ELECTRON CRYSTALLOGRAPHY OF ORGANIC PIGMENTS.

Geraldine Boyce

A thesis submitted for the degree of Doctor of Philosophy

Department of Chemistry

University of Glasgow

 October 1997

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ABSTRACT

The principle aim of this thesis is the detailing of the development and subsequent use of electron crystallographic techniques which employ the maximum entropy approach.

An account is given of the electron microscope as a crystallographic instrument, along with the necessary theory involved. Also, an overview of the development of electron crystallography, as a whole, is given. This progresses to a description of the maximum entropy methodology and how use can be made of electron diffraction data in *ab initio* phasing techniques. Details are also given of the utilisation of image derived phases in the determination of structural information. Extensive examples are given of the use of the maximum entropy program MICE, as applied to a variety of structural problems.

A particular area of interest covered by this thesis is regarding the solid state structure of organic pigments. A detailed structure review of both β-naphthol and acetoacetanilide pigments was undertaken. Information gained from this review was used as a starting point for the attempted structural elucidation of a related pigment, Barium Lake Red C. Details are given of the synthesis, electron microscope studies and subsequent *ab initio* phasing procedures applied in the determination of structural information on Barium Lake Red C.

The final sections of this thesis detail electron crystallographic analyses of three quite different structures. Common to all was the use of maximum entropy methods, both for *ab initio* phasing and use of image derived phases.

Overall, it is shown that electron crystallographic structure analyses using maximum entropy methods are successful using electron diffraction data and do provide distinct structural information even when significant perturbations to the data exist.

The important thing in science is not so much to obtain new facts, as to discover new ways of thinking about them.

Sir W.L. Bragg.

ACKNOWLEDGMENTS

First and foremost, I must thank both of my supervisors, Dr. John Fryer and Dr. Chris Gilmore, not only for their knowledge and expertise in their respective fields of electron microscopy and theoretical crystallography, but also for their guidance, encouragement and sense of humour, when required.

Thanks are also due to Ciba Specialty Chemicals for financial sponsorship throughout this research and in particular, my supervisors at Paisley, Dr. Ian Fraser and latterly, Dr. Greig Chisholm.

Technical advice and support in microscopy was received from David Thom and Vicky Yeats, while computing advice was thankfully available from Stuart Mackay and William Nicholson - all greatly appreciated.

Dr. Doug Dorset must be thanked for allowing me to carry out some further microscopy work at The Hauptman-Woodward Institute, Buffalo, USA,

NATO provided funding for attendance at an International School for Crystallography, with The Royal Society of Chemistry and The Institute of Physics providing financial support for attendance at other conferences.

Finally, a thank-you to friends and fellow Alchemists' who, over my time in the Chemistry department, have provided the necessary social support! Especially, Marisa, Gerry, Andy, Neil and Kirsty - another crystallographer who never understood columns either.

TABLE OF CONTENTS

1.0. ORGANIC PIGMENTS.

2.0. ELECTRON MICROSCOPY.

3.0. ELECTRON CRYSTALLOGRAPHY.

6.0. ELECTRON CRYSTALLOGRAPHIC STUDIES OF BARIUM LAKE RED C.

7.0. STUDIES OF DYNAMICAL EFFECTS IN SUBSTITUTED PHTHALOCYANINE MOLECULES.

8.0. ELECTRON CRYSTALLOGRAPHIC STUDIES OF AN ALUMINIUM PHTHALOCYANINE POLYMER SYSTEM.

LIST OF FIGURES

1.0. ORGANIC PIGMENTS.

2.0. ELECTRON MICROSCOPY.

3.0. ELECTRON CRYSTALLOGRAPHY

4.0. MAXIMUM ENTROPY AND LIKELIHOOD METHODS IN CRYSTALLOGRAPHY.

5.0. REVIEW OF PIGMENT CRYSTAL STRUCTURES.

8.0. ELECTRON CRYSTALLOGRAPHIC STUDIES OF AN ALUMINIUM PHTHALOCYANINE POLYMER SYSTEM.

9.0. ELECTRON CRYSTALLOGRAPHY OF POLY[1,6-DI(N-CARBAZOLYL)-2,4-HEXADIYNE].

LIST OF TABLES

5.0. REVIEW OF PIGMENT CRYSTAL STRUCTURES.

6.0. ELECTRON CRYSTALLOGRAPHIC STUDIES OF BARIUM LAKE RED C.

7.0. STUDIES OF DYNAMICAL EFFECTS IN SUBSTITUTED PHTHALOCYANINE MOLECULES.

8.0. ELECTRON CRYSTALLOGRAPHIC STUDIES OF AN ALUMINIUM PHTHALOCYANINE POLYMER SYSTEM.

9.0. ELECTRON CRYSTALLOGRAPHY OF POLY[1,6-DI(N-CARBAZOLYL)-2,4-HEXADIYNE].

LIST OF PLATES

6.0. ELECTRON CRYSTALLOGRAPHIC STUDIES OF BARIUM LAKE RED C.

1.0. ORGANIC PIGMENTS.

1.1. A Perspective.

The origins of colourants and their applications can be traced back to prehistoric man, who used colourants of animal and vegetable origin to dye furs, textiles and other materials. Even the hieroglyphs of Ancient Egypt contain a thorough description of the extraction and application of natural colourants.

Prior to the discovery of Mauveine in 1856 by William Perkin[1], colour had been obtained from natural sources such as indigo and cochineal.[2] Perkin can be credited as being the pioneer of the organic chemical industry, not only because of his discovery of Mauveine, but because he set to work to manufacture and sell it.

Rapid advances in colour chemistry were initiated by the discovery of diazo compounds and their derivatives. The colour potential of this class of compounds and their ease of preparation led to the development of azo colours, which represent the largest fraction of organic colourants.

The basic level of classification of a colourant is either inorganic or more importantly in this context, organic. The most important differentiation of colourants is that they are either dyes or pigments.

In general, colourants are either soluble (dyes) or insoluble (pigments). Pigments are coloured or fluorescent particulate solids that are usually insoluble in and essentially are physically or chemically unaffected by the vehicle or medium in which they are incorporated. They alter appearance either by selective absorption or by scattering of light. Among organic pigments, azo compounds constitute the largest group both with respect to the number of different chemical structures and to the total volume of commercial production.

The purpose of this introductory chapter is to provide background to some aspects of pigment chemistry and later in chapter 5, a review is given on the extensive work that has been carried out on the solid state structures of pigments related to the ones examined in this research

1.2. Pigment Properties.

Before undertaking a detailed discussion of the chemistry used to describe azo pigments, it is of fundamental importance to examine some aspects of pigment properties.

By definition, pigments retain a crystallite structure throughout the colouring process. Therefore, consideration has to be made of crystal properties and not just simple molecular chemistry. Thus, it can be stated that pigment properties are not dependent entirely on molecular chemistry present.

The interactions of the pigment with its specified medium are mediated through features including particle size, shape and the nature of the crystal surface. It follows that pigment properties are a result of both molecular chemistry and the crystallographic arrangement of these molecules.

Unlike dye chemistry, the relationship between chemical structure and colour for organic pigments is not as well defined and is found to be subject to influence from a greater number of factors.

At the level of an individual molecule, variation in substituents can lead to improved properties for the pigment through simple electronic effects, in the case of halogen functionality or through physical effects including extensive hydrogen bonding or an increase in packing volume of molecules.

Fastness properties of pigments are principally conditional on molecular structure, although particle size can also influence lightfastness. Larger particles have smaller specific surface area and since light energy initially destroys outer molecular layers, larger particles will have better lightfastness.

Most organic molecules exhibit a small solubility in polar organic solvents and some pigments contain structural features designed to enhance solvent fastness. It is thought that presence of functionality that leads to the formation of intermolecular or intramolecular hydrogen bonding can augment fastness to light, heat and solvents, hence another substituent effect.

Colour strength of a pigment is dependent on the amount of reflected light, thus it is important to develop as large a surface area as possible. Therefore, in considering pigment crystallites, the degree of particle dispersion, aggregation and flocculation is very critical for general pigment properties such as strength, transparency and rheology.

This is by no means an exhaustive summary of pigment properties, but merely an attempt to illustrate the point that pigment properties are in fact influenced by a variety of factors, and that to gain any understanding of these properties, some knowledge on the crystallographic arrangement of the molecules within the particles is a prerequisite.

The next section of this introductory chapter offers a treatise on some aspects of the basic chemistry of azo pigments, the major class of organic pigment of which some of the pigments investigated during the course of this research belong.

1.3. Azo Pigments.

1.3.1. Background.

The success of azo pigments is due to the simplicity of their synthesis by diazotisation and azo coupling and to the almost innumerable possibilities presented by variation of diazo compounds and coupling components. Most commercially important azo pigments are formed by successive diazotisation and coupling. A primary aromatic amine is diazotised to give a diazonium salt, which is then coupled with an appropriate coupling component to form the azo compound.

The azo pigments contain one or more chromophore of the form (-N=N-). These pigments can be subdivided into two groups, pigment dyes and precipitated azos. Pigments include products that are insoluble in the aqueous reaction medium directly on formation and hence require no metal ions or other means to effect precipitation. Precipitated azos include products with salt forming groups, principally sulphonic or carboxylic acid which precipitate. Salts most commonly used include those of calcium, barium or manganese.

The physical and chemical characteristics that define the performance of a commercial pigment in a vehicle system include its chemical composition, stability, particle size and shape, degree of dispersion, crystal geometry and especially surface character. Not all of these properties are dependent entirely on the molecular chemistry of the pigment, but depend for instance on the form and size of its particles. The chromophore, the intrinsic chemical unit in the pigment crystal imparts colour and may also be important in properties such as solubility, heat fastness and light stability.

Characteristic properties of azo pigments are partial solubility in organic solvents and vehicles, poor bleed resistance, good acid and alkali resistance and good lightfastness.

Colouring applications of azo pigments are widespread. Traditional uses such as inks and paints still account for a large percentage of the pigments produced. These generally require large quantities of pigment to achieve the desired shade. The colouring of plastics e.g. Polyvinylchloride (PVC) require proportionally less pigmentary material. Other areas of usage include coatings and man made fibre industries.

1.3.2. Synthesis of Azo Pigments.

As was discussed briefly in the previous section, azo pigments are formed by the coupling of a diazotised amine with a variety of coupling components. In general, the overall reaction can be stated as in Figure 1.1.

Figure 1.1 Overview of diazotisation reaction.

1.3.3. Diazotisation

The reaction between a primary amine and nitrous acid in an aqueous medium results in the formation of a diazonium salt [4]. For aromatic amines, the reaction is very general. Halogen, nitro, alkyl, aldehyde and sulphonic acid groups do not interfere. Despite the fact that diazotisation takes place in acid solution, the actual species attacked is not the salt of the amine, but the small amount of free amine present [5]. In dilute acid, the actual attacking species is N_2O_3, which acts as a carrier of NO^+. Evidence is that the reaction is second order in nitrous acid and, at sufficiently low pH, the amine does not appear in the rate expression.[6] The mechanism as elucidated by Hughes, Ingold and Ridd [7] is given in Figure 1.2.

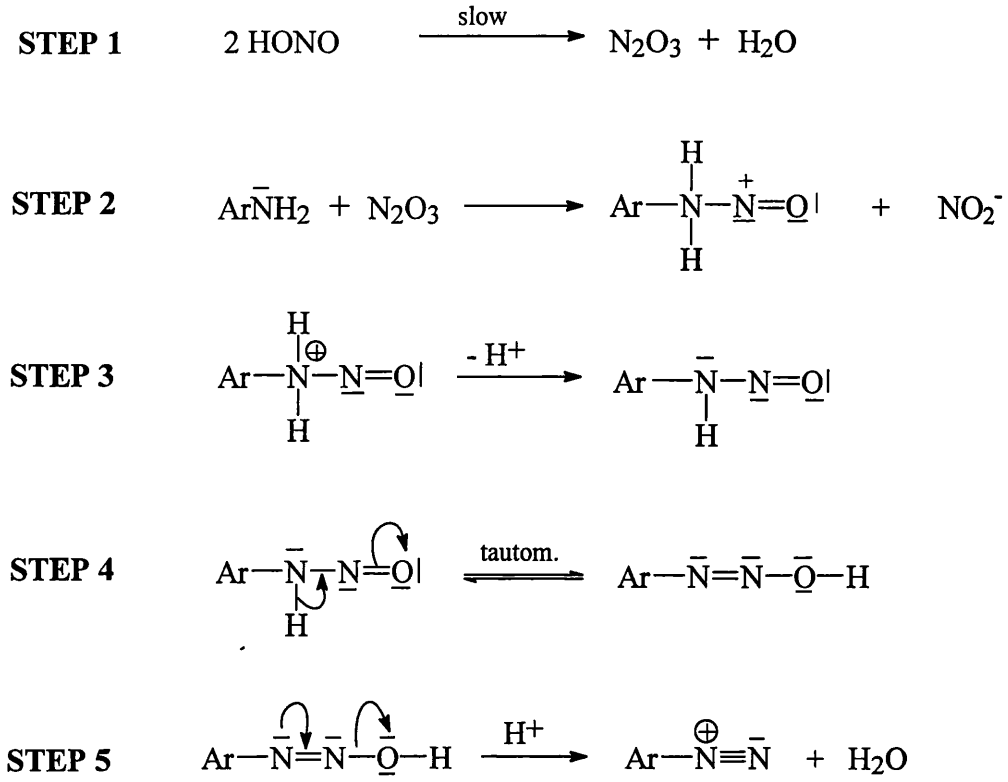


Figure 1.2. Mechanism of diazotisation.

Nitrosation of the amino group is the essential step in diazotisation.

Unlike aliphatic diazonium salts, the aromatic product is relatively stable at low temperatures. This is explained by the resonance interaction between the nitrogens and the ring. (Figure 1.3.)

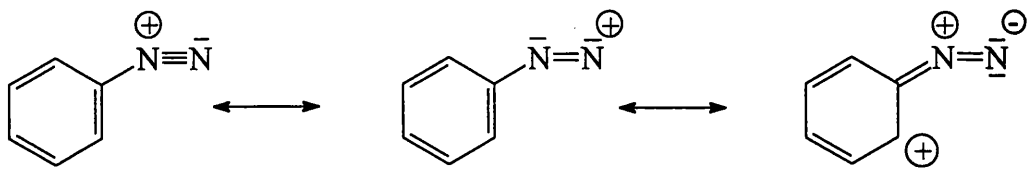


Figure 1.3. Resonance of nitrogen atoms in benzene diazonium.

The diazonium ion may be described as a Lewis acid, in which the β-nitrogen has electrophilic character. Under the correct conditions, diazonium ions attack

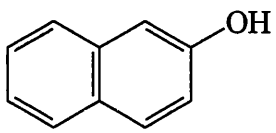
nucleophilic coupling components at a position with high electron density. This is the azo coupling reaction. Activation by electron releasing groups indicates that coupling is electrophilic aromatic substitution.

1.3.4. Coupling Reaction

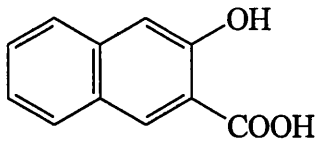
The coupling component must, in general, contain a powerfully electron releasing group. Common coupling components are amines and phenols. Due to the size of the attacking species, substitution is mostly *para* to the activating group, unless that position is already occupied, in which case ortho substitution takes place. The pH of the solution is important for both amines and phenols. For amines, the solution may be neutral or mildly acidic. This pH strikes a balance between maximising the quantity of free amine available, while keeping the diazonium ion stable.

Phenols must be coupled in slightly alkaline solution, where they are converted to the more reactive phenoxide ions, because phenols themselves are not active enough for the reaction.

2-Naphthol



3–Hydroxy-2-naphthoic Acid



Acetoacetanilide

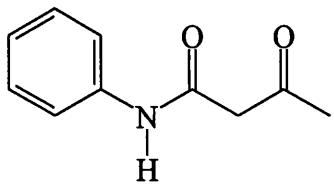


Figure 1.4. Examples of coupling components.

1.4. Classification of Azo Pigments.

In order to review the multiplicity of azo pigments that have already been synthesised, it is useful to classify these pigments according to their chemical structure. However, it should be kept in mind that pigments of identical structure need not necessarily have the same applicational properties. Thus, the presentation of pigments in such classes can serve only as an initial but fairly reliable guide. This will be discussed in the next section and examples of the relevant classes of azo pigment will be illustrated.

1.4.1. Azo Yellows

Hansa Yellows are mono azo pigments, a representative of which is Pigment Yellow 3 (Figure 1.5.). The Hansa Yellows are intense pigments that are semi-opaque and are resistant to both acid and alkali.

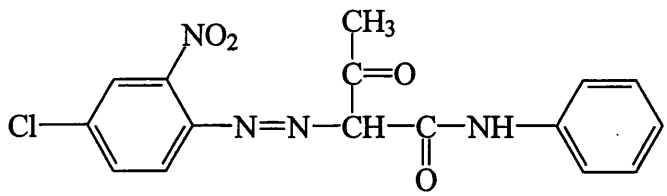


Figure 1.5. Pigment Yellow 3.

1.4.2. Diarylide Yellows

Related quasi-dimers of the Hansa Yellows are the Diarylides. The same chemistry is involved, except that the first component amine is a disubstituted diaminobiphenyl. Upon diazotisation this yields two diazonium salt groupings. The latter couples with two molecules of the second component to yield the disazo yellow. An example pigment is Yellow 12 (Figure 1.6.), which commercially belongs to the top five of all organic pigments as it is relatively cheap.

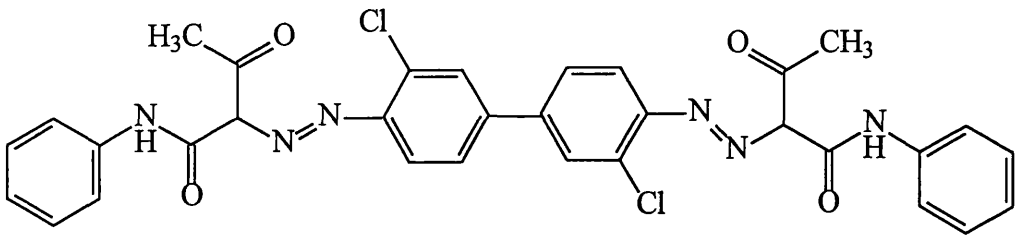


Figure 1.6. Pigment Yellow 12.

1.4.3. Toluidine Reds

Pigment Red 3 (Figure 1.7.) is a typical example of the Toluidine Red pigment class. It is one of the most popular red pigments for industrial enamels. Pigments of this class are available in several different shades, but are being displaced by other more durable reds.

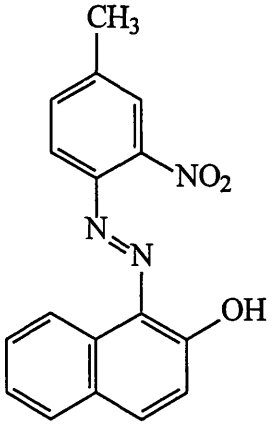


Figure 1.7. Pigment Red 3.

1.4.4. Para Reds

Para Red (Pigment Red 1) was the first commercial azo pigment and it was obtained from diazotised *p*-nitroaniline and 2-naphthol as early as the end of the 19th century.

Use of this class of pigment is severely limited by poor bleed resistance. Structure of this pigment is given in Figure 1.8.

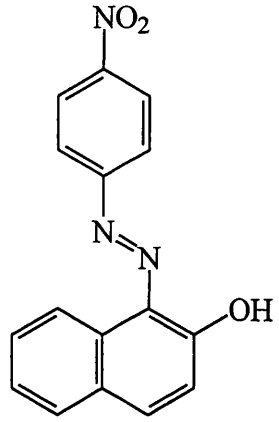


Figure 1.8. Pigment Red 1.

1.4.5. Lithol Reds

These are among the more important of the precipitated azo pigment dyes. They comprise a family of the sodium, barium, calcium and strontium salts of the coupling product from diazotised 2-naphthylamine-1-sulphonic acid and 2-naphthol. Shown in Figure 1.9., is the barium salt of Pigment Red 49.

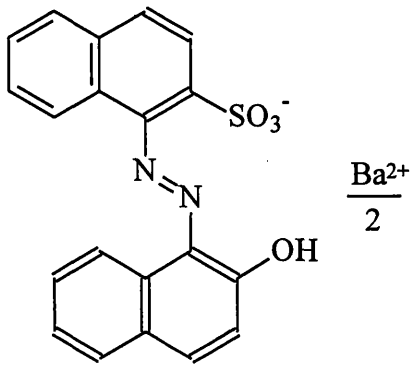


Figure 1.9. Pigment Red 49.

1.4.6. Laked β-Naphthol Pigments

Synthesis of these pigments involves formation of the soluble sodium salt by alkaline coupling then the pigment is produced by laking, i.e. precipitating the soluble salt with calcium or barium chloride. One of the more commonly used of this class is Lake Red C (Pigment Red 53:1). This is produced by coupling diazotised 2-amino-5-chloro-4-methylbenzene sulphonic acid with 2-naphthol and then laking with barium chloride. Structure is illustrated in Figure 1.10.

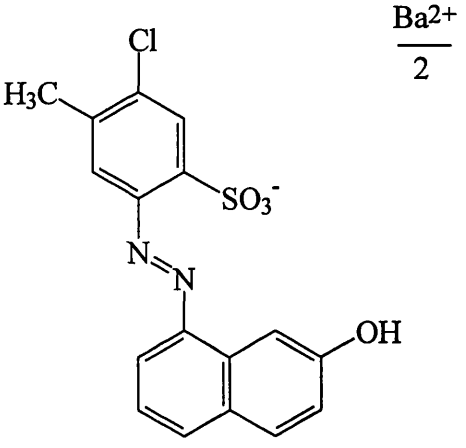


Figure 1.10. Pigment Red 53:1.

1.4.7. BON Reds

The BON Reds derive their name from the use of β-oxynaphthoic acid as the second coupling component in the coupling of various diazotised amines containing salt forming groups. Precipitation is effected by metal ions e.g. calcium, barium or manganese. A typical example is Pigment Red 57.

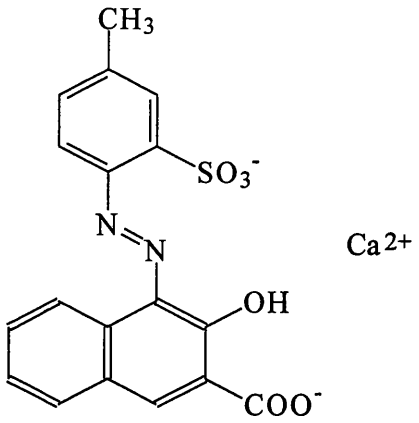


Figure 1.11. Pigment Red 57.

Having considered both generalities of pigment properties and some more detailed chemistry of azo pigments, it becomes obvious that a detailed picture of the crystallographic arrangement of pigment molecules must be obtained. This knowledge must be gained before any attempt to correlate pigment structure to property is attempted or even improvement of those properties by pigment manufacturers is undertaken.

Previous attempts to attain this crystallographic information about azo pigments via conventional X-ray methods have proved difficult due to the small crystallite size present. However, earlier work by Duckett[8], Miller[9] and Connell[10] utilising techniques of electron microscopy and latterly electron crystallography have shown that this methodology could provide crystallographic information on pigments of the azo class. Therefore, one of the aims of this work is to utilise developments in the field of electron crystallography to obtain structural information about some organic pigments.

This leads to the next sections where a discussion of the use of the electron microscope as a crystallographic instrument is given, along with a review of the development of electron crystallography as a powerful technique for elucidation of structural information.

2.0. ELECTRON MICROSCOPY

2.1. Electron Microscopy - An Introduction.

Much has been written about the development of the theory describing the wave nature of electrons, but in terms of electron microscopy, the most important statement is that of De Broglie.[11] In his theory, De Broglie combined two universal principles that linked fast moving electrons with short wavelength, these being the quantum theory of Planck[12] and the theory of relativity of Einstein.[13]

De Broglie postulated that any fast moving particle would have an electromagnetic radiation associated with it. If the particles are electrons, with a velocity one third the speed of light, the wavelength is approximately 0.05Å. This is 100 000 times shorter than the wavelength of green light, so if electrons are utilised, resolution is greatly enhanced over that of the best light microscope. Thus the De Broglie relationship was derived.

$$\lambda = \frac{h}{mv} \qquad \text{Equation 2.1.}$$

where λ=wavelength, h=Planck's constant, m=mass of an electron and v=velocity of an electron.

This theoretical analysis was confirmed experimentally by Davisson and Germer[14] and by G.P. Thomson[15], with the observation of diffraction phenomena.

It is not clear from the published literature who it was who had the original idea of using electron beams to form images of objects at resolution higher than that of the conventional light microscope. Electron optics were thoroughly investigated by Busch[16,17] and he showed that axial magnetic fields refracted electron beams in a way geometrically analogous to refraction of light beams by glass lenses.

This research encouraged Knoll and Ruska to construct the first true electron microscope. A description of an instrument designed for high resolution work was published by Ruska in 1934[18].

Subsequent development of both instruments and techniques saw the rapid adoption of electron microscopy as a method for obtaining structural information about a wide variety of materials ranging from those of a biological nature through to materials of chemical interest.

The purpose of this chapter is to provide a description of the transmission electron microscope and to outline some important theories that are relevant to later discussions on the development of electron crystallography.

2.2. The Electron Microscope.

The conventional transmission electron microscope has an electron-optical arrangement which is in essence the same as the optical arrangement in a light microscope. In an electron microscope the magnification scheme is more elaborate and of course, the electron source, the specimen and the whole assembly of lenses are in an evacuated column. The electrons are emitted most commonly from a hot tungsten filament and are accelerated through a large potential difference, V, of the order of 10^5 volts and so acquire kinetic energy, eV, where e is the electron charge. The velocities of the electrons are not negligible in comparison with the velocity of light, so that a relativistic correction has to be made. The wavelength of an electron is thus given by Equation 2.2.

$$\lambda = h\left\{2m_o\, eV\left(1+\frac{eV}{2m_o c^2}\right)\right\}^{-\frac{1}{2}} \qquad \text{Equation 2.2.}$$

where m_o is the rest mass of an electron and c is speed of light.

The arrangement of the parts of the microscope are shown schematically in Figure 2.1. The electrons emitted from the electron gun are immediately controlled by a sequence of two electromagnetic lenses, the condenser lenses, which determine the area of the specimen illuminated by the electron beam and the divergence of the electron beam incident on the sample.

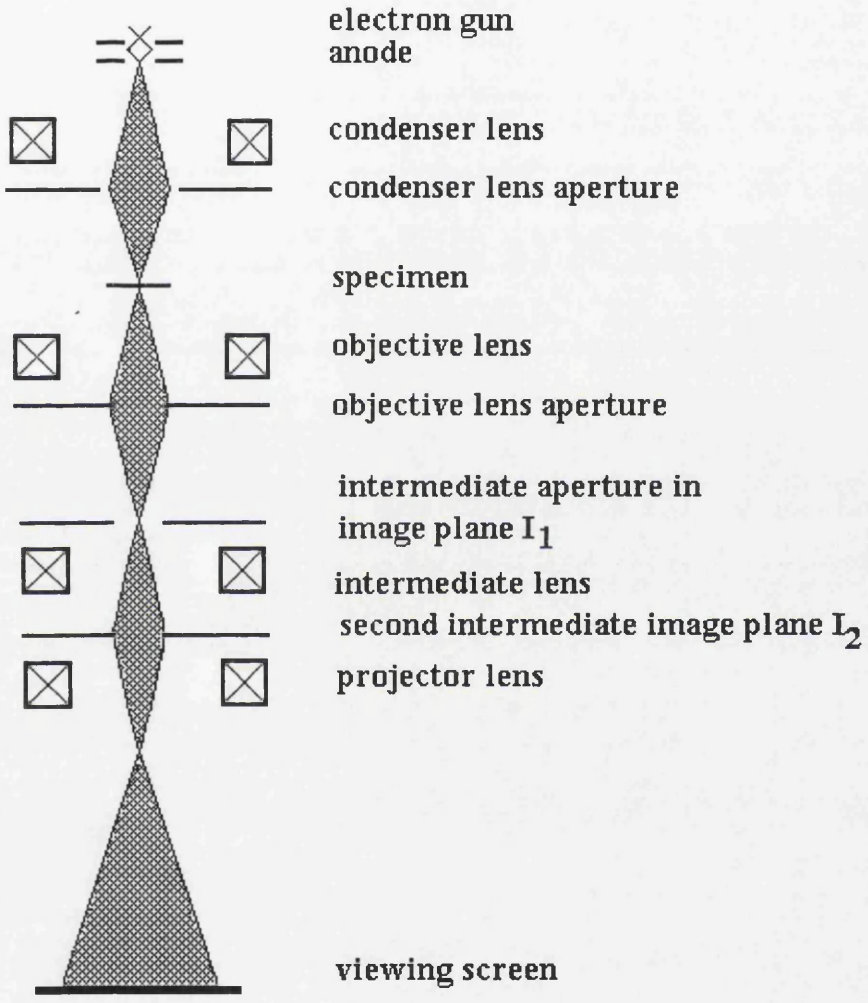


Figure 2.1. Descriptive diagram of an electron microscope.

Since the specimen is necessarily very thin, of the order of 1000Å, it must be supported in a special holder, and since the microscope interior is at high vacuum, the holder is introduced through an airlock. The specimen stage in which the holder fits, is situated within the bore of the next electromagnetic lens, the objective lens and is provided with the drives necessary to both move and tilt the specimen, so that

different areas can be studied at controllable orientations with respect to the microscope axis and the incident electron beam.

The objective lens forms an image of the specimen at image plane 1, I_1 in Figure 2.1, which is known as the first intermediate image. A sequence of two projector lenses further magnify this image produced by the objective lens. The first projector lens produces an image of I_1 at I_2, the second intermediate image and the second projector lens produces an image of I_2 on a fluorescent viewing screen. The final image may be recorded on a photographic film placed beneath the viewing screen or by an electronic recording device such as a CCD camera. The magnification of the final image can be varied in the range from 1000x to as much as 100 000x by varying the focal lengths of the two projector lenses, with the variation being achieved by adjustment of the energising current supplied to the lens.

If the image formed by the electron microscope were a perfect image, then any point in the image would correspond to a point in the specimen. This is not the situation in practice because magnetic lenses suffer from defects. The arrangement of the microscope is such that defects in the objective lens have a greater effect on the final image than do defects in the projector lenses. The principal effects to manifest themselves are astigmatism, chromatic aberration and spherical aberration. Astigmatism can be eliminated in the design and construction of the magnetic lenses. Chromatic aberration occurs when there is an energy spread in the electron beam, as well as on the magnetic field produced by the lens. Chromatic aberration can be minimised by stabilisation of the energizing current of the lenses and of the electron source.

An important defect of magnetic lenses is spherical aberration. If a lens is free from spherical aberration, all the electrons leaving the object from the point S will reach the image plane at point I (Figure 2.2) When spherical aberration is present, the rays passing through the outer parts of the lens will be bent excessively and so brought to a point closer to the lens than the point, I.

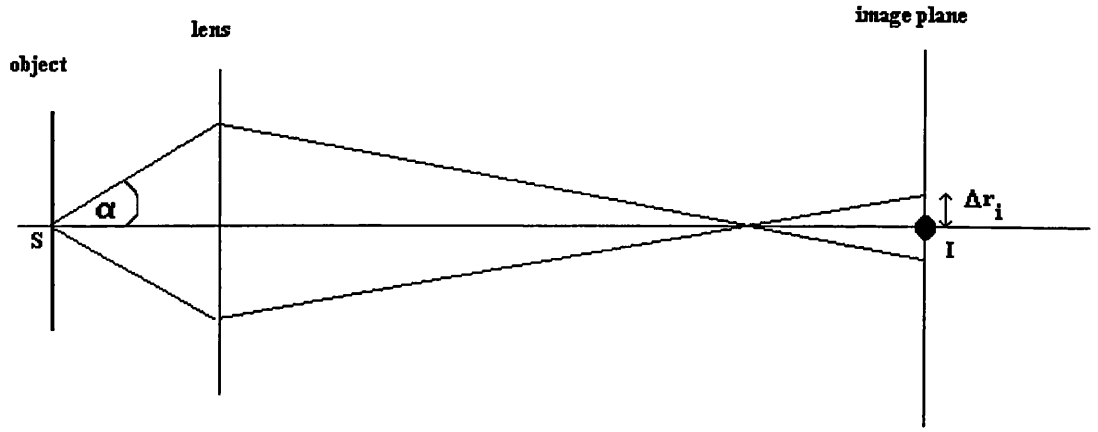


Figure 2.2. Illustration of spherical aberration.

Having given a physical description of the microscope, now the interactions of electrons with the specimen have to be considered and following from this, an understanding of the theoretical nature of image formation and diffraction within the microscope has to be gained.

2.3. Interaction of Electrons with Sample.

When considering the theory behind the interactions of electrons with the specimen, it is essential to refer to the comprehensive works of Hirsch et al[19] and Buseck et al[20] These texts offer a thorough discourse on electron interactions and as such only a cursory description of the theoretical basis will be discussed here.

Relative to other radiation, electrons interact strongly with matter. The electrons are scattered by the electrostatic potential, $\varphi(r)$ which plays the same role as electron density $\rho(r)$ does in X-ray scattering.

In considering the interactions of electrons with matter, various approximations exist which provide more simple mathematical descriptions of the scattering events that are occurring. Each of the approximations has a useful range of validity and this will be explored.

The scattering process is readily understood by the analogy of interaction of light with a transparent object. For electrons, there is a change in refractive index where the wave enters a region of different electrostatic potential, which changes the velocity and hence the wavelength of the electron waves.

2.4. Phase Object Approximation.

As a first approximation to understanding the interaction of electrons and specimen, assume that electrons passing through a thin specimen suffer a phase change that depends on the potential distribution along a straight line through the crystal. Thus if a potential distribution in the object is represented by a function, $\varphi(xyz)$, a plane wave transmitted through the object in z direction suffers a phase change that is a function of the xy coordinates proportional to the projection of the potential in the z direction, as shown in Equation 2.3.

$$\varphi(xy) = \int \varphi(xyz)\, dz \qquad \text{Equation 2.3.}$$

The phase change of the wave relative to a wave transmitted through a vacuum $(\varphi=0)$ is given by the product of $\varphi(xy)$ and an interaction constant, σ, where $\sigma = \frac{2\pi me\lambda}{h^2}$. This defines the strength of interaction with matter.

The effect on an incident wave of the phase change is given by multiplying the incident wave amplitude by a transmission function,

$$q(xy) = \exp[-i\sigma\varphi(xy)] \qquad \text{Equation 2.4.}$$

This equation gives the Phase Object Approximation. It is an approximation in that it ignores sideways scattering of waves.

Phase object approximation is useful because it emphasises the non-linear nature of electron scattering. For a single heavy atom, the phase differences from the centre to the outside of the atom in projection may exceed π. If the electron wave passes successively through several atoms in a row, as in a crystal, the phase changes add and non-linearity of the function given in Equation 2.4. ensures that scattering amplitudes are not added linearly.

For thicker specimens, the phase object approximation fails as the whole concept of transmission function becomes invalid in that the effect of a specimen on an incident wave cannot be represented by a multiplicative function.

For suitably thin specimens the wavefunction at the exit surface can be described as a 2D function, $\psi_1(xy)$, which can be represented as the product of the incident wave amplitude $|\psi_0(xy)|$ and the transmission function $q(xy)$. The exit wave has variations in phase and amplitude related to the specimen structure. Due to these variations, scattered waves diverge into the aperture of the objective lens. The aims of the imaging process are then to produce a magnified representation of $\psi_1(xy)$ and to record an intensity distribution from which some information about the specimen structure can be derived.

2.5. Weak Phase Object Approximation.

A further simplification of the phase object approximation allows for the derivation of the weak phase object approximation. This approximation assumes that the total phase change is small and is only valid when, for the maximum phase change due to the object, $\sigma\varphi(r) \leq 1$. Thus scattering is now expressed by,

$$q(xy) = 1 - i\sigma\varphi(xy) \qquad \text{Equation 2.5.}$$

The exit wave function is therefore linearly related to the projected potential of the crystal. The constant, 1, is then representative of the directly transmitted wave, which is unaffected by the object. The scattering function is given by $i\sigma\varphi(xy)$, which gives

rise to distribution of scattering amplitude in the back focal plane. For weak elastic scattering, the scattered wave is always 90° out of phase with the incident beam. The interaction constant σ, previously defined, decreases slowly with incident electron energy. The most important factor is the distribution of the projected potential which is dependent upon the atomic number, Z, of the atoms present. Another important factor to be considered is the number of atoms superimposed in the beam direction.

As described, the condition of validity of the weak phase object approximation is $\sigma\varphi(r) \leq 1$. This condition is not satisfied for single, very heavy atoms. At the centre of the projection of a uranium atom $\sigma\varphi(r)$ may be of the order of 1. For a light atom such as carbon, nitrogen or oxygen, the maximum is of the order of 0.1 However, in a specimen 50 to 100Å thick, especially in a crystal, 10 or more atoms may be aligned in the beam direction. In this situation values of $\sigma\varphi(r)$ for the projection through the line of atoms may greatly exceed unity. Hence the validity of the weak phase object approximation is strictly limited.

The simple approximations given above describing the interactions of electrons with the specimen and the subsequent formation of image have provided a framework that has allowed further development of imaging theory. Description of the specifics of imaging theory and its areas of application are beyond the remit of this thesis and, again, consultation with the aforementioned texts provides an extensive treatment.

The next section of theoretical discussion on electron microscopy details the more pertinent theories of electron diffraction, after first giving a description of some of the more important physical and geometrical aspects.

2.6. Electron Diffraction.

After consideration of the physical description of the electron microscope given in section 2.2, it can be seen that the final image is formed at the viewing screen. However, due to the existent lens arrangement, each lens within the column has its own image and back focal plane. Manipulation of the lens system and insertion of appropriate apertures allows for the collection of electron diffraction patterns on the viewing screen and subsequently on photographic film or CCD camera. In particular, if an aperture is inserted at the image plane of the objective lens, thus facilitating the magnification of a limited area of the sample, the diffraction pattern observed will also originate from this area. This technique can provide methods of obtaining diffraction patterns from discrete crystals or indeed specific areas within a given specimen.

2.7. Formation of Electron Diffraction Pattern.

In accordance with the well developed theories of X-ray diffraction, observation of electron diffraction patterns occurs under fulfillment of the Bragg condition.

Bragg's law[21] states that for a given set of planes (*hkl*), the reflected monochromatic radiation, in this instance electrons, only occurs at certain angles that are determined by the wavelength of the electrons and the perpendicular distance between adjacent planes. The relationship between these three variables is the Bragg equation.

$$2d \sin\theta = n\lambda \qquad \text{Equation 2.6.}$$

where d=interplanar spacing, θ=Bragg angle, n=integer and λ=wavelength of incident electrons.

The small wavelength of high energy electrons considerably simplifies the geometric theory of electron diffraction patterns and by using the Bragg relationship, the concept of the reciprocal lattice can be arrived at.

The reciprocal lattice is defined by axial vectors $\mathbf{a}^*,\mathbf{b}^*,\mathbf{c}^*$ which have lengths a^{-1}, b^{-1} and c^{-1} where a,b,c are conventional unit cell dimensions. Thus, the condition for a strong Bragg diffracted beam being produced from an incident beam vector, $\mathbf{s}_0$, is that the diffraction vector drawn from the origin of reciprocal space should end on a reciprocal lattice point, defined by a vector, $\mathbf{h}=h\mathbf{a}^*+k\mathbf{b}^*+l\mathbf{c}^*$, where *hkl* are Miller indices. Therefore Bragg's law can be written more completely as,

$$\mathbf{s} \equiv \mathbf{s}_1-\mathbf{s}_0=\mathbf{h} \equiv h\mathbf{a}^*+k\mathbf{b}^*+l\mathbf{c}^* \qquad \text{Equation 2.7.}$$

A graphical representation of the above equation is given by the Ewald sphere construction which is shown in Figure 2.3.

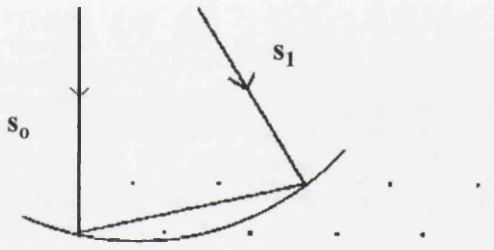


Figure 2.3. Representation of Ewald sphere.

For a given incident beam direction, the vector $\mathbf{s}_0$ of length λ^{-1} is drawn from a point L to the origin of the reciprocal lattice. Then, a sphere of radius λ^{-1} (the Ewald sphere) is drawn about the point L. If this sphere cuts the *hkl* reciprocal lattice point, then the vector from L to the reciprocal lattice point will be the vector $\mathbf{s}_1$ for a diffracted beam.

For electrons of wavelength of about 0.05Å, the corresponding Ewald sphere, λ^{-1}, is very large compared with the distance between reciprocal lattice points, a^{-1} (for a real-space, unit cell dimension, a). The curvature of the Ewald sphere is small and as such, segments are approximately planar and it is therefore possible to record simultaneously the set of reflections belonging to the zero order plane of the reciprocal lattice.

The electron diffraction pattern can be thought of as an image of the central plane cross section of the reciprocal lattice. Thus, utilising this simple relationship, a geometrical relationship between the distance, r, of reflections in the pattern from the central spot with $d_{hkl} = H_{hkl}^{-1}$ is shown in Figure 2.4.

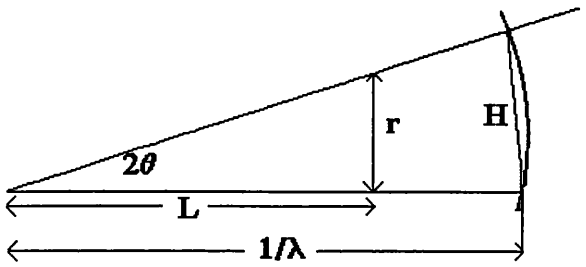


Figure 2.4. - Adapted from Fundamentals of Crystals[22], p339.

Thus,

$$H_{hkl}/\lambda^{-1} = \frac{r}{L}, \qquad rd_{hkl} = L\lambda, \qquad r_{hkl} = HL\lambda$$

where L is the distance from the specimen to the photographic plate and λ is the wavelength. In other words, the reciprocal lattice cross section is directly represented on the electron diffraction pattern to a scale $L\lambda$.

Crystals are defined by the periodic repetition of groups of atoms in three dimensions with periodicities described in terms of a real-space lattice of points defined by unit cell axes a,b,c and interaxial angles α,β,γ. This real-space pattern of repetition manifests itself within the symmetry of diffraction patterns. This information, combined with the relation of geometry in real space to reciprocal space can provide information about unit cell constants and the symmetry therein.

2.8. Atomic Scattering Factors for Electrons.

After examination of aspects of the formation of electron diffraction patterns, where the assumption was made according to Bragg's Law, of reflection from planes, consideration has to be made of how areas of atomic presence scatter the incident beam, or more correctly the factors affecting atomic scattering.

As previously described, electrons are scattered by potential fields thus the nucleus and electrons of a given atom are involved in electron scattering phenomena. Normally, the potential distribution is expressed in terms of the contribution of individual atoms centered at the positions, $\mathbf{r}=\mathbf{r}_i$, leading to,

$$\varphi(\mathbf{r}) = \sum_i \varphi_i(\mathbf{r} - \mathbf{r}_i) \qquad \text{Equation 2.8.}$$

As the crystal is periodic, a Fourier series can be used to represent this periodicity and as such, the potential distribution can be represented by a Fourier transform, V_i of the $\varphi_i(\mathbf{r})$,

$$V(\mathbf{h}) = \sum_i V_i(\mathbf{h})\, exp(2\pi\, i\mathbf{h}\cdot\mathbf{r}_i) \qquad \text{Equation 2.9.}$$

where **h** represents reciprocal lattice vectors.

The function $\varphi_i(\mathbf{r})$ may be identified with the potential distributions for individual isolated atoms or ions, with the usual spread due to thermal motion. It has to be noted however, that interatomic binding and interactions of ions are neglected and these may have important effects on diffraction intensities.

The Fourier transforms of individual atoms may be given as,

$$V_i(s) = f_i^B(s)/K \qquad \text{Equation 2.10.}$$

where $s=4\pi\lambda^{-1}\sin\theta$, f^B = Born electron scattering amplitudes as conventionally defined in units of Å, θ is half the scattering angle and $K=\sigma/\lambda$.

Values of $f^B(s)$ are obtained from the atomic potentials, φ_0 for isolated, spherically symmetrical atoms or ions using the relation,

$$f^B(s) = 4\pi K \int_0^\infty r^2 \varphi(\mathbf{r}) \sin sr/sr \, d\mathbf{r} \qquad \text{Equation 2.11.}$$

By using known relationships between potential and charge density distribution, the Mott-Bethe formula[23] was derived to describe $f^B(s)$ in terms of the previously derived atomic scattering factors for X-rays , $f_x(s)$,

$$f^B(s) = 2\pi \frac{me^2}{h^2\varepsilon_0} \{Z - f_x(s)\}/s^s \qquad \text{Equation 2.12.}$$

where ε_0 is the permittivity of a vacuum or another representation is,

$$f^B(s) = 0.023934\ \lambda^2 \{Z - f_x(s)\}/\sin^2\theta \qquad \text{Equation 2.13.}$$

where λ is in Å, $f^B(s)$ in Å and $f_x(s)$ in electron units.

Values of atomic scattering amplitudes for electrons are listed in International Tables for Crystallography Volume C,[24] calculated by Doyle and Turner[25].

It can be seen that the scattering amplitude thus varies with wavelength, λ and scattering angle θ. For $\sin\theta/\lambda$ greater than 0.4Å^{-1}, scattering amplitude increases regularly with atomic number, while for $\sin\theta/\lambda$ less than about 0.4Å^{-1}, scattering amplitude does not vary in a regular manner. For elements of the first three periods of the periodic table there is a tendency for scattering amplitude at small angles to decrease with increasing atomic number.

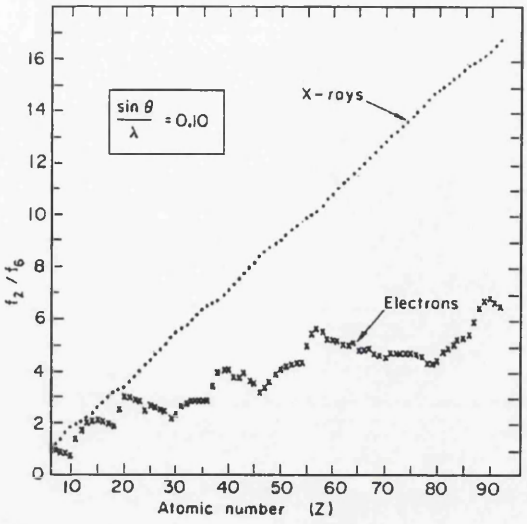


Figure 2.5. Comparison of atomic scattering factors.

Since the wavelength of electrons is very much less than that of X-rays, θ will be much smaller for a given value of sinθ/λ. As a consequence of this, the scattering factor for electrons falls off very rapidly, and therefore diffracted beams have appreciable intensity only at rather small Bragg angles.

As with the interactions of electrons with the specimen, certain approximations can be made regarding electron diffraction. The use of approximations provides relatively simple relationships that allow for determination of diffraction geometry, as seen with the concept of the Ewald sphere being regarded as planar. Also, approximations enable estimates to be made of the relative intensities in diffraction patterns.

2.9. Kinematical Scattering.

In essence, the kinematical theory of electron diffraction assumes that for sufficiently thin crystals, the intensity of the incident beam is not significantly reduced in its passage through the crystal and that a diffracted beam is not re-scattered.

In the previous treatment of the geometry of electron diffraction, it was utilisation of kinematical approximations that allowed derivation of the stated geometrical

relationships. Kinematical theory of electron diffraction also allows for assumptions to be made regarding the intensity of diffracted beams.

Information relating to the relative positions of atoms within a crystal is contained in the intensities of diffraction spots. Kinematical diffraction allows for the derivation of a simple relationship between these intensities and the required atomic positions.

Bragg reflections from a crystal will have the scattered amplitudes from all real space lattice points adding in phase. If there is an identical atom at each point in the lattice, the scattering from each atom is given by the atomic scattering factor, $f(\mathbf{u})$ so that the total scattering in the direction defined by a reciprocal lattice point is proportional to $Nf(\mathbf{u})$, if there are N atoms present. Then, the intensity scattered in this direction defined by $\mathbf{u}=\mathbf{h}$ is proportional to $N^2 f^2(\mathbf{u})$.

For most crystals there is in fact more than one atom associated with each real space lattice point. The scattering amplitudes from all atoms in the unit cell must be summed, with their phase factors corresponding to the atom position, $\mathbf{r_j}$, with respect to the unit cell origin. Therefore, instead of an amplitude, $f(\mathbf{u})$ from each lattice point, there is a contribution from all atoms in the unit cell given by,

$$\Phi(\mathbf{u}) = \sum_i f_j(\mathbf{u}) \exp(2\pi i\, \mathbf{u} \cdot \mathbf{r_j}) \qquad \text{Equation 2.14.}$$

The total integrated intensity for a Bragg reflection then becomes proportional to $N|\Phi(u)|^2$. The quantity $|\Phi(u)|$ is the structure amplitude. The intensity is that of the $\mathbf{h}$-reciprocal lattice point and as such,

$$\Phi_{\mathbf{h}} = \sum_i f_j(\mathbf{h}) \exp\left[2\pi i\left(h\frac{x_i}{a} + k\frac{y_i}{b} + l\frac{z_i}{c}\right)\right] \qquad \text{Equation 2.15.}$$

and therefore,

$$I_{hkl} \propto |\Phi(hkl)|^2 \qquad \text{Equation 2.16.}$$

As electrons are scattered by the potential distribution of the atoms, $\varphi(\mathbf{r})$, the atomic scattering factor for a single atom is given by,

$$f(\mathbf{u}) = \Im\varphi(\mathbf{r}) \qquad \text{Equation 2.17.}$$

where $\Im$ represents a Fourier transform integral.

Similarly, the structure amplitude $\Phi(\mathbf{u})$ is given by the Fourier transform of the potential distribution in the whole unit cell. This Fourier transform integral can be inverted leading to the potential distribution in the unit cell being obtainable from the scattered amplitude,

$$\varphi(\mathbf{r}) = \Im^{-1}\Phi(\mathbf{u}) = \int \Phi(\mathbf{u}) \exp(-2\pi i \mathbf{u}\cdot\mathbf{r}) d\mathbf{u} \qquad \text{Equation 2.18.}$$

The integral can be replaced by the summation,

$$\varphi(\mathbf{r}) = \sum_{hkl} \Phi(hkl) \exp\left[-2\pi i\left(\frac{hx}{a} + \frac{ky}{b} + \frac{lz}{c}\right)\right] \qquad \text{Equation 2.19.}$$

Equation 2.19. is identical to the expression relating the electron density distribution, $\rho(\mathbf{r})$ to the structure amplitudes, F_{hkl}, for X-ray diffraction. For the X-ray case, this relationship forms the basis of structure analysis from diffraction data. In an analogous manner, Equation 2.19. provides the foundation for structure elucidation using electron diffraction data.

As with X-ray diffraction, the outstanding problem with this approximation is that it is the intensities, proportional to $|\Phi_{\mathbf{h}}|^2$, rather than the amplitudes, $\Phi_{\mathbf{h}}$ that are derived directly from kinematical diffraction data. Thus, in the recording of intensities the phase of structure amplitudes $\Phi_{\mathbf{h}}$ are lost, illustrating the fundamental crystallographic phase problem.

2.10. Limitations of Kinematical Approximation.

The kinematical approximation is based upon the assumption that scattering amplitudes are directly proportional to the three dimensional Fourier transform of the potential distribution, $\varphi(r)$,

$$V(u) = \int \varphi(\mathbf{r}) \exp\{2\pi i \mathbf{u} \cdot \mathbf{r}\} d\mathbf{r} \qquad \text{Equation 2.20.}$$

so that the potential $\varphi(r)$ takes the place of $\rho(r)$, the charge density distribution, relevant for X-ray scattering.

Thus, it can be stated that the kinematical approximation is described as a single scattering event since it is first order with respect to the scattering potential. This approximation is attractive for structural analysis in electron microscopy due to the linear relationship of transmitted wave function to the object structure. The important property that this linearity provides is that the approximation is invertible, in the sense that the object structure can be retrieved directly from the transmitted wave function.

The validity of the kinematical approximation as a basis for structure analysis is limited. The domain of validity for quantitative structure analysis of organic crystals was evaluated by Jap and Glaeser[26] in terms of crystal thickness, structural resolution and incident electron energy. The thickness for which the approximation gives reasonably accurate zone axis patterns from single crystals is the order of 100Å for 100keV electrons and increases approximately as σ^{-1} for higher energies.

For the validity domains calculated in the above work, the cases studied had their potential projected along a crystal axis, however, the single scattering approximation may have a larger domain of validity for projections that are not parallel to the zone axis.

2.11. Dynamical Diffraction.

Having considered what are very generalised approximations about the scattering of electrons by solids, along with the limitations inherent in these approximations, the formalism has to be further developed to one in which the diffraction process is more thoroughly understood.

Although different theories have been postulated, they are basically equivalent in that they must produce the same results. Each theory provides a different insight into the diffraction process and can allow convenient descriptions of particular special cases.

Bethe[27] set up the wave equation for electrons in the periodic potential field of a crystal and then applied the boundary conditions appropriate for a crystal surface with incident and diffracted waves in a vacuum.

MacGillavry[28] further developed the general theory of Bethe, with a simplification by limiting the number of diffracted beams to one, so that only the interactions of the incident and diffracted beam are considered. However, Dorset[29] has stated that while this procedure can be used to detect the presence of dynamical scattering, the two-beam model also can be shown to be inadequate for explaining how individual intensities are influenced by multiple scattering. In consequence, this theory is inadequate for explaining electron scattering from single crystals. Failure of this two-beam theory led to the later development of the many-beam formalism as described by Hirsch et al.[19]

The alternative approach is to consider the progressive modification of wave amplitudes or diffracted beam amplitudes as the electron beam passes through the crystal. Various theoretical concepts are based on this, including the multislice proposal of Cowley and Moodie[30], where scattering is by planar sections of the specimen. Another theoretical direction was taken by Howie and Whelan[31] in using differential equations to give the progressive modification of diffracted beam amplitudes by the scattering processes.

Other approaches involve Bethe-type, Bloch wave formulations using a scattering matrix, introduced by Sturkey[32], which was the basis of the first attempts at many-beam dynamical intensity calculations.

Of the above mentioned theoretical approaches only the Bloch wave formulation and Multislice theory will be discussed. For a more detailed discourse on the theories involved refer to the coverage in the texts of Hirsch et al[19] and Buseck et al[20].

2.12. The Bloch Wave Formulation.

Following from the work of Bethe, solutions are sought for the Schrödinger equation for electron waves in the periodic potential field of a crystal. A solution consists of an incident beam, represented by wave vector, $\mathbf{k_0}$ directed to the reciprocal lattice origin and many diffracted beams represented by wave vectors, $\mathbf{k_h}$, directed to the reciprocal lattice points. All of these wave vectors originate at the same point, *L*. The collection of incident and diffracted waves that constitute one solution of the wave equation is known as a Bloch wave.

The wave vectors $\mathbf{k_h}$ do not have the same lengths and there is no restriction that requires all diffracted beams to have the same wavelength as seen in kinematical theory. Instead, with each wave vector, $\mathbf{k_h}$, there is an associated amplitude $\Psi_\mathbf{h}$ which will be large if the **h**- reciprocal lattice point lies close to the Ewald sphere.

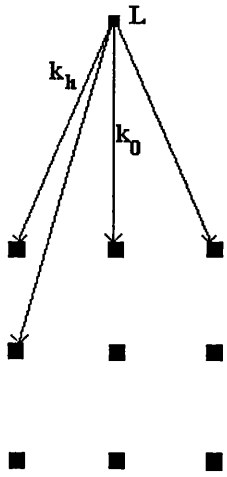


Figure 2.6. Illustration of Bloch wave description.

As the direction of the incident wave vector, $\mathbf{k}_0$, varies, so does its length, depending on the diffraction effects. Hence, instead of being a sphere of radius $2\pi/\lambda$, the locus of point L is a surface, known as the dispersion surface.

In general, there are many solutions to be found to the wave equation and it follows that there will be many Bloch waves within a crystal. In practice, only a finite number of Bloch waves need be considered.

A wave function satisfying the wave equation in a periodic potential field must have the periodicity of the crystal lattice and so the Bloch wave takes the form,

$$\psi(\mathbf{r}) = \sum_{\mathbf{h}} \Psi_{\mathbf{h}} \exp(2\pi i \mathbf{h}\cdot\mathbf{r}) \cdot \exp(i\,\mathbf{k}_0\cdot\mathbf{r}) = \sum_{\mathbf{h}} \Psi_{\mathbf{h}} \exp(i\,\mathbf{k}_{\mathbf{h}}\cdot\mathbf{r}) \qquad \text{Equation 2.21.}$$

where $\mathbf{k}_0$ is the incident wave vector and $\mathbf{k}_h = \mathbf{k}_0 + 2\pi\mathbf{h}$ where $\mathbf{h}$ is a reciprocal lattice vector.

To determine what Bloch waves are generated in a crystal and more importantly, what their relative amplitudes are for a given wave incident on a crystal, the boundary conditions for the entrance face must be applied. The requirement is that the

projection onto the crystal boundary of any incident wave vector in the crystal must be equal to the projection on the boundary of the wave vector, **K** of the incident wave in a vacuum.

For each Bloch wave there is a corresponding point L^i and a corresponding surface that is the locus of point L^i. All points L^i must lie on a line perpendicular to the surface. Since the positions of the points L^i define all the wave vectors for all the Bloch waves, the problem then is to solve the wave equation to deduce the relative strengths of the various Bloch waves and the amplitudes of the diffracted waves within each Bloch wave. The usual form is to employ a matrix equation with eigenvalues and eigenvectors representing strength and amplitudes of Bloch waves for the solution. This theoretical approach can form the basis of methods of computation of diffraction intensities and amplitudes.

2.13. Multislice Theory of Dynamical Electron Diffraction.

Having briefly mentioned the two-beam theory of dynamical diffraction and the reasons for its non-inclusion in common descriptions of dynamical diffraction, a description is now given of the multislice theory of Cowley and Moodie.[30]

As previously described, the reciprocal lattice is sampled by an Ewald sphere of very large radius and as such, the interaction of all simultaneously diffracted beams must be taken into account. More simply, not only has the association of a diffracted beam to the incident beam to be considered, but also the interactions and association of a diffracted beam with all other diffracted beams. Therefore, a relevant dynamical scattering theory has to account for interactions of all beams involved in the diffraction phenomena. Several different accounts of this *n*-beam dynamical theory have been given.[30,31,33,34].

It is often most convenient to employ the multislice method of Cowley and Moodie for calculating dynamical structure factors and phases.

As the incident electron wave passes through a crystal, it can be considered to have been modified in two ways. The phase of the wave is affected by traversing the varying potential, $\varphi(\mathbf{r})$ of the crystal and the wave is spread by Fresnel diffraction or propagation effects. In this approach, the two effects are separated, dealt with individually and then recombined.

The phase changes are considered to be concentrated on a series of planes spaced a distance of Δz apart.

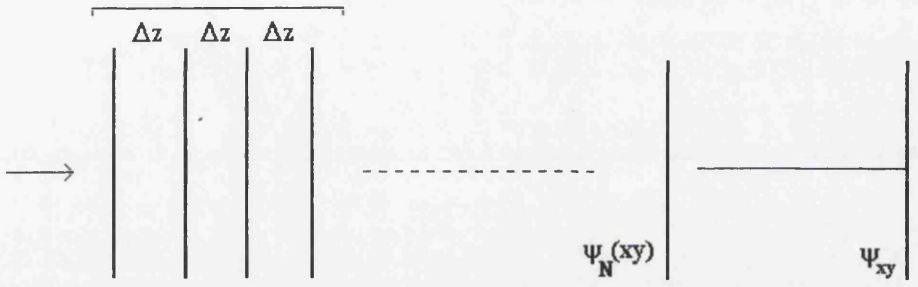


Figure 2.7. Description of Multislice theory.

For each thickness, Δz, the phase change due to the potential $\varphi(\mathbf{r})$ is considered to take place on one plane and is found by

$$q_n(xy) = \exp[-i\sigma\varphi_n(xy)] \qquad \text{Equation 2.22.}$$

where

$$\varphi_n(xy) = \int_{Z_n}^{Z_n+\Delta z} \varphi(\mathbf{r})\, \mathbf{dz} \qquad \text{Equation 2.23.}$$

where $q_n(xy)$ is the transmission function.

The Fresnel diffraction effects are taking place between planes and are represented by convolution with the propagation function $p_n(xy)$. As the small angle approximation holds for electron scattering, the propagation function is

$$p(xy) = \exp\left[-\pi i \left(x^2 + y^2 \right) / \Lambda z \cdot \lambda \right] \qquad \text{Equation 2.24.}$$

Then, the wave function after the n^{th} slice is given in terms of the wave function after the (*n-1*) slice as,

$$\varphi_n(xy) = \left[\varphi_{n-1}(xy) * p_{n-1}(xy) \right] \cdot q_n(xy) \qquad \text{Equation 2.25.}$$

The wave function at the exit face of the specimen can then be obtained by performing the series of alternating convolutions and multiplications given in Equation 2.25. for the successive slices of the specimen, starting from the incident wave function.

As before, when the scattering object is a crystal, the potential distributions $\varphi_n(xy)$ for the individual slices are periodic in two dimensions and as such the transmission function is also periodic.

$$q_n(xy) = \sum_{hk} Q_n(hk) \exp\left[-2\pi i \left(\frac{hx}{a} + \frac{ky}{b} \right) \right] \qquad \text{Equation 2.26.}$$

If, in the limiting case, slice thickness tends to zero, the first order approximation to $q(xy)$ can be made

$$q_n(xy) = 1 - i\sigma\varphi(xyz_n)\Delta z \qquad \text{Equation 2.27.}$$

and

$$= 1 - i\Delta z \cdot \sigma \sum_h \sum_k \sum_l \Phi(hkl) \exp\left[2\pi i \left(\frac{hx}{a} + \frac{ky}{b} \right) \right] \exp\left(2\pi i \frac{l}{c} z_n \right) \qquad \text{Equation 2.28.}$$

For a periodic object such as a crystal it is more convenient to deal with the set of discrete structure amplitudes and as such the exit wave function can be represented by

$$\Psi_n(uv) = \left[\Psi_{n-1}(uv) \cdot P_n(uv)\right] * Q_n(uv) \qquad \text{Equation 2.29.}$$

where

$$\Psi_n(uv) = \sum_h \sum_k \sum_l (hk)\delta\left(u - \frac{h}{a}, v - \frac{k}{b}\right) \qquad \text{Equation 2.30.}$$

Cowley and Moodie showed that the iteration of Equation 2.29 through N slices of the crystal could be expressed in a form analogous to that of a Born series. Thus, they had arrived at a general solution of the many-beam, dynamical diffraction problem in terms of a doubly infinite series. Summing the series in various ways gave a single series from which, under the appropriate assumptions, it is possible to derive the kinematical approximation, the two-beam approximation, the phase object approximation and some other less useful approximations.

As a basis for the calculation of diffracted amplitudes using this multislice approximation, the assumption is made of a finite number of thin slices. The diffracted amplitudes are calculated by iteration of Equation 2.27, once for each slice. The number of iteration thus depends on the crystal thickness.

2.14. Multiple Scattering.

Previous sections have considered several theories and approximations regarding electron diffraction. An understanding was gained of the non-simplistic relationships that can exist between a given structure and the intensity information present within a diffraction pattern.

Other factors exist that have the propensity to cause deviations from the above discussed theoretical constraints. Multiple scattering is one such effect that can cause perturbations to intensity information.

The presence of multiple electron scattering is favoured by the existence of a layered structure. This layered structure allows the strongly excited beams from upper crystalline layers, which are incoherently related to each other, to act as ancillary primary beams for the lower layers of the crystal.

Observation of space group forbidden reflections in electron diffraction patterns was commonly thought to be due to *n*-beam dynamical scattering. In fact, appearance of these forbidden reflections is due to secondary or multiple scattering[29] and increased thickness of the crystal.

The pronounced effect of multiple scattering, besides concealment of actual symmetry, is the spurious increase of data resolution. Multiple scattering involves self-convolution of intensities such that all reflections in the diffraction pattern are involved. To correct for this would involve deconvolution of the whole data set, which would be a difficult task. Thus, the best correction for multiple scattering is taking all steps to ensure its avoidance.[35]

2.15. Radiation Damage.

The final effect to be considered, with respect to perturbations of structural information obtainable from the electron microscope, is that of radiation damage. In essence, the degradation of samples on exposure to the electron beam is the fundamental limitation to the collection of structural information. This is especially true when the aim of the electron microscope study of a given material is to obtain high resolution images.

The major problem in electron microscope studies of organic pigments, of which this research is especially concerned with, is their radiation sensitivity.[36] Degradation of the crystal in the electron beam can affect both the size and shape of the particle under investigation, as well as eliminating information about the internal order of the particle.

Discussions of the specifics of the processes involved in radiation damage of organic materials have been given extensive treatment by several researchers.[37,38,39,40]

Although, consideration has to be made of the existence of such processes, it is beyond the brief of this thesis to become involved in any further discourse. In this section, only an introduction to radiation damage and more importantly, discussion of the methods employed to alleviate its effects will be given.

A proportion of an electron beam incident on the specimen undergoes inelastic interactions with that specimen. Subsequently, energy is transferred from the electron to the target atom. This energy can be dissipated by a variety of means including ionisation, formation of excited molecular states or even radical reactions. Many of these processes will cause bond rupture within the sample.

The excited molecular species produced by this primary radiation damage can then undergo secondary chemical reactions which can involve neighbouring molecules in a complex series of steps. Thus, the consequences of these inelastic events are that the damage caused is normally of a chemical origin.[41] The rates of these chemical reactions and their effect upon the molecular and crystal structure of the crystal, will vary from specimen to specimen as they are dependent on the chemical identity of the sample.

Damage is evident as loss of structure in high resolution imaging, and the feature of damage which is immediately apparent for diffraction is that the diffraction pattern fades from the outside towards the centre. Also, radiation damage is seen to be present at lower electron accelerating voltages, therefore damage is voltage dependent. The rate of loss of structure, as seen with diffraction patterns illustrates the point that at higher resolution, the problem of radiation damage increases, as fine detail is more susceptible to damage.

Methods employed to improve resistance to damage include reduction of temperature.[40,42,43,44] This has the effect of stabilising the crystalline order of the

damaged sample. However, this may not stop the initial radiation damage but can slow down the diffusion of the chemically reactive products of initial damage.[45]

Another method of increasing resistance to electron beam damage is that of encapsulation.[38,46] In simple terms this may involve protection of the specimen by deposition of a protective film of carbon.[47] The protection offered by the carbon is a function of the nature of the peripheral atom of the specimen molecule and not related to the structure of the parent molecule.

Employment of minimum exposure techniques[48] can also aid in the reduction of radiation damage to the specimen. The essence of minimum exposure lies in attempting to record images or diffraction patterns from a specimen as close as possible to its original state. Two variants of this method exist. The first is to deflect the electron beam away from the normally exposed area and carry out focusing on the specimen in this deflected position before returning to normal illumination conditions, along with recording of image at the same time. Alternatively, focusing is carried out as normal, then the beam is either deflected or blanked above the specimen. shifting the specimen then brings an unexposed area into the beam path and then unblanking of the beam and exposure to specimen begin simultaneously with image capture.[49]

Finally, use of a sensitive photographic emulsion, such as X-ray film allows for the use of comparatively small electron exposure times, again reducing time of exposure of sample to the damaging electron beam.

2.16. Conclusions.

Extensive discussion has been made of the theoretical fundamentals of electron diffraction. Also, an awareness has been achieved of some of the causes of deviation of observed electron diffraction intensities from a straightforward representation of the unit cell. Only by taking all of these points into consideration and by ensuring use of experimental conditions that favour collection of quasi-kinematical intensity

information, can one proceed with any confidence towards the elucidation of structural information using electron diffraction data.

3.0. ELECTRON CRYSTALLOGRAPHY.

3.1. Introduction.

"*Electron crystallography by its broadest definition includes any use of electron diffraction, electron scattering or direct imaging in which the analysis of data is essentially crystallographic in its execution or objectives.*[50] *This has led to its application in both biological and chemical domains.*"

In molecular biology, crystallographic image processing is a standard technique for structure determination of proteins and other biological macromolecules by electron microscopy. The first practical use of the electron microscope for the three-dimensional crystallographic structure determination of a biological specimen was described by DeRosier and Klug.[51] A generalised approach of merging electron diffraction data from different tilt angles to obtain the 3D structure of bacteriorhodopsin was made by Henderson and Unwin.[52] More recent developments include the structures of such macromolecules as tubulin[53] and Aquaporin-1[54]. Thus, it can be seen that the terminology employed has a very broad frame of reference. Within the context of this thesis, a more precise definition of electron crystallography is required.

A further clarification is that of structural electron crystallography. This is characterised as the quantitative use of electron scattering data to determine the average structure of a crystal. With reference to the crystallographic phase problem and within the confines of this definition, an additional demarcation exists between strict *ab initio* phase determination and use of phases derived from electron microscope images.

3.2. Development of Electron Crystallography.

Early structure determinations were led by Russian researchers following from the characterisation of paraffin layer packing from electron diffraction data by Rigamonti.[55] Initial work by the Russian groups was based on intensity data from polycrystalline textures along with single crystal data. Among those organic structures studied, diketopiperazine[56], urea[57], thiourea[58,59] and copper salts of some amino acids [60, 61] were featured. Reasonable results were achieved with this approach using light atom structures, but difficulties were experienced when electron diffraction data from heavier atom structures were utilised.[62,63]

As a test of validity of structure analyses of this kind, comparisons were made between observed diffraction amplitudes and those calculated from a packing model based on structure solution from previous X-ray studies. No obvious correspondence was found leading to skepticism of the methodology from conventional X-ray crystallographers.[64]

In recent years there has been a resurgence in the use of electron diffraction data for structure elucidation of small molecules, mainly due to the championing of the subject by Dorset.[65] Initial work involved the techniques of traditional direct methods, which has stimulated the interest of other crystallographic research groups bringing with them new crystallographic techniques of phase determination to be applied to electron diffraction data. In fact, it has now been stated that electron crystallography is indeed a valid technique for quantitative *ab initio* structure analysis.[66]

At this point, a description of the theoretical framework of electron diffraction has been given, along with a perspective on the development of the subject. The next treatise is a theoretical consideration of phase determination and crystal structure analysis. Therefore a section that links these two areas is needed to describe the retrieval of intensity information from electron diffraction patterns required for subsequent structure elucidation.

3.3. Extracting Intensities from Electron Diffraction Patterns.

Before any use is made of electron scattering data in structure analysis, there has to be a quantification of the intensity information present in electron diffraction patterns. In the past this was achieved by digitisation of electron diffraction patterns using a microdensitometer. Corrections were made for background, non-linearity of the densitometer and curvature of the Ewald sphere. This methodology led Baldwin and Henderson to develop a program that would measure and evaluate electron diffraction patterns from two dimensional crystals of the membrane protein, bacteriorhodopsin.[67]

A new program has emerged for quantitatively evaluating electron diffraction patterns via digitisation using CCD cameras and conventional flatbed scanners. This facility is called ELD[68], which is part of the CRISP[69] suite of programs for crystallographic image processing. A digitised image of an electron diffraction pattern is imported into ELD either from disk, as a saved image or directly from CCD camera allied with a frame grabber.

Before the intensities of the electron diffraction peaks can be extracted, the lattice must be found with sufficient accuracy. The lattice is defined by the user, by specifying the indices of two diffraction maxima. The criteria for reflections chosen for lattice refinement is that they are neither too weak or strong and that they are without neighbouring reflections that are too strong. All other strong diffraction points are found automatically by the program and after two cycles of least-squares refinement, the exact lattice is established.

Once the lattice vectors have been found and refined, intensities are integrated for each peak found on the pattern from the exact positions predicted by the refined lattice vectors and (0,0) position. Peak heights are determined from a curve-fitting procedure, allowing ELD to estimate intensities quite accurately over a large dynamic range. The list of reflections found along with their extracted intensities can be saved to file for

subsequent use. A summary of the procedure involved in extraction of electron diffraction intensities within the ELD program is given in Figure 3.1.

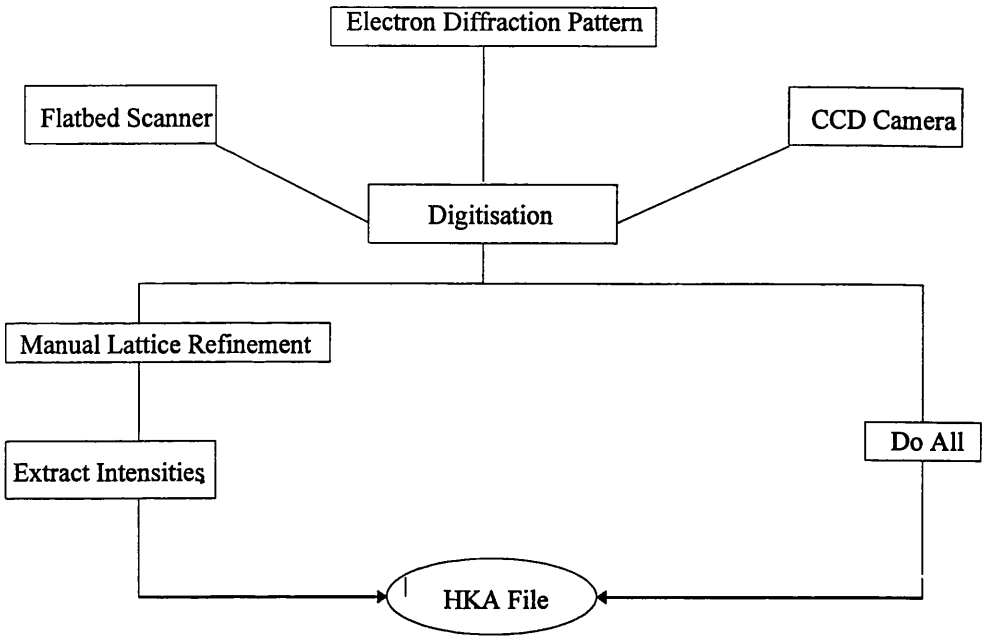


Figure 3.1. Outline of use of ELD.

Thus, suitable retrieval of intensity information has been achieved and as was stated in the theoretical preamble, before structure reconstruction can be realised, the fundamental problem of missing phase information has to be overcome.

3.4. Solution of the Phase Problem.

Electron crystallography has mirrored the development of X-ray crystallography earlier this century. This is also true for procedures of overcoming the phase problem, where there has been a crossover of some of the conventional X-ray methods for phase determination to electron crystallography. These 'Direct Methods' will be mentioned and a discussion given of recently developed phase determining methods, along with a description of phase retrieval from images which is peculiar only to electron microscopy.

3.5. Phase Retrieval from Images.

It has been stated that an advantage that electron crystallography has over X-ray crystallography is that the focusing of electrons by electromagnetic lenses allows for the acquisition of images - but why should this be an advantage?

Quite simply, the reasoning for this is with regard to the crystallographic phase problem. A low resolution electron microscope image, under certain conditions and employing certain approximations, can be directly related to the projected potential of the crystal. It follows that the Fourier transform of the image will provide a section of the reciprocal space and as such structure factors and hence, amplitudes and phases.

This can only be found to be true if certain factors are considered. Assumption is made of the existence of conditions where the weak phase object approximation or single scattering approximation are valid. Also, correction for the contrast transfer function, which modulates the amplitudes and phases, is taken into account. It must also hold that the low resolution electron microscope images have been obtained at approximately Scherzer defocus, in that phases from images are shifted by 180° with respect to phases of structure factors for electrons.

Admittedly, this is a very generalised conjecture of quite complex theories involving image formation, but as was stated previously, the aim of this thesis is not to provide a rigorous discourse on image theory but to show the generalised applications of such detailed theory. Again for an extended approach, recourse to the appropriate text[70] is recommended.

Application of this school of thought regarding phase retrieval have been successful in structural terms, initially in biophysical areas[71] and subsequently to inorganic materials.[72] Another approach has been taken, whereby phases derived from low resolution images by methods described are used as the start point for phase extension

via a variety of direct phasing routes.[73,74,75] It is these routes which will be explained in ensuing sections.

3.6. Trial and Error methods.

In attempting *ab initio* structure determination from diffraction data, it is possible that by using a trial and error guess of the atomic or molecular packing, a correct structure solution could be arrived at.

By using trial positions, it is possible to calculate the structure factors, F(**h**). Use of the crystallographic residual or *R*-factor allows for a measure of correctness of the trial structure.

$$R = \frac{\sum_{\mathbf{h}} \left\| F_{\mathbf{h}}^{O} \right| - k \left| F_{\mathbf{h}}^{C} \right\|}{\sum_{\mathbf{h}} \left| F_{\mathbf{h}}^{O} \right|} \qquad \text{Equation 3.1.}$$

where $k = \sum_{\mathbf{h}} \left| F_{\mathbf{h}}^{O} \right| \Big/ \sum_{\mathbf{h}} \left| F_{\mathbf{h}}^{C} \right|$ Equation 3.2.

Trial and error methods consist of a systematic trial of all structural hypotheses compatible with the known physical and chemical properties of the crystal being considered. Utilisation of the *R*-factor allows a check to be made as to whether the structure amplitudes calculated for the model are in qualitative agreement with the observed magnitudes.

As shown by Wilson[76], atoms placed randomly in the unit cell give a crystallographic residual value of 0.83 and 0.59 for centrosymmetric and non-centrosymmetric space groups respectively. It has also been shown that if a *R*-factor value of 0.45 is calculated, then the model under consideration has distinct merit. However, if a residual value of 0.35 is reached, then a model has been proposed that can be refined to the correct structure. Finally, if a *R*-factor value of 0.25 is calculated then the

atomic positions found should be within 0.1 Å of their correct value. Therefore, when a minimum value of the crystallographic residual has been achieved, structure refinement has been realised.

The difficulty in using this trial and error approach is that for any unit cell, there is an infinite number of possible packing arrangements for a given molecule, that will fit the unit cell. However, not all of these possibilities are energetically favourable.[77]

Another approach, especially in the study of polymer chains, has been to utilise molecular geometry suggested from an X-ray structure to search for the best fit to the data. Monomer units define the conformationally invariant part of the polymer repeat. Therefore, structure search using electron diffraction data involves investigation of the conformational changes around linkage bonds.[78,79]

Structures of flexible aromatic molecules have been determined using a procedure whereby molecular packing and conformation found by energy minimisation techniques, have been verified by direct methods.[80,81]

3.7. Patterson Synthesis.

Use of the Patterson function was the major means of determining crystallographic phases in the early days of X-ray crystallography.[82]

$$P(\mathbf{uvw}) = V^{-1}\sum |F|^2(hkl)\cos 2\pi\ (h\mathbf{u} + k\mathbf{v} + l\mathbf{w}) \qquad \text{Equation 3.3.}$$

Maxima in the Patterson function are found to correspond to all possible interatomic vectors within the unit cell. The height of each peak will be proportional to the product of the atomic numbers of the atoms connected by the vector **u**, multiplied by the multiplicity of the same vector.

The Patterson function will have the same periodicity as the electron density and therefore the size of the unit cell will be identical. However, the number of peaks in the Patterson map is much greater than in the corresponding electron density map, in that N atoms in the unit cell give rise to N^2 peaks in the Patterson map.

The Patterson function has also been employed for structure analyses in electron crystallography, including the determination of methylene chain packing in alkane derivatives.[83] Likely molecular orientation assignation can be aided by the use of the Patterson map, as seen in the electron crystallographic investigation of phospholipids.[84]

A likely drawback of the use of Patterson maps in electron crystallography is the poor detectability of heavy atom positions[29] due to the range of electron scattering magnitudes being compressed relative to that for X-ray form factors.[25]

3.8. Direct Methods of Crystallographic Phase Determination.

Direct methods of X-ray crystallography are defined as the class of methods which attempt, by mathematical means, to derive the phases of the structure factors using only intensity information.[85] These direct methods have been shown to be applicable to phase determination in electron crystallographic analyses. An overview will be given of different aspects of conventional direct methods.

In general the phase and amplitude of a wave are independent quantities and in order to understand how it is possible to relate these two quantities, two important properties of the electron density function have to be considered. The first is that obviously the electron density must be positive at every point within the crystal. The second property is that the electron density is composed of discrete atoms, thus the constraint of atomicity.

Historically, the first mathematical relationships capable of giving phase information were obtained in the form of inequalities by Harker and Kasper[86] and then further developed by Karle and Hauptman.[87]

Three separate papers were published in Acta Crystallographica by Sayre[88], Cochran[89], and Zachariasen[90] which provided a mathematical relationship linking the phases of a centrosymmetric crystal where the phases are restricted to 0 and π. Thus, the phase $\phi_{\mathbf{h}}$ can be thought of as a sign, $s_{\mathbf{h}}$ of the structure factor where $\phi_{\mathbf{h}}=0$ corresponds to $s_{\mathbf{h}}=+1$ and $\phi_{\mathbf{h}}=\pi$ corresponds to $s_{\mathbf{h}}=-1$. For this reason, the relationship given in these papers is referred to as the 'sign relationship'.

This triplet product sign relationship can be expressed in the form given in Equation 3.4.

$$s(\mathbf{h})s(\mathbf{k})s(\mathbf{h}+\mathbf{k}) \approx +1 \qquad \text{Equation 3.4.}$$

where $\approx$ means 'probably equal'. Another way of expressing this relationship is,

$$s(\mathbf{h}+\mathbf{k}) \approx s(\mathbf{h})s(\mathbf{k}) \qquad \text{Equation 3.5.}$$

such that if the signs of two structure factors are known, then the probable sign of the third can be obtained. The probability of the relationship depends on the product of the three structure amplitudes and where all three are large, the probability can be close to or even equal unity.

In 1953 Hauptman and Karle further developed this relationship to allow the basic concepts and probabilistic foundations of direct methods to be established.[91] The key to the solution of non-centrosymmetric structures was provided by Cochran in 1955.[92] Here, a relationship was introduced for general phases

$$\phi_{\mathbf{h}1}+\phi_{\mathbf{h}2}+\phi_{\mathbf{h}3} \approx 0 \text{ (modulo } 2\pi) \qquad \text{Equation 3.6.}$$

where $\approx$ now means 'is concentrated about', vector index, **h** represents *h,k,l* and $\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0$ is the null vector. Therefore, if the structure factors are all large, then the sum of the three phases will tend to be close to 0, 2π, 4π, etc.

As statistical or probabilistic methods are to be applied to structure factor amplitudes, which themselves are derived from experimental intensity information, a correction has to be made for the fall off of scattering factors with diffraction angle.

$$F(hkl)=\frac{1}{Lp}\sqrt{I(hkl)} \qquad \text{Equation 3.7.}$$

where *Lp* is the Lorentz polarisation correction factor.

Measured structure factors are found to decrease with increasing diffraction angle. To fulfill the criteria of atomicity, where the structure is composed of point atoms at rest, use has to be made of the normalised structure factor, $E_{\underline{h}}$, defined in Equation 3.8.

$$\left|E_{\underline{h}}\right|^2=\frac{K\left|F_{\underline{h}}\right|^2}{\varepsilon_{\underline{h}}\sum_{j=1}^{N}f_j^2e^{-2B\sin^2\theta/\lambda^2}} \qquad \text{Equation 3.8.}$$

where f_j is the scattering factor, λ is wavelength, $\varepsilon_{\underline{h}}$ is a statistical weight that compensates for special classes of reflection, *K* is a scale factor which places $F_{\underline{h}}$ on an absolute scale and *B* is the isotropic temperature factor where,

$$B=8\pi^2\langle u\rangle^2 \qquad \text{Equation 3.9.}$$

where $\langle u\rangle^2$ is the mean squared amplitude of atomic vibration. Both *B* and *K* in the above equation can be derived from a Wilson plot.[93]

It is easier to understand how this derivation occurs, if the normalised structure factor is expressed in a slightly different way.

Another expression for the normalised structure factor is given in Equation 3.10.

$$|E_{\underline{h}}|^2 = I_{\underline{h}}^{obs} \Big/ \varepsilon \sum_i f_i'^2 \qquad \text{Equation 3.10.}$$

where ε is as defined above and f' is as defined below,

$$f' = f \exp\left(-B_{iso} \sin^2\theta / \lambda^2\right) \qquad \text{Equation 3.11.}$$

A plot of the natural logarithm of the right hand side of Equation 3.11., against median values of $\sin^2\theta/\lambda^2$ will approximate to a straight line. A least squares line can then be calculated and B is obtainable from the slope and K from the intercept at $\theta=0°$.

Normalised structure factors are found by definition to satisfy the relationship, $\langle |E|^2 \rangle = 1$.

By using normalised structure factors, the Cochran phase relationship can be represented by the quantity,

$$\Phi_3(\mathbf{h},\mathbf{k}) = \phi(\mathbf{h}) + \phi(\mathbf{k}) + \phi(\bar{\mathbf{h}} + \bar{\mathbf{k}}) \qquad \text{Equation 3.12.}$$

which has the probability distribution,

$$P[\Phi_3(\mathbf{h},\mathbf{k})] = \frac{\exp\{-\kappa(\mathbf{h},\mathbf{k})\cos[\Phi_3(\mathbf{h},\mathbf{k})]}{2\pi I_0[\kappa(\mathbf{h},\mathbf{k})]} \qquad \text{Equation 3.13.}$$

where $\kappa = 2N^{-1/2}\left|E(\mathbf{h})E(\mathbf{k})E(\bar{\mathbf{h}}+\bar{\mathbf{k}})\right|$

Another way to express Equation 3.12. by changing the form of the indices is

$$\phi(\mathbf{h}) \approx \phi(\mathbf{k}) + \phi(\mathbf{h}-\mathbf{k}) \qquad \text{Equation 3.14.}$$

then the probability distribution is for values of $\phi_{\mathbf{h}}$.

The relationship given in Equation 3.14. gives a probable value for $\phi_{\mathbf{h}}$ when there is a pair of known phases $\phi(\mathbf{k})$ and $\phi(\mathbf{h}-\mathbf{k})$. This was developed upon by Karle and Hauptman[94] who provided the tangent formula. Where there are several such triplets for a given $\phi(\mathbf{h})$, the tangent formula given in Equation 3.15, allows for an estimate for $\phi(\mathbf{h})$ when several pairs of known phases are available.

$$\tan[\phi(\mathbf{h})] \approx \frac{\sum_{\mathbf{k}} \kappa(\mathbf{h},\mathbf{k}) \sin[\phi(\mathbf{k}) + \phi(\mathbf{h}-\mathbf{k})]}{\sum_{\mathbf{k}} \kappa(\mathbf{h},\mathbf{k}) \cos[\phi(\mathbf{k}) + \phi(\mathbf{h}-\mathbf{k})]} \qquad \text{Equation 3.15.}$$

It is this tangent formula which is the basis for some of the most commonly used computer programs for *ab initio* phase determination in X-ray crystallography, an example of which is SHELX.[95]

3.9. Multisolution Methods in Crystallographic Phase Determination.

The basic idea of multisolution methods is that instead of using symbols for phase assignment, starting phases are given approximate numerical values. Initial errors present due to these values are found not to spoil subsequent structure solutions.

The starting set of phases, defined by a convergence procedure, will include the origin, enantiomorph fixing reflections if necessary and a limited number of other phases necessary to initiate the phase expansion process via the tangent formula.

The number of possible combinations of numerical phase values developed by the tangent formula grows very rapidly. Only by limiting the number of starting set reflections is it possible to keep calculations computationally feasible. This is greatly helped by use of the so-called magic integers[96] which allow a considerable reduction of the number of combinations with a minimum increase in phase error.

Most multisolution programs such as MULTAN[97] are based mainly on the use of triplets estimated by Cochran's formula. It is possible using multisolutional methods to solve structures with up to 70 atoms in the asymmetric unit. However, two main limitations exist regarding these methods. The first is that the initial steps of the phase expansion are very critical due to the restricted number of starting set phases. This effect will be amplified by the chain nature of the phase expansion process by means of the tangent formula.

3.10. Figures of Merit.

The phase determination process invariably leads to more than one solution. Given several sets of phases, it would be time consuming to compute and attempt to interpret all the corresponding electron density maps to see which yield the correct structure. Instead, it is easier to compute some appropriate functions called figures of merit which allow an *a priori* estimate of the correctness of each phase set.

For a figure of merit to be independent of the phasing technique, it should use phase relationships that have not been used as a phasing technique. This leads to a situation where a relationship may be strong enough for calculation of a figure of merit, but is also good enough to derive phases. Generally a compromise is reached regarding these two situations. Several figures of merit are in use in traditional direct methods and following are brief descriptions of the four most commonly used.

3.10.1. ABSFOM.

ABSFOM is a measure of the internal consistency of the triplet invariants used in estimating the phases.

$$ABSFOM = \sum_{h}(\alpha_{h} - \alpha_{Rh}) \Big/ \sum_{h}(\alpha_{Eh} - \alpha_{Rh}) \qquad \text{Equation 3.16.}$$

where $$\alpha_{Eh}{}^{2} = \sum_{k} A_{hk}^{2} + \sum_{k}\sum_{k'} A_{hk} A_{hk'} \frac{I_1(A_{hk}) I_1(A_{hk'})}{I_1(A_{hk}) I_1(A_{hk'})} \qquad \text{Equation 3.17.}$$

$$\alpha_{Rh} = \left(\sum_{h} A_{hk}^{2} \right)^{\frac{1}{2}} \qquad \text{Equation 3.18.}$$

and $$A_{hk} = 2\left|E_{h} E_{k} E_{-h-k}\right| \Big/ \sqrt{N} \qquad \text{Equation 3.19.}$$

For random phases ABSFOM is zero. Typical values considered as giving a good indication about the phases set are values of ABSFOM of 1.0-1.3.

3.10.2. RESID.

RESID is a *R*-factor between estimated and actual values of α and since it is dependent on α, it must also rely on tangent refinement. RESID can also be defined on E-magnitude.

$$R_{\alpha} = 100 \times \sum_{h} \left|\alpha_{h} - \alpha_{Eh}\right| \Big/ \sum_{h} \alpha_{Eh} \qquad \text{Equation 3.20.}$$

A correct phase set should have a residual value less than 20%, although structures containing a large number of atoms in the unit cell may have correct phase sets with R_{α} in excess of this figure.

3.10.3. PSI-ZERO.

PSI-ZERO, ψ_0, is a measure of fit of small E-magnitudes. This figure of merit is largely independent of the phasing methods and is sensitive to atomic positions.

$$\Psi_0 = \frac{\sum_h \left| \sum_k E_k E_{h-k} \right|}{\sum_h \sqrt{\left(\sum_k \left| E_k E_{h-k} \right|^2 \right)}} \qquad \text{Equation 3.21.}$$

A value of ψ_0 less than one is usually indicative of a correct phase set. However, if all other figures of merit are good but a value of ψ_0 greater than one is obtained, then this indicates a correct fragment in the wrong position.

3.10.4. NQEST.

NQEST is summed over all negative quartets. Like ψ_0, NQEST uses the information contained in the small E-magnitudes although the information is used in a different way, thus making the two figures of merit independent of each other. This dependence on small E-magnitudes also makes NQEST largely independent of the phasing procedure. NQEST can lie anywhere in the range -1.0 to +1.0, with the most negative values likely to correspond to correct phase sets.

$$NQEST = \frac{\sum_{h,k,l,m} W_{hklm} \cos(\phi_h + \phi_k + \phi_l + \phi_m)}{\sum_{h,k,l,m} W_{hklm}} \qquad \text{Equation 3.22.}$$

where, for centrosymmetric reflections $W_{hklm} = |1 - 2P^+|$ Equation 3.23.

and for non-centrosymmetrics reflections $W_{hklm} = \frac{1}{\sigma^2}$ Equation 3.24.

and σ^2 is the variance of the quartet probability distribution.

3.11. Summing Up.

As was stated in the introduction to this section, the aim was to provide an overview of direct methods and therefore reference has to be made to the appropriate texts for a more thorough discussion of the relevant theoretical descriptions of direct methods.[98,99]

For detailed examples of the application of some of these conventional direct methods to phase determination in electron crystallography, again reference has to be made to the monograph of Dorset.[29]

As the research detailed in this thesis has applied maximum entropy and likelihood procedures to the determination of crystallographic phase information a full treatment of the theory will be given in the next section, combined with a discussion of the applications of this methodology to electron crystallography.

4.0 MAXIMUM ENTROPY AND LIKELIHOOD METHODS IN CRYSTALLOGRAPHY.

4.1 Limitations of Conventional Direct Methods.

Direct methods of crystallographic phase determination have been found to be an extremely powerful tool to the small molecule crystallographer, as illustrated by their use in routine structure solution to resolution of around 1.1 - 1.2 Å. However, these well developed methods can fail in certain circumstances.

Traditional direct methods require data at atomic resolution coupled with a complete sampling of reciprocal space. This criteria is justification alone for direct methods to be inapplicable to projection electron diffraction data. Problems are also encountered with powder and fibre diffraction data and also with certain aspects of macromolecular crystallography. Maximum entropy methods do not offer a completely different approach to that of direct methods, it is an improvement on conventional methods in terms of the probability theory on which they are based.

Before describing the maximum entropy formalism, it is pertinent to review in a simplistic way the assumptions of conventional direct methods, with regard to probability.

4.2. Probability Distributions in Direct Methods.

Direct methods are based on the idea of a probability distribution describing where atoms might be located in the, as yet unknown, structure contained in the unit cell. A probable distribution of atomic positions can be approximated to a conventional electron density map, where probabilities and not electron density are present at any given point.

Initially, this distribution is assumed to be uniform as we have no knowledge about the locus of atoms. The probable distribution of atomic positions changes as phases and other aspects of the structural arrangement become known. Owing to the mathematical relationship that exists between atomic positions and structure factors, any knowledge about one of these changes the probable distribution of the other. Thus, as phases become known, the probable distribution of atomic positions changes and so, too, does the probable distribution of the remaining phases. When enough phase information has been accumulated, the probability distribution of the remaining phases becomes so sharp as to allow determination of their values with great accuracy, and thus the crystal structure is solved.

The initial assumption that the probability distribution is uniform, is the safest one to make when no information is available regarding the atomic positions. However, once diffraction data has been collected, phases can be chosen for two or three strong reflections allowing the origin of the unit cell to be fixed. Doing so changes the uniform atomic probability distribution law irreversibly and now some regions of the unit cell are more likely than others to contain atoms. This new probability distribution is said to be conditional, in the sense that it is only valid given the specific phase choices that have already been made. Associated with this is also a new conditional probability distribution for the remaining unknown phases.

The problem of choosing new probability distribution laws for the random atomic positions and unknown phases, once some reflections have been phased is the quintessential phase problem of crystallography.

One approximation to the probable distribution of atomic positions is the Fourier synthesis calculated from the structure factors of the phased reflections. However, for a realistic map to be obtained, reasonable estimates for the phases and amplitudes of the missing reflections has to be introduced. Structure solution will be realised when the probable distribution of atomic positions and the electron density map converge to very similar functions.

4.3. Shortcomings of Conventional Direct Methods.

A fundamental limitation of conventional direct methods is that it assumes that atoms are distributed randomly, independently and uniformly in the asymmetric unit. This is quite obviously a false assumption as atomic positions are not random entities, but are in fact governed by strict structural criteria regarding bond lengths and geometries. As discussed by Bricogne[100], the limit theorems of probability theory used to approximate joint probability distributions have turned out to be unsuitable for large structure factor amplitudes. Traditionally, the Gram-Charlier or Edgeworth series[101,102] or formal expressions related to them have been used. Thus, use of these series based upon the above assumption of uniformity yield approximations to the probability distribution which are good only for small structure factor amplitudes. Therefore, the situation can arise where the approximate conditional distribution of phases are most accurate where least informative and least accurate where potentially most informative.[103] The difficulty is that there is no tractable, unique expression for the probability distribution into which the measured values of large structure factor moduli could be substituted, which would then yield the conditional distribution of the corresponding phases.

By considering these two limitations, it can be seen that there are intrinsic failings in the probabilistic foundations of direct methods, and as such, these can be overcome by utilisation of more powerful analytical techniques.

4.4 Bayesian Statistics.

Maximum entropy methods are based upon Bayesian statistics which themselves are derived from Bayes' theorem.[104] In its simplest form, this theorem can be stated as given in Equation 4.1.

$$P(E|F) \propto P(E)P(F|E) \qquad \text{Equation 4.1.}$$

where E and F are two events and P denotes probability.

The three components of the above equation are termed the *posterior*, the *prior* and the *likelihood* respectively. $P(F|E)$ is the posterior and is calculated after measurement and it is the probability of event F occurring given that event E has already occurred. $P(E)$ is the prior and is knowledge before measurement and can be thought of as the probability of event F occurring. Finally, $P(E|F)$ is the likelihood and consults the measured data and thus it is the probability of event E being true if event F has happened.

In this Bayesian approach, prior knowledge is first defined, then likelihood is used to consult the experimental data and then the two are used to produce a posterior probability which reflects modifications to the prior which arise from the data. Thus, the posterior generated can be used as the new prior and the whole process repeated in a cyclical way until likelihood remains unchanged. This is described in Figure 4.1.

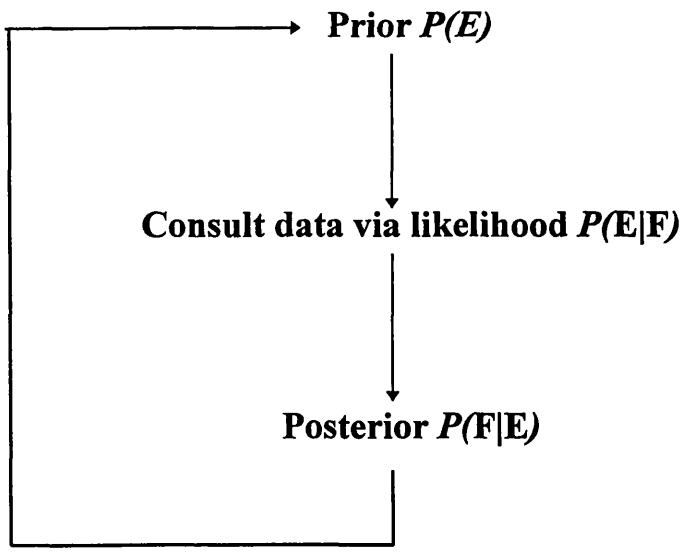


Figure 4.1. Bayes' theorem.

When Bayesian statistics have been applied to situations of scientific inference, concerns have been expressed regarding the possibility of bias towards the solutions

that are sought, in that a bias of the prior could lead to a bias in the final result. Daniel expressed these concerns in [105]

> 'Construction of a solution is easy; the difficult part is discovering what features of that solution follow inexorably from the data and which ones are a result of the model. Model fitting is easy, common and dangerous.'

Bias is reduced by placing constraints upon the Bayesian prior. The constraints placed upon the prior are:

> (i) The prior must always be capable of reproducing what we know to within experimental error.
>
> (ii) The prior must be maximally non-committal towards that which we do not know. This can be expressed another way, in that the prior must carry minimal information content about that which it does not know.

This can be achieved through implementation of maximum entropy methods.

4.5. The Maximum Entropy Principle.

Consider now a discrete probability distribution for which we know a set of N normalised probabilities $P=(p_1,p_2,.....p_n)$. Such a distribution has an entropy, S, as above or information content, I, given by,

$$S = -I = -\sum_{j=1}^{N} p_j \log p_j \qquad \text{Equation 4.2.}$$

Note that logarithm can be to any base.

Consider then, m constraints expressed as expectation values, $\langle A_r \rangle$ which represent knowledge derived from experiment. These constraints can be written as,

$$\sum_{j=1}^{N} A_{ri} p_i = \langle A_r \rangle \qquad \text{Equation 4.3.}$$

where r=1,2,.....m

From this we want to derive a set of probabilities {P}. The situation is often mathematically indeterminate, in that N is much greater than m. The maximum entropy principle consists of maximising the entropy in Equation 4.2. subject to the constraints of Equation 4.3. to determine the set of probabilities {P}.

Equation 4.2. was first derived by Shannon[106] working in the field of information theory and is applicable in the discrete case. For an continuous function such as electron density, Equation 4.2. can be written as an integral.

$$S = -\int p(\underline{x}) \log \frac{p(\underline{x})}{m(\underline{x})}\, m(\underline{x}) \qquad \text{Equation 4.4.}$$

where m(x) is a prior which can be uniform.

Thus, for a probability distribution to be maximally non-committal , its entropy must be a maximum, subject to the constraints of reproducing what is known. It can then be seen that if a prior is chosen via maximum entropy methods, it has a minimum bias and is maximally non-committal towards the missing data and therefore is an optimal prior for use within Bayes theorem.

Use of maximum entropy methods combined with Bayesian statistics can be viewed as having two benefits with regard to scientific inference;

(i) Bayes' theorem allows for the modulation of prior probabilities as extra evidence arrives.

(ii) Maximum entropy methods indicate how to assign the prior probabilities initially.

Having gained an overview of both Bayesian statistics and maximum entropy methods, the next section details their relevance pertaining to the crystallographic phase problem.

4.6. Application of Maximum Entropy and Bayesian Statistics in Crystallography.

Bricogne[100,103,107,108,109] has detailed in a series of papers the applicability of the concepts outlined above, to the crystallographic phase problem. He has employed the Saddlepoint approximation[110,111] which overcomes the limitations of the aforementioned Edgeworth series. Use of the saddlepoint approximation always yields optimal estimates of joint probability distributions involving large structure amplitudes. This is equivalent to requiring that the distribution of random atomic positions should be updated whenever phase assumptions are made, so as to retain maximum entropy under the constraints embodied in those assumptions.

Therefore, Bricogne demonstrated that the results obtainable from maximum entropy can also be derived without recourse to the maximum entropy formalism by use of the saddlepoint approximation and that in essence, the two techniques are equivalent. Application of this revised statistical analysis of the phase problem does not necessitate use of distributions such as triplets and quartets to generate phase information, although it is related to such methods. Before a description of the implementation of the theory involved, it is necessary to describe normalisation of data for use within maximum entropy calculations.

4.7. Normalisation of Data for Maximum Entropy Calculations.

It has been found that for maximum entropy calculations it is better to use unitary structure factors, U_H, rather that normalised structure factors, E_H, of conventional direct methods.

$$|U_{\underline{h}}| = \sqrt{\varepsilon_{\underline{h}}} \frac{|E_{\underline{h}}|}{\sqrt{N}} \qquad \text{Equation 4.5.}$$

where $|E_{\underline{h}}|$ is the normalised structure magnitude and N is the number of atoms in the unit cell.

From the above equation the maximum value of $|U_{\underline{h}}|$ is 1.0. Also, the distribution of $|U_{\underline{h}}|$ is independent of scattering angle, θ. Thus observed $|F_{\underline{h}}|$ values have been modified so that they correspond to a hypothetical structure in which actual atoms with more or less diffuse electron densities have been replaced by point atoms.[112]

4.8. Strategy for Implementation of Maximum Entropy Calculations.

Having dealt with the rationale of the theory, it can be stated that maximum entropy methods view an unknown crystal structure as being made up of atoms of known chemical identity but unknown atomic positions. These positions are considered as being random, with an initially uniform distribution in the asymmetric unit. Thus, structure determination consists of the gradual removal of that randomness.[113]

Implementation of the principles involved has been described extensively.[114,115,116,117] The MICE program (Maximum Entropy In a Crystallographic Environment) is the practical realisation of part of the Bricogne formalism, and it is to this program that all references about calculations allude to.

4.8.1. The Basis Set.

Solution to the phase problem for a specific set of data does not always start with a complete lack of knowledge. Following on from normalisation of data, a set of unitary structure factors are obtained, $|U_{\underline{h}}|^{obs}$ There is also a set of phase angles, $\phi_{\underline{h}}$, most of which are unknown. In order to fully define a structure in real space, an origin must be defined to enable a fixed frame of reference for atomic coordinates to be obtained. This is achieved in reciprocal space by assigning phases to a limited number of $|U_{\underline{h}}|^{obs}$. For ab initio structure solution, because of the rules of origin definition, some phases can be assigned subject to certain rules, depending on the space group chosen. A criterion that must be fulfilled is that the origin defining reflections must contain no semi-invariants and be linearly independent of each other. This means that the combinations of two or three reflections must not be structure invariants or semi-invariant. Phase information can also be derived, in the case of electron diffraction, from the Fourier transform of the corresponding image.

Partitioning of data can occur at this point and the phased reflections define the basis set, *{H}*, with the remaining unphased reflections being assigned to the non-basis or disjoint set, *{K}*.

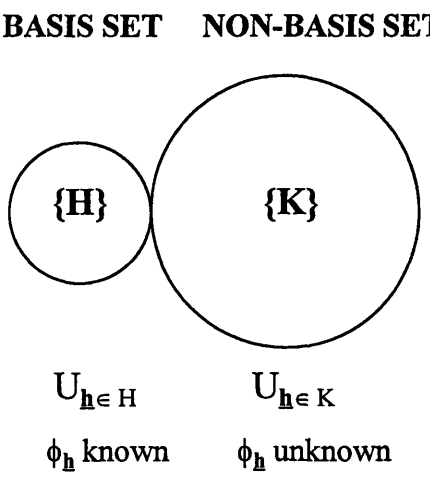


Figure 4.2. Partitioning of data within MICE.

The reflections in the basis set *{H}* are now used as constraints in the generation of a maximum entropy prior, termed $q^{ME}(\mathbf{x})$.

4.8.2. Entropy Maximisation.

Before generation of a maximum entropy prior, the reflections in the basis set are used to calculate an initial electron density map, $\rho(\mathbf{x})$, by Fourier transform. This map contains only information that has been input and is therefore heavily biased. As in conventional direct methods, areas of negative electron density are forbidden in the maximum entropy formalism and so all areas of the map which are negative are set to zero to form a new map, $\rho'(\mathbf{x})$. This new map is still heavily biased and so its entropy is maximised subject to the constraint that the Fourier transform of the map $\rho'(\mathbf{x})$ must contain the basis set reflections to within experimental error of those input to the original calculation. Thus, in the language employed by Bricogne et al. in the theory, the prior $q^{ME}(\mathbf{x})$ must reproduce the known phases and amplitudes but be maximally non-committal with respect to unknown structure factors.

The Fourier transform of the $q^{ME}(\mathbf{x})$ map contains not only the phases of the basis set reflections used in the initial calculation but also contains phase information on $U_{\mathbf{h}} \in \mathbf{K}$. This generation of estimates of amplitudes and phases for reflections in $\{\mathbf{K}\}$ is called *extrapolation*, shown graphically in Figure 4.3.

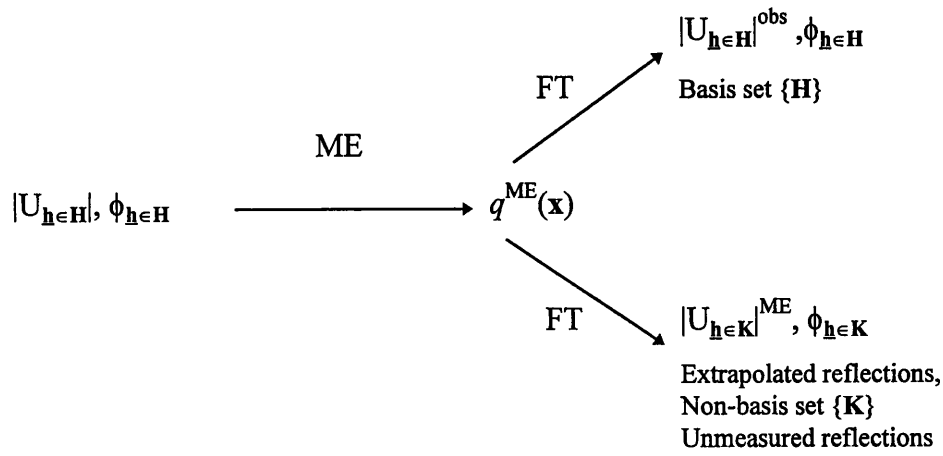


Figure 4.3. Extrapolation of phases and amplitudes.

4.8.3. The Phasing Tree.

When the basis set is small, comprising only 2 or 3 reflections, the extrapolation is weak and most reflections will not have reliably extrapolated phases. While it is possible to add the strongly extrapolated reflections to the basis set, difficulties can be encountered as entropy optimisation can become trapped into a local minimum. Thus some other technique has to be employed to move reflections from the non-basis set $\{\mathbf{K}\}$ to the basis set $\{\mathbf{H}\}$.

To proceed, therefore several reflections with large U-magnitude need to be given permuted phases, just as in conventional direct methods, giving rise to a multisolution environment. These new permuted reflections are added to the basis set and are treated as additional constraints in entropy maximisation. The reflections are chosen on the premise that they are those about which the current $q^{ME}(\mathbf{x})$ knows least. This implies that the next $q^{ME}(\mathbf{x})$ map will have maximum information gain relative to the previous map. The reflections about which the $q^{ME}(\mathbf{x})$ map knows least are also the reflections least coupled to those already known, thus preventing islands of highly correlated reflections being determined.

Use of this method of phase permutation gives rise to a node for each choice of phase for each permuted reflection and so build a phasing tree, as illustrated in Figure 4.4. for a centrosymmetric situation.

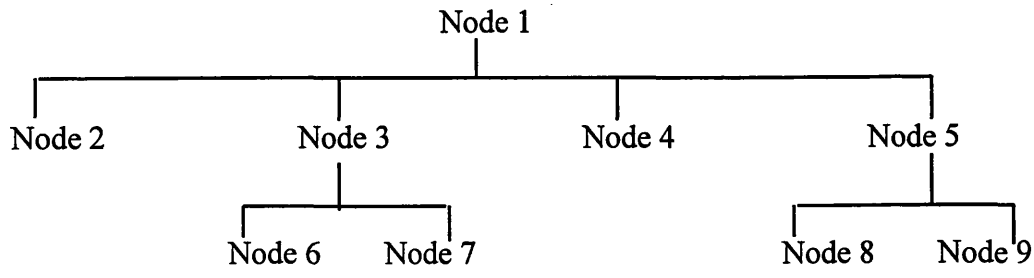


Figure 4.4. Phasing tree.

Clearly construction of such a phasing tree could prove to be computationally cumbersome, since each node has to be subjected to entropy maximisation. To determine which nodes are most promising to continue the phasing procedure, the other component of Bayesian statistics, likelihood, is employed.

4.8.4. Likelihood as a Figure of Merit.

In the early stages of phase determination likelihood has proved to be a reliable figure of merit, and as such is used as a criteria for phase set discrimination. Likelihood measures the degree to which the phase choices made are able to anticipate correctly, through maximum entropy extrapolation, some of the information present in the non-basis set. Thus, likelihood can be thought of as a figure of merit determining the future viability of a given node.

Likelihood is not used in its native form, but rather calculated relative to the null hypothesis, L_0, where for all the extrapolated reflections, $|U_{\underline{h}}|^{ME}=0$. Thus, likelihood is in the form of a ratio, or as it is used in MICE, as the log likelihood gain (LLG).

For the acentric extrapolated reflections, ($\underline{\mathbf{h}} \in \mathbf{K}$), log likelihood gain is defined as

$$\Lambda_{\mathbf{k}} = \frac{|U_{\mathbf{k}}|^{obs}}{\varepsilon_{\mathbf{k}}\Sigma_a + \sigma_{\mathbf{k}}^2} \exp\left\{-\frac{1}{2}\frac{\left(|U_{\mathbf{k}}|^{obs}\right)^2 + |U_{\mathbf{k}}^{ME}|^2}{\varepsilon_{\mathbf{k}}\Sigma_a + \sigma_{\mathbf{k}}^2}\right\} I_0\left(\frac{|U_{\mathbf{k}}|^{obs}|U_{\mathbf{k}}^{ME}|}{\varepsilon_{\mathbf{k}}\Sigma_a + \sigma_{\mathbf{k}}^2}\right) \quad \text{Equation 4.6.}$$

where $\varepsilon_{\mathbf{k}}$ is the statistical weight of reflection $\mathbf{k}$, $\sigma_{\mathbf{k}}^2$ the variance of $|U_{\mathbf{k}}|^{obs}$, Σ_a a refinable measure of unit cell contents $\Sigma \approx 1/(2N)$ for N point atoms in the unit cell and I_0 is a zero order Bessel function.

For centric reflections, log likelihood gain is as shown in Equation 4.7.

$$\Lambda_{\mathbf{k}} = \frac{2|U_{\mathbf{k}}|^{obs}}{\pi\left(2\varepsilon_{\mathbf{k}}\Sigma+\sigma_{\mathbf{k}}^2\right)}\exp\left\{-\frac{1}{2}\frac{\left(|U_{\mathbf{k}}|^{obs}\right)^2+|U_{\mathbf{k}}^{ME}|^2}{2\varepsilon_{\mathbf{k}}\Sigma+\sigma_{\mathbf{k}}^2}\right\}\cosh\left(\frac{|U_{\mathbf{k}}|^{obs}|U_{\mathbf{k}}^{ME}|}{2\varepsilon_{\mathbf{k}}\Sigma+\sigma_{\mathbf{k}}^2}\right) \quad \text{Equation 4.7.}$$

As the likelihood equations given were derived using the diagonal form of likelihood, they are only sensitive to the extrapolated moduli, $|U_{\underline{h}}|^{ME}$ and not to the associated extrapolated phases.

A corresponding null hypothesis is defined for the situation of null extrapolation, $|U_{\mathbf{k}}^{ME}| = 0$, which gives the Gaussian distribution of Wilson statistics. This is defined for acentric reflections as,

$$\Lambda_{\mathbf{k}}^{0} = \frac{|U_{\mathbf{k}}|^{obs}}{\varepsilon_{\mathbf{k}}\Sigma+\sigma_{\mathbf{k}}^2}\exp\left\{-\frac{1}{2}\frac{\left(|U_{\mathbf{k}}|^{obs}\right)^2}{\varepsilon_{\mathbf{k}}\Sigma+\sigma_{\mathbf{k}}^2}\right\} \quad \text{Equation 4.8.}$$

Then, if $L_{\mathbf{k}}$ is defined as,

$$L_{\mathbf{k}} = \log\frac{\Lambda_{\mathbf{k}}}{\Lambda_{\mathbf{k}}^{0}} \quad \text{Equation 4.9.}$$

global log likelihood gain (LLG), can now be defined as,

$$LLG = \sum_{\mathbf{k}} L_{\mathbf{k}} \quad \text{Equation 4.10.}$$

Log likelihood gains are analysed for phase indication using the Student t-test.[118] The simplest test involves the detection of the main effect associated with the sign of a single phase. The log likelihood gain averages μ^+ and μ^- with associated variances, V^+ and V^- are computed for those nodes in which the sign of the phase under test is + or - respectively. The use of the *t*-test enables a sign choice to be derived with an associate significance level.

$$t = \frac{\left|\mu^{+} - \mu^{-}\right|}{\sqrt{V^{+} + V^{-}}} \qquad \text{Equation 4.11.}$$

This calculation is repeated for all the single phase indications and then is extended to combinations of two or three phases. In general, only relationships with associated significance levels >2% are used, but this is sometimes relaxed with sparse data sets.

Only those solutions which are consistent with the *t*-test results are kept, and further reduced if necessary to 8 in a given level. Further reflections are then permuted and a new level of nodes generated. This procedure is continued until most structure factors have significant phase indications.

4.8.5. Centroid Maps.

The maximum entropy distributions associated with the various nodes are not electron density maps in the traditional sense. A $q^{ME}(\mathbf{x})$ maps displays many of the features normally associated with maps, in that peaks correspond to atomic positions. The $q^{ME}(\mathbf{x})$ map is subjected to a Sim filter[119,120] thus producing a centroid map.

For **k** acentric the Fourier coefficients are:

$$|U_{\mathbf{k}}|^{obs}\left[I_1(X_{\mathbf{k}})/I_0(X_{\mathbf{k}})\right]\exp(i\varphi_{\mathbf{k}}^{ME}) \qquad \text{Equation 4.12.}$$

where:

$$X_{\mathbf{k}} = (2N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{obs}\left|U_{\mathbf{k}}^{ME}\right| \qquad \text{Equation 4.13.}$$

For **k** centric, these coefficients become:

$$|U_{\mathbf{k}}|^{obs}\tanh(X_{\mathbf{k}})\exp(i\varphi_{\mathbf{k}}^{ME}) \qquad \text{Equation 4.14.}$$

with:

$$X_{\mathbf{k}} = (N/\varepsilon_{\mathbf{k}})|U_{\mathbf{k}}|^{obs}\left|U_{\mathbf{k}}^{ME}\right| \qquad \text{Equation 4.15.}$$

In the centroid map, both reflections from the basis set {**H**} and the extrapolated reflections {**K**} are used. Phase angles are those extrapolated by the maximum entropy procedure. These centroid maps are then examined in a conventional manner.

4.9. Application of Maximum Entropy Methods using MICE.

As described, maximum entropy methods are not subject to the same constraints as conventional direct methods. This leads to application of said methods in crystallographic situations where perhaps data is of low resolution or there is incomplete sampling of reciprocal space. This has particular relevance in dealing with powder diffraction, macromolecular crystallography and of course, electron diffraction data from both small molecules and macromolecules.[121]

4.9.1 Powder Diffraction and MICE.

Powder diffraction patterns are a one-dimensional representation of a three-dimensional pattern that has been reduced by spherical averaging. This results in the possibility of reflections that would be otherwise separate, overlapping - a phenomena which increases with Bragg angle and unit cell dimensions. The problem of peak overlap in the powder pattern is the factor which can limit the complexity of structure that can be solved. The maximum entropy approach has been shown to be applicable to powder work, and as such, MICE has been adapted to deal with such datasets.[118,122]

The theoretical basis for the application of maximum entropy and likelihood to powder diffraction has been outlined by Bricogne.[109] The approach here differs from conventional avenues to ab initio structure solution from powder data, where overlapped reflections are ignored or have their intensities equally partitioned. Maximum entropy methods utilise all the intensity data in the normalisation procedure and the log likelihood gain is calculated using both overlapped and non-overlapped reflections. The overlapped reflections are also used in the final centroid maps, thus playing a contributory role in the entire phasing procedure.

The methodology detailed has been employed to solve structures from powder data, ab initio, for a variety of inorganic compounds including the salts, $KAlP_2O_7$[122] and Mg_3BN_3 [118] and anhydrous lithium triflate, $LiCF_3SO_3$[123]. A more complex structure solved by these means was the clathrasil, Sigma-2.[118] Further application to organic structures has been achieved through determination of formylurea[124] and 1,3,4,6-tetrathiapentalene-2,5-dione.[125]

The examples given demonstrate the feasibility of applying maximum entropy methods to the determination of previously unknown structures directly from powder diffraction data.

4.9.2. Macromolecular X-ray Crystallography and MICE.

Some of the earliest applications of maximum entropy techniques were made in protein crystallography, as Bricogne illustrated in the extrapolation of phases from 3Å resolution to that of 1.5Å, using experimental data coupled with correct phases.[100] MICE has been used, with adaptation, in situations where additional constraints arise from the molecular envelope in conjunction with solvent flattening.[126] The previously unknown protein TrpRS (*Bacillus stearothermophilus* tryptophanyl-tRNA synthetase) was solved, again utilising MICE. In this instance, phase permutation was required to overcome problems that were encountered in the previous attempts at structure elucidation using conventional direct methods.[127] Again, demonstration of the power of maximum entropy methods in the face of problems shunned by conventional direct methods.

4.9.3. Protein Electron Crystallography and MICE.

Certain aspects of macromolecular electron crystallography involving the processing of low-dose high resolution electron micrographs from tilted two dimensional data have been discussed previously. These methods have proved successful in certain cases, but ultimately their usefulness is limited by the resolution of the experimental

images. However, a varied approach to this problem was offered in the study of two dimensional purple membrane data from *Halobacterium Halobium.*[128] The route taken for the maximum entropy calculations, in this instance involved using phases obtained from the Fourier transform of a suitable image as the basis set.[117,129] Thus, the full apparatus of phase permutation, phasing trees and likelihood ranking was not employed. Use was simply made of the centroid maps produced by phase extrapolation. Overall, this phase extension provides resolution enhancement as seen on examination of the centroid maps produced.

A commonly held belief is that *ab initio* phase determination is not practical in most cases of protein electron crystallography, due to the structural complexity of the molecules. However, *ab initio* structure determination of two membrane protein structures in projection was achieved using maximum entropy and likelihood.[130] Phase determination was carried out for the Omp F porin from the outer membrane of *E. coli.* and for Halorhodopsin. Accurate phase information was found for the most likely solutions, thus enabling potential maps to be calculated. These maps were found to contain most of the essential structural features of the macromolecules. The essential difference in this work, compared to previous examples employing maximum entropy methods is that no use was made of image derived phases as a starting set for phase extension. Also, there was no need to use envelopes or electron density histograms though they may add more to the power of the method.

4.9.4. Small Molecules, Electron Diffraction and MICE.

As with macromolecular electron crystallography, initial use of maximum entropy methods as applied to small molecules, was for phase extension purposes. The first molecule to be studied was perchlorocoronene, using projection data.[73] Experimental images of resolution of 3.2Å were obtained, together with a diffraction pattern extending to 1.0Å resolution. Fourier transform of the image yielded four phases which were used as the basis set for entropy maximisation procedures. Maps obtained

were compared to results from a previous X-ray study[131] and were found to produce a solution that, in projection, was identical.

Validation of data sets and indeed structure solutions of organic molecules determined in the early days of electron crystallography by the group led by Vainshtein, has also been achieved through application of the maximum entropy formalism. Structure solution of both diketopiperazine and $CuCl_2.3Cu(OH)_2$ was accomplished by *ab initio* means using MICE.

MICE has also been utilised in a study, whereby the unknown structure of 4-[4-(N,N-dimethylamino)benzylidene]pyrazolidine-3,5-dione was solved in projection.[80] Although structure solution was from a very sparse data set, excellent agreement was seen between the maximum entropy solution and independent model building combined with high resolution electron microscopy studies.

In an analogous manner, the structure of [9,9'-bianthryl]-10-carbonitrile was solved by the same group.[81] Again, a sparse data set was used and the results obtained from maximum entropy methods were independently verified by model building and image simulation techniques. Thus, giving credence to the belief that *ab initio* structure solution using electron diffraction data does indeed have a distinct viability.

4.10. Maximum Entropy Methods in Crystallography - Perspective.

In common with many of the radical departures from conventional beliefs that have been proposed throughout the course of the development of science as a whole, maximum entropy methods in crystallography were treated with a certain amount of guardedness and not a little disdain.

In a wide ranging review on the role of direct methods in X-ray crystallography, Woolfson stated,

> 'It seems that maximum entropy is adding nothing completely new to the crystallographic scene and since it involves a great deal of effort, perhaps nothing useful.'

Having detailed aspects of the role and application of maximum entropy methods in crystallography, it can be concluded that these methods do indeed add something fundamentally new to crystallographers searching for ways of overcoming the phase problem and that overall, it is worth the effort.

4.11. Concluding Remarks.

Electron crystallography, in the context of this body of research, has brought together many differing aspects of structural science. It is hoped that this broad theoretical description has highlighted the reasoning behind use of such methods and illustrated the common thread of gaining structural information to whatever degree about a variety of molecules. The methodology has been described ranging from simple electron diffraction through to complex crystallographic concepts.

The subsequent sections of this thesis bring an account of all of these methods and techniques as applied to pigment structures, both in further investigation of previously studied molecules and attempted *ab initio* structure solution of an unknown structure.

5.0. REVIEW OF PIGMENT CRYSTAL STRUCTURES

As was discussed in the introduction, the properties of organic pigments are strongly influenced by the chemical structure and the physical characteristics of the crystals. The importance of particle size, shape and knowledge of the crystallographic arrangement of the pigment molecules within the crystal are crucial in understanding the relationship between structure and property.

Crystal structure analyses have been carried out on several pigments of the β-Naphthol and Acetoacetanilide classes. There is a diversity of functionality between these distinct classes, but there are important structural features that are consistent in both. Preliminary reviews have been undertaken and these identified certain structural trends[132,133]. Since that work was carried out, more structures have been solved and the aim of this work is to include these structures in a further, more detailed review of the structural architecture of these pigments.

Analysis of these structures and trends therein could lead to further work involving the design of surface active molecules that could modify the physical properties of the pigment, such as solubility, dispersibility and fastness.

5.1. Structural Review - Background.

Structural information was derived from many sources including the structural literature, Chemical Abstracts and most importantly, the Cambridge Structural Database.[134,135,136] Crystallographic information could be taken from CSD and used in computer based molecular visualisation packages, namely *CERIUS*^{2}.[137]

CERIUS^2 allows for the investigation, manipulation and analysis of 3D representations of molecules. Areas that can be investigated include mode of packing of molecules, extent of hydrogen bonding and surface functionality.

5.2. Cambridge Structural Database.

Computers have had an unprecedented role in the advancement of modern crystallography. This has led to the facile management of data either in its collection, retrieval or analysis. The rapid development of structural crystallography has created a vast amount of structural information that obviously needed some sort of comprehensive organisation.

Crystallographic databases have had the common aim of recording not only the relevant biographical and chemical information needed for searching and for access to the original literature, but also the often extensive primary numerical results. These primary results include cell dimensions and symmetry, atomic coordinates and thermal parameters. All this information provides a basis from which a wealth of secondary information can be derived and used by anyone interested in chemical structure.

The main crystallographic database is the Cambridge Structural Database. It stores the primary results of full three-dimensional X-ray and neutron diffraction studies of organics, organometallics and of metal complexes having organic ligands. The database is fully retrospective and is updated regularly. The CSD can be thought of as a reservoir of precise coordinate based information on molecular structure. This is information which is accessible via search queries composed, primarily in chemical terms. Entry is provided to the relevant literature and numerical experimental data for individual crystal structures. However, the literature is already replete with detailed descriptions of individual structures, so the most important function of CSD is to provide the basis for systematic and detailed analyses of sets of related compounds.

Crystallographic data can be employed in studies which result in a systematic statement about molecular shape, molecular dimensions or of the intramolecular and intermolecular interactions.

5.3. Inferred Trends from Crystallographic Databases.

5.3.1. Bond Lengths.

The determination of molecular geometry is of vital importance to the understanding of chemical structure and bonding. This led to the first compilation of results from X-ray analysis[138] giving interatomic distances and configurations in molecules and ions. Further studies followed, including summary tables of bond lengths between carbon and other elements in Vol. 3 of International Tables for Crystallography.[139]

CSD has been used to prepare a table of average bond lengths in organic compounds.[140] The table lists average lengths for bonds involving the elements H, B, C, N, O, F, Si, P, S, Cl, As, Se, Br, Te and I. Classification of bonds used in the table is based on common functional groups, rings, ring systems and coordination spheres. Within the context of this study, the bond lengths of most interest are those that define either the azo or hydrazone tautomer of a particular pigment, that is, N-N, C-N and C-O distances. Listed in Table 5.1. are the values obtained from CSD and those normally accepted values quoted in International Tables[141] are listed in Table 5.2.

	Single	Double
N-N	1.401	1.240
C-N	1.465	1.279
C-O	1.432	1.221

Table 5.1. CSD Average Bond Lengths (Å)

	Single	Double
N-N	1.44	1.24
C-N	1.47	1.32
C-O	1.43	1.23

Table 5.2. International Tables Average Bond Lengths (Å)

The tabulation of results in this manner, was an attempt by the authors to obtain the average geometries of functional groups, rigid rings and the low energy conformations of flexible rings. It is hoped that the systematic survey could be extended to derive information about distances, angles, directionality and environment dependence of hydrogen bonds and non-bonded interactions.

5.3.2. Space Group Frequencies for Organic Compounds.

Another study[142] was undertaken which used a different database, NBS Crystal Data Identification File.[143] The frequency of occurrence has been calculated for each of the 230 different space groups. The space group frequencies were calculated for organic materials. Most of the entries correspond to entries that results from full structure determinations.

Since it was possible to tabulate frequencies for 29 059 organic compounds, the results should be representative of their distribution in nature. The five most common space groups and their percentage occurrence are given in Table 5.3.

Space Group	% Occurrence
$P2_1/c$	36.0
P1	13.7
$P2_12_12_1$	11.6
$P2_1$	6.7
$C2/_C$	6.6

Table 5.3. Space Group occurrence.

From examination of the relevant results, it can be concluded that 75% of the compounds studied have been described in only 5 space groups and 90% of the structures in 16 space groups. The space group frequencies table can be used in routine structure work. The frequencies may indicate that a space group determination should be rechecked if the substance has been characterised by a rare or previously unoccupied space group.

It may be possible to develop theories which could explain why certain space groups are rare or even uninhabited. Within the remit of this research on pigments, other researchers[144] have tried to correlate molecular shape or physical properties of a given pigment with the probability that the compound falls into a certain space group or class of space groups.

5.3.3. Hydrogen Bond Geometry in Organic Crystals.[145]

In 1920, Latimer and Rodebush suggested that a free pair of electrons on one water molecule might be able to exert sufficient force on a hydrogen held by a pair of electrons on another water molecule to bind the two molecules together.[146,147] This speculation generated many years of research into the hydrogen bond.

Even although many crystal structures are determined with the specific intention of studying the hydrogen bond arrangement, there are still many aspects of hydrogen bonding which are still contentious.

The hydrogen bond is largely an electrostatic phenomenon. Consequently, the length of a hydrogen bond is highly dependent on the nature of the donor and acceptor atoms. A survey was undertaken of 1509 (N-H....O=C) hydrogen bonds observed by X-ray or Neutron diffraction in 889 organic crystal structures. Again, only the hydrogen bond distances of most relevance in this work will be considered. The distances examined were (N....O) and (H....O) of the (N-H....O=C) bonding. Mean values of the hydrogen bonds in question are given in Tables 5.4. and 5.5.

Type	(H....O) Distance (Å)
Intramolecular	1.988
Intermolecular	1.913

Table 5.4. (H....O) Distance (Å)

	(N....O) Distance (Å)
Intramolecular	2.755
Intermolecular	2.892

Table 5.5. (N....O) Distance (Å)

The value for intermolecular distance for (N....O) corresponds favourably to the value given by Pimental and McClellan of 2.90Å.[148] The (H....O) and (N....O) distances of intramolecular bonds tend to be longer and shorter respectively than those of intermolecular bonds. Although hydrogen bond lengths are affected by the nature of the donor and acceptor atoms, there is little doubt that most of the observed variation in hydrogen bonding distances is due to crystal packing effects.

The values listed here can be used to determine whether intermolecular or intramolecular hydrogen bonding exists within the pigment structures studied. It is thought that hydrogen bonding is essential for keeping the molecules planar, which is thought to be crucial to pigment properties. The above three studies show how analysis of the vast array of structural data held within various crystallographic databases can be analysed for the trends therein.

5.4. Crystal Structures of β-Naphthol Pigments.

From their initial discovery, it was thought that β-Naphthol pigments existed as the hydroxy-azo tautomer and this is the form that they are normally represented as. This view was challenged by Liebermann[149] in 1883, who postulated that the hydroxy proton of 1-phenylazo-2-naphthol was labile and thus capable of bonding with a nitrogen atom of the azo group. Shifting of this proton gives rise to the keto-hydrazone tautomer shown in Figure 5.1.

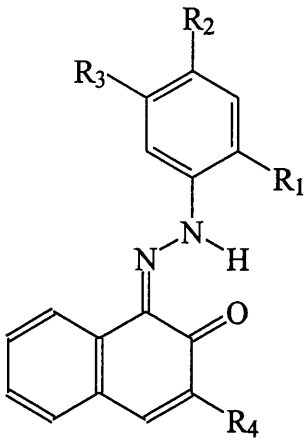


Figure 5.1. Keto-hydrazone tautomer.

Further evidence to support this came from Zincke and Bindewald [150] who found that they got the same product from either coupling benzene diazonium ion with 1-naphthol or by condensing phenylhydrazine with 1,4-naphthoquinone. Later studies [151,152] involving ^{13}C NMR, UV/Vis and IR spectroscopy have all proved the existence of the ketohydrazone tautomer.

Brown and Yadav[153] have suggested that this hydrazone configuration was in agreement with the concept of extensive π-electron resonance throughout the molecule, which promotes geometric planarity and in consequence, stability and light fastness of pigments. Partial double bonds would also suggest a degree of delocalisation.

Conclusive evidence about the tautomeric form in which a given pigment exists can be decided from certain bond lengths within the structure. As the azo form and hydrazone form only differ with regard to certain bonds, it is worthwhile to compare the lengths of these bonds in each pigment.

If the azo tautomer is found to exist, then the azo bond (-N=N-) is a double bond, with adjacent bonds single, (-C-N) and hydroxyl, (C-OH). If, however, it is the hydrazone tautomer , then these bonds are single azo, (-N-N-) and double bonds for both (C=N)

and (C=O). The values for these bond lengths can be accurately obtained from the analysis of the results of a crystal structure solution.

5.4.1. Structure Retrieval and Examination.

Extensive literature searching followed by use of Cambridge Structural Database [Version 5.09, April 1995] furnished X-ray crystal structure results for 17 pigments of the β-Naphthol class and 17 of the Acetoacetanilide class. The protocol employed to search CSD is outlined in Figure 5.2. The file created by CSD, *filename.fdat* contains structural information including atomic coordinates, element types, crystal cell parameters, symmetry operators and bonding connectivity. As described, it was possible to use this file created from successful searches within CSD for use in other applications, namely *VISTA* [154] and *CERIUS*^{2}.

VISTA stands for *Vi*sualisation of *STA*tistics and allows interactive examination of the numerical results of searches carried out within the database. The program allows the production of histograms, scattergrams and principle component analysis plots.

CERIUS^2 was used to aid molecular visualisation. With the geometry tools contained in this suite of programs, it was possible to closely examine the 3D structure of the molecular models being studied. The geometry analysis tools do not alter the stored coordinates of the model, they only add extra information or change the view being observed. Possible information that can be retrieved includes measurement of distances, angles, calculation of close contacts between non-bonded atoms and calculation of contact surfaces for a molecule, crystal or surface.

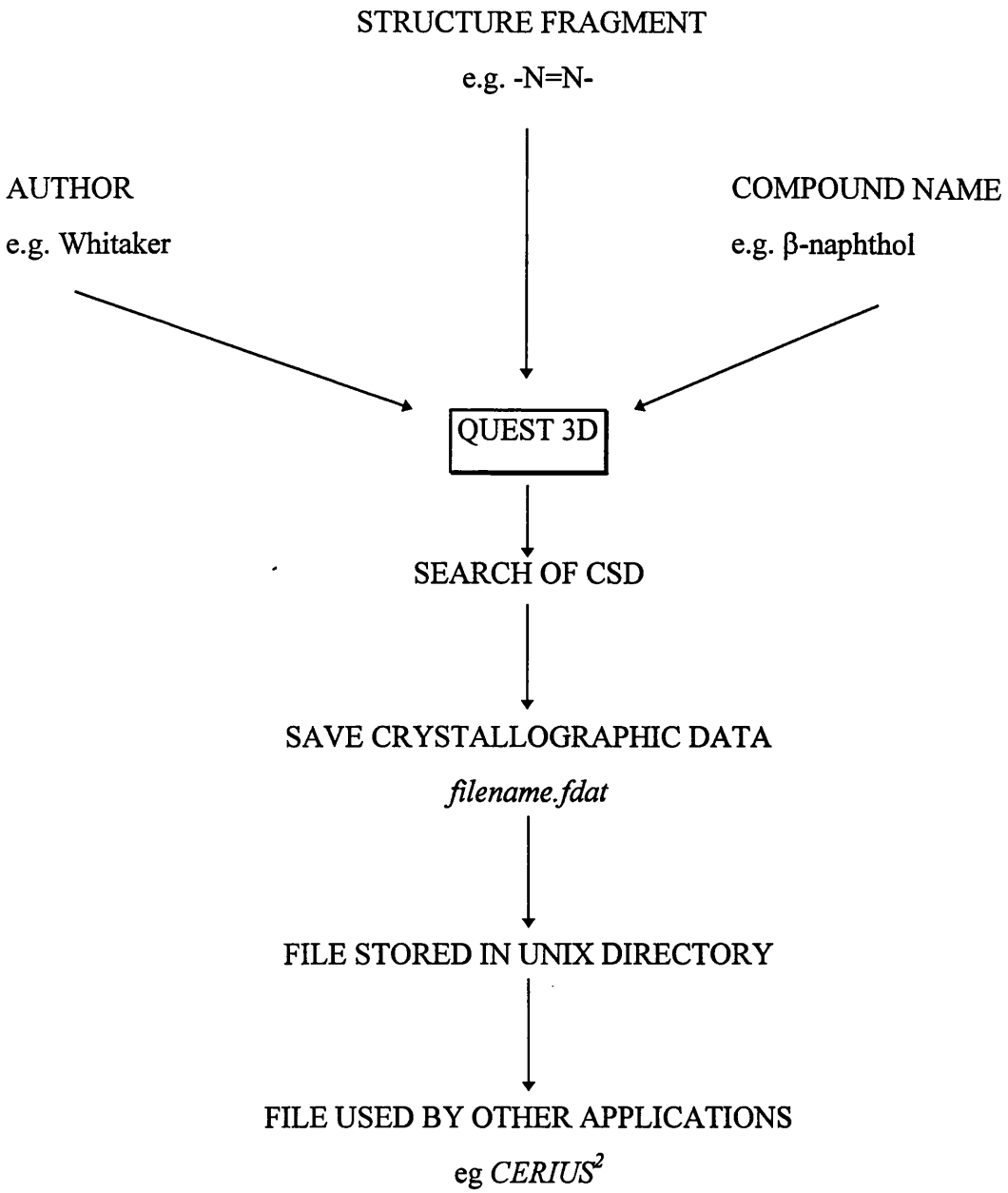
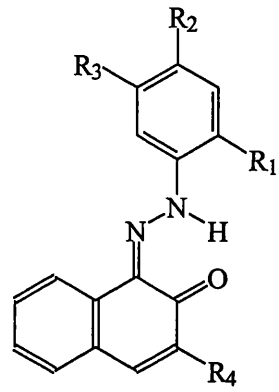


Figure 5.2. CSD Search Protocol.

5.4.2. Structure Review of β-Naphthol Pigments.

As mentioned above numerical results from 17 crystal structures of β-Naphthol pigments were compiled and a systematic survey undertaken of certain structural characteristics. Individual molecular structures of each pigment found, along with CSD reference code and original literature reference are given in Table 5.6. Various structural parameters were tabulated and examined using VISTA, the results of which are now discussed.



CSD Refcode	R_1	R_2	R_3	R_4	Ref No.
ANACMO	Cl	H	Cl	see Fig. 5.11.	155
CBANAP	H	H	H	H	156
CPZNXA	Cl	H	Cl	see Fig. 5.11.	157
CPZPAN	Cl	H	Cl	see Fig 5.11.	158
JARPEX	H	H	H	H	159
JARPEX01	H	H	H	H	161
JARPEX02	H	H	H	H	160
JARPIB	H	NMe	H	H	161
JARPIB01	H	NMe_2	H	H	161
JATJIX	OCH_3	H	H	H	162
MNIPZN	NO_2	CH_3	H	H	161
NBZANO	H	NO_2	H	H	162
NBZANO11	H	NO_2	H	H	163
NQNCPH	NO_2	Cl	H	H	164
NQNCPH01	NO_2	Cl	H	H	165
PAMBOO	OH	H	H	H	166
TOAZNI	H	H	CH_3	H	167

Table 5.6. CSD refcodes and Molecular Formula of selected β-Naphthol pigments

Figure 5.3. shows a plot obtained from VISTA of the distribution of (N-N) bond lengths observed in 17 β-Naphthol pigments. Mean (N-N) bond length was found to be 1.308(9) Å. As was shown in Table 5.1, average (N-N) single and double bonds are 1.401 Å and 1.240 Å respectively. Therefore, the mean value obtained in this survey

is intermediate between double and single bond order, thus a certain degree of delocalisation is present.

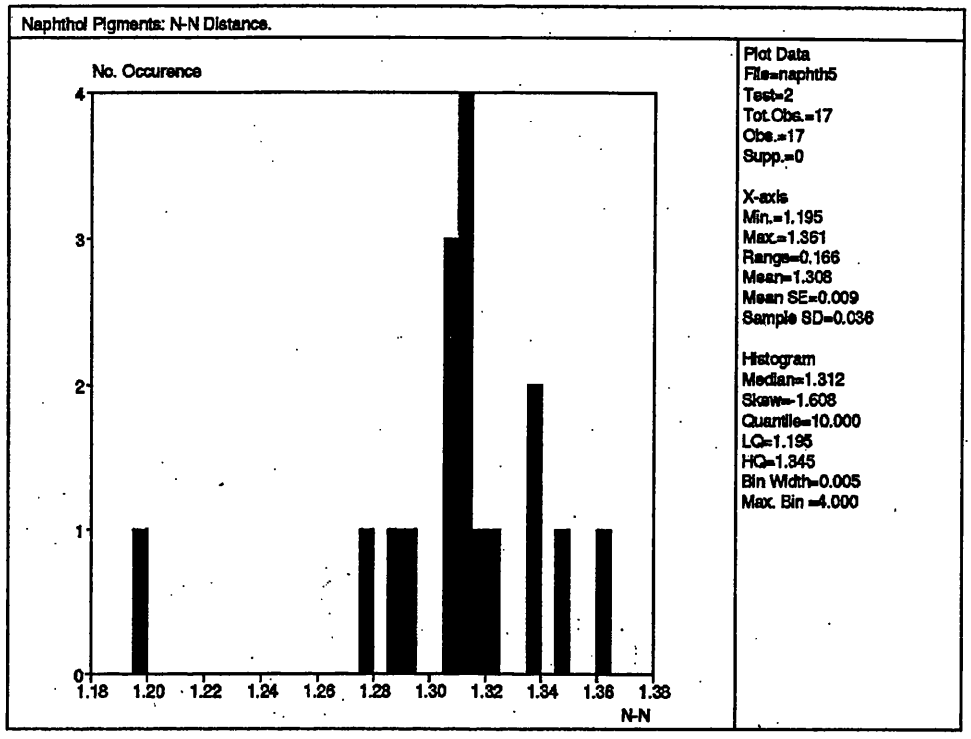


Figure 5.3. N-N Distance.

Other bonds whose order will help in the determination of tautomeric form present, include (C=N). Figure 5.4. displays the VISTA obtained plot of the distribution of (C=N) distance, which has a mean value of 1.342(6) Å, thus indicating the presence of (C=N) double bond as the normally accepted values quoted in International Tables for crystallography give (C-N) as 1.47 Å and (C=N) as 1.32 Å. This bond length is indicative of the existence of the keto-hydrazone tautomer.

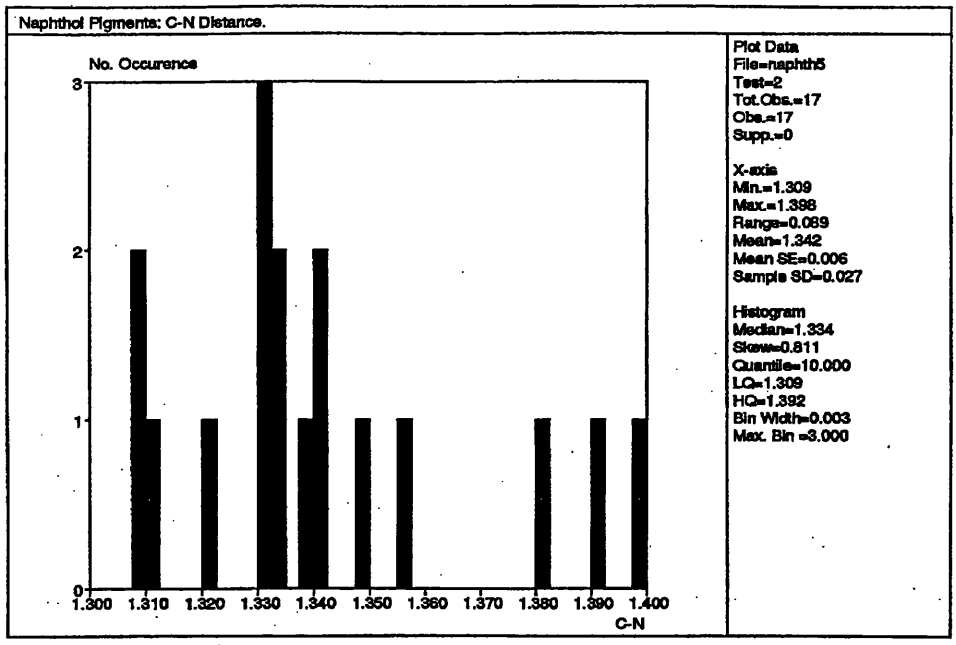


Figure 5.4. C-N distance.

Existence of this tautomeric form is further confirmed by (C=O) bond length. Figure 5.5. contains another VISTA plot of the distribution of (C=O) bond lengths, with a calculated mean value of 1.267(8)Å.

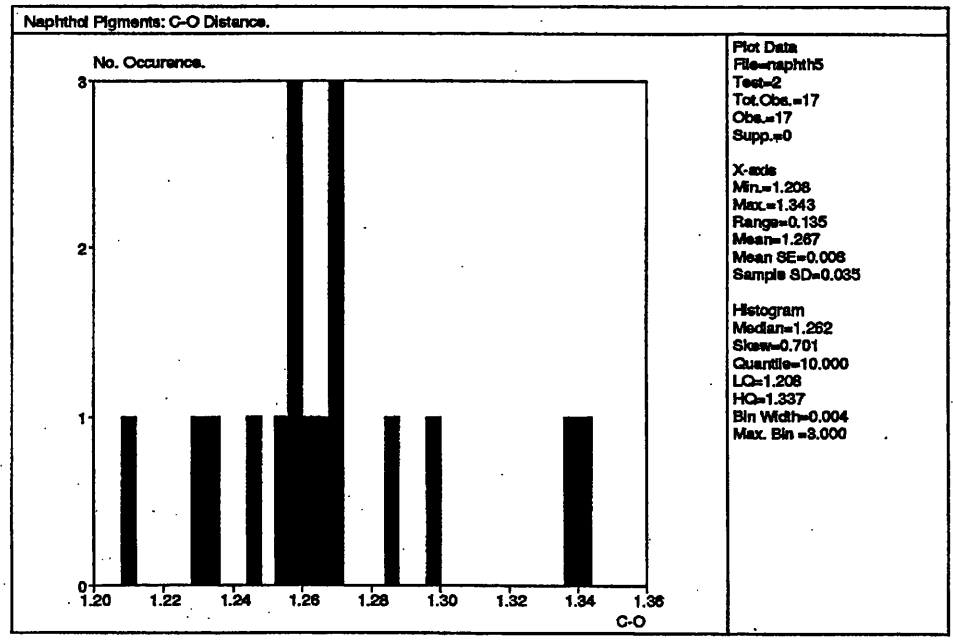


Figure 5.5. C-O Distance.

Comparison with accepted values from International Tables for crystallography of 1.43Å and 1.23Å for single and double (C-O) bonds respectively, leads to the decision that the (C-O) bond present has double bond character. From bond length information alone, the presented evidence is in favour of the presence of the hydrazone tautomer.

In addition, the angles that the bonds subtend at the nitrogen atoms are not 180° as implied by a linear modelling of the azo linkage, but have an approximate value of 120°. Conclusive evidence of this is derived from a plot of distribution of angles at the two nitrogen atoms, shown in Figure 5.6. Mean bond angle value of 118.060° was obtained. Strictly speaking, this means that the molecule could be either *cis-* or *trans-*, but in all cases investigated, the *trans*-configuration of the keto-hydrazone tautomer was seen to prevail.

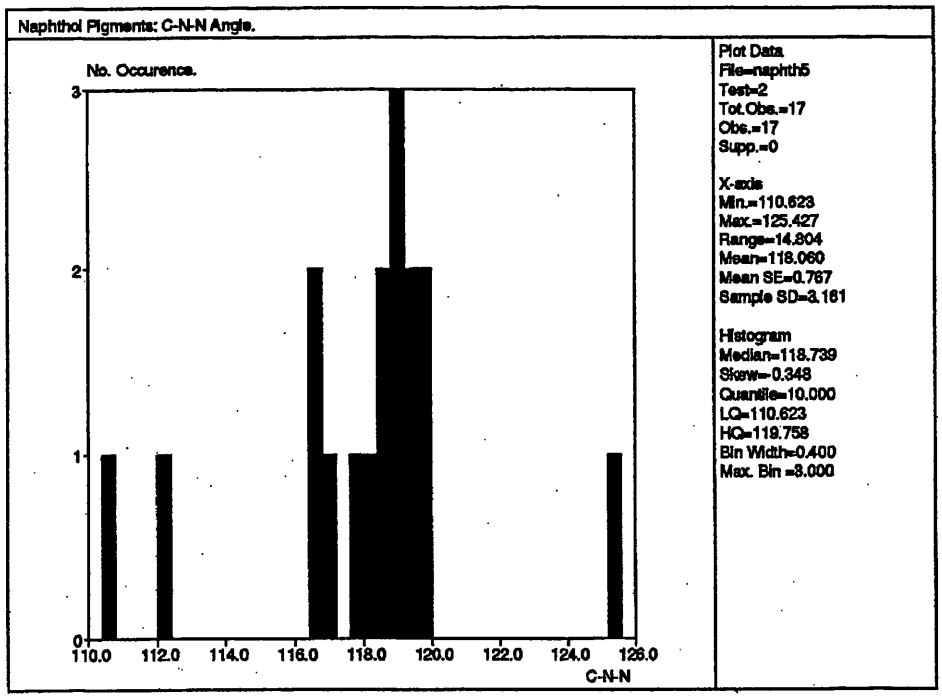


Figure 5.6. C-N-N Angle.

Further investigation of the structural architecture arising from this primary geometrical information leads to consideration of non-bonded interactions. Values for

non-bonded distances can indicate whether intramolecular and intermolecular hydrogen bonding is present within the structure.

The presence of a hydrogen atom attached to one of the azo nitrogens allows for possible intermolecular hydrogen bonding to the naphthol oxygen, as demonstrated in Figure 5.7.

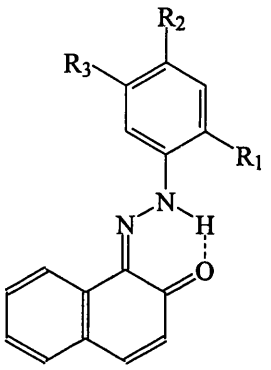


Figure 5.7. Possible intramolecular hydrogen bonding.

Again, using VISTA, it was possible to tabulate intramolecular (N....O) distances. The mean value of 2.561(14)Å was obtained from the histogram shown in Figure 5.8.

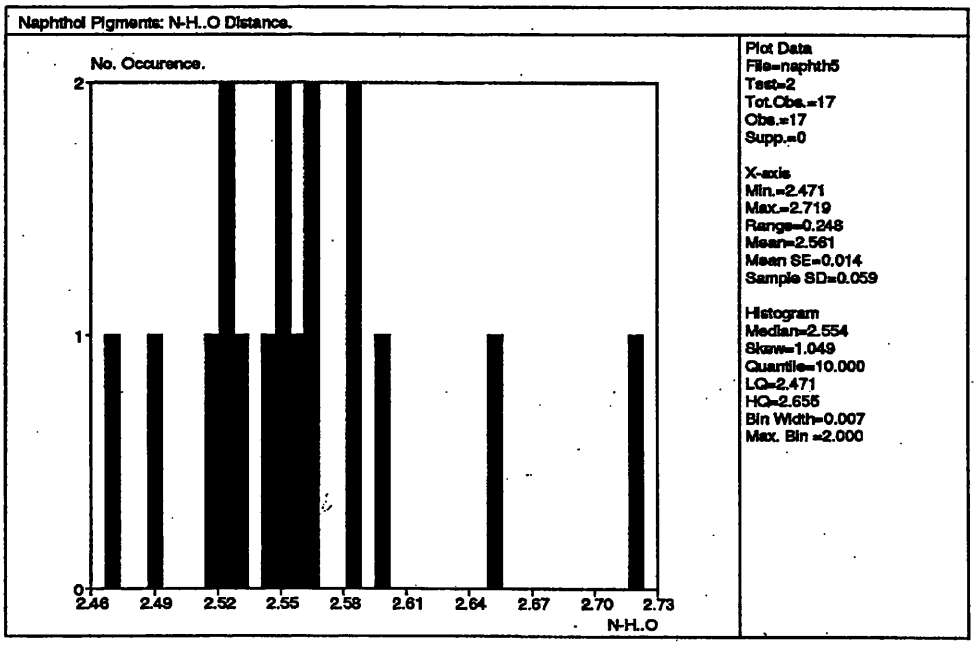


Figure 5.8. N-H....O Distance.

Comparison with tabulated hydrogen bonding distances[168] suggest that there is hydrogen bonding between N(1) and naphthol oxygen in all cases, as the average (N-H....O) distance is about 2.9Å

Intramolecular hydrogen bonding in these molecules has two effects, firstly it tends to hold the azo group in the same plane as the naphthalene moiety. Secondly it will tend to lock the structure in the *trans* configuration, thus allowing a hydrogen bonding interaction, which would otherwise not be possible if a *cis* conformation was adopted.

Note however that 2 pigments of those studied, namely CBANAP and TOAZNI have slightly longer (N-H....O) distances of 2.719Å and 2.655Å respectively. This is due to the fact that both of these pigments have two organic fragments coordinated to a metal atom, although the distances involved still qualify the interaction as a hydrogen bonding one. An interesting feature to note is that certain functionality at R(1) of the phenyl residue can potentially form another intramolecular hydrogen bonding interaction. The most common side groups, and those exhibiting hydrogen bonding potential are Cl and NO_2. In all cases investigated, this side group, -R, is on the same side of the azo bond as the naphthol oxygen.

Tabulation of the distances from side group, R=Cl, to N(1) of the azo bond are given in Table 5.7., along with their found mean value.

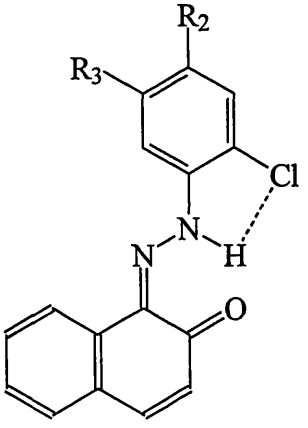


Figure 5.9. H....Cl Intramolecular Hydrogen Bonding.

Refcode	Cl....H-N(1) Å
ANACMO	2.924
CPZNXA	2.893
CPZPAN	3.026
Mean Value	2.948(40)

Table 5.7. N-H....Cl Distances (Å)

The above tabulated distances are too short for a non-bonded pair of atoms and comparison with the known average N-H....Cl bond of 3.21Å, confirms that there is further hydrogen bonding present.

When R=NO_2, values for (N-H....O) distances have been obtained and are given in Table 5.8.

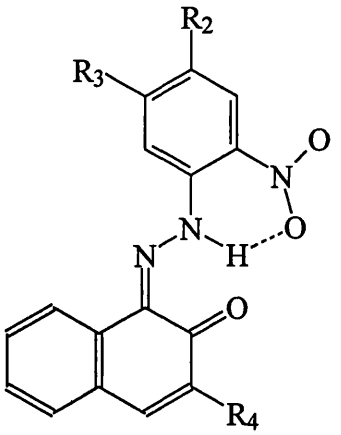


Figure 5.10. (N-H.....O) intramolecular interactions.

Refcode	N-H....O(NO_2) Å
MNIPZN	2.619
NQNCPH	2.608
NQNCPH01	2.619
Mean Value	2.615(6)

Table 5.8. N-H....O Distances (Å)

Again, remembering that average N-H....O distance is about 2.9Å, all of the three above interactions can be considered to be intramolecular hydrogen bonds.

Thus, it can be concluded that if there is a side group, in this case Cl or NO_2, in the *ortho* position of the phenyl connected to the azo group that has hydrogen bonding capabilities, then a bifurcated hydrogen bond exists between the keto oxygen and this side group. Furthermore, this is another driving force for molecular planarity.

Three of the pigments studied have functionality present at R_4. In all cases, this is a derived amido side chain of the form, CONHPh, with the subsidiary phenyl group having some additional functional groups, shown in Figure 5.11.

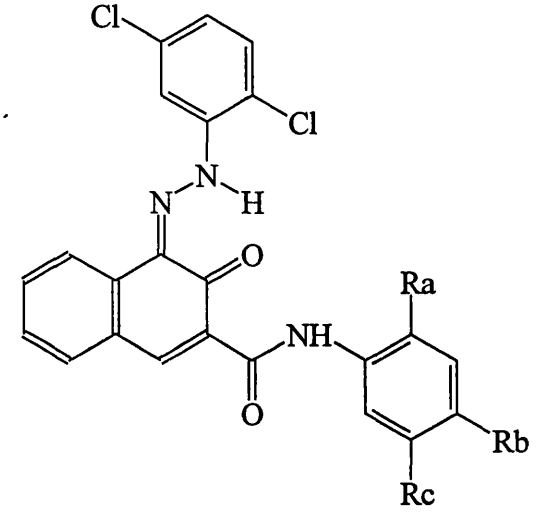


Figure 5.11. Amido Functionality.

Refcode	R_a	R_b	R_c
ANACMO	OCH_3	Cl	OCH_3
CPZNXA	OCH_3	H	H
CPZPAN	H	H	H

Table 5.9. Functionality.

Utilising previous structural evidence for intramolecular hydrogen bonding, these pigments with the amido side groups should have the propensity to hydrogen bond, but only if the nitrogen of the amido group is sufficiently close to the naphthol oxygen.

These distances were measured and are given in Table 5.10, along with the obtained mean value. Values show that hydrogen bonding does exist between these two atoms.

Refcode	(amido) N-H....O Distance Å
ANACMO	2.281
CPZNXA	2.307
CPZPAN	2.285
Mean Value	2.291(8)

Table 5.10. N-H(amido)....O Distance (Å)

In continuing the investigation of intramolecular hydrogen bonding, two pigments, CPZNXA and CPZPAN whose molecular structure has previously been illustrated, contain in the subsidiary amido group, an *ortho*-methoxy group that could possibly have a bifurcated hydrogen bond with the amido NH as shown in Figure 5.12.

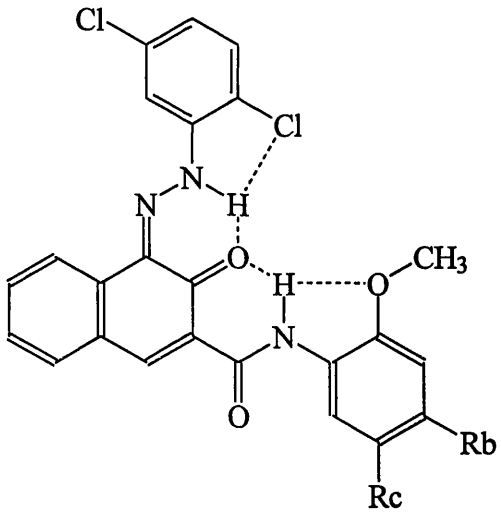


Figure 5.12. Bifurcated hydrogen bonding.

Again, using VISTA the values obtained for N-H....O(methoxy) distance for ANACMO and CPZNXA are 2.549 Å and 2.554 Å respectively thus indicating another bifurcated hydrogen bond as normally accepted value is around 2.9Å.

Thus it is possible to have an intramolecular network of up to 4 hydrogen bonds present in the structure, depending on the functionality present, which all contribute to the planarity of these molecules.

Investigation of the 3D arrangement of these β-naphthol pigments, finds that they all conform with regard to the stacking of molecules. They stack above one another, thus giving rise to a columnar structure. The column nature of the crystal structure develops parallel to the shortest unit cell dimension. Although the molecules stacking in this manner may do so such that the relationship between adjacent molecules may be anti-parallel or parallel, a common feature is for the development of a 'herring-bone' type packing. Figure 5.13. depicts a view of JARPIB01. As can be seen, when viewed along certain directions, the orientation of the molecules is found to alternate such that the molecules in adjacent columns are inverted and rotated with respect to each other.

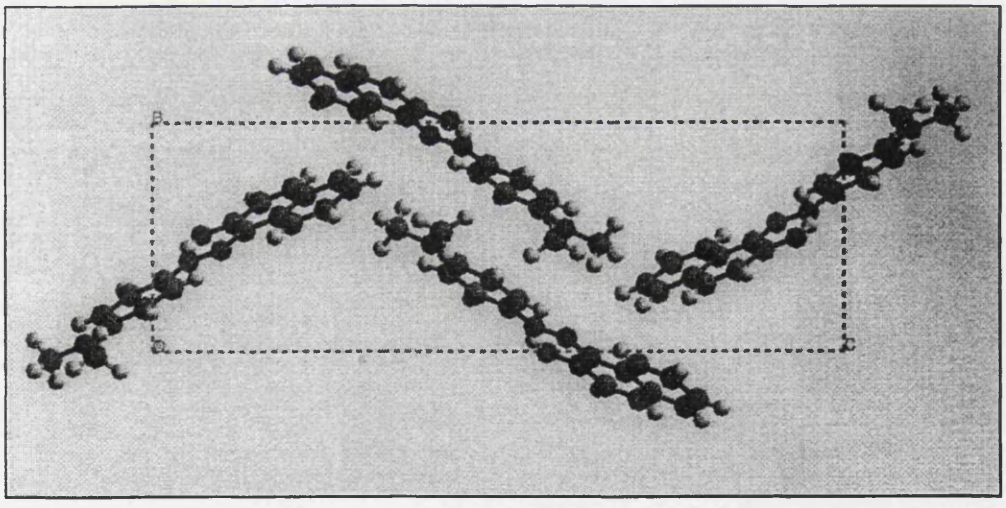


Figure 5.13. Packing arrangement in JARPIB01.

Previous studies of β-naphthol pigments, based on more limited structural information concluded that no intermolecular hydrogen bonding was found to exist. It was thought that molecules of this type were only linked by weak van der Waals forces, and indeed for some of the structures studied here, this still holds true. However, contrary to the above, the more extensive studies presented here show that intermolecular hydrogen bonding can, and does exist in some of the pigments studied.

Using *CERIUS*[2] it was possible to monitor all possible close contacts between adjacent molecules. A cut-off distance of 3.28 Å was imposed, following from the evidence presented by Sutor[168] that C-H....O hydrogen bonding could occur at distances of up to 3.28 Å. Table 5.14. lists the (H....O) contact distances found for each of the β-naphthol pigments under investigation.

Refcode	(H...O) Distance						
ANACMO	-	-	-	-	-	-	-
CBANAP	-	-	-	-	-	-	-
CPZNXA	-	-	-	-	-	-	-
CPZPAN	-	-	-	-	-	-	-
JARPEX	2.65	3.07	3.13				
JARPEX01	2.58	2.97	3.09				
JARPEX02	-	-	-	-	-	-	-
JARPIB	-	-	-	-	-	-	-
JARPIB01	-	-	-	-	-	-	-
JATJIX	-	-	-	-	-	-	-
MNIPZN	2.69	2.89	2.92	3.07	3.14	-	-
NBZANO	-	-	-	-	-	-	-
NBZANO11	2.48	2.96	3.04	-	-	-	-
NQNCPH	2.51	2.68	2.81	2.98	3.02	-	-
NQNCPH01	2.67	2.75	2.81	2.91	3.12	3.16	-
PAMBOO	2.28	2.69	2.80	2.83	2.84	2.89	2.90
TOAZNI	-	-	-	-	-	-	-

Table 5.11. (C-H.....O) Intermolecular contact distances.

PAMBOO, which has hydroxyl functionality *ortho* to the azo linkage is found to possess the most extensive intermolecular hydrogen bonding. Other pigments exhibiting this intermolecular interaction , although to a lesser degree, are MNIPZN, NBZANO11, NQNCPH and NQNCPH01. All are found to have nitro functionality, again in the *ortho* position, but also in the position *para* to the azo linkage. This highlights the fact that NO_2 groups present in pigment molecules have a propensity to hydrogen bond, both intramolecularly and intermolecularly. It is believed that extensive hydrogen bonding interactions are beneficial for the required pigmentary properties, and as such inclusion of nitro functionality, either in *ortho* or *para* position, is recommended.

An interesting point to note is that possible (C-H....O) interactions were identified in both JARPEX and JARPEX01. Neither of these pigments have recognised

functionality capable of hydrogen bonding, as each possible position on the phenyl ring is occupied by hydrogen. Therefore, hydrogen bonding in this case is via the naphthol oxygen and is derived only from the basic structural template.

This extensive examination of crystal structures representative of the β-naphthol class has identified several structural features that are common to all of the pigments studied, regardless of functionality. Also other features important with respect to pigment chemistry have been investigated, especially hydrogen bonding characteristics and identification of functionality that can aid this important structural feature.

Overall, these pigments are found to exist not as the azo-tautomer, but as the keto-hydrazone form. Depending on the functionality present, an extensive network of up to 4 hydrogen bonds can develop within the molecule, aiding planarity. Finally, intermolecular hydrogen bonding does exist under certain conditions of functionality within the stacked columns of molecules.

5.5. Structure Review of Acetoacetanilide Pigments.

Azo Yellows, or pigments belonging to the acetoacetanilide class have the general structural formula shown in Figure 5.14.

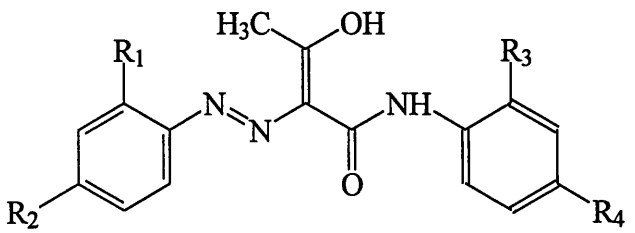


Figure 5.14. Generalised structure of Acetoacetanilides.

As discussed, pigments of this class are typically orange and yellow pigments, with their coupling components being acetoacetanilide derivatives.

Acetoacetanilide pigments share many of the structural features discussed previously for β-naphthol pigments, especially the tautomerism at the azo bond. There exists the possibility for azo-enol form or keto-hydrazone form.

Again, study of the results of crystal structure analyses can help in the determination of which tautomeric form each given pigment exists as, along with finding structural trends that could give a clue to the relationship that exists between the structure of pigments and their physical characteristics.

17 completed crystal structures of pigments in this class were found using the search procedure previously outlined. Individual Cambridge Database refcodes, molecular structure and primary literature reference are now given in Table 5.12.

Refcode	R_1	R_2	R_3	R_4	Reference
CEWGOA	CH_3	H	CH_3	H	169
CEWGUG	OCH_3	H	OCH_3	H	171
CIVYIP01	H	H	H	H	170
CIVZEM10	CH_3	H	H	H	171
DANDIF	NO_2	H	CH_3	Cl	172
DUXXID	NO_2	Cl	H	H	173
FOVNOT	NO_2	H	H	H	174
FUCTOM	NO_2	CH_3	OCH_3	H	175
FUCTUS	OCH_3	NO_2	OCH_3	H	176
NCPAAA01	NO_2	Cl	H	H	177
QQQAZG02	NO_2	CH_3	H	H	155
QQQAZG03	NO_2	CH_3	H	H	178
QQQAZG21	NO_2	CH_3	H	H	179
QQQAZG31	NO_2	CH_3	H	H	180
QQQEKP10	NO_2	Cl	Cl	H	181
QQQELA03	H	NO_2	H	H	182
SANZEM	H	Cl	H	H	183

Table 5.12. CSD Refcodes for acetoacetanilide pigments.

Further examination of these crystal structures using *VISTA* and *CERIUS*[2] has allowed tabulation of various structural parameters in an analogous manner to that of β-naphthol pigments.

Bond lengths to be considered when assigning the tautomeric form of the pigment are those adjacent to and including the azo bond. In the acetoacetanilide case, the

situation is more complicated due to the fact that there are two hydrogen atoms that maybe involved in tautomerism and these give rise to several possible configurations, some of which are given in Figure 5.15.

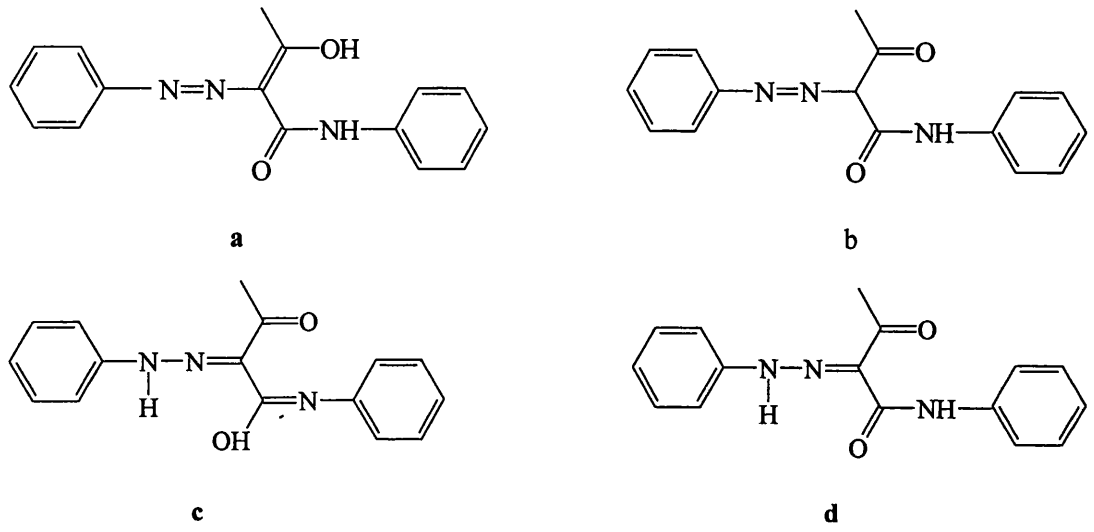


Figure 5.15. Possible tautomers of acetoacetanilide pigments.

In X-ray crystal structure analyses, there are two ways of detecting a tautomer. One is to determine the hydrogen atoms positions and the other is to determine the various bond lengths. As these lengths will determine whether bonds are single or double in character, accurate information on bond length will help to assign bond character. When determining hydrogen atom positions, it must be remembered that hydrogen is a much weaker scatterer of X-rays than carbon, oxygen or nitrogen, so much so, that it may not be possible to find hydrogen atoms in the presence of heavier atoms. Subsequently, not all of the crystal structures studied have located the positions of the relevant hydrogen atoms, therefore, bond lengths will be used to assign tautomeric form.

As was mentioned, examination of the appropriate bond lengths adjacent to the azo bond indicate the given tautomeric form. Figure 5.16. shows a more detailed picture of the bonds involved (shown in red) and whose distance will be tabulated and discussed later.

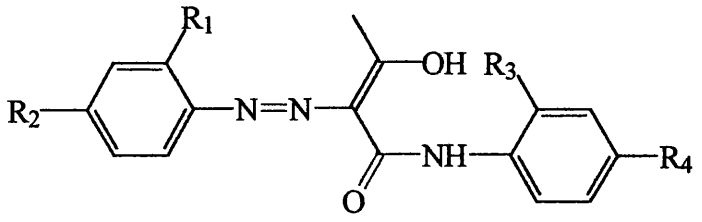


Figure 5.16. Bond distances of interest.

The bond of primary interest is the azo linkage. *VISTA* obtained plot for N(1)-N(2) distance is given in Figure 5.17. A mean N-N distance of 1.319(2)Å was found and comparison with tabulated double and single values shows that this bond is intermediate between double and single values of 1.240Å and 1.401Å respectively. This suggests some degree of delocalisation.

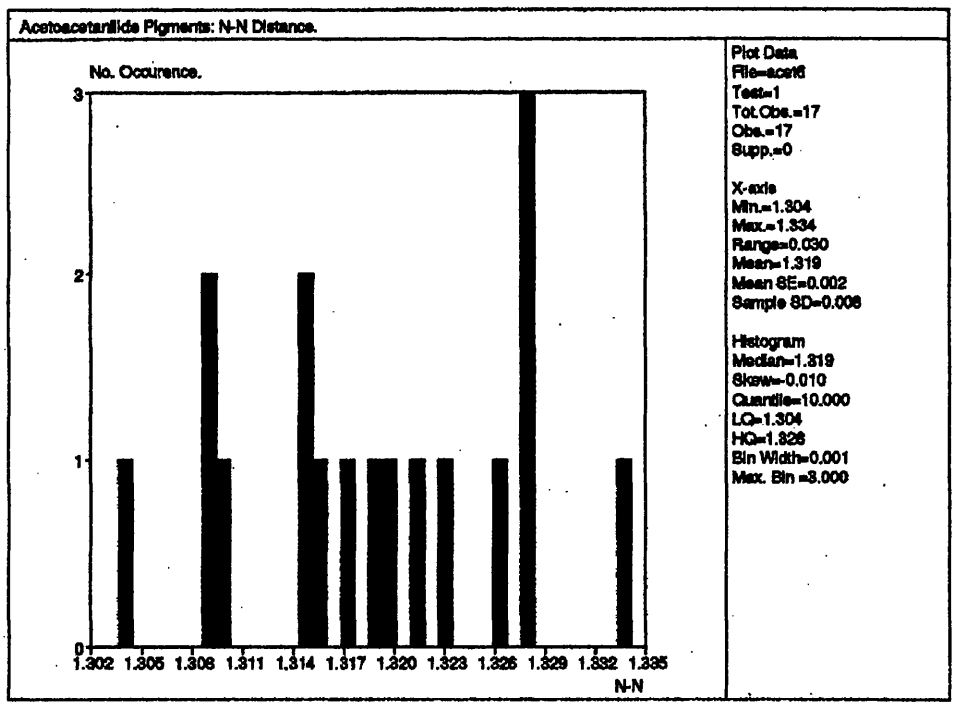


Figure 5.17. N-N Distance.

The C-N(2) bond adjacent to the azo linkage can again be either single or double bond in character depending on the conformation adopted. Expected lengths for double and single bonds are 1.32Å and 1.47Å respectively. Using *VISTA* for analysis

The C-N(2) bond adjacent to the azo linkage can again be either single or double bond in character depending on the conformation adopted. Expected lengths for double and single bonds are 1.32Å and 1.47Å respectively. Using *VISTA* for analysis of (C-N) bond lengths allows for the production of histogram shown in Figure 5.18. A mean value of 1.313(2)Å was obtained, which is indicative of double bond character.

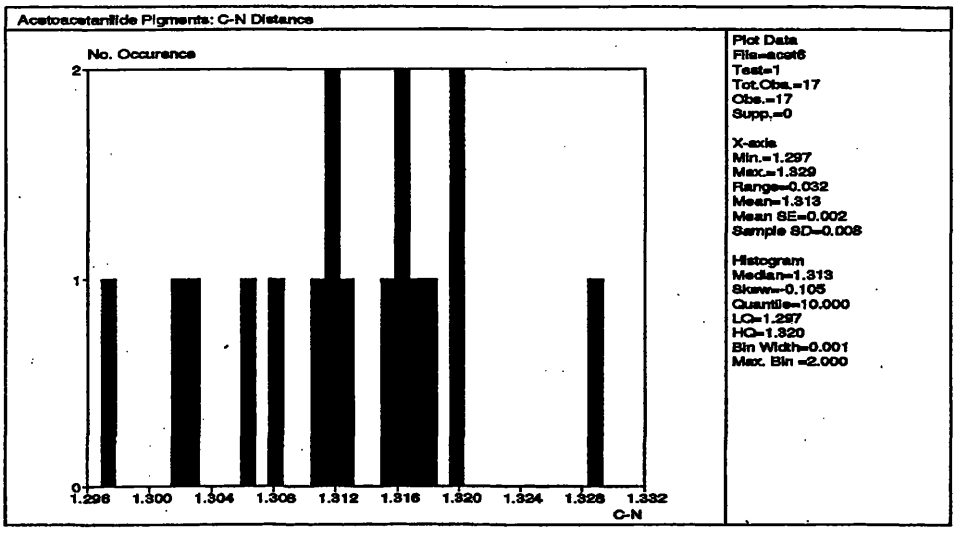


Figure 5.18. C-N Distance.

The next bond to be considered, as shown in Figure 5.16., is [C-(O1)]. This was found to have a mean value of 1.224(2)Å, as demonstrated in the histogram in Figure 5.19., which again indicates the presence of a double bond when compared to the tabulated value for (C=O) of 1.23Å.

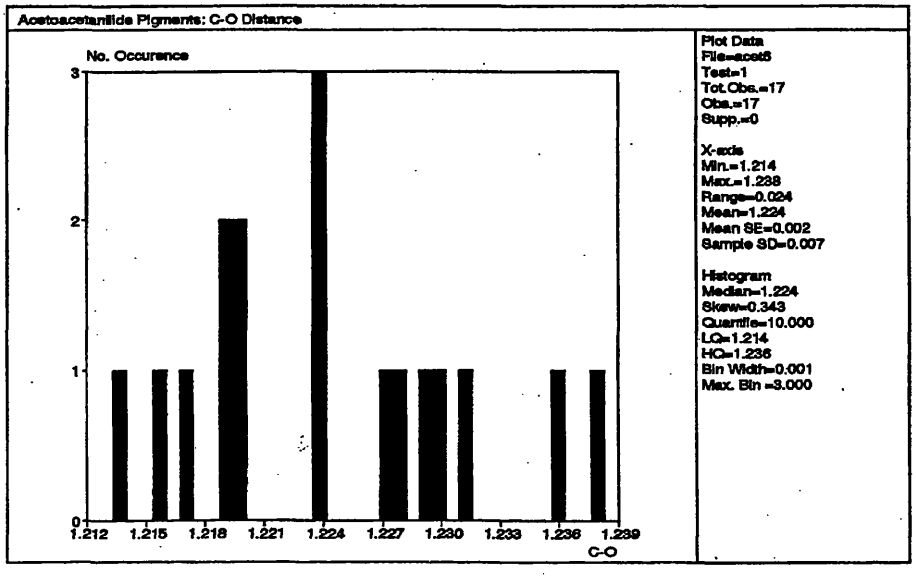


Figure 5.19. [C-O(1)] Distance

In an analogous manner, [C-O(2)] bond length was found to have a mean value of 1.233(2)Å, as obtained from Figure 5.20. Again, this is confirmation of the existence of double bond character when compared to the expected (C=O) values given above.

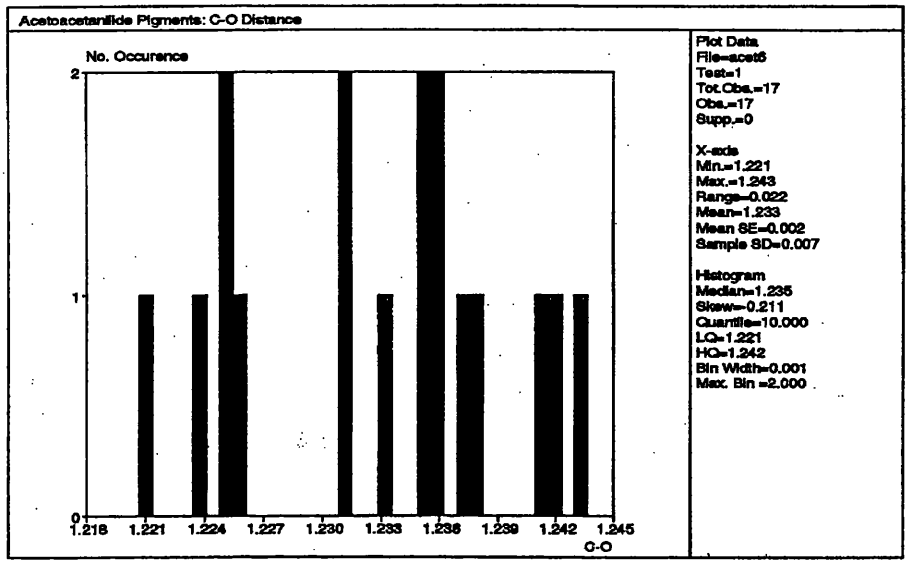


Figure 5.20. C-O(2) Distance.

The final bond to be investigated is [C-N(3)], which if following the above convention that suggests hydrazone conformation, should have close to single bond character. In fact, mean (C-N) length was found to be 1.348(1)Å, which is longer than the accepted partial double bond value of 1.32Å, but still shorter that single bond of 1.47Å. Again giving credence to the postulated concept of delocalisation within these structures. The presence of these bond lengths as outlined above confirms the hydrazone conformation, along with it would appear, the presence of some partial double bonds, in that their lengths lie between accepted single and double bond values.

It has to be noted that the distributions present in the histograms are not complete and do not have a Gaussian distribution and as a consequence of this, the standard deviations given could be unreliable.

Once primary bond lengths were considered, non-bonded or more specifically hydrogen bonding interactions within these molecules were examined.

Common to all of the 17 Acetoacetanilide pigments studied are two (N....O) interactions which arise due to the adoption of the hydrazone conformation. The atoms involved and the positions of these interactions are shown in Figure 5.21.

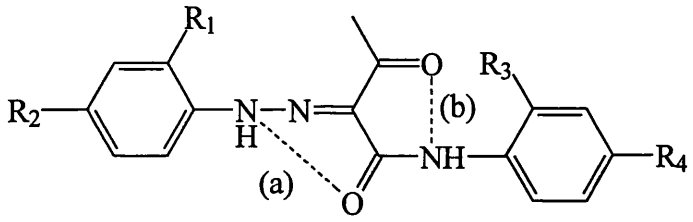


Figure 5.21. Possible Hydrogen bonding interactions.

Mean values obtained for interaction (a) were obtained from the *VISTA* histogram in Figure 5.22., and was found to be 2.584(5)Å. Comparison with tabulated hydrogen bonding distance for (N....O) of 2.9Å proves that this is in fact a non-bonded interaction.

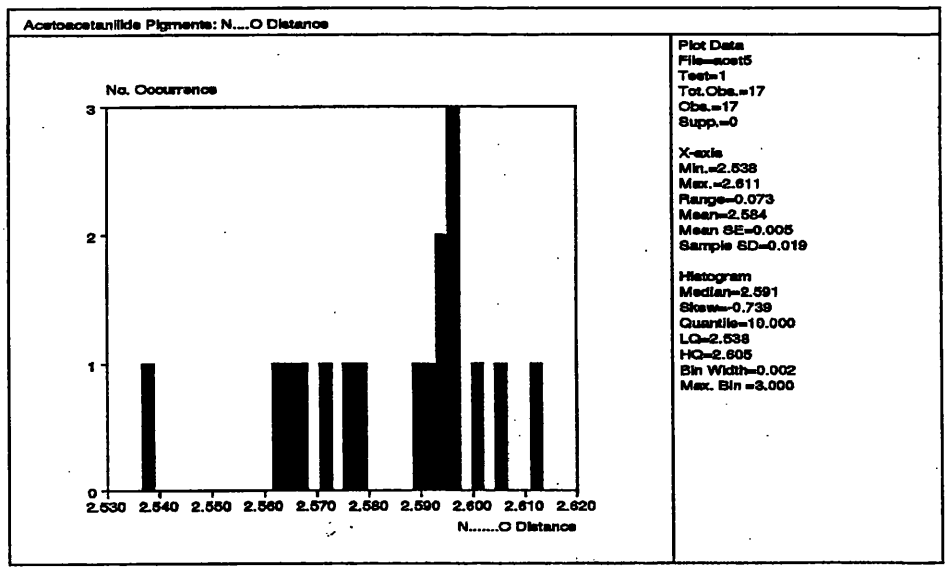


Figure 5.22. N-H....O Distance.

Examination of interaction (b) furnishes a mean non-bonded (N....O) interaction distance of 2.658(3)Å as shown in Figure 5.23.

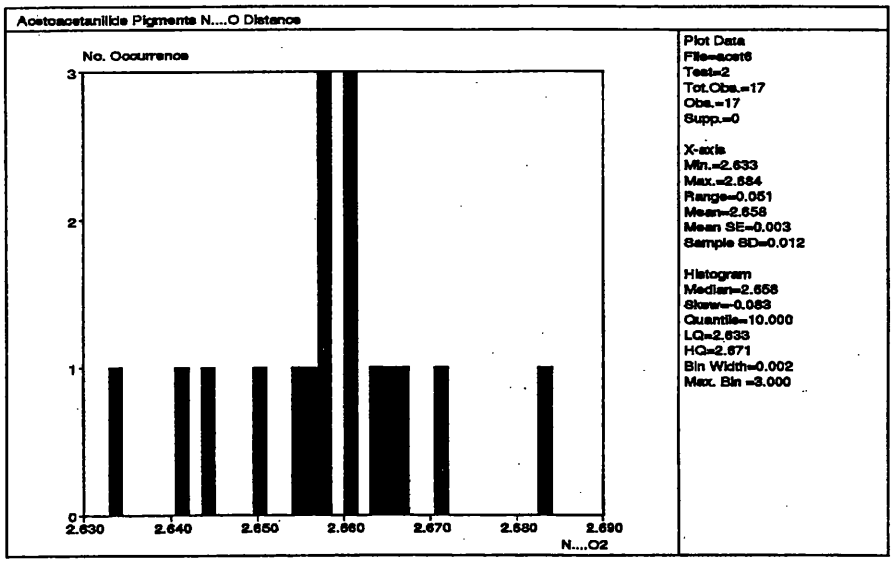


Figure 5.23. N-H....O Distance.

For some of the pigments under investigation there is functionality present in the *ortho-* position of one or both of the phenyl groups. If these functional groups are oxygen containing, then further intramolecular hydrogen bonding can be seen to exist. The most common side groups involved are NO_2 and OCH_3.

When R_1 or R_3 is OCH_3, Figure 5.24. shows the extent of possible (N....O) interactions, along with the found mean values.(Table 5.13.)

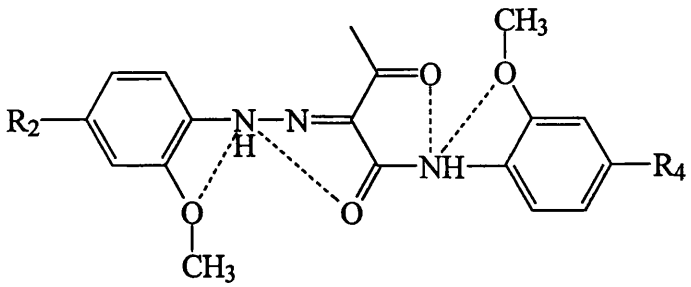


Figure 5.24. Possible (N-H....O) Interactions.

Refcode	N1....O(OCH_3)Å	N3....O(OCH_3)Å
CEWGUG	2.599	2.599
FUCTUS	2.623	2.568
Mean Value	2.611(1)	2.583(2)

Table 5.13. (N....O) Distances Å

Again, all of these values indicate a (N....O) non-bonded interaction, thus a network of hydrogen bonding can be seen to exist.

When the functionality present is R_1=NO_2, a similar trend can be seen to exist. Table 5.14. contains tabulated (N....O) distances as obtained from conformational analysis using *VISTA.*

Refcode	[N1-H.....O(NO_2)]Å
DANDIF	2.625
DUXXID	2.603
FOVNOT	2.603
FUCTOM	2.600
NCPAAA01	2.617
QQQAZG02	2.604
QQQAZG03	2.604
QQQAZG21	2.608
QQQAZG31	2.599
QQQEKP10	2.598

Table 5.14. (N-H....O) intermolecular interactions.

As with previous findings, the values shown here indicate the presence of (N-H....O) intramolecular hydrogen bonding.

Overall, it can be concluded that when there is an oxygen-containing side group, in this instance methoxy or nitro, occupying the position *ortho* to the hydrazone linkage, then this hydrogen bond is bifurcated. The important point to note is that this occurs regardless of whether the functionality is nitro or methoxy.

Most of the pigments studied do not have functionality present at the *ortho* position on the second phenyl ring. However, CEWGUG, FUCTOM and FUCTUS have methoxy functionality in this position and as indicated, this gives rise to another bifurcated hydrogen bond. Therefore, there exists the possibility of a network of hydrogen bonding throughout the molecule thus holding parts of the molecule planar, which again is a prerequisite for good pigment properties.

Having completed an investigation of the important structural features of individual molecules, it follows that the arrangement and relationship of molecules in 3D as found in a crystal have to be examined.

Initial investigations of the determined crystal structures found that, in common with β-naphthol pigments, the almost planar molecules are stacked above one another giving a columnar structure.

In all cases examined, the relationship between adjacent molecules within the columns is of an anti-parallel nature. Figure 5.25. illustrates the most common packing motif with a view of CEWGUG shown.

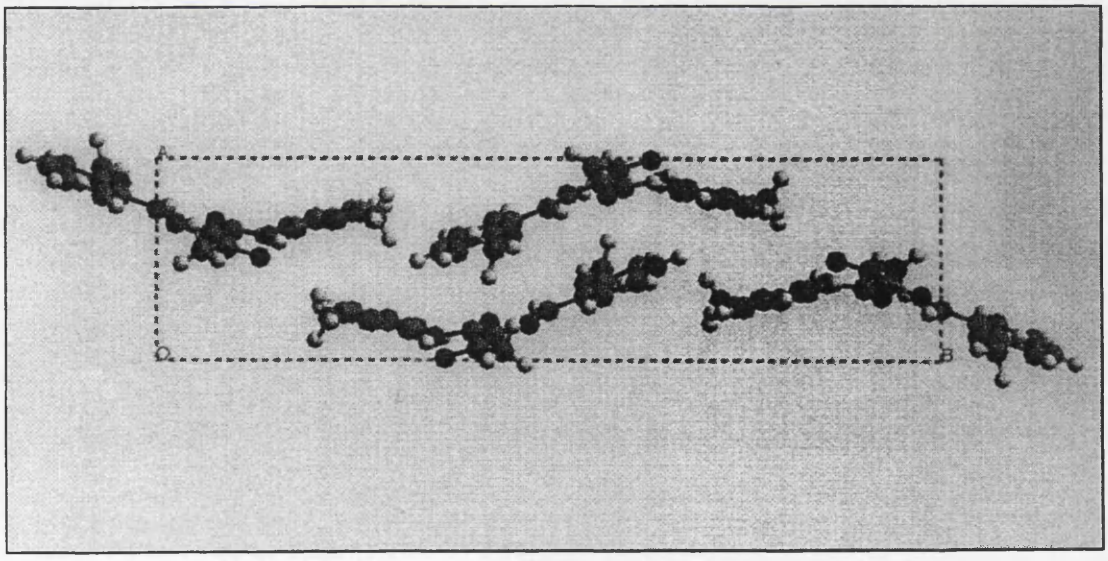


Figure 5.25. Packing arrangement in CEWGUG.

Note that this is similar to the 'herring-bone' packing arrangement identified in the previously studied β-naphthol pigments. However, QQQEKP10 was found to have a different three-dimensional array of molecules, in that the molecules are found in slightly undulating layers, shown in Figure 5.26.

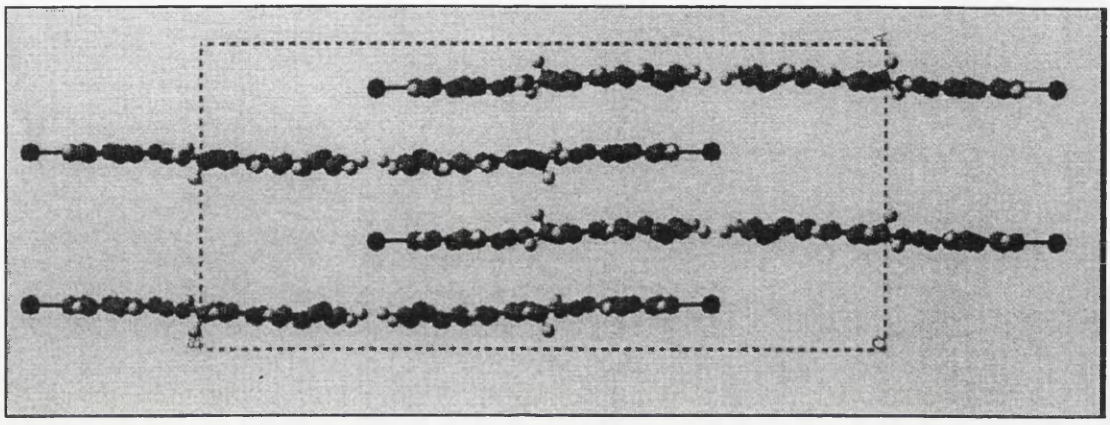


Figure 5.26. Layered structure of QQQEKP10.

As for packing of columns of molecules, it is found to be parallel to the shortest unit cell dimension.

Pointed out previously was the extensive intramolecular hydrogen bonding found in all of the pigments studied, but is this also seen for intermolecular interactions? Use was made of *CERIUS*[2], with a cut-off distance of 3.28 Å imposed, as it was expected that no hydrogen bonding interaction exists beyond this distance. Each pigment was studied in turn and Table 5.15. lists the values obtained for (C-H....O) intermolecular interactions.

In contrast to the β-naphthol pigments, intermolecular hydrogen bonding is found to be extensive throughout this acetoacetanilide series of pigments. All values found are less than 3.28 Å, thus indicating the presence of hydrogen bonding interactions.

Refcode	(C-H....O) Å					
CEWGOA	2.92	3.06	-	-	-	-
CEWGUG	2.57	2.83	2.89	3.01	-	-
CIVYIP01-	-	-	-	-	-	-
CIVZEM10	2.54	2.58	2.68	2.77	2.78	-
DANDIF	2.93	3.04	-	-	-	-
DUXXID	2.58	2.71	2.75	-	-	-
FOVNOT	2.54	2.88	2.89	2.91	3.17	3.23
FUCTOM	2.58	2.95	3.09	3.25	-	-
FUCTUS	2.54	-	-	-	-	-
NCPAAA01	2.81	2.82	2.88	2.92	3.21	-
QQQAZG02	2.53	2.67	2.68	2.70	2.95	3.21
QQQAZG03	2.60	2.71	2.80	3.02	3.10	-
QQQAZG21	2.57	2.70	2.71	2.95	-	-
QQQAZG31	2.55	2.73	2.80	2.97	3.06	-
QQQEKP10	2.82	2.90	-	-	-	-
QQQELA03	2.59	-	-	-	-	-
SANZEM	-	-	-	-	-	-

Table 5.15. (C-H....O) Intermolecular interactions.

If the functionality present has the propensity to hydrogen bond, as with methoxy and nitro groups, then it is found that intermolecular hydrogen bonding is common. This is further evidence for the inclusion of such functionality thus giving rise to cohesive intermolecular interactions which are of particular importance to the pigment chemist.

Thus, the extensive findings presented here provide further evidence for certain structural features that are of paramount importance for the success of these molecules as pigments.

In conclusion, these pigments exist in the solid state as the hydrazone tautomer. Intramolecular hydrogen bonding is a common structural aspect and regardless of the functionality present, there is always at least two of those interactions per molecule. However, subject to the appropriate functionality of the phenyl rings, this may rise to up to four per molecule. Finally, a columnar stacking array is most common and extensive intermolecular hydrogen bonding can exist, again conditional of the side-groups present.

6.0. ELECTRON CRYSTALLOGRAPHIC STUDIES OF BARIUM LAKE RED C.

6.1. Introduction.

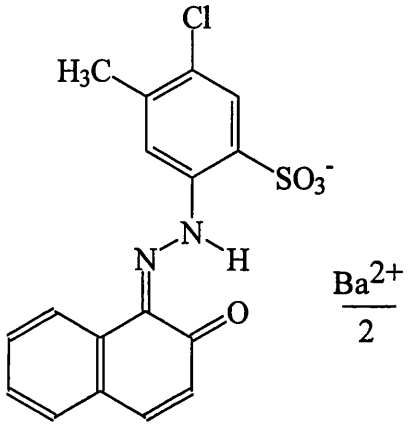


Figure 6.1. Barium Lake Red C.

At the present time Barium Lake Red C (Pigment Red 53:1)[184,185,186] is one of the most commonly used organic pigments. It is produced in a range of particle sizes, with a result that its qualities can range from transparent to opaque. As a group, this family of pigments are characterised by fairly poor lightfastness and indifferent chemical fastness. They do however exhibit good solvent and migration resistance. They are generally inexpensive pigments and are tinctorially strong, thus finding their greatest usage in printing inks. Surprisingly, Barium Lake Red C exhibits good heat stability and is occasionally used in plastic applications where lightfastness is not a crucial factor.

Barium Lake Red C is formed by the coupling of β-naphthol with 2-chloro-5-amino-toluene-4-sulphonic acid. Pigment is laked or precipitated with barium chloride. The choice of laking agent determines the shade of the pigment, extending from yellowish red to bluish red. The most important colour is the yellowish red of the barium salt.

As crystallite size is of the range 0.01 - 0.1 μm, it is not possible to carry out structure elucidation by conventional X-ray methods and so structure determination has to be achieved by other means.

6.2. Experimental Synthesis of Barium Lake Red C.

The following procedure describing the experimental synthesis of Barium Lake Red C is adapted from an industrial process. Synthesis of the compound was carried out during time spent at the research laboratories of Ciba Specialty Chemicals, Paisley.

6.2.1. Diazotisation.

In the diazotisation vessel 2-chloro-5-amino-toluene-4-sulphonic acid (43.05g, 0.194 mol) was dispersed in water (320 cm^3). Added to this solution was concentrated ammonia solution (10.3g), while continuously stirring the mixture to effect complete dissolution. Once fully dissolved the solution was checked to be alkaline in nature. Hydrochloric acid (62g, 1.701mol) was added and then temperature of solution was lowered to below 3°C by the addition of ice. This temperature was maintained throughout the course of the reaction. In a separate beaker sodium nitrate (13.2g, 0.191mol) was dissolved in water (25cm^3). This solution was then added gradually to the diazo solution. Solution was then stirred continuously over a period of 45 minutes.

6.2.2. Coupling Component.

β-naphthol (28.04g, 0.195mol) was dispersed in water (270ml) and sodium hydroxide (17.9g, 0.447mol) was added. Once β-naphthol was completely dissolved, the temperature of the reaction was reduced to less than 5°C by the addition of ice.

6.2.3. Direct Coupling.

In the direct coupling route, the coupling reaction occurs in the coupling component vessel. The diazo component was pumped at a continuous rate into the coupling component vessel, which is being constantly stirred. This was continued until pH was seen to fall from about 10 down to 4. At this point diazo addition was stopped and pH of the solution raised to 7 with the addition of dilute ammonia. Coupling was then continued with diazo, while maintaining pH 7 by the simultaneous addition of dilute ammonia. Reaction mixture was then stirred for a further 5 minutes.

6.2.4. Laking.

In a separate beaker barium chloride (55.7g, 0.268 mol) in water (150 cm^3) were stirred and heated to 80°C. Pigment slurry was then heated to the same temperature by the passing of steam directly into the beaker. Upon reaching this temperature, barium chloride solution was added to the pigment slurry and the resulting mixture stirred for a further 5 minutes. Temperature was then raised to 90°C and maintained for 10 minutes. Temperature was then reduced to 70°C by the addition of ice. Pigment slurry was the filtered and washed through a large Buchner funnel until conductivity of filtrate was found to be less than 200μS. Pigment was then dried in an oven overnight at 70°C. The dried pigment was then weighed (70.78g, 71% yield)

6.3. Elemental Analysis of Barium Lake Red C.

Previous studies of similar pigments[8,9,10] had proposed dimer structures, in that the relationship between metal and organic moieties was in the ratio 2:1. Due to the presence of Ba^{2+} and only one negatively charged functionality in Barium Lake Red C, this ratio of metal to organic component was also expected.

This was confirmed by simple elemental analysis, where the ratio of organic to barium was indeed found to be in the ratio 2:1. Thus, agreeing with previous studies and

suggesting the presence of a dimeric relationship between barium and organic diazo component.

6.4. Determination of Tautomeric Form.

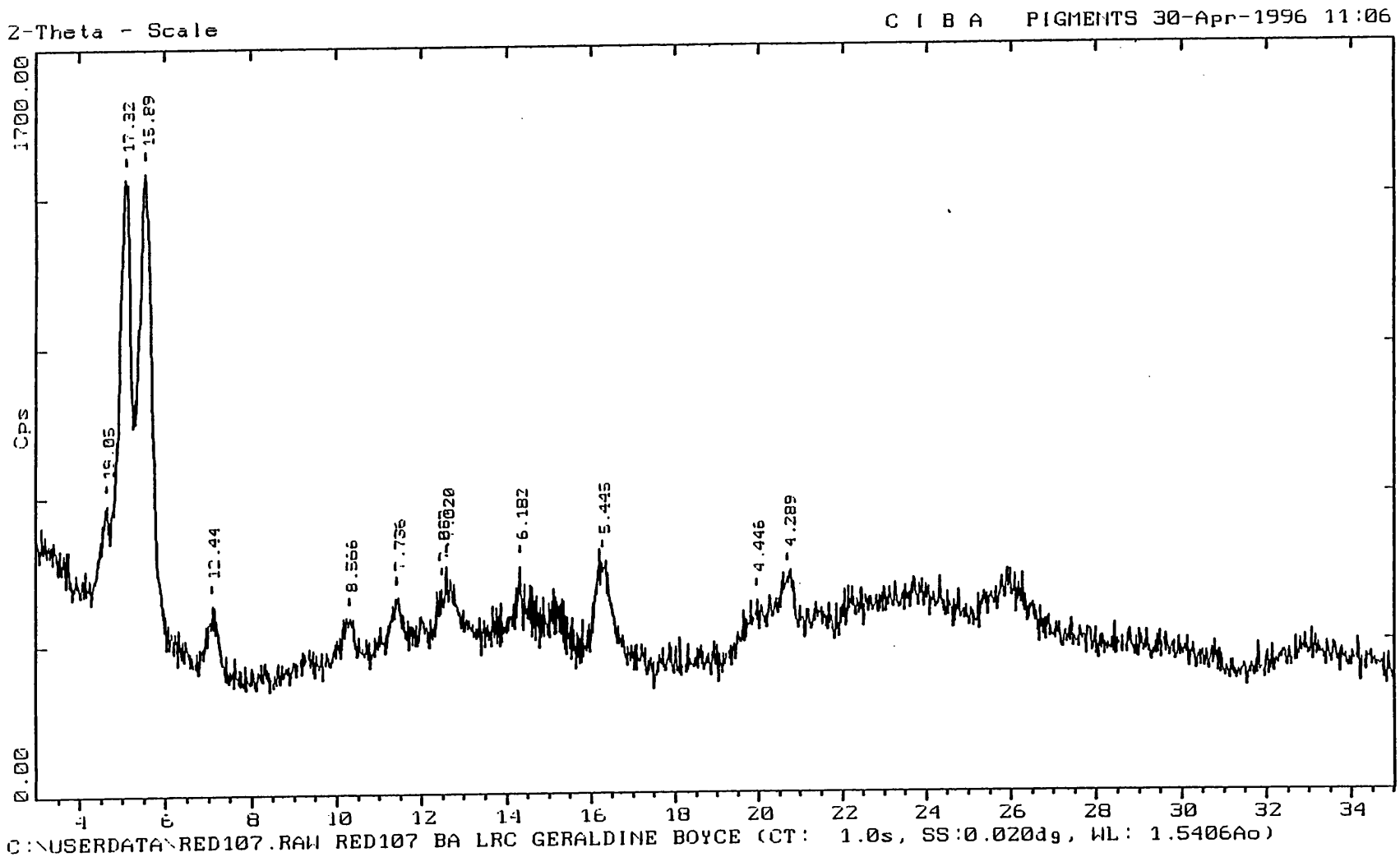
One of the most important structural features to be discussed in the previous chapter that reviewed crystal structures of some β-naphthol pigments, was that of keto-hydrazone tautomerism. It was noted that all pigments investigated were found to exist, not as the expected hydroxy-azo tautomer, but as the related keto-hydrazone tautomer. As was shown, the existence of a particular tautomer can be determined through examination of bond lengths in crystal structures. However, infrared spectroscopy studies can also aid in the elucidation of tautomeric form.

Each tautomer will have identifiable stretches in infrared spectroscopy. The hydroxy-azo tautomer will be recognisable due to the presence of a strong hydroxyl stretch, whereas the presence of both a carbonyl stretch and N-H band will indicate the existence of the keto-hydrazone tautomer.

IR spectrum was obtained for Barium Lake Red C. The presence of a carbonyl stretch at $1500cm^{-1}$ and the amido stretch at $3400cm^{-1}$ are both indicative of the keto-hydrazone tautomer, thus agreeing with those pigments studied previously in chapter 5.

6.5. X-ray Powder Diffraction.

X-ray powder diffraction spectra of Barium Lake Red C were obtained from Ciba Specialty Chemicals, Paisley. Studies were carried out using a Siemens D5000 Powder Diffractometer with Cu radiation of wavelength 1.5406 Å, sampling at 0.02° per second. Figure 6.2. shows the spectra obtained for a sample of Barium Lake Red C, with appropriate indexing and spacings found indicated.


Z-Theta - Scale
C I B A PIGMENTS 30-Apr-1996 11:06
1700.00
0.00
Cps
17.32
15.89
16.05
12.44
8.566
7.756
6.182
5.445
4.446
4.289
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
C:\USERDATA\RED107.RAW RED107 BA LRC GERALDINE BOYCE (CT: 1.0s, SS:0.020dg, WL: 1.5406Ao)

6.6. Electron Microscopy of Barium Lake Red C.

6.6.1. Sample Preparation.

The aim of sample preparation for electron microscopy is to obtain mainly thin, well dispersed crystals throughout the sample grid. Previous studies on azo pigments had found that vapour epitaxy methods used to deposit thin films of the pigment were unsuitable due to the pigment undergoing degradation upon heating. As a result, these methods were not employed in these studies.

Initially, the pigment was ground using an agate mortar and pestle. The ground pigment was then added to a test tube containing a mixture of ethanol and water (ratio 1:10) to give a fine suspension. A test tube was then placed in an ultrasonic bath for a period of 10 minutes. Using a micropipette, a drop of this suspension was pipetted on to carbon coated copper grids which were then dried in an oven overnight at 40°C.

6.6.2. Specimen Examination.

All microscopy was carried out using a JEOL 1200EX electron microscope, operating with an accelerating voltage of 120kV. This microscope is equipped with a side entry specimen stage, fitted with a goniometer allowing tilting of specimen along both the x and y axes. As the material under investigation is very susceptible to radiation damage within the electron beam, techniques were employed to limit the exposure of the crystals to the damaging electron beam. Various methods of reducing radiation damage were discussed in section 2.14.

During the course of these studies, the technique used was one where focusing is carried out as normal followed by the electron beam being deflected above the specimen. Shifting of the specimen then brings a previously unexposed area of the sample into view. Return of the electron beam to the specimen is simultaneous with image capture using the in-built camera, thus reducing the degree of radiation damage.

All images were captured on X-ray film (Kodak Industrex 100 CX), thus ensuring use of relatively small electron exposure times due to the inherent increased sensitivity of the film.

6.7. Electron Microscope Images of Barium Lake Red C.

The low magnification micrograph of crystals of Barium Lake Red C given in Plate 6.1. show an even distribution of crystals. Pigment is found to consist of mainly flat plate-like crystals, along with some shorter, thicker crystals. The crystallites are found to have average dimensions, in projection, of 5 μm and 20μm . Some of the crystals are found to have rounded edges which are characteristic of the manifestations of radiation damage.

Plate 6.2. shows the achieved resolution of a large lattice spacing of around 18 Å observable in some of the thin, plate-like crystals.

6.8. Electron Diffraction Patterns of Barium Lake Red C.

Upon examination of single crystals of Barium Lake Red C by selected area diffraction, two patterns were routinely observed. Plate 6.3. shows projection 1, the most commonly obtained projection. This was found to originate from the thinner plate-like crystals.

Plate 6.4. shows the less frequently obtained projection 2, which is found to occur from the observed thicker crystals, which themselves are found to be present in lesser numbers than the above plate-like crystals.

In an attempt to obtain further projection diffraction patterns, tilting experiments were carried out both in Glasgow and at The Hauptman-Woodward Institute, Buffalo. Very

Plate 6.1. Low magnification micrograph of Barium Lake Red C

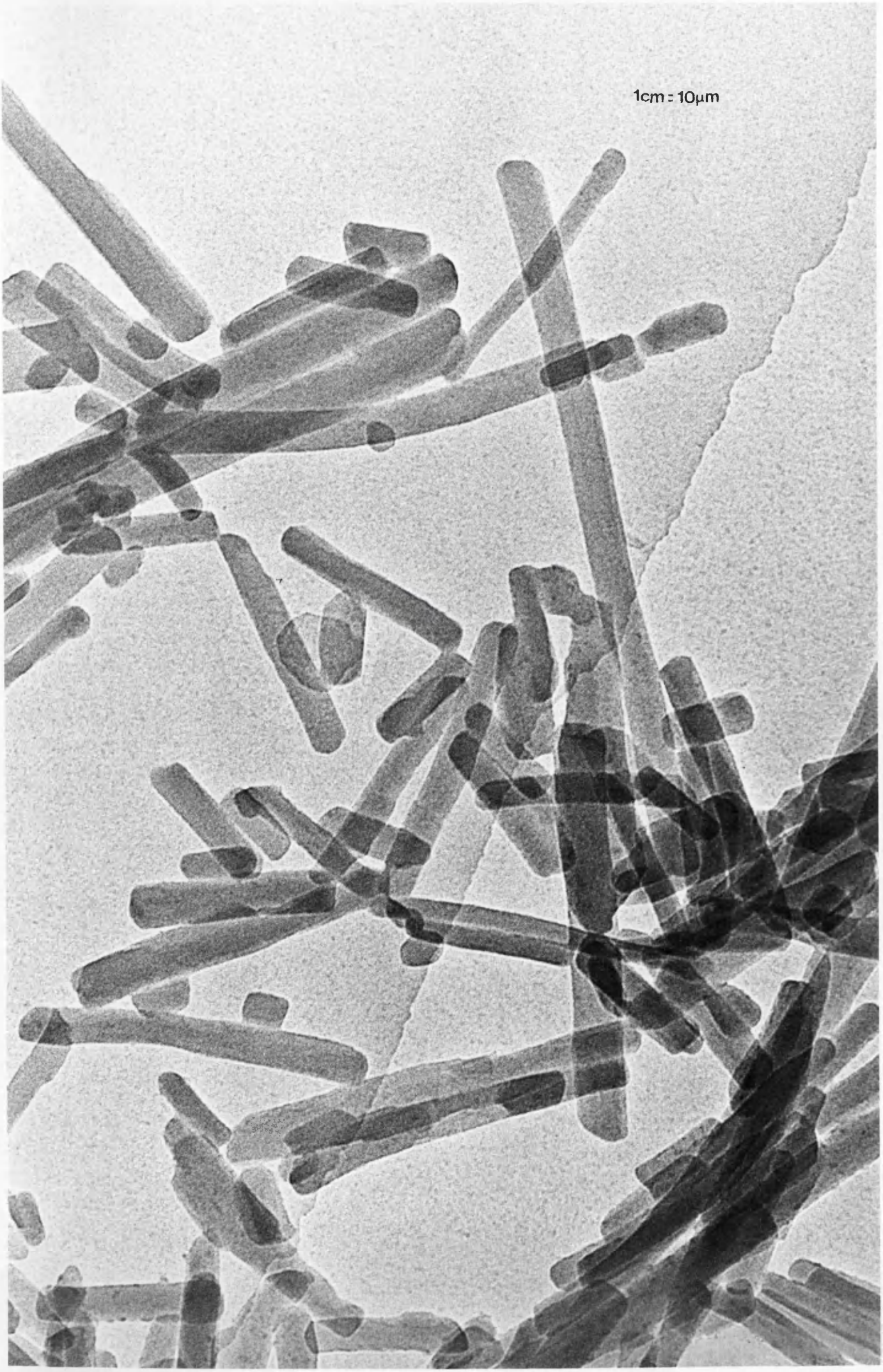

1cm = 10μm

Plate 6.2. Micrograph of Barium Lake Red C showing lattice spacing

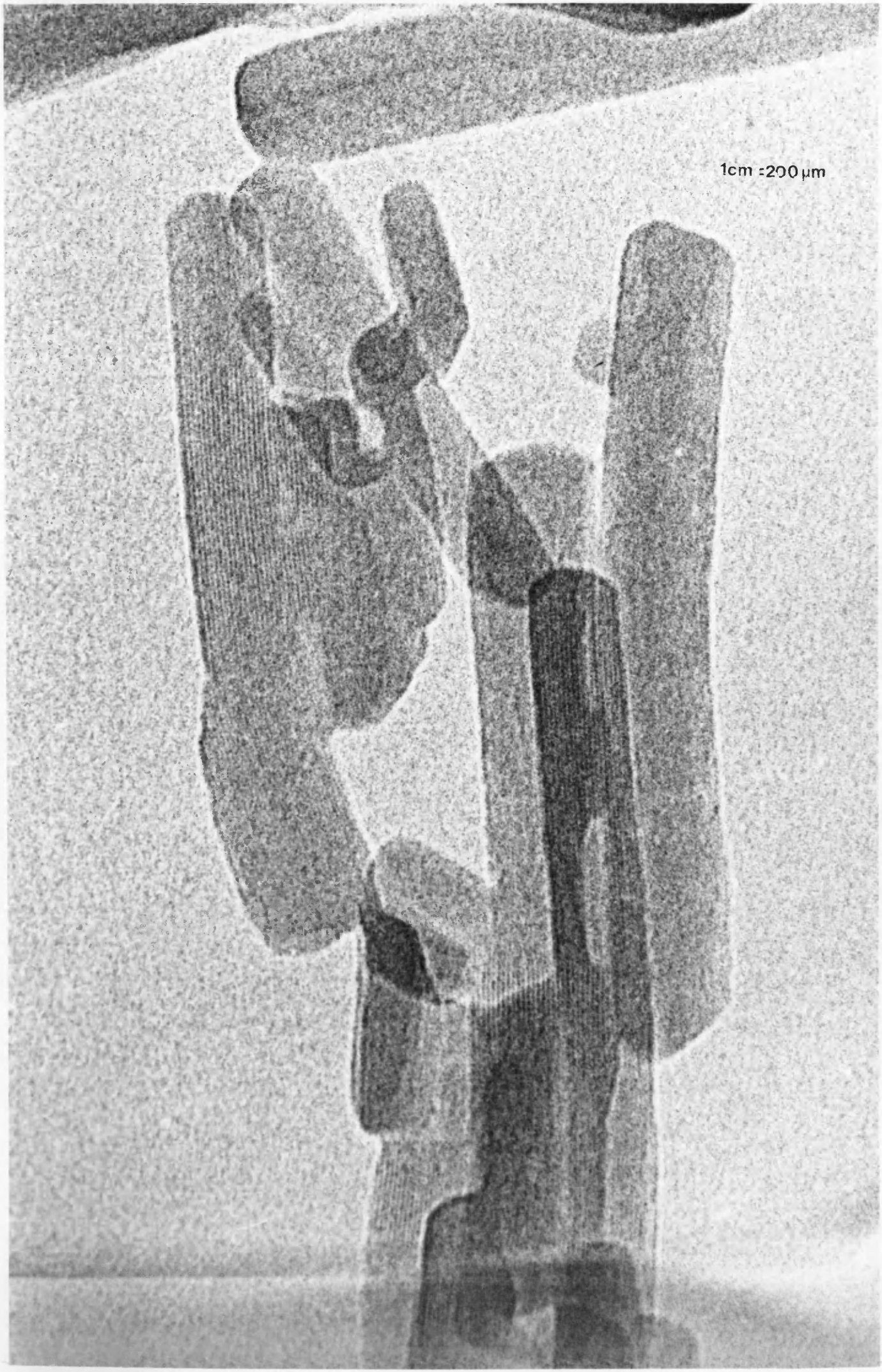

1cm =200 µm

Plate 6.3. Electron diffraction pattern from projection 1

Plate 6.4. Electron diffraction pattern from projection 2

little success was achieved, with the limiting factor being the amount of exposure that the crystals could survive in the damaging electron beam.

Using electron microscope operating constants, the *d*-spacings present in each diffraction pattern could be determined. It was found that the two projections had a *d*-spacing in common. Comparison with the previously obtained X-ray powder diffraction pattern, found that the most common spacings were existent in both the X-ray and electron diffraction patterns. Also, by taking into account the systematic absences that are present in the X-ray powder pattern and in projection 1, the monoclinic unit cell dimensions were determined to be a = 34.98 Å, b = 10.92 Å, c = 8.74 Å and $\beta = 94.2°$.

Using standard crystallographic relationships regarding density and unit cell parameters, Z, the number of molecules per unit cell could be determined. This was found to be equivalent to 4 of the postulated dimeric units.

6.9. Intensity Retrieval and Data Processing.

As described previously in section 3.3, it is possible to retrieve intensity information directly from electron micrographs using digitisation techniques followed by use of ELD[68]- software for quantification of diffraction patterns. Obtained from this are a list of indexed reflections and their corresponding intensities.

Many diffraction patterns were collected for each projection, thus enforcing the criteria of self-consistency as stated by Dorset.[29] Quantification of diffraction intensities for the best of these diffraction patterns of each projection was carried out in the manner as described above. Then, by utilisation of the SG MERGE facility within the *maXus*[187] suite of programs for structure determination, it was possible to correctly scale data and merge equivalent reflections. *R*-factor on merging was R_{init}= 0.122 and R_{init}= 0.169 for projections 1 and 2 respectively. By conventional X-ray structure analysis standards these values are quite poor, but upon consideration of the

possible perturbations that can affect the quality of electron diffraction data, they are deemed acceptable for use in electron crystallographic analysis.

At this point in the electron crystallographic analysis of Barium Lake Red C, intensity information for several diffraction patterns of each projection have been successfully combined giving rise to suitable single data sets for each projection. Now follows a description of the attempted *ab initio* phase determination for each projection using maximum entropy methods.

6.10. Ab initio Phasing using Maximum Entropy Methods.

6.10.1. *hk*0 Projection.

In accordance with the unit cell parameters detailed previously, projection 1 can be considered as the *hk0* projection due to the spacings found. Examination of the diffraction pattern and the systematic absences therein led to the assignment of plane group *c2mm*, due to the presence of reflection conditions, $h+k=2n$ and centering of the lattice.

From the outset of these calculations, it was expected that the heavier barium atoms would be the dominant structural feature. Comparison of scattering factors for electrons of barium and the other atoms present found that barium scatters electrons at least twice as strongly as the next heaviest atom, based upon values for neutral atoms. This was confirmed by the generation of a Patterson map where peaks due to vectors between barium atoms were found to overshadow any other structural features inherent in the Patterson map.

The theoretical basis of *ab initio* phase determination by maximum entropy methods was described extensively in chapter 4, here this theory is implemented through use of the MICE program.[114, 115, 116, 117.]

The diffraction data were normalised using MITHRIL to give the unitary structure magnitudes ($|U_h|^{obs}$) and their associated standard deviations using electron scattering factors. The overall isotropic temperature factor B was found to be 1.94. After normalisation, 21 unique reflections remained for use within MICE. Table 6.1. tabulates relevant reflection numbers, h, k, unitary structure factors and resolution, d in Å.

No.	h	k	$\|U_h\|^{obs}$	d (Å)
1	16	0	0.264	2.18
2	18	0	0.243	1.94
3	12	0	0.170	2.91
4	15	1	0.157	2.27
5	13	1	0.135	2.61
6	10	0	0.132	3.49
7	20	0	0.106	1.74
8	14	0	0.104	2.49
9	8	0	0.102	4.36
10	4	2	0.089	4.63
11	9	1	0.082	3.65
12	2	2	0.079	5.21
13	10	2	0.077	2.94
14	19	1	0.076	1.81
15	7	1	0.074	4.53
16	0	2	0.070	5.46
17	6	0	0.066	5.81
18	12	2	0.064	2.57
19	11	1	0.061	3.05
20	5	1	0.554	5.88
21	17	1	0.553	2.02

Table 6.1. Reflection number, h, k, $|U_h|^{obs}$ and resolution, d in Å.

As this projection is C-centred, only one reflection was required for origin definition, thus forming the basis set {H}. Reflection used for origin definition was number 4, (15,1) with U-magnitude of 0.157. The root node of the phasing tree was then generated by carrying out constrained entropy maximisation, in which the phases and amplitudes of the origin phases were used as constraints in the production of a maximum entropy distribution $q^{ME}(\mathbf{x})$.

Seven strong reflections were selected and their phases permuted, thus generating 2^7 nodes. The phasing tree was now found to extend from node 1 through to node 129. Each node was subjected to a constrained entropy maximisation as previously

outlined. Subsequent analysis of nodes allowed the log likelihood gain (LLG) to be computed for each node, thus providing a ranking of nodes. The first level of the phasing tree is summarised in Table 6.2.

Node	Entropy	LLG	Node	Entropy	LLG	Node	Entropy	LLG
1	-0.022	0.00	44	-0.075	-0.09	87	-0.075	-0.08
2	-0.072	-0.09	45	-0.078	-0.05	88	-0.081	-0.15
3	-0.076	-0.02	46	-0.086	0.00	89	-0.076	-0.06
4	-0.082	-0.06	47	-0.077	-0.10	90	-0.076	-0.04
5	-0.080	-0.12	48	-0.074	-0.11	91	-0.075	-0.13
6	-0.084	-0.13	49	-0.075	-0.08	92	-0.075	-0.09
7	-0.069	-0.04	50	-0.079	-0.17	93	-0.074	-0.04
8	-0.075	0.00	51	-0.073	-0.06	94	-0.069	-0.01
9	-0.074	-0.07	52	-0.075	-0.04	95	-0.077	-0.11
10	-0.073	-0.05	53	-0.077	-0.10	96	-0.072	-0.08
11	-0.073	-0.04	54	-0.076	-0.11	97	-0.074	-0.05
12	-0.071	-0.05	55	-0.065	-0.02	98	-0.077	-0.09
13	-0.075	-0.11	56	-0.076	-0.05	99	-0.075	-0.13
14	-0.079	-0.09	57	-0.074	-0.12	100	-0.072	-0.08
15	-0.079	-0.06	58	-0.076	-0.13	101	-0.078	0.01
16	-0.078	-0.03	59	-0.083	-0.09	102	-0.075	0.02
17	-0.073	-0.09	60	-0.079	-0.02	103	-0.074	-0.06
18	-0.079	-0.09	61	-0.076	-0.11	104	-0.072	-0.12
19	-0.074	-0.13	62	-0.074	-0.08	105	-0.079	-0.08
20	-0.072	-0.06	63	-0.080	-0.04	106	-0.076	-0.14
21	-0.080	0.03	64	-0.059	-0.02	107	-0.074	-0.04
22	-0.080	0.01	65	-0.070	-0.12	108	-0.076	-0.04
23	-0.073	-0.05	66	-0.077	-0.14	109	-0.073	0.00
24	-0.073	-0.13	67	-0.079	-0.07	110	-0.075	0.02
25	-0.080	-0.07	68	-0.081	-0.06	111	-0.074	-0.07
26	-0.078	-0.07	69	-0.075	-0.08	112	-0.077	-0.12
27	-0.072	-0.13	70	-0.071	-0.07	113	-0.073	-0.09
28	-0.077	-0.06	71	-0.077	-0.03	114	-0.072	-0.09
29	-0.072	0.00	72	-0.079	-0.06	115	-0.075	0.00
30	-0.071	0.01	73	-0.078	-0.12	116	-0.064	-0.04
31	-0.078	-0.06	74	-0.0076	-0.11	117	-0.080	-0.14
32	-0.074	-0.12	75	-0.081	-0.08	118	-0.077	-0.12
33	-0.077	-0.09	76	-0.076	-0.01	119	-0.076	-0.03
34	-0.076	-0.09	77	-0.073	-0.08	120	-0.076	-0.01
35	-0.076	-0.12	78	-0.070	-0.07	121	-0.069	-0.09
36	-0.072	-0.11	79	-0.081	-0.06	122	-0.071	-0.08
37	-0.080	-0.03	80	-0.077	-0.04	123	-0.074	-0.04
38	-0.063	-0.02	81	-0.073	-0.11	124	-0.077	-0.04
39	-0.071	-0.08	82	-0.077	-0.06	125	-0.076	-0.12
40	-0.071	-0.10	83	-0.072	-0.10	126	-0.074	-0.10
41	-0.078	-0.06	84	-0.075	-0.11	127	-0.077	-0.07
42	-0.078	-0.04	85	-0.083	-0.04	128	-0.070	-0.01
43	-0.074	-0.11	86	-0.071	-0.02	129	-0.076	-0.09

Table 6.2. First level of phasing tree from MICE.

The top 8 nodes from analysis and likelihood ranking had centroid maps generated. Initial inspection of these maps offered very little structural information, even in terms

of identification of molecular orientation within the unit cell. As a result, a second level of the phasing tree was generated.

In common with other structure analyses using maximum entropy methods, generation of the second level of the phasing tree required selection of more reflections to undergo phase permutation. Five further reflections were selected, thus after phase permutation, the second level of the phasing tree extended to 321 nodes. As before, nodes were subjected to constrained entropy maximisation followed by analysis and ranking using likelihood. Table 6.3. lists the top nodes for level 2 and it is found that these are predominately generated from node 102.

Node	From node	Entropy	LLG
165	102	-0.086	0.22
167	102	-0.095	0.24
169	102	-0.085	0.23
185	102	-0.083	0.28
197	110	-0.085	0.26

Table 6.3. Top nodes from level 2 of the phasing tree.

Again, centroid maps were computed for these nodes and Figure 6.3. depicts the maps obtained for node 185. As expected the main structural features to appear are the barium atom positions. However, at this stage, the very crowded unit cell does not allow for any other molecular fragments to be located. Therefore, the phasing tree was extended to a further, third level.

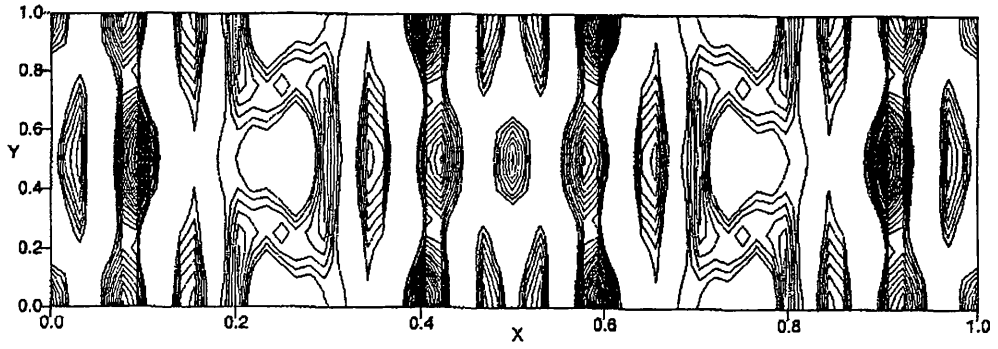


Figure 6.3. Highest node from second level of MICE.

A third level phasing tree required selection of reflections and permutation of phases as before. Five reflections were selected, these being numbered 5, 15, 16, 18 and 21. Phasing tree now grew to 481 nodes. Entropy maximisation followed by analysis and generation of likelihood valued produced a ranking of nodes, the top 5 of which are listed in Table 6.4.

Node	From node	Entropy	LLG
328	185	-0.151	3.37
332	185	-0.148	2.26
340	185	-0.147	2.03
342	185	-0.138	1.66
348	185	-0.141	1.71

Table 6.4. Top nodes from level 3 using MICE.

Figure 6.4. shows the centroid map generated for the highest node of level 3, node 328. This was the best possible centroid map, or phase solution that was obtainable with this data.

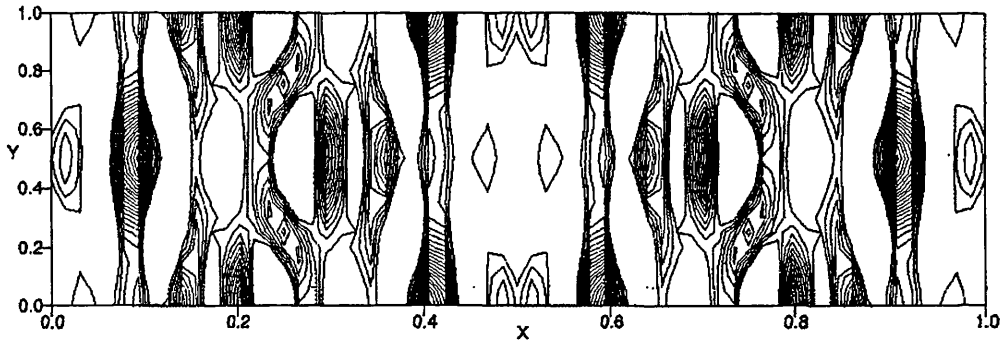


Figure 6.4. Centroid map obtained for node 328.

Barium atom positions are immediately apparent and are found to be the overriding aspect of this centroid map. However, it is now possible to see the development of other localised areas corresponding to molecular features. It can be assumed that due to the crowded nature of the unit cell, some molecules are being viewed face-on, explained by the observation of localised areas of molecular density. Attempts to generate any further levels in the phasing environment failed as LLG values were

found to collapse, an indication that generation of three levels of the phasing tree was as much as this very sparse and limited data set would allow.

Using MICE it was possible to extract projection coordinates for the barium positions identified from the centroid map. Table 6.5. lists the found coordinates along with their expected occupancy values. Previously, from using standard crystallographic relationships, unit cell contents of 4 dimeric units were proposed. Some conformations of these postulated contents is given here by the identification of the barium positions that correspond to a total barium content in the unit cell of 4.

Atom	x	y	Occupancy
Ba 1	0.974	0.500	1
Ba 2	0.092	0.500	1
Ba 3	0.416	0.000	0.5
Ba 4	0.599	0.000	0.5
Ba 5	0.416	1.000	0.5
Ba 6	0.599	1.000	0.5

Table 6.5. Projection coordinates for barium atoms.

Various attempts were made to increase the structural content achievable from this data set by alteration of different parameters within MICE. Ultimately, this met with little success and it was concluded that due to the very limited data combined with the disrupting presence of the heavy barium atoms, perhaps location of barium positions was all that could be achieved in this study. However, it was hoped that structural information could be retrieved from similar maximum entropy calculations on data from the other projection and at the very least, a correlation made between the barium positions in each projection.

6.10.2. 0*kl* Projection.

Following from assignment of projection 1 as being *hk*0, the data under investigation here has the axis of length 10.92Å in common and as such can be though of as being the 0*kl* projection.

From examination of the diffraction pattern, plane group *pm* can be used to describe the symmetry present. As was explained earlier, the single data set used here was created from the merging of several diffraction patterns that were obtained for this projection.

Following the protocol outlined for the calculations on projection 1, Patterson maps were generated for this data set. The effect of the barium atoms has been discussed, with respect to their electron scattering ability. In accordance with the previous work, the vectors between barium atoms were the outstanding feature of the Patterson map. This was taken as an indication that for the subsequent *ab initio* phasing by maximum entropy means, the centroid maps generated should also have the barium atoms as the most identifiable feature.

The diffraction data were normalised using MITHRIL to give the unitary structure factors ($|U_h|^{obs}$) and their associated standard deviations using electron scattering factors. The overall isotropic temperature factor B was found to be 0.815. This was a favourable value for electron diffraction data as the temperature factor is often found to be negative. After normalisation 94 unique reflections remained. Obviously this is a larger dataset than that which was used for projection 1 and this is mainly due to the quality of the respective diffraction patterns. Table 6.6. lists data for the reflections used with reflection number, h, k, unitary structure factors and resolution, d, in Å tabulated.

Subsequent maximum entropy calculations were initiated by definition of an origin. Two reflections were required in this instance and they were numbers 9 and 33, both of which were given phase angles of 0°. These reflections defined the basis set {H} and from this the root node of the phasing tree was generated. This node was then subjected to constrained entropy maximisation, where the amplitudes and phases of the origin defining reflections were used as constraints to produce a maximum entropy distribution $q^{ME}(\mathbf{x})$.

Reflections were chosen to have their phase permuted, thus generating more nodes on the phasing tree and subsequently these nodes had to undergo entropy maximisation procedures. Analysis of nodes was achieved through utilisation of *t*-tests, The phasing tree generated at this level for projection 2 is summarised in Table 6.7.

No.	k	l	$\|U_h\|^{obs}$	d(Å)	No.	k	l	$\|U_h\|^{obs}$	d(Å)
1	0	8	0.187	1.36	48	2	4	0.082	2.31
2	15	4	0.181	0.57	49	5	3	0.082	1.57
3	13	8	0.171	0.60	50	2	3	0.081	2.79
4	13	3	0.168	0.66	51	10	3	0.080	0.85
5	4	0	0.160	2.18	52	2	8	0.078	1.30
6	14	2	0.158	0.62	53	0	9	0.077	1.21
7	0	10	0.143	1.09	54	6	5	0.077	1.21
8	0	4	0.142	2.73	55	7	2	0.076	1.21
9	8	1	0.139	1.08	56	7	4	0.074	1.13
10	8	3	0.138	1.04	57	10	2	0.073	0.86
11	13	7	0.138	0.62	58	5	6	0.073	1.26
12	6	0	0.133	1.45	59	8	5	0.073	0.98
13	13	5	0.132	0.64	60	1	5	0.071	2.12
14	13	6	0.131	0.63	61	9	3	0.070	0.94
15	8	0	0.130	1.09	62	0	5	0.070	2.18
16	14	3	0.127	0.61	63	12	4	0.070	0.70
17	11	8	0.124	0.69	64	2	7	0.069	1.47
18	1	8	0.123	1.35	65	5	5	0.068	1.36
19	12	5	0.117	0.69	66	5	0	0.067	1.74
20	4	3	0.115	1.87	67	9	0	0.067	0.97
21	4	2	0.115	2.02	68	4	5	0.067	1.54
22	4	1	0.113	2.14	69	6	6	0.065	1.14
23	6	1	0.113	1.44	70	7	0	0.063	1.25
24	8	2	0.111	1.07	71	2	5	0.063	1.95
25	9	2	0.110	0.95	72	9	4	0.060	0.91
26	6	3	0.108	1.35	73	3	5	0.060	1.75
27	13	0	0.107	0.67	74	9	9	0.059	0.76
28	12	3	0.103	0.71	75	7	5	0.059	1.08
29	7	1	0.103	1.24	76	0	6	0.059	1.82
30	3	3	0.102	2.27	77	7	7	0.058	0.97
31	10	1	0.100	0.87	78	0	7	0.058	1.56
32	12	6	0.098	0.67	79	1	10	0.058	1.08
33	5	1	0.099	1.72	80	4	7	0.058	1.27
34	9	1	0.097	0.96	81	4	6	0.057	1.40
35	6	2	0.095	1.40	82	3	6	0.057	1.54
36	4	4	0.095	1.70	83	1	11	0.057	0.99
37	5	2	0.095	1.66	84	1	9	0.056	1.20
38	6	4	0.093	1.28	85	5	4	0.056	1.47
39	0	3	0.089	3.64	86	3	1	0.056	2.81
40	7	3	0.087	1.18	87	2	9	0.055	1.17
41	11	5	0.086	0.74	88	2	6	0.054	1.68
42	8	4	0.086	1.01	89	8	6	0.054	0.93
43	11	4	0.085	0.76	90	1	4	0.052	2.61
44	13	1	0.084	0.67	91	1	6	0.052	1.78
45	3	9	0.083	1.12	92	3	4	0.051	1.99
46	1	3	0.083	3.36	93	9	5	0.050	0.89
47	3	2	0.083	2.57	94	7	6	0.050	1.03

Table 6.6. Reflection number, *k*, *l*, unitary structure factor $|U_h|^{obs}$ and *d* in Å.

Node	Entropy	LLG	Node	Entropy	LLG	Node	Entropy	LLG
1	0.00	0.00	44	-0.050	0.01	87	-0.032	-0.07
2	-0032	-0.09	45	-0.037	-0.02	88	-0.031	-0.08
3	-0.034	-0.04	46	-0.034	-0.04	89	-0.034	-0.04
4	-0.035	-0.02	47	-0.031	-0.08	90	-0.035	-0.04
5	-0.031	-0.07	48	-0.032	-0.07	91	-0.036	-0.02
6	-0.035	-0.02	49	-0.035	-0.02	92	-0.047	0.01
7	-0.034	-0.04	50	-0.041	-0.03	93	-0.035	-0.02
8	-0.037	-0.02	51	-0.035	-0.04	94	-0.034	-0.04
9	-0.046	0.01	52	-0.036	-0.02	95	-0.035	-0.02
10	-0.031	-0.07	53	-0.046	0.01	96	-0.032	-0.07
11	-0.032	-0.08	54	-0.035	-0.02	97	-0.031	-0.08
12	-0.034	-0.04	55	-0.033	-0.04	98	-0.031	-0.08
13	-0.034	-0.02	56	-0.032	-0.08	99	-0.032	-0.07
14	-0.051	0.01	57	-0.031	-0.07	100	-0.036	-0.02
15	-0.035	-0.02	58	-0.041	0.01	101	-0.034	-0.04
16	-0.035	-0.04	59	-0.037	-0.02	102	-0.035	-0.02
17	-0.039	-0.02	60	-0.034	-0.04	103	-0.050	0.01
18	-0.035	-0.02	61	-0.040	-0.02	104	-0.036	-0.02
19	-0.032	-0.08	62	-0.030	-0.06	105	-0.035	-0.04
20	-0.031	-0.08	63	-0.035	-0.02	106	-0.034	-0.04
21	-0.034	-0.04	64	-0.034	-0.04	107	-0.031	-0.08
22	-0.036	-0.02	65	-0.032	-0.08	108	-0.032	-0.08
23	-0.052	0.01	66	-0.040	-0.02	109	-0.035	-0.02
24	-0.035	-0.02	67	-0.034	-0.04	110	-0.035	-0.04
25	-0.035	-0.04	68	-0.037	-0.02	111	-0.035	-0.02
26	-0.034	-0.04	69	-0.041	0.01	112	-0.052	0.01
27	-0.036	-0.02	70	-0.032	-0.08	113	-0.036	-0.02
28	-0.032	-0.07	71	-0.034	-0.04	114	-0.034	-0.02
29	-0.031	-0.08	72	-0.035	-0.02	115	-0.034	-0.04
30	-0.035	-0.04	73	-0.030	-0.06	116	-0.032	-0.08
31	-0.036	-0.02	74	-0.046	0.01	117	-0.031	-0.07
32	-0.050	0.01	75	-0.036	-0.02	118	-0.039	-0.02
33	-0.035	-0.02	76	-0.035	-0.04	119	-0.035	-0.04
34	-0.035	-0.02	77	-0.041	-0.03	120	-0.035	-0.02
35	-0.047	0.01	78	-0.031	-0.07	121	-0.051	0.01
36	-0.036	-0.02	79	-0.032	-0.08	122	-0.031	-0.07
37	-0.035	-0.04	80	-0.033	-0.04	123	-0.035	-0.02
38	-0.031	-0.08	81	-0.035	-0.02	124	-0.034	-0.04
39	-0.032	-0.07	82	-0.037	-0.02	125	-0.032	-0.09
40	-0.035	-0.02	83	-0.050	0.01	126	-0.046	0.01
41	-0.034	-0.04	84	-0.035	-0.02	127	-0.037	-0.02
42	-0.035	-0.04	85	-0.035	-0.04	128	-0.034	-0.04
43	-0.035	-0.02	86	-0.035	-0.02	129	-0.035	-0.02

Table 6.7. First level of the maximum entropy phasing tree.

Examination of the listing of nodes shows mostly negative LLG values, with only a few nodes having the same positive value of 0.01. No identifiable solution is reached at this point of the calculations and so further phasing has to be undertaken.

A second level was generated by keeping the nodes with positive LLG values and selecting new reflections to undergo phase permutation, thus extending the phasing tree to 513 nodes. Again, constrained entropy maximisation was repeated, followed by analysis and ranking of nodes. Table 6.8. lists the highest nodes for level 2, along with their entropy and LLG values.

Node	From node	Entropy	LLG
160	112	-0.049	0.036
162	112	-0.045	0.036
177	112	-0.050	0.031
272	126	-0.048	0.039
336	121	-0.047	0.034
394	23	-0.047	0.038
418	23	-0.046	0.039

Table 6.8. Top nodes from second level of phasing tree.

Centroid maps were generated and as expected the barium atomic positions were very prominent. However, in comparison with projection 1, more detail is present regarding other atomic presence. As no unique solution had manifested itself, a further third level of phasing was undertaken.

Again, further reflections were selected to have phases permuted and the resulting nodes were subjected to constrained entropy maximisation. Analysis by use of *t*-tests and likelihood values provided a ranking of nodes. The top 5 nodes are listed in Table 6.9. and it can be seen that three nodes have the same LLG values and two of these have the same entropy value. Generation of centroid maps found that the maps were very similar and indeed, were found to have the barium atoms in the same positions. Figure 6.5 illustrates the centroid map obtained for node 698, which is representative of those generated for nodes 678 and 694.

Node	From node	Entropy	LLG
678	394	-0.049	0.05
682	394	-0.049	0.04
694	394	-0.049	0.05
698	394	-0.048	0.05
702	394	-0.049	0.04

Table 6.9. Top nodes from level 3 of phasing tree.

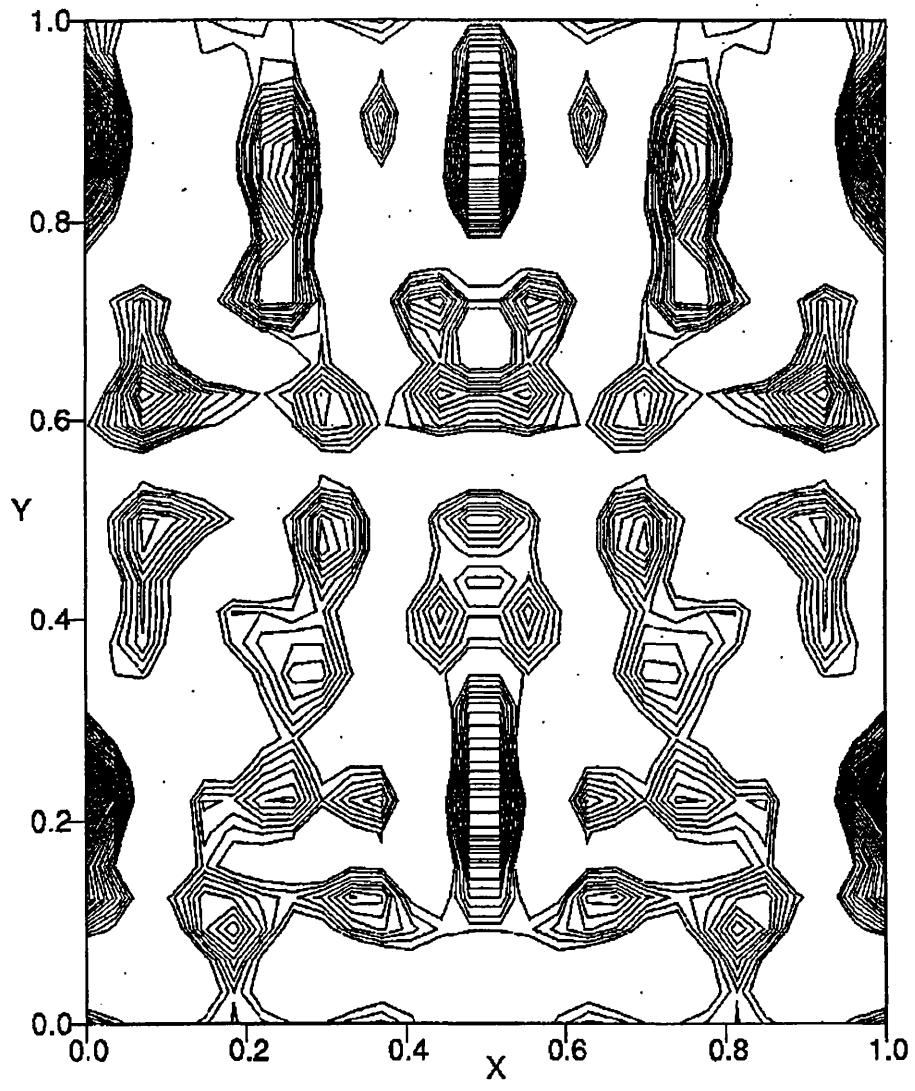


Figure 6.5. Centroid map generated for node 698.

It was pleasing to obtain from MICE coordinates for 6 barium positions, totaling an occupancy of 4, thus giving credence to the postulated unit cell contents. Table 6.10. lists these obtained projection coordinates along with relevant occupancy values.

Atom	x	y	Occupancy
Ba 1	0.500	0.874	1
Ba 2	0.500	0.220	1
Ba 3	0.000	0.218	0.5
Ba 4	0.000	0.875	0.5
Ba 5	1.000	0.218	0.5
Ba 6	1.000	0.875	0.5

Table 6.10. Barium coordinates for projection 2.

Attempts were made to improve the content of the structural information available from the data using maximum entropy methods, however, the sparseness and quality of the data set were perhaps the limiting factors, and no improvement could be achieved.

6.10.3. Comparison of Projection 1 and 2.

As previously indicated, projections 1 and 2 share a common axis and as such it is possible to correlate information obtained from each individual projection. Projection coordinates of the barium atoms were found to have some common positions. Combinations of the individual projection information led to the realisation of three dimensional coordinates for the barium atoms, as listed in Table 6.11.

Atom	x	y	z
Ba 1	0.974	0.500	0.874
Ba 2	0.092	0.500	0.220
Ba 3	0.416	0.000	0.218
Ba 4	0.599	0.000	0.875
Ba 5	0.416	1.000	0.218
Ba 6	0.599	1.000	0.875

Table 6.11. Coordinates for barium atomic positions.

Using the *CERIUS*[2] molecular visualisation package, it was possible to place the barium atoms at their found positions within the proposed unit cell, as illustrated in Figure 6.6.

Investigation of possible Ba-Ba contact distances found that the minimum distance to be 8.01 Å, which was deemed acceptable after comparison with previously solved structures containing barium atoms.

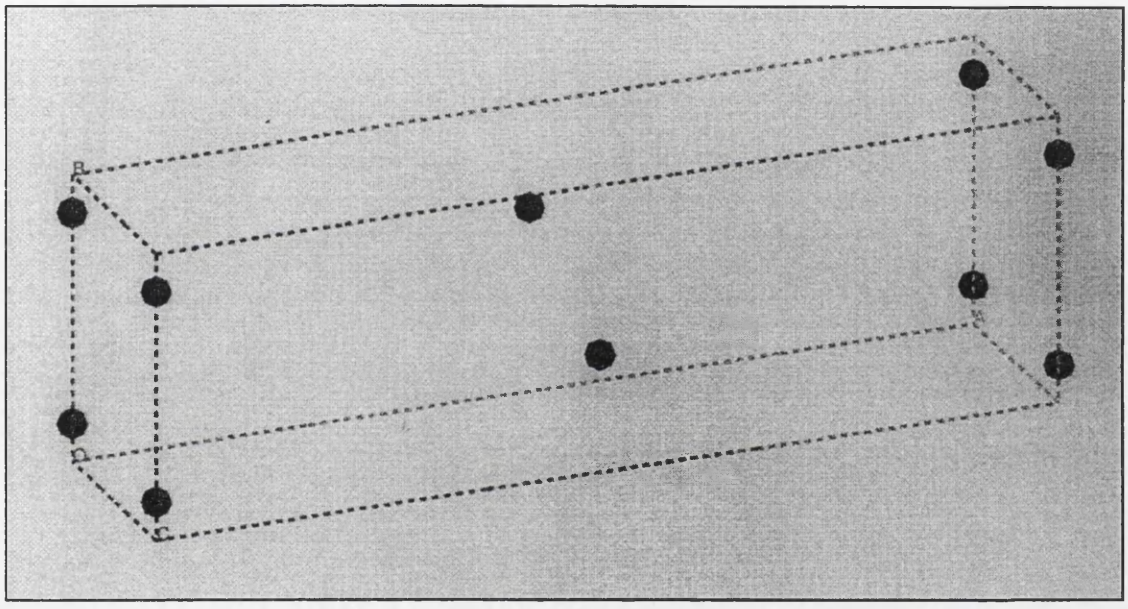


Figure 6.6. Barium atom positions.

6.10.4. Combining of Projection Data.

Many other electron crystallographic structure analyses have merged extensive projection data to form a 3D data set. In this study, as much structural information as was possible had been obtained from the individual datasets. Therefore it was decided to combine projection information, themselves composed of combined datasets, to form a 3D dataset. It is apparent that this is far from a complete sampling of reciprocal space, but given the limited information obtained from projection data, it was deemed a suitable course of action.

The two sets of projection data were appropriately combined and duplicate reflections were rejected. Normalisation procedures using MITHRIL were followed to give unitary structure factors ($|U_{\mathbf{h}}|^{obs}$) and their associated standard deviations using electron scattering factors. 61 unique reflections remained after normalisation. Overall isotropic temperature factor B was calculated to be 1.506 $Å^2$, this being an acceptable value for electron diffraction data. Examination of the symmetry present in both projections led to the assignment of symmetry appropriate to space group C2/m.

Maximum entropy calculations were carried out using MICE, as before. Two suitable reflections were chosen to define the origin, obeying origin defining rules appropriate to the symmetry used: The reflections chosen were 12(0,6,1) and 18(15,1,0), with U-magnitudes of 0.178 and 0.156 respectively. The root node of the phasing tree was then generated by carrying out constrained entropy maximisation, in which the phases and amplitudes of the origin reflections were used as constraints to produce a maximum entropy distribution $q^{ME}(\mathbf{x})$.

The phasing tree was then extended by the selection of seven strong reflections to undergo phase permutation. This gave rise to a further 128 nodes and each node then underwent constrained entropy maximisation, as before. Analysis of the nodes was achieved through use of t-tests and LLG values were computed for each. The top 8 nodes of the phasing tree were identified and centroid maps were produced for each. Table 6.12. lists these nodes, while Figure 6.7. depicts the centroid map obtained for node 67.

Node	Entropy	LLG
59	-0.185	0.079
60	-0.179	0.052
63	-0.189	0.050
64	-0.194	0.074
67	-0.185	0.079
68	-0.179	0.052
71	-0.189	0.050
72	-0.194	0.074

Table 6.12. Top 8 nodes from first level of phasing tree.

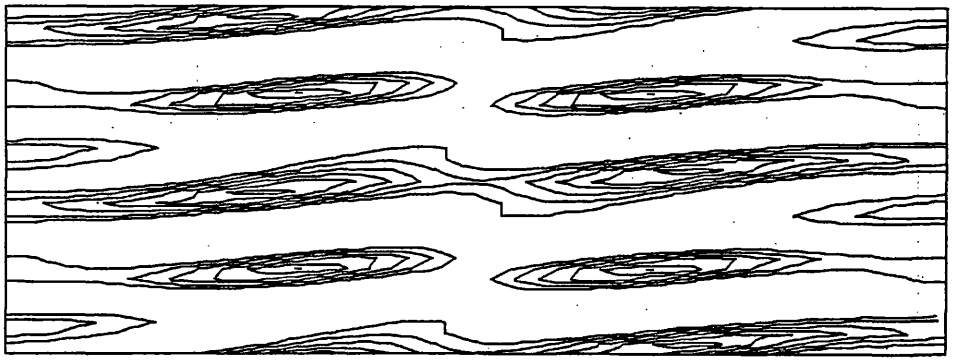


Figure 6.7. Centroid map from node 67.

All of the 8 centroid maps generated were very similar, in that distinct areas of molecular presence could be identified.

A second level of the phasing tree was generated by keeping the top 8 nodes from level 1 and permuting phases for 5 more reflections, these being numbered 4, 16, 22, 23 and 32. Entropy maximisation procedures were repeated, followed by analysis and calculation of LLG values. Again, centroid maps were generated and Figure 6.8. illustrates the centroid map obtained for node 215.

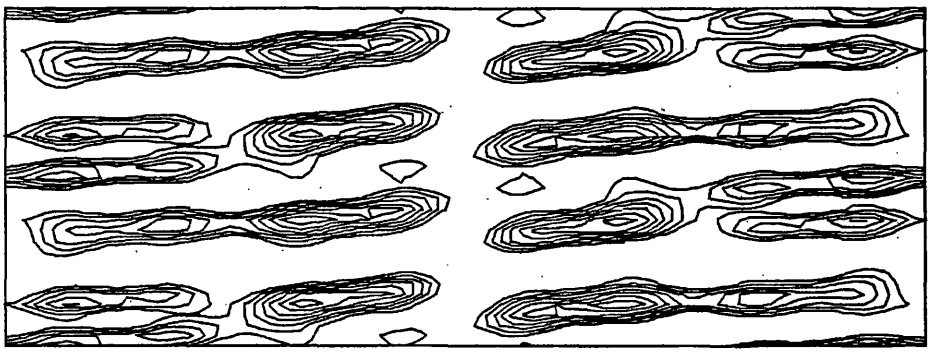


Figure 6.8. Representative centroid map of level 2 of phasing tree.

At this stage of the maximