$$K = \frac{\langle |F_o|^2 \rangle}{\langle |F|^2 \rangle}$$

$$K = \frac{\langle |F_o|^2 \rangle}{e^{2B\sin^2 \theta / \lambda^2} \sum_{j=1}^{N} f_j^2}$$

we get,

$$\ln \frac{<|F_o|^2>}{\sum_{j=1}^{N} f_j^2} = -2B\sin^2 \theta + \ln k$$

The reciprocal space is divided into a number of annular zones, each zone containing a considerable number of reflections (Woolfson, 1961). The variation of $\sin \theta / \lambda$ across each zone is small. Both $<|F_o|^2>$ and $\sum_{j=1}^{N} f_j^2$ correspond to the $<\sin^2 \theta / \lambda^2>$ for the particular annular zone. The graph of $\ln <|F_o|^2> \sum_{j=1}^{N} f_j^2$ against $\sin^2 \theta / \lambda^2$ called as Wilson plot is drawn and a value of $B$ is estimated from the slope of the resulting line (Rogers, 1965). With the scaling of the intensity data, the value of an overall isotropic temperature factor will be realized. Then an attempt is made to postulate a preliminary model of the structure. To different procedures used in the present work are given below.

**Structure determination: The phase problem**

Crystallographically available data consists only structure factor magnitudes but not phases. An attempt to find solutions to the phase problem (Lipson & Cochran, 1957; Bunn, 1961) started around 1934. Through the phase problem can be solved in different ways, two major methods viz., Patterson method and direct method are commonly used.

**Application of Patterson function to solve structures containing heavy atoms**

Patterson showed that when phaseless $|F|^2$ ‘s instead of $F''$ ‘s are used as the coefficients of a Fourier series, the function obtained known as Patterson function $P(uvw)$ will be given by,
and a peak in the Patterson map corresponds to an interatomic vector in the structure. i.e., at \( u = x_1 - x_2 \), \( v = y_1 - y_2 \), \( w = z_1 - z_2 \). Moreover, the height of a peak is proportional to the product of the atomic numbers of the corresponding pair of atoms. If the unit cell of a crystal contains ‘N’ atoms, the Patterson map will have \( N^2 \) peaks due to \( N^2 \) interatomic vectors. Of these, \( N(N-1) \) peaks will be distinct and a large peak will appear at the origin.

The peaks in a Patterson map are inherently broader than those in a Fourier map for finite size of atoms. Instead of \( N \) peaks in a unit cell of a Fourier map, there is a crowding of \( N(N-1) \) peaks in the same volume of the Patterson map enhancing the probability of superposition. However, when the structure contains one or more heavy atoms, the peaks due to interactions of the heavy atom stand out among other peaks of lower magnitude making interpretation comparatively easy. The Patterson technique is therefore best applicable to heavy atom structures.

There are special planes and lines in the Patterson map containing peaks due to the interactions between pairs of atoms related by symmetry called Harker planes and Harker lines. Thus peaks on the Harker line and Harker plane facilitate to the determination of the coordinates of heavy atoms. For example, in the case of space group \( P2_1/n \), peaks due to pairs of atoms related by the glide plane along the line \((\frac{1}{2}, v, \frac{1}{2})\).

Generally the positions of the heavy atoms are determined from the Patterson map, together with the positions of some other atoms. And a few peaks corresponding to heavy-to-light interatomic vectors can be correctly interpreted. The Fourier map reveals some more atomic positions and this process is repeated until the complete asymmetric unit is obtained. Sometimes, when only a few atoms remain undetected, it is worthwhile to complete a difference Fourier map, by summing a Fourier series with \( ||F_o| - |F_c|| \) as coefficients, phased with phases of \( F_c \). In such a map, all atoms which have been correctly included in the calculation of \( F_c \) will be absent and those that have been erroneously included will stand out as peaks of negative height.
Difference Fourier map are generally used for correcting the atomic coordinates which have too much error to be corrected by least-squares method and also for locating hydrogen positions when the parameters of non-hydrogen atoms have been sufficiently refined. Sometimes, the number of atoms that are located initially is too small, so that the sum of their atomic numbers is much less than one-third of the total atomic number of the asymmetric unit. In such cases computation of the Fourier map as mentioned above is not recommended. Sim (1960) showed that under these circumstances, if the coefficients of the Fourier series are suitably weighted, peaks corresponding to the atom not included in the calculation of $F_c$ would show up in the map. The weighting function given by him is

$$W = \frac{I_1(x)}{I_0(x)}$$

where $I_1$ and $I_0$ are respectively the first and zero order Bessel functions of the first kind as

$$X = 2 \frac{|F_0|}{|F_H|} \sum_{j} f_j^2$$

where $F_H$ is the structure amplitudes calculated on the basis of located atoms and $L$ includes only the atoms which are yet to be located. Repeated computation of weighted Fourier synthesis reveals the location of the atoms in the entire asymmetric unit.

For structures containing only light atoms, the Patterson technique does not prove to be very helpful. An entirely different approach will be used for such structures.

**Direct method**

During a diffraction experiment only the magnitudes of $F$ is observed, the corresponding phase $\phi$ being lost. This is however true only for any single reflection. Because, it is well known that with a complete set of intensity data, the collection of $|F|$ values does contain enough information regarding the phases. Thus, it is possible to determine the phase directly from the set of $|F|$ values measured with reasonable accuracy—hence the name direct method.
The electron density at a point with position vector \( \mathbf{r} \) can be expressed as

\[
\rho(\mathbf{r}) = \frac{1}{V} \int \text{all } \mathbf{k} \text{ } F_{\text{rel}}(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}} e^{-i\mathbf{k} \cdot \mathbf{r}} d(\mathbf{k})
\]

And the quantity \( F(\mathbf{k}) \) is given by

\[
F(\mathbf{k}) = \int \rho(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}} d\mathbf{r}
\]

where the integration is carried over the unit cell volume. For a unit cell having \( n \) discrete atoms this becomes,

\[
F(\mathbf{k}) = \sum_{n} f_{j} e^{i\mathbf{k} \cdot \mathbf{r}_j}
\]

The Phase of this quantity is given by

\[
\tan \phi(\mathbf{k}) = \frac{\sum_{n} f_{j} \sin(\mathbf{k} \cdot \mathbf{r})}{\sum_{n} \cos(\mathbf{k} \cdot \mathbf{r})}
\]

This is an indirect method of estimating the values of \( \phi(\mathbf{k})'s \).

**Unitary and normalized structure factors**

While searching for relations between the phases and the magnitudes of \( |F| \), we must remember that apart from their dependence on the position of the atoms (Sayre, 1956), \( |F|'s \) are also affected by the finite size of the atoms and thermal vibration contributes due to these factors are evident for our purpose and the magnitudes of \( |F| \) may advantageously stripped off of these two effects by suitable modification, so that the modified values correspond to the idealized situation of a structure built up by stationary point atoms.

One way of doing this is to express \( F_h \) as a fraction of its maximum value, which is

\[
\frac{\sum_{j=1}^{N} f_{j}.h}{\sum_{j=1}^{N} f_{j}.h e^{-R_j^2}}
\]

where vector \( \mathbf{h} \) has been used for \( \mathbf{hkl} \) and \( f_{j}.h \) indicates the value of \( f_j \) at \( s_{(hkl)} \) corresponding to \( \mathbf{h} \) and \( f_j \) being the scattering factor of the atom.
Thus we get the unitary structure factor \( U_h = \frac{F_h}{\sum_{j=1}^{N} f_j h} \) the values of which evidently lie between \(-1\) and \(+1\).

Another way of achieving the same is as follows. Consider the fraction of \( |F_h|^2 \) and its average value \( \langle |F_h|^2 \rangle \). We have already mentioned earlier that \( |F_h|^2 = \sum_{j=1}^{N} f_j^2 \), but it has been observed that the presence of symmetry axes and planes, average intensity of special class of reflections (those with one or two indices zero or with two equal indices etc.) is enhanced by a factor \( \varepsilon \) with respect to that of the general hkl reflections. Depending on the nature of the symmetry element, \( \varepsilon \) may have values like 2, 4, etc., even 12. To take care of this fact, \( |F_h|^2 \) is divided by \( \varepsilon_h \), for general reflections and takes different integer values for special class of reflections as necessitated by the symmetry elements present. Thus we write,

\[
|E_h|^2 = \frac{|F_h|^2}{\varepsilon_h \sum_{j=1}^{N} f_j^2}
\]

and

\[
E_h = \frac{F_h}{(\varepsilon_h \sum_{j=1}^{N} f_j^2)^{1/2}}
\]

Since the atomic form factor curves are of identical shape for all atoms, we may assume
\( f_j = Z_j \vec{f} \), where \( \vec{f} \) is a common shape vector and hence,

\[
E_h = \varepsilon_h^{-1/2} \sum_{j=1}^{N} (Z_j / \sigma_m^{1/2}) e^{(2\pi hr)}
\]

where,

\[
\sigma_m = \sum_{j=1}^{N} Z_j^m
\]

\( E_h \) is called the normalized structure factor because of its property

\[ \langle |E_h|^2 \rangle = 1. \]
Structure invariants and semi-invariants

Quantities which are invariant under any arbitrary change of the origin are called structure invariants. It may be easily verified that if $h+k+l=0$ then $\phi_h+\phi_k+\phi_l$ is a structure invariant. The magnitude of $|F_h|$ is obviously independent of our choice of origin for describing the structure. But the phases $\phi_h$ are very much dependent on origin specification.

In all the spacegroup except P1, a point with some advantageous symmetry environment is chosen as origin. For example, in the centrosymmetric space group $P\bar{1}$, the origin is at a center of symmetry. In the noncentrosymmetric space group $P2_12_12_1$, origin is at a point half-way between the three pairs of non-intersecting screw axis. Now with the origin chosen in the respective ways as stated, we find that in case of both the space groups mentioned the following eight points in the unit cell. $(0,0,0)$, $(\frac{1}{2},0,0)$, $(0,\frac{1}{2},0)$, $(0,0,\frac{1}{2})$, $(0,\frac{1}{2},\frac{1}{2})$, $(\frac{1}{2},0,\frac{1}{2})$, $(\frac{1}{2},\frac{1}{2},0)$, $(\frac{1}{2},\frac{1}{2},\frac{1}{2})$ have identical symmetry environments. Change of origin from one to another of such points may be called as allowed shift of origin. Quantities which remain unaltered under an allowed shift or origin are called structure semi-invariants. Since the structure factor magnitudes of $|F_h|$ are structure invariants, they are expected to provide information regarding other structure invariants and semi-invariants only and not about individual phases in general. This fact will further lead to some relations involving the magnitudes and phases.

The fact that in general the magnitudes of $\phi_h$ vary with our choice of origin even among the allowed alternatives for a space group calls for a procedure for fixing the origin towards the beginning of our phase determination procedure. This is done by assigning arbitrary phase to a requisite number (depending on the particular space group; maximum 3) of reflections with indices of suitably chosen parity.

In case of non-centrosymmetric structure the enantiomorph is also fixed. Because, a change from one enantiomorph to the other causes a change in the sign of the phases and consequently in those of linear combinations which are structure invariants. To fix enantiomorph the phase of reflection with suitably chosen indices are restricted. Often within the range from 0 to $\pi$ and from $-\frac{\pi}{2}$ to $+\frac{\pi}{2}$.
Refinement of the structure

Once the model structure has been proposed by one of the methods explained earlier, it is necessary to refine the preliminary coordinates by some process of refinement. The model obtained is however an approximate one and the positional and thermal parameters describing the model together with a scale factor for the structure factors calculated on this basis have to be refined, so that the actual structure may be approached as closely as possible within the limits of experimental accuracy.

An agreement between the calculated $F_c$ and the observed $F_o$ structure factor indicates the degree of refinement. The most common method of assessing the agreement is by calculating the residual index $R$ or reliability index, which is defined as

$$R_1 = \frac{\sum |F_o| - |F_c|}{\sum |F_o|}$$

the summations being over all the reflections. Also, another form of the residual used is

$$wR_2 = \left(\frac{\sum \omega (F_o^2 - |F_c|^2)^2}{\sum \omega |F_o|^2}\right)^{1/2}$$

where $\omega = \frac{1}{\sigma^2(F_o)}$, $\sigma(F_o)$ being the standard deviation of $F_o$.

Thus $R$ measures the relative discrepancies between $F_c$ and $F_o$, so that the lower the value of $R$, the better is the agreement. In an ideal case $R = 0$.

The simplest approach to structure involves merely the combination of the observed structure amplitudes with the calculated phases and use of these as the coefficients in a Fourier series (Stout & Gensen, 1968). The resulting Fourier synthesis should not only return the atoms of the input model, but it should improve the coordinates of their location and reveal new atoms. This process is repeated until the non-hydrogen atoms are located. This process is called cyclic Fourier refinement.

Further refinement of structure is done using the method of least-squares. The function minimized is

$$R = \sum \omega_{hkl} \left( |F_o|^2 - |F_c|^2 \right)^2$$

where $\omega_{hkl}$ is the weight of the observation. All possible least-squares equations are constructed and solved by the matrix method.
Difference Fourier synthesis

After refining the structure using any one of the model, the remaining lighter atoms and hydrogen atoms can be located by the method of difference Fourier synthesis and expressed in the form

$$\Delta \rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_i (|F_i| - |F_c|) e^{i \alpha_c} e^{-2\pi i (hx+ky+lz)}$$

Where $\alpha_c$ is the phase of $F_c$ and $x$, $y$ and $z$ are coordinates of a point in the unit cell.

From the value of maximum $\Delta \rho(xyz)$ and minimum $\Delta \rho(xyz)$, the missing atoms or misplaced atoms or the atom which has been considered as a different atom can be identified. In general $\rho(xyz)$ for hydrogen atoms are not calculated and are fixed at chemical acceptable positions.

Weighting scheme

In the process of refinement, every observation should have a weight (Cruickshank, 1965) $\omega_i$ associated with it, reflecting the reliability of that particular observation. The proper weight to be assigned to an observation is

$$\omega_i = \frac{1}{\sigma_i^2}$$

Where $\sigma_i$ is the standard deviation. The reciprocal of the variance is indeed a measure of its reliability and the function minimized is

$$\sum \omega (|F_o| - |F_c|)^2$$

where $\sigma$ is the standard deviation and it can be shown that

$$\sigma = \sqrt{N}$$

where $N$ is number of counts. In diffractometer while measuring the X-ray reflection left background and right background counts are considered and the standard deviation turns out to be

$$\sigma = \frac{k}{2 \sqrt{LP}} \left( \frac{N_T + N_{LBG} + N_{RBG} + (0.01N_{net})^2}{N_{net}} \right)^{1/2}$$

where $k$ is scale constant, $L$ is the Lorentz factor, $P$ is the polarization factor and $N_{net} = N_T - N_{LBG} - N_{RBG}$. 
Some results derived from the refined structure

The most important results that can be derived from the structure are the distances between bonded atoms and the angles between pairs of bonds. It is also necessary to find the equation of the mean plane passing through a group of atoms supposed to lie on a plane, to check the deviations of the atoms in such a group from the plane and perhaps also to calculate the distance of some other atoms from it.

In our computer programs for all such calculations, the fractional coordinates of any atom \( X = (x,y,z) \) with respect to crystallographic axes are transformed into coordinates \( X' = (x',y',z') \) in Å, with respect to the orthogonal coordinates whose \( x \)-axis is coincident with the crystallographic \( a \)-axis, \( y \)-axis lies in \( ab \)-plane and \( z \)-axis along \( c^* \)-direction. These transformed coordinates are given by

\[
X' = UX
\]

where \( U \) is transformation matrix, given by

\[
U = \begin{pmatrix}
a & b \cos \gamma & c \cos \beta \\
0 & b \sin \gamma & -c \cos \alpha \sin \beta \\
0 & 0 & c \sin \alpha \sin \beta
\end{pmatrix}
\]

Bond distances, angles and their standard deviations

Calculation of interatomic distances and angles using the orthogonal coordinate system is too simple to be described here. If \( \sigma_x, \sigma_y \) and \( \sigma_z \) are the estimated deviations of the fractional coordinates \( x, y \) and \( z \) of an atom, then the matrix \( P_o \) with respect to crystallographic axes is

\[
P_o = \begin{pmatrix}
\sigma_x^2 & \sigma_x \sigma_y \cos \gamma^* & \sigma_x \sigma_z \cos \beta^* \\
\sigma_x \sigma_y \cos \gamma^* & \sigma_y^2 & \sigma_y \sigma_z \cos \alpha^* \\
\sigma_x \sigma_z \cos \beta^* & \sigma_y \sigma_z \cos \alpha^* & \sigma_z^2
\end{pmatrix}
\]

The variance matrix \( P_s \) with respect to orthogonal axes, as defined above is given by

\[
P_s = UP_o \tilde{U}
\]

Where \( \tilde{ } \) denotes matrix transposition. For an atom with variance matrix \( P_s \), the variance in the direction with direction cosines \( L = (l,m,n) \) is

\[
\sigma^2 = LP_s \tilde{L}
\]
Now, for a distance $r_{AB}$ between atoms A and B having respective variance $\sigma^2_A$ and $\sigma^2_B$ in the direction of bond, the variance is given by

$$\sigma^2(r_{AB}) = \sigma^2_A + \sigma^2_B$$

In case of distance $r_{AA}$ of symmetry related atoms

$$\sigma^2(r_{AA}) = 4\sigma^2_A.$$

For the angle $\psi$ between atoms A, B and A,C of respective lengths $r_{AB}$, $r_{AC}$, the variance (Cruickshank and Robertson, 1953) is given by

$$\sigma^2(\psi) = \frac{\sigma^2_B}{r_{AB}^2} + \sigma^2_A \left( \frac{1}{r_{AB}^2} - \frac{2\cos\psi}{r_{AB}^2 r_{AC}^2} + \frac{1}{r_{AC}^2} \right) + \frac{\sigma^2_C}{r_{AC}^2},$$

where $\sigma^2_B$ and $\sigma^2_C$ are the variance of atoms B and C in the direction at right angle to $r_{AB}$ and $r_{AC}$ respectively and $\sigma^2_A$ is the variance of A in the direction of the center of the circle passing through B, A and C. When A and B are related by symmetry,

$$\sigma^2(\psi) = \sigma^2_A \left( \frac{1}{(1/2r_{AB})^2} - \frac{2\cos\psi}{1/2r_{AB} r_{AC}} + \frac{1}{r_{AC}^2} \right) + \frac{\sigma^2_C}{r_{AC}^2}.$$

In our calculation of standard deviation of bond angles it is assumed that the variances of the atomic coordinates are spherically symmetric,

i.e., $\sigma^2 = \frac{1}{3} (\sigma^2_{xx} + \sigma^2_{yy} + \sigma^2_{zz}).$

**Mean planes**

The equation of a plane in orthogonal coordinate system at a distance $p$ from the origin is

$$lX + mY + nZ - p = 0$$

where l, m, n are direction cosines of the normal to the plane. The distance of a point $(X_i, Y_i, Z_i)$ from the plane being

$$P_i = lX_i + mY_i + nZ_i - p$$

The equation of the mean plane through a group of $N$ atoms is obtained by minimizing the quantity,

$$R = \sum_{i=1}^{N} \omega_i P_i \quad \text{where} \quad \omega_i = \frac{1}{\sigma_i^2}, \quad \frac{3}{\sigma_{xi}^2 + \sigma_{yi}^2 + \sigma_{zi}^2}.$$
The condition,
\[ \frac{\delta R}{\delta P} = 0 \]
gives
\[ l \Sigma \omega_i X_i + m \Sigma \omega_i Y_i + n \Sigma \omega_i Z_i - p \Sigma \omega_i = 0 \]
or
\[ l \overline{X} + m \overline{Y} + n \overline{Z} - p = 0 \]
where
\[ \overline{X} = \frac{\Sigma \omega_i X_i}{\Sigma \omega_i}, \quad \overline{Y} = \frac{\Sigma \omega_i Y_i}{\Sigma \omega_i}, \quad \overline{Z} = \frac{\Sigma \omega_i Z_i}{\Sigma \omega_i}, \]
showing that the mean plane passes through the centroid \((\overline{X}, \overline{Y}, \overline{Z})\) of the group of N atoms. Writing now,
\[ R = \Sigma \omega_i \left[ X_i - \overline{X} \right] + m(Y_i - \overline{Y}) + n(Z_i - \overline{Z}) \]
the condition \(\frac{\delta R}{\delta l} = \frac{\delta R}{\delta m} = \frac{\delta R}{\delta n} = 0\) gives three linear homogeneous equations in \(l, m, n\), which are solved by a procedure described by Blow (1960), remembering that \(l^2+m^2+n^2+1\). Thus equation for the mean plane referred to orthogonal axes as,
\[ l \overline{X} + m \overline{Y} + n \overline{Z} = 0 \]
The deviation \(P_i\) (i=1 to N) of the N atoms from the mean plane may be calculated and we can also calculate
\[ X^2 = \sum_{i=1}^{N} \frac{P_i^2}{\sigma^2(P_i)}, \quad \text{where,} \quad \sigma^2(P_i) = l^2 \sigma^2(X_i) + m^2 \sigma^2(Y_i) + n^2 \sigma^2(Z_i) \]

**Torsion angles**

Torsion angles of ring system or chain of atoms are the dihedral angles between planes formed by consecutive bonds. For a sequence of atoms A,B,C,D the torsion angle of bond B→C is the angle between the positive normals to the planes (B→A,B→C) and (C→B,C→D), where the positive direction of a normal is defined to be that which forms a right-handed set with the two vectors defining the plane.

The sign of torsion angle is defined as positive for bond B→C, if the plane containing atoms A,B,C has to be rotated clockwise, looking in the direction B→C, to make its positive normal coincident with that of the plane containing atoms B,C,D, so that the atoms in the front eclipses those in the rear.
Crystallographic studies on the structure, symmetry and conformation of molecules of medicinal and biological interest form the basis of this chapter. The knowledge of stereochemistry of the molecules is also derived from single crystal X-ray diffraction studies. The double-helical structure of DNA, the planarity of peptide bond, the triple helical structure of collagen, the globular conformation of enzymes and antibodies—all owe their discoveries to X-ray diffraction studies. Practically all the drugs known today are the result of either accidental discoveries or from trial and error process. The X-ray studies thus play an important role in the design of appropriate drugs.

EXPERIMENTAL

Crystallization

The process of crystallization is ordering among randomly arranged ions, atoms, or molecules to take up regular positions and shape in the solid state. It involves the phenomenon of nucleation and it may be considered to be in dynamic equilibrium between particles in the fluid phase and solid phase from saturated solutions. Several techniques are available for crystallization of small molecules such as slow evaporation, slow cooling, diffusion methods etc.

Single crystals of all the compounds presented in this thesis are obtained from slow evaporation technique only. Crystals are examined under a polarizing microscope before mounting in the goniometer head for data collection.

Crystal Mounting Facilities
**Intensity Data Collection**

In practice three different methods are available to determine the structure of a compound, namely, X-ray diffraction technique in solid state, NMR studies in liquid state and quantum chemical or molecular mechanics calculations in isolated state. Of these, X-ray diffraction is the most powerful technique used to elucidate the three-dimensional structure of molecules. In the present studies, data were collected using different diffractometers in USA, MALAYSIA and SOUTH AFRICA and are given below.

Bruker APEX II CCD; Siemens (Bruker) SMART CCD X-ray diffractometer; Bruker SMART CCD area-detector; Bruker-Nonius Kappa CCD; Oxford Diffraction Gemini diffractometer; Oxford Diffraction XCALIBUR diffractometer; Bruker Kappa APEX II diffractometer. The pictures of diffractometers are shown below.

**Data Reduction**

The raw data collected from the diffractometer suffers from physical and geometrical errors and hence cannot be used for structure determination straightaway. The intensity data have to be corrected for Lorentz, polarization and absorption effects.

**Bruker APEX II CCD**

BRUKER APEX II Diffractometer is based on Platform Goniometer and equipped with APEX II CCD Detector, Oxford Cryosystem 700 Plus low temperature Device. Instrument producing data of excellent quality for both routine structure determination and for high accuracy and high resolution data. Instrument presently equipped with Mo-radiation.
Siemens (Bruker) SMART CCD X-ray diffractometer

The Siemens (Bruker) SMART CCD X-ray diffractometer is a top-of-the-line single crystal diffractometer. Since its introduction, the CCD area detector has revolutionized the field of small molecule crystallography, reducing data collection times by orders of magnitude. This instrument is equipped with a programmable low temperature apparatus and structural determinations are typically performed at -100°C. Data are reduced/integrated offline on a locally networked PC and work-up then proceeds to structure solution using the latest version of the SHELX software suite. XSEED software is implemented for interfacing with SHELX, graphical production of results, and a variety of other useful tasks.
Oxford Diffraction Gemini diffractometer

This four circle CCD (EOS detector) dual source Mo/Cu diffractometer is equipped with two low temperature devices, namely, a nitrogen cooler, CRYOJET, allowing a routine working temperature of 90 K, and a helium cooler, HELIJET, allowing to reach a temperature as low as 15 K.

Oxford Diffraction XCALIBUR diffractometer

This four circle CCD (sapphire detector) diffractometer is equipped with two low temperature devices, namely, a nitrogen cooler, CRYOJET, allowing a routine working temperature of 90 K, and a helium cooler, HELIJET, allowing to reach a temperature as low as 15 K.
3. Bruker Kappa APEX II diffractometer:

This modern four circle CCD (sensitive and fast CCD APEX II detector) diffractometer is equipped with a cryostream nitrogen allowing to go down to a working temperature of 90 K.

The Lorentz and polarization corrections are must for every case, since the reflection intensity varies with the reflection angle, whereas the absorption correction has to be applied depending upon the nature of the compound and the radiation used, i.e., depending on the linear absorption coefficient value (Stout and Jensen, 1968). The space groups of the crystals are determined from the systematic absences of the reflections and by intensity statistics. If space group ambiguity arises then the contents of the unit cell, the number of molecules present in the cell, the distribution of intensity and other relevant details are to be analyzed in depth.

Structure solution
In general, the structure factor is represented as

\[ F_{hkl} = \sum f_j \exp[2\pi i (hx_j + ky_j + lz_j)] \]

where \( f_j \) is the atomic scattering factor or form factor for the \( j \)th atom.

In other words,

\[ F_{hkl} = |F_{hkl}| e^{i\phi(hkl)} \]
Here, \(|F_{hkl}|\) is the structure amplitude and \(\phi(hkl)\) is the associated phase. The structure amplitude can be obtained directly from the positive square root of the observed intensity. But there is no means to get the associated phase values experimentally in order to compute the electron density,

\[
\rho(x,y,z) = \frac{1}{V} \sum \sum F_{hkl} \exp[-2\pi i (hx_j + ky_j + lz_j)],
\]

where, \(V\) is the volume of the unit cell.

One needs both the structure amplitudes and phases to arrive at the electron density. The non-availability of the phases to compute the electron density is called the phase problem in crystallography. Several methods are used to tackle the phase problem, and some of them are:

i) Direct Methods

ii) Heavy atom method

iii) Isomorphous replacement method and

iv) Anomalous dispersion method

The above methods can be successfully applied to locate the approximate positions of all the atoms (trial structure of a molecule) in the unit cell. Direct methods have been used to solve the structures presented in this thesis. A brief summary of the stepwise procedure involved in the above method is given below.

**Direct Methods**

Direct methods are used to calculate the phases directly by simple mathematical procedure from a single set of X-ray intensities. The basic postulates of direct methods are the positivity (the electron density is positive everywhere) and the atomicity (the atoms are spherically symmetric). The structure amplitudes and phases are linked with electron density through Fourier transformation. A mathematical constraint on the function \(\rho(x)\) imposes a corresponding constraint on the structure factor. The constraint however does not hamper the evaluation of \(\phi(hkl)\) directly. The various steps involved in the direct methods are:

**Step I** Conversion of observed structure factors \(|F_{hkl}|\) to normalized structure factor \(|E_{hkl}|\) which are independent of \(\theta\).
Step II  Setting up of phase relations using triple phase relations (triplets) and four phase relations (quartets).

Step III  Selection of a few reflections, the phases of which are assigned apriori.

Step IV  Phase propagation and refinement using tangent formula (Karle and Hauptman, 1950).

Step V  Calculation of best phase sets and expressing the reliability of the phases in terms of Combined Figure of Merit (CFOM).

Step VI  Calculation of electron density map (E-map) with $|E_{hkl}|$ as the Fourier Coefficient.

**Structure Refinement**

The data obtained from the diffraction experiment are a set of structure factor magnitudes $|F_O|$, $|F_C|$’s are the calculated structure factor magnitudes for the proposed model. It is necessary to find the atomic coordinates $(x_j, y_j, z_j)$ and thermal factors that give maximal agreement with the observed structure factors. Hence, refinement on the position and the thermal parameter for each atom is essential to get best possible fit with the experimental data. SHELXL 93 (Sheldrick, 1993) is used for the full-matrix least-squares refinement. The refinement is based on $F_O^2$ since it is impossible to refine $F$, which would involve taking the square root of a negative number for reflections with negative intensity ($F_O^2$). The Residual factor defining the correctness of the model is given by

$$ R_1 = \frac{\sum |F_O| - |F_C|}{\sum |F_O|} $$

where the summation is over all the observed reflections. Lower $R_1$-value, greater the accuracy of the model. A suitable weighing scheme is applied at the end of the refinement procedure and the weighted $R_2$-factor (intensity based) $wR_2$ is given by

$$ wR_2 = \left( \frac{\sum w(|F_O| - |F_C|)^2}{\sum w(|F_O|^2)} \right)^{1/2} $$

where

$$ w = \frac{1}{\sigma(F_O^2) + (K_1*P)^2 + K_2*P} $$

$K_1$ and $K_2$ are constants and $P = (F_O^2 + 2F_C^2)/3$
Goodness-of-fit ($S$) is defined as

$$S = \left( \frac{\sum w(|F_o| - |F_c|)^2}{m - n} \right)^{1/2}$$

Where $m$ is the number of reflections and $n$ is the total number of parameters or variables.

**Calculation of Geometrical Parameters**

Crystal structure determination provides us the unit cell contents and fractional atomic coordinates of all the atoms and their associated thermal displacement parameters. The geometrical parameters such as bond lengths, bond angles and torsion angles can be obtained from the coordinates of the relevant atoms.

For a triclinic lattice, the distance between the two points in fractional atomic coordinates $(x_1, y_1, z_1)$ and $(x_2, y_2, z_2)$ is given by the law of cosines in three dimensional geometry as

$$L = \sqrt{a^2 + b^2 + c^2 - 2ab \cos \alpha - 2bc \cos \beta - 2ca \cos \gamma}$$

where $a$, $b$, $c$, $\alpha$, $\beta$, and $\gamma$ are the unit cell parameters and $\Delta x = x_1 - x_2$, $\Delta y = y_1 - y_2$, $\Delta z = z_1 - z_2$. The above equation can be applied for any crystal system to calculate the bond lengths. Bond length values are useful to identify the nature of chemical bonds (triple, double, partially double or single bond) present in the molecule.

Bond angle formed by the three atoms A, B and C where the angle is subtended by the bonds AB and AC can be calculated using the formula,

$$\theta = \cos^{-1} \left( \frac{(AB)^2 + (AC)^2 - (BC)^2}{2(AB)(AC)} \right)$$

Bond angles are useful to find the type of hybridization of a particular atom.

Torsion angle is the angle between the two planes designed as XYZ and YZU by the four atoms X, Y, Z and U.
Torsion angles are calculated using the following equation.

$$\chi = \left(\frac{N_1 \cdot N_2}{|N_1| |N_2|}\right)$$

where $N_1$ and $N_2$ are normals to the XYZ and YZU planes respectively. Torsion angles are useful to understand the orientation / confirmation of the plane / ring with another plane / ring formed by the various groups of atoms in a molecule.

**Inter and Intramolecular Interactions**

In crystalline state, the molecules are stabilized by intramolecular and intermolecular interactions like hydrogen bonds, van der Waal’s forces and possibly some short contacts between the two atoms. Hydrogen bonding is the specific type of non-bonded interaction between two electronegative atoms (donor and acceptor) where the hydrogen atom is bonded to them. The usual convention for the representation of the hydrogen bond is $D\cdash H\cdots A$ where $D$ is the donor and $A$ is the acceptor. The crystal structures presented in the present work are found to have O–H⋯O, C–H⋯N, N–H⋯O and C–H⋯O types of hydrogen bonds.

The existence of C–H⋯O bonds in crystals is evident from the study of Taylor and Kennard (1982) and Desiraju (1991; 1996). The ability of a C-H group to act as a proton donor depends on the hybridization [$C\text{-}(sp)$-$H\text{-}C\text{(sp$^3$)}$-$H\text{-}C\text{(sp$^3$)}$-$H$] and increases with the number of adjacent withdrawing groups (Steiner, 1996).

Weak attractive forces between uncharged atoms or molecules are collectively referred to as van der Wall’s forces. These forces arise from the electrostatic attraction of the nuclei of one molecule by the electrons of a different molecule. The repulsion arising between the electrons of two molecules as well as the nuclei of two molecules counteract the electrostatic attractions, but there is always a small net attractive force. The van der Wall’s forces are short-range forces i.e., they are significant only when the molecules are very close to one another.

**Computations**

Data collection: X-AREA (Stoe & Cie, 2001); CrysAlis CCD (Oxford Diffraction, 2004); SMART (Bruker, 1998); COLLECT (Hooft, 1998); cell refinement: X-AREA; CrysAlis RED (Oxford Diffraction, 2004); SAINT -Plus (Bruker, 2001); DENZO (Otwinowski & Minor, 1997); data reduction: X-AREA;
CrysAlis RED; SAINT -Plus; DENZO, SCALEPACK and COLLECT; program(s) used to solve structure: SHELXS97 (Sheldrick, 1990); SHELXTL/PC (Sheldrick, 1990); SHELX97 (Sheldrick, 1997); SHELXS97 (Sheldrick, 1997); program(s) used to refine structure: SHELXL97 (Sheldrick, 1997); SHELXTL/PC; molecular graphics: XP in SHELXTL-Plus (Sheldrick, 1991); SHELXTL/PC and MERCURY (Version 1.2.1; Bruno et al., 2002); ORTEP-3 (Farrugia, 1997) and PLATON (Spek, 2003); software used to prepare material for publication: SHELXL97; PLATON (Spek, 2003); WinGX (Farrugia, 1999) and PARST (Nardelli, 1995).

All the calculations were performed using Pentium PC computing systems.

Crystal engineering is “… the understanding of intermolecular interactions in the context of crystal packing and in the utilization of such understanding in the design of new solids with desired physical and chemical properties.” by G. R. Desiraju

**CRYSTAL EXPLORER**

Crystal Explorer (Wolff et al., 2007) has become an essential tool in CE. It explores the packing modes and intermolecular interactions in molecular crystals using Hirshfeld surfaces (Hirshfeld, 1977) from which information about all the intermolecular interactions can be encoded (Spackman & McKinnon, 2002). The Hirshfeld surface reflects the position of atoms and molecules in the crystal structure thus giving vital information on intermolecular interactions between the molecules (McKinnon et al., 1998). The Hirshfeld surface uses sophisticated interactive graphics (SIG) to encode information about all the intermolecular interactions and it is quite time consuming to compute these surfaces. The 2-D fingerprint plot has been developed to include all intermolecular interactions without using SIG software (Spackman & McKinnon, 2002). The 2-D fingerprint plot gives an elaborate visual account of the type of intermolecular interactions experienced in the crystal structure i.e. hydrogen bonding, close and distant van der Waals contacts, C-H…π interactions and π…π interactions on 2-D axes.

**CIF**

Crystallographic Information File (CIF) is a standard text file file format for representing crystallographic information, promulgated by the International Union of Crystallography (IUCr). CIF was developed by the IUCr Working Party on
Crystallographic Information in an effort sponsored by the IUCr Commission on Crystallographic Data and the IUCr Commission on Journals. The file format was initially published by Hall, Allen, and Brown and has since been revised, most recently version 1.1. Full specifications for the format are available at the IUCr website. Many computer programs for molecular viewing are compatible with this format, including Jmol.

**CIF, CIFTAB and Electronic Publication**

**CIF archive format**

The CIF format represents a major step forward in the archiving, publication and communication of crystallographic data. At last it is possible to publish crystal structures and incorporate structural data into the crystallographic databases without the expensive and error-prone retyping of tables by hand. CIF format also provides a convenient method of transferring data from one program system to another. The ACTA instruction instructs SHELXL to write two CIF-format files: name.fcf contains the reflection data and 'name.cif' all other data. These files contain all the items needed for archiving the structure; those answers not known to SHELXL (e.g. the color of the crystal) are left as a question mark. In general the final 'name.cif' file should be edited using CIFTAB or any text editor to replace most of these question marks. The file is then suitable for deposition in the CSD (organic) and ICSD (inorganic crystal structure) databases.

For publication of a routine structure determination via electronic mail it will normally be necessary to add the authors' names, title, text etc., which may also be done in CIF-format; this is followed by the edited contents of one or more .cif files each describing one structure (or possibly the same structure at different temperatures etc.). In general SHELXL provides all the CIF identifiers required by Acta Cryst. except those that begin with '_publ'. Further details are given below, and an example of a paper submitted to Acta Cryst. in this way may be found in the file example.cif (it has been brought up to date for the 1997 requirements for authors; whether it would pass the new stricter quality controls is another matter!). SHELXL users are strongly recommended to familiarize themselves with the definitive paper by the I.U.Cr. Commission on Crystallographic Data by Hall, Allen & Brown (1991), and with the
current Acta Crystallographica Instructions for Authors. Since the archiving of macromolecular data in CIF format is still being debated, SHELXL only creates a standard ‘small molecule’ CIF file, suitable for Acta Cryst. etc.; a macromolecular CIF file is likely to contain much more information. However the LIST 6 instruction in the new version of SHELXL does produce a CIF format reflection data file suitable for archiving with the PDB. This file also contains all the information necessary for the calculation of electron density maps, though as yet it appears that no standard macromolecular graphics package is able to read CIF format. Macromolecular coordinates etc. should be deposited in PDB format; SHELXPRO provides the necessary facilities for extending the .pdb file produced by SHELXL so that it can be used as a template for deposition.

**WINGX Version 1.80.05**

WinGX is a MS-Windows system of programs for solving, refining and analysing single crystal X-ray diffraction data for small molecules. It provides a consistent and user-friendly GUI for some of the best publically available crystallographic programs, and has interfaces to other popular programs such as SHELX-97 and SirWare programs (SIR-97, SIR-2004).

WinGX uses SHELX format ASCII files, and contains the following facilities and features

1. Utility programs to write initial model files, name.INS and STRUCT.CIF. Utility programs to operate on these files and import/export other file formats.
2. Programs for data reduction of Enraf Nonius CAD4 and Siemens P4 diffractometer files. Support for KappaCCD, Bruker SMART/APEX, Rigaku data and a GUI for SORTAV.
3. Graphical representation of the reciprocal lattice and reflection profiles.
4. Numerous absorption corrections, including analytical, gaussian quadrature, spherical-crystal, psi-scans, DIFABS, XABS2 and MULTISCAN (Blessings method for multiple equivalents).
5. Structure solution by DIRDIF-08, SUPERFLIP, SHELXS86, PATSEE and SIR-92, and interfaces to SIR-97/SIR-2004/SHELXS97.
6. Structure refinement with SHELXL-97 (externally available)
7. Fourier maps, contouring, XHYDEX location of hydrogen atoms.
8. Several graphics programs, including CAMERON, ORTEP-3, STRUPLo, PLUToN. Interfaces to SCHAKAL and RasMOL, and the ray-tracing programs POV-Ray and Raster3D. See examples of POV-Ray rendered pictures here.
9. Programs for analysis of crystallographic results, including PLAToN, PARST, THMA11
10. Publication of CIF based files and tables, validation of CIF files using IUCr CIF checking algorithms & PLAToN checking tools. Interfaces to word processing packages e.g. MS-Word.
11. User modifiable interfaces to other MS-DOS or MS-Windows programs.
12. A very useful set of introductory TUTORIALS for WinGX has been made available at the CCP14 site (thanks to Lachlan Cranswick). The DOOBRY test data set for these tutorials is available here.
13. WinGX may be used across networks. All WinGX programs accept either DOS or UNIX style ASCII files.

SHELXS97

The first version of SHELX was written at the end of the 1960's. The gradual emergence of a relatively portable FORTRAN subset enabled it to be distributed (in compressed form including test data as one box of punched cards) in 1976. SHELX-76 survived unchanged - the extremely compact globally optimized code proved difficult to modify - until major advances in direct methods theory made an update of the structure solution part necessary (SHELXS-86). Rewriting and validating the least-squares refinement part proved more difficult but was finally achieved with SHELXL-93. SHELXS-86 and SHELXL-93 were as far as possible upwards compatible with SHELX-76 (for example the format of the reflection data file was unchanged) and are now employed in well over 50% of all small-molecule structure determinations. A commercial version including interactive reciprocal and real space graphics is available in the form of the Siemens SHELXTL system.

A further release of SHELX in the current millenium was never intended, but the increased (mis) use of the programs by macromolecular crystallographers, and some changes to CIF format, have unfortunately made it necessary to release this new
version of the complete package as SHELX-97. This also provided an opportunity to update the structure solution algorithms, to fix various bugs, and to improve the documentation. Various beta-test versions were made available to selected guinea-pigs in 1996; these should be referred to as SHELX-96, and the final release (in 1997) as SHELX-97.

For the latest news the SHELX homepage at http://shelx.uni-ac.gwdg.de/ should be consulted.

Programs

SHELX-97 - contains the following six executable programs:
SHELXS - Structure solution by Patterson and direct methods.
SHELXL - Structure refinement (SHELXH for refinement of very large structures).
CIFTAB - Tables for publication via CIF format.
SHELXA - Post-absorption corrections (for emergency use only).
SHELXPRO - Protein interface to SHELX.
SHELXWAT - Automatic water divining for macromolecules.

The structure solution program SHELXS now includes more powerful direct methods (Sheldrick, 1990) and the use of the Patterson vector superposition method (Sheldrick et al., 1993) - completely different to the naive Patterson interpretation algorithm used in SHELXS-86 - for the automatic location of heavy atoms. This new Patterson interpretation routine is not only effective for small structures, but is also useful for the location of heavy atom sites from isomorphous or anomalous DF data of macromolecules.

MERCURY 3.0

Crystal structure visualisation, exploration and analysis made easy.

Mercury offers a comprehensive range of tools for structure visualisation, the exploration of crystal packing and statistical analysis of ConQuest geometric searches.
A base version of Mercury is available as a free download. Basic Mercury features include:

- a full range of structure display styles, including displacement ellipsoids (N.B. displacement ellipsoids will be displayed for CIFs or SHELX res files which contain Uequiv or Uij values only)
- the ability to measure and display distances, angles and torsion angles involving atoms, centroids and planes
- a number of output options for saving displays
- functionality for creating and displaying centroids, least-squares mean planes and Miller planes
- location and display of intermolecular and/or intramolecular hydrogen bonds, short nonbonded contacts, and user-specified types of contacts
- the ability to build and visualise a network of intermolecular contacts
- the ability to calculate, display and save the powder diffraction pattern for the structure on view
- the ability to display unit cell axes, the contents of any number of unit cells in any direction, or a slice through a crystal in any direction

Licenced Mercury features include:

- options for displaying multiple structures simultaneously and performing least-squares overlay of pairs of structures
- the ability to manually edit structures, or to automatically assign bond-types, standardise bond-types to CSD conventions and to add missing hydrogen atoms
- the ability to calculate graph sets
- the ability to view structures in stereo assuming the correct hardware is available (not Mac OS X)
- an interface to MOPAC (2007 and newer) which allows users to perform a limited number of molecule-based gas phase calculations
- Calculation of inter-molecular potentials using Gavezotti’s UNI atom pair potentials
- Gasteiger charge calculation
- display of space-group symmetry elements
• calculation and display of voids (free space in crystal structures) based either on contact surface or solvent accessible surface
• links to libraries of of intra and inter-molecular geometry (Mogul and IsoStar)
• statistical analysis tools that can be used to:
  o analyse the results of ConQuest substructure searches where geometric parameters (e.g. bond lengths, angles, torsions etc) have been defined in the query.
  o explore data interactively using a spreadsheet linked to Mercury's 3D view
  o plot data e.g. as histograms and scatter plots
  o calculate descriptive statistics, correlation matrices and hypothesis (significance) tests
• a CSD Solid Form module which features:
  o a knowledge-based informatics toolset for improving effectiveness, quality and risk assessment in the development of solid formulations of drugs, agrochemicals and molecular materials.

ORTEP-3

**Ortep-3 for Windows** is a MS-Windows version of the current release of ORTEP-III (1.0.3), which incorporates a Graphical User Interface (GUI) to make production of thermal ellipsoid plots much easier. Most of the commonly used features of ORTEP-III are directly available from the GUI. Loading a coordinate file will result in a default view of the molecule immediately, and no knowledge of the inner workings of ORTEP is required to produce excellent publication quality output.

The main features of **Ortep-3 for Windows** are:

• **Ortep-3 for Windows** can directly read many of the common ASCII crystallographic file formats which hold information on the anisotropic displacement parameters. Currently supported formats are SHELX, GX, CIF, SPF, CRYSTALS, CSSR-XR, CSD-FDAT, GSAS, Sybyl MOL, XYZ, PDB, Rietica-(LHPM) and Fullprof.

• **Ortep-3 for Windows** will also read any legal ORTEP-III instruction file.

• Up to 999 atoms may be present in the asymmetric unit. Covalent radii for the first 94 elements are stored internally, and may be modified by the user.

• Several style-templates are supplied, but the graphical representations of thermal ellipsoids for any specific atom, or the bonds between any specified pairs of atoms can be individually chosen. Over 120 different colours are stored internally and up to 200 colours may be defined. All graphical objects may be drawn in any colour.

• The viewpoint can be rotated by dragging the left mouse button, in the standard Windows fashion.
• A mouse labelling routine is provided by the GUI. Any number of selected atoms may be labelled, and any available Windows font may be used for the labels.

• Graphic output in the following formats is supported: HPGL and PostScript, BMP. Direct printing to an on-line printer is also supported. It is also possible to prepare drawings of the correct size for direct reproduction in journals (many leading journals are now requesting this).

• Ray-traced output is available using the Raster3D or POV-Ray formats. Several representations are available from the POV-Ray interface, including thermal ellipsoids (standard ORTEP, RMSD and MSD surfaces) and van der Waals plots.

• A simple text-editor is provided, so that input files may be modified without leaving the program.

• Symmetry expansion of the asymmetric unit to give complete connected molecules may be carried out easily.

• Automatic unit cell packing diagrams.

PLATON

PLATON is a versatile SHELX97 compatible multipurpose crystallographic tool. PLATON is an easy to use interactive graphics program used for viewing and analyzing molecular structures. It was developed by Spek.

Most PLATON features complement those available in the excellent and widely distributed SHELX97 package for crystal structure determination and refinement.

Historically (1980), PLATON started out as a program for the automated calculation of derived geometrical data (i.e. bond distances, bond angles and torsion angles, including su's (esd's)) for structures refined with SHELX76. Since then, that basic function evolved into an automatic function to calculate everything of possible interest (both intra- and intermolecular) for a given structural parameter set (Instruction: CALC).

PARST

PARST is a system of FORTRAN routines for calculating the molecular parameters from the results of crystal structure analyses. It was developed by Nardelli. PARST is used for calculating the conformational parameters (bond angle, torsion angle, hydrogen bonds, etc) of all molecules.

* * * * *
Section 1.2

Introduction to hydrogen bonding

Hydrogen bonds (HBs) are the most important ‘weak’ interactions encountered in solid, liquid and gas phases. They define the crystal packing of many organic and organo-metallic molecules, the 3D structure of biological macromolecules, as well as modulate the reactivity of different groups within a molecule. Hydrogen bonds (HBs) can be defined as an attractive interaction between two molecular moieties (two molecules or two parts of the same molecule) in which at least one of them contains a hydrogen atom that plays a fundamental role in the interaction. In the rank of interactions among atoms, the HB falls between chemical bonds (as covalent bonds) and nonbonding interactions such as van der Waals interactions. In general, a HB is characterized by: (i) a weak to medium interaction energy (Hibbert & Emsley, 1990); (ii) a considerable interpenetration of the isolated electronic clouds of the two moieties involved; (iii) a certain electron transfer between the two moieties, and (iv) a preferred geometry (MacDonald & Whitesides, 1994; Bernstein et al., 1994; Koch & Popelier, 1995). Hibbert and Emsley, 1990 define three kinds of HBs depending on the interaction energy values obtained. Thus, HBs with energies between -2.4 and -12 kcal mol$^{-1}$ (1 cal = 4.184 J) are defined as weak HBs, those with energies between -12 and -24 kcal mol$^{-1}$ are defined as strong HBs and those with energies more negative than -24 kcal mol$^{-1}$ are considered very strong HBs. The classical HBs correspond to those formed by two hetero atoms, A and B, with a hydrogen atom bonded to one of them and located approximately in between (A–H···B). In general, there has been considered an electrostatic attraction between the positive end of the bond dipole of A–H and the centre of negative charge on B (generally a lone pair of electrons). Usually, the A–H moiety is defined either as an ‘electron acceptor’ or as a ‘hydrogen bond donor (HBD)’ and the B moiety as an ‘electron donor’ or a ‘hydrogen bond acceptor (HBA)’. Therefore, most published works of HBs are the type O–H···B or N–H···B in which the HB acceptor ‘B’ posses N, O or F lone pairs responsible for the HB formation. These classical HBs have been generalized in other directions such as: (i) HBs with unconventional H donors such as C–H, (ii) HBs with unconventional H acceptors as p-bonded functional groups, halogens or C atoms, and (iii) dihydrogen bonds A–H···H–B.
A hydrogen bond is a special type of dipole-dipole bond that exists between an electronegative atom and a hydrogen atom bonded to another electronegative atom. This type of bond always involves a hydrogen atom. That explains the name. Hydrogen bonds can occur between molecules (intermolecularly), or within different parts of a single molecule (intramolecularly). The typical hydrogen bond is stronger than van der Waals forces, but weaker than covalent, ionic and metallic bonds. Hydrogen bonds can be classified into two types; Intermolecular hydrogen bond and Intramolecular hydrogen bond.

**Intermolecular hydrogen bond.**

This occurs when the hydrogen bonding is between H-atom of one molecule and an atom of the electronegative element of another molecule. For example (i) hydrogen bond between the molecules of hydrogen fluoride. (ii) hydrogen bond in alcohol or water molecules.

![Intermolecular Hydrogen Bond Diagram](image)

Intermolecular hydrogen bond results into association of molecules. Hence, it usually increases the melting point, boiling point, viscosity, surface tension, solubility, etc.

**Intramolecular hydrogen bond** - This bond is formed between the hydrogen atom and an atom of the electronegative element (F, O, N), of the same molecule. Intramolecular hydrogen bond results in the cyclization of the molecules and prevents their association. Consequently, the effect of this bond on the physical properties is negligible. For example, intramolecular hydrogen bonds are present in molecules such as onitrophenol, o-nitrobenzoic acid, etc.
Hydrogen bonding plays a key role in structure and function of proteins including features such as protein folding, local architecture, protein-ligand recognition, enzymatic activity, protein hydration and molecular dynamics. These classical hydrogen bonds are well established and have been studied in detail (Baker and Hubbard, 1984; Jeffrey and Saenger, 1991). However, a set of somewhat weaker interactions has also been recognized to play an important role in protein structure and stability (Desiraju and Steiner, 1999). This set includes N-H···π, O-H···π, C-H···O and C-H···π. The overall stabilization energy of these interactions is much smaller (not more than a few Kcal/mol), but is still significant.

Hydrogen bonds are the most important intermolecular interactions. In biochemistry, pharmacy, crystallography and supramolecular chemistry, as well as in molecular recognition and self-organisation, this kind of bonding is playing a significant role. Weak H-bonds have recently received considerable interest (Jeffrey & Saenger, 1994; Jeffrey, 1997 and Desiraju & Steiner, 1999).

The hydrogen bond is a unique phenomenon in structural chemistry and biology. Its fundamental importance lies in its role in molecular association. Its functional importance stems from both thermodynamic and kinetic reasons. In superamolecular chemistry, the hydrogen bond is able to control and direct the structures of molecular assemblies because it is sufficiently strong and sufficiently directional. This control is both reliable and reproducible and extends to the most delicate of architectures. In mechanistic biology, it is of vital importance because it lies in energy rang intermediate between van der waals interactions and covalent bonds. This energy range is one that permits hydrogen bonds to both associate and dissociate quickly at ambient temperatures. This twin ability renders the interaction
well suited to achieving specificity of recognition within short time spans, necessary condition for biological reactions that must take place around room temperature. For these reasons, the subject of hydrogen bonding is of major interest and remains relevant with each new phase in the kaleidoscope of chemical and biological research (Desiraju & Steiner, 2001).

In crystalline state, the molecules are stabilized by intramolecular and intermolecular interactions like hydrogen bonds, van der Waal’s forces and possibly some short contacts between the two atoms. Hydrogen bonding is the specific type of non-bonded interaction between two electronegative atoms (donor and acceptor) where the hydrogen atom is bonded to them. The usual convention for the representation of the hydrogen bond is D—H···A where D is the donor and A is the acceptor. The crystal structures presented in the present work are found to have O—H···O, O—H···N, N—H···O and C—H···O types of hydrogen bonds.

The existence of C—H···O bonds in crystals is evident from the study of Taylor and Kennard (1982) and Desiraju (1991; 1996). The ability of a C—H group to act as a proton donor depends on the hybridization [C(sp)—H>C(sp2)—H>C(sp3)—H] and increases with the number of adjacent withdrawing groups (Steiner, 1996; 2003).

Weak attractive forces between uncharged atoms or molecules are collectively referred to as van der Wall’s forces. These forces arise from the electrostatic attraction of the nuclei of one molecule by the electrons of a different molecule. The repulsion arising between the electrons of two molecules as well as the nuclei of two molecules counteract the electrostatic attractions, but there is always a small net attractive force. The van der Wall’s forces are short-range forces i.e., they are significant only when the molecules are very close to one another.

C-H···X interaction

In recent years it has been established that a C-H group can be a hydrogen bond donor. Although considered weak in nature, the bond C-H···X (X: O, π) termed as C-H interaction is known to be distributed widely among protein structures. These interactions have shown to be of greater importance in proteins, and polypeptides, protein-protein, protein-ligand and drug-binding interactions.
C-H...O interaction C-H...π interaction

C-H...O interactions form 20-25% of the total number of hydrogen bonds constituting the second most important group (Weiss, 2001). Important early work indicated a role for C-H...O interactions in protein structures. Steiner and Saenger (1993) reported the presence of C-H...O bonds completing the coordination of the buried water molecules in papain. Derewenda et al., (1994) presented evidence of functional significance in serine hydrolases of a hydrogen bond involving the C-H group of the active site H is and a neighboring main-chain carbonyl oxygen. The past studies have revealed that Cα-H...O hydrogen bonds occur almost ubiquitously in beta-sheets (Fabiola et al. 1997), but they also appear to occur frequently in alphahelical proteins (Chakrabarti & Chakrabarti, 1998).

Identification of Cα-H...O interactions

The C-H...O interactions in the dataset have been identified using the program HBexplore (Lindauer et al., 1996) that requires structural information in the Protein Data Bank (PDB) format (Bernstein et al., 1977). The set I criteria with geometrical parameters (distance H···O < 2.5Å, Cα···O < 3.9Å and Cα-H···O > 90°) have been used. The program identifies all potential hydrogen bonds according to geometrical criteria and generates an output. The output lists the name, position number of donor and acceptor residues along with geometrical parameters and subgroup of the identified H-bond set e.g. backbone-backbone bond.
C-H...π interaction

Hydrogen bonds, C-H···π with π-acceptor constitute yet another considerable fraction. Steiner and Koellner, 2001 described hydrogen bonds in proteins involving aromatic acceptors, and Brandl et al. 2001 exhaustively surveyed the occurrence of interactions involving all possible C-H groups (Cα, Caliphatic-H and Caromatic-H) as donors and all possible side chain π systems as acceptors. The cases in which C-H···π interactions have been described in proteins include the formation of complexes of proteins with ligands or cofactors such as the heme group (Nishio et al., 1998) and design of serine proteases inhibitors (Shimohigashi et al., 1996; 1999). Previous studies have also shown that C-H···π interactions are even responsible for the stabilization of structural elements such as alpha or 310 helices or non-proline cis peptide bonds (Brandl et al., 2001).

H-bonds are known to span a large energy range and Jeffrey (1997) classifies them in the three groups of weak, moderate, and strong H-bonds according to whether their energies are in the ranges 1–4, 4–15 and 15–45 kcal mol⁻¹. Empirically, energies and geometries are found to be mutually interrelated and encompassed between two extremes: (i) weak, long, dissymmetric, proton–off-centred, and mostly bent D–H···A bonds of essentially electrostatic nature; and (ii) strong, short, symmetric, proton-centred, and linear D–H···A bonds reducible to three-centre–four-electron interactions of covalent nature.

Hydrogen bonding plays an important role in the conformation, interaction, and recognition of both organic and biological structures (Jeffrey & Saenger, 1991; Jeffrey, 1997; Nangia, 2002). Therefore, the evaluation of new hydrogen bonding patterns has attracted considerable interest in several decades. Strong or conventional O–H···O, O–H···N, N–H···O and N–H···N hydrogen bonds (5-15 kcal/mol) have been well established in as early as the 1940s (Pauling & Delbruck, 1940; Gellman, 1998; Seebach & Matthews, 1997; Licini et al., 2005; Goodman et al., 2007; Li et al., 2008; Wu et al., 2008; Saraogi & Hamilton, 2009; Gong, 2001; Huc, 2004; Li et al., 2006). In the context of weak hydrogen bonding, the landmark study by Taylor and Kennard has stimulated increasing interest in the C–H···O hydrogen bond (Taylor & Kennard, 1982). Systematic studies by Desiraju and others reveal that this weak interaction is equally important as the conventional analogues in both small-molecule

Table 1.1 - Bond length (pm) and bond energy (kJ/mol)

<table>
<thead>
<tr>
<th>Bond</th>
<th>Length</th>
<th>Energy</th>
<th>Bond</th>
<th>Length</th>
<th>Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>H—H</td>
<td>74</td>
<td>436</td>
<td>H—C</td>
<td>109</td>
<td>413</td>
</tr>
<tr>
<td>C—C</td>
<td>154</td>
<td>348</td>
<td>H—N</td>
<td>101</td>
<td>391</td>
</tr>
<tr>
<td>N—N</td>
<td>145</td>
<td>170</td>
<td>H—O</td>
<td>96</td>
<td>366</td>
</tr>
<tr>
<td>O—O</td>
<td>148</td>
<td>145</td>
<td>H—F</td>
<td>92</td>
<td>568</td>
</tr>
<tr>
<td>F—F</td>
<td>142</td>
<td>158</td>
<td>H—Cl</td>
<td>127</td>
<td>432</td>
</tr>
<tr>
<td>Cl—Cl</td>
<td>199</td>
<td>243</td>
<td>H—Br</td>
<td>141</td>
<td>366</td>
</tr>
<tr>
<td>Br—Br</td>
<td>228</td>
<td>193</td>
<td>H—I</td>
<td>161</td>
<td>298</td>
</tr>
<tr>
<td>I—I</td>
<td>267</td>
<td>151</td>
<td>C=C</td>
<td>154</td>
<td>348</td>
</tr>
<tr>
<td>C=C</td>
<td>154</td>
<td>348</td>
<td>C≡C</td>
<td>134</td>
<td>614</td>
</tr>
<tr>
<td>C≡N</td>
<td>147</td>
<td>308</td>
<td>C≡C</td>
<td>120</td>
<td>839</td>
</tr>
<tr>
<td>C—O</td>
<td>143</td>
<td>360</td>
<td>C≡S</td>
<td>182</td>
<td>272</td>
</tr>
<tr>
<td>O—O</td>
<td>148</td>
<td>145</td>
<td>C≡F</td>
<td>135</td>
<td>488</td>
</tr>
<tr>
<td>O=O</td>
<td>121</td>
<td>498</td>
<td>C≡Cl</td>
<td>177</td>
<td>330</td>
</tr>
<tr>
<td>C—Br</td>
<td>194</td>
<td>288</td>
<td>N=N</td>
<td>145</td>
<td>170</td>
</tr>
<tr>
<td>C—I</td>
<td>214</td>
<td>216</td>
<td>N=N</td>
<td>110</td>
<td>945</td>
</tr>
</tbody>
</table>

It is well-known that two strong N—H···O or N—H···N hydrogen bonds in aromatic amides could combine together to form stable intramolecular three-center hydrogen bonds, in which one amide H is shared by two acceptor O or N atoms or one acceptor O or N is shared by two amide H atoms (5). Hydrogen bonds can vary in strength from very weak (1-2 kJ mol⁻¹) to extremely strong (>155 kJ mol⁻¹), typical values include:

- F—H···F (155 kJ/mol or 40 kcal/mol)
- O—H···N (29 kJ/mol or 6.9 kcal/mol)
- O—H···O (21 kJ/mol or 5.0 kcal/mol)
- N—H···N (13 kJ/mol or 3.1 kcal/mol)
- N—H···O (8 kJ/mol or 1.9 kcal/mol)

The length of hydrogen bonds depends on bond strength, temperature, and pressure. The bond strength itself is dependent on temperature, pressure, bond angle, and environment (usually characterized by local dielectric constant). The typical length of a hydrogen bond in water is 197 pm.
Hydrogen bonding plays a vital role in chemistry, nature and in all biological systems (Jeffrey, 1997). They are essential as they help in the stabilization of protein structures within the skin, muscles and other animal tissues (Brown et al., 2003). Hydrogen bond properties differ relative to the groups of atoms surrounding them (Jeffrey, 1997), and these can be characterized into three types; namely strong, moderate and weak interactions (Table 1.1):

**Table 1.2 - Characteristics of strong, moderate and weak hydrogen bonds (Jeffrey, 1997).**

<table>
<thead>
<tr>
<th>Interaction type</th>
<th>Strong</th>
<th>Moderate</th>
<th>Weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examples</td>
<td>F-H-F</td>
<td>O-H-O</td>
<td>O-H-O</td>
</tr>
<tr>
<td></td>
<td>P-O-O</td>
<td>N-H-O-C</td>
<td>C-O</td>
</tr>
<tr>
<td>Bond lengths in Å</td>
<td>12-1.5</td>
<td>1.5-2.2</td>
<td>&gt;2.2</td>
</tr>
<tr>
<td></td>
<td>22-2.5</td>
<td>2.5-3.2</td>
<td>&gt;3.2</td>
</tr>
</tbody>
</table>

Due to the large number of different types of hydrogen bonding interactions occurring amongst molecules, a methodology in the form of graph set notation was proposed by Etter (1990) (Fig. 1.1). This methodology takes into account the formation of certain hydrogen bonding patterns without considering the chemical nature of the compounds involved (Bernstein et al., 1995, Jeffrey, 1997). This methodology allows all hydrogen bonding patterns to be broken down and indentified in terms of chains (C), rings (R), finite complexes (D) and intramolecular interactions (S), as indicated by Fig. 1.15 (Bernstein et al., 1995, Jeffrey, 1997). In addition to this, the number of hydrogen bond acceptor (a) and donor (d) sites can be described along with the number of symmetry independent hydrogen bonds (n) (Jeffrey, 1997), hence leading to the formation of a graph-set designation $G_{a,d}(n)$ (Jeffrey, 1997).
Fig. 1.1 - Graph set notation of various hydrogen bonding molecules (Etter, 1991).

**\(\pi...\pi\) interactions**

\(\pi...\pi\) interactions allows for the partial overlapping of aromatic rings (Hunter et al., 1990). This formation is possible due to these organic molecules being planar (Barooah et al., 2003), hence allowing for overlapping of \(p\)-orbitals in \(\pi\)-conjugated systems to occur. These systems do not directly overlap each other due to repulsion interactions between the rings and are aligned in an offset orientation relative to each other. The strength of these interactions are usually strong and of a short range order, and depends on the stacking arrangements of the molecules (Hunter et al., 1990; Gabriel et al., 2002; Zheng et al., 2005).

Etter et al proposed several hydrogen bonding rules, one of which states: “the best hydrogen-bond donor and the best hydrogen-bond acceptor will preferentially form hydrogen bonds to one another” (Etter, 1990; 1991). The Etter rules describe anticipated hydrogen patterns for several well studied functional groups and have been used as a working model for the hierarchical application of these rules to the non-covalent synthesis of supramolecular structures.

Attempts to classify hydrogen bonded motifs were also addressed by Etter based upon a system of graph-set notation. 138 Hydrogen bonded patterns are described as chains (C), dimers (D), rings (R), or intramolecular hydrogen bonds (S). The number of donors (d) and acceptors (a) used in each motif are assigned as subscripts and superscripts respectively and the number of atoms involved in the
pattern is indicated in parenthesis. For example, $R^2_2(8)$ graph set notation denotes an eight-membered ring with two hydrogen bond acceptors and two hydrogen bond donors, and is exemplified by the carboxylic acid dimer (Fig. 1.2a). The utility of the graph-set notation lay in the evaluation of the frequency of a given hydrogen bonding pattern. However, information with respect to the types of proton donors and proton acceptors engaged in the pattern are not provided. Therefore the frequency of supramolecular synthon composed of specific hydrogen bond donors and acceptors cannot be addressed via the graph-set notation system.

![Image of supramolecular synthons]

**Fig. 1.2** – Examples of supramolecular synthons: a) carboxylic acid homosynthon and b) carboxylic acid···pyridine heterosynthon

Scientists have been looking at hydrogen bonding interactions dating back to the 1920s when Latimer and Rodebush investigated and postulated the hydrogen bonding phenomenon in water using varying temperatures. Pauling in the 1940s then postulated that hydrogen bonding is fairly ionic in nature and forms between only the most electronegative atoms. He then mentioned that the hydrogen bond is of great significance in determining the properties of substances and today we can bear witness to the truth of that statement. Pimentel and McClellan during the 1960s defined a hydrogen bond between a functional group A-H and an atom, or a group, B in the same or different molecule when

1. there is evidence of bond formation be it association or chelation
2. there is evidence that this new bond linking A-H and B specifically involves the hydrogen atom already bonded to A.

This definition does not take into account the chemical nature of the participants, including the polarities, and net charges are widely unspecified. The definition has no restrictions on interaction geometry; it only makes the assumption that a hydrogen atom must be involved. It took about 40 years for the advancement of this statement thus showing the controversy around the hydrogen bond. Steiner gave a definition of hydrogen bonding which states the following:
An \textit{A-H\cdots B} interaction is called a “hydrogen bond”, if 1. it constitutes a local bond, and 2. \textit{AH} acts as a proton donor to \textit{B}.

This definition has widely been accepted as the general chemical definition for hydrogen bonding as it addresses the acid/base properties of \textit{A-H} (donor) and \textit{B} (acceptor), which in principle can be understood by the incipient proton-transfer reaction from \textit{A-H} to \textit{B}. There are many specialized definitions of hydrogen bonding which are based on certain properties that are studied using particular techniques.

The terminology on hydrogen bonding is not uniformly used within the literature; therefore some of the terms used here will be explained. In a particular hydrogen bond \textit{A-H\cdots B}, the group \textit{A-H} is called the proton donor and \textit{B} is called the proton acceptor. Some authors refer to it as electron-acceptors and electron-donors. Due to the hydrogen bond having such a long range a single donor can interact with up to three acceptors simultaneously (Fig. 1.3).

\textbf{Fig. 1.3} – The different types of hydrogen bonding a) a single hydrogen bond, b) bifurcated hydrogen bond c) trifurcated hydrogen bond.

The energy of hydrogen bonds in the solid-state cannot be measured with any standard crystallographic technique, but it can be explained by its geometry, which makes it possible to be calculate energies computationally. The typical hydrogen bond geometry between a donor (X-H) and acceptor (A), is related by an angle $\theta$ (Fig. 1.2) which ranges between 160 - 180°. Since hydrogen bonds have such a long range it raises a question: Can hydrogen bonding be characterized into “strong”, “moderate” and “weak” hydrogen bonds? According to Jeffery, this is possible (see Table 1.1).

\textbf{Fig. 1.4} – Schematic of a typical hydrogen bonding geometry between donor and acceptor
GRAPH-SET NOTATION

Many research groups, dating back to the 1920s, have attempted to categorize hydrogen bonding systems based on the topology of the atoms involved in hydrogen bonding. Etter was one of the pioneers to explain the topology of hydrogen atoms involved in the different kinds of hydrogen bonding. Etter used graph theory for categorizing hydrogen bonding motifs in the organic solid-state using graph set notation. Graph set notation describes the topology of the atoms involved in hydrogen bonding with great detail. This graph set notation categorizes hydrogen bonds into four different types. The modes of hydrogen bonding are classified into four designators (G): self bonding (S) (intermolecular interaction within the same molecule), chain (C), discrete (D) and ring (R) and they are depicted as follows:

\[ G_d^a (r), \]

where

- \( G \) = pattern designator
- \( r \) = the degree (total number of atoms involved in the hydrogen bonding)
- \( a \) = number of hydrogen bond acceptors
- \( d \) = number of hydrogen bond donors

Using this nomenclature various hydrogen bonding motifs can be classified into the above pattern designators. Fig. 1.5 shows the graph set notation for different hydrogen bonds.

**Fig. 1.5** – Examples of hydrogen bonding motifs assigned with their respective graph-set Notations
Use of Graph Sets

All hydrogen-bond patterns that were obtained are described and compared using the recently developed graph set notation (Etter, 1990; Etter et al., 1990; Bernstein et al., 1995). Explanation of the graph set terminology and examples of its use in numerous situations may be found in the references listed in the 4th footnote. The following is a brief summary of some of the fundamentals of the graph set terminology. Graph sets describe the topological nature of the hydrogen-bond pattern and the numbers of donors and acceptors involved while highlighting the common features of molecular aggregates that are not addressed by the empirical formulae or by the symmetry considerations of the crystallographic space group. A graph set is specified as $G^{a_d}(r)$, where `G' is the pattern designator, `a' and `d' are the numbers of acceptors and donors, respectively, that are used in forming the hydrogen-bond pattern, and the degree `r' is the total number of atoms involved in the pattern.

![Diagram](image)

**Scheme 1.**

The pattern designator (G), describes the pattern of hydrogen bonding and can be one of four types; S, C, R, and D. As illustrated in Scheme 1, S (or self) denotes an intramolecular hydrogen bond, C refers to infinite intermolecular hydrogen-bonded chains, R refers to intermolecular rings, and D refers to non-cyclic intermolecular diads and other finite hydrogen-bonded sets. The subscript and superscript are omitted from the graph set when both values are unity, as is the case in the examples of S, C, and D described later. The degree (r) is also omitted, as seen in D, when the only atoms involved in the pattern are the donor and acceptor atoms.

A hydrogen-bond pattern containing one unique type of hydrogen bond (distinguished from other types of hydrogen bonds by the chemical nature and/or the symmetry relation of the donor and acceptor atoms used in the hydrogen bond) is
referred to as a motif. Although each example in Scheme 1 contains only one motif, a graph set may define a single motif or it may define a pattern containing two or more motifs; both are in this paper. A listing of all the hydrogen-bond motifs present in a given structure is known as the first level network. Hydrogen-bond patterns containing two different motifs are called second level networks.

The concept of different levels of networks is illustrated in Figure 4c where two sulfonamide molecules are hydrogen bonded together. In this example there are two unique types of hydrogen bonds or motifs: one in which the amido proton of the upper molecule hydrogen bonds to the aromatic nitrogen of the lower molecule and one in which the amido proton of the lower molecule hydrogen bonds to the sulfonyl oxygen of the upper molecule. Each of these motifs may be described by the graph set D, as is always the case when two unique entities are joined by a hydrogen bond. The First level network, simply a listing of these two motifs, would be N1.DD. Combining of these two motifs in a second level network results in a ring involving two donors (both amido), two acceptors (one amido and one sulfonyl), and a total of eight atoms. Consequently the second level network is defined as N2. R\textsubscript{2} (8). In this paper we show how graph sets can be used to systematically compare hydrogen-bond patterns observed in many crystal structures, highlighting their similarities and differences.

**WEAK HYDROGEN BONDING AND THE VAN DER WAALS INTERACTION**

It has been well documented that the C-H group acts as a weak hydrogen bond donor (Desiraju, 1991; Steiner, 1997). All atoms and molecules attract one another as a result of transient dipole-dipole interactions. Molecules need not have a net charge to participate in a dipole interaction. This transient dipole-dipole interaction is known as a van der Waals interaction. This is a weak interaction and its effects are based on the distance between the atoms/molecules involved (Kuchel & Racston, 1998).

What is the difference between a hydrogen bond and a van der Waals interaction? This is a fundamental question when dealing with intermolecular interaction classifications. A paper published by Cotton et al., (1997) states that a typical C-H···O/N hydrogen bond represents nothing more than a classical van der Waals interaction, but this statement was later challenged by Steiner and Desiraju (Steiner & Desiraju, 1998). Steiner and Desiraju did a complete CSD survey and their
results show that the fundamental difference between the hydrogen bond and van der Waals interactions lies in the difference in directionality of these interactions. The hydrogen bond is directional and very close to linear geometry; the van der Waals contacts are fairly isotropic, thus the bond is independent of the angle θ. This difference allows us to distinguish between the two contacts.

**C-H…π and π…π INTERACTIONS**

C-H…π interactions between a hydrogen atom and the π-system of an aromatic ring only recently gained interest in the literature (Malone et al., 1997), but the geometry of this interaction is poorly understood. The C-H…π interaction is considered to be a weak hydrogen bonding interaction where the π-system acts as a hydrogen bond acceptor. Hanton observed that structures containing no hydrogen bonding acceptors will make use of π-systems of aromatic ring as acceptors. It is therefore important for us to understand how this intermolecular interaction influences the mode of packing in the crystal. The aromatic type interaction has two basic orientations: the face-to-face (stacked) or edge-to-face (Malone et al., 1997) shown in **Fig. 1.6**. The face-to-face stacking appears to be more favourable if the molecules are different and share similar isoelectronic systems. The edge-to face is normally found in structures of small aromatic molecules (Steed & Atwood, 2000). The distance between the parallel and off-set packed aromatics rings (**Fig. 1.6**) is approximately 3.3 – 3.8 Å.

**Fig. 1.6** – The types of aromatic π-π interactions: (a) face-to-face (interplanar distance about 3.3-3.8 Å) and (b) edge-to-face orientations. (c) The repulsion between negatively charged π-electron clouds of facially oriented aromatic rings.

* * * * *
References


Kennard, O., Simanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer
Shape Software (2003). ATOMS. Shape Software, 525 Hidden Valley Road, Kingsport, Tennessee, USA.
Sheldrick, G.M., SADABS. (1996). Program for Empirical Absorption Correction, University of Gottingen, Germany

* * * * *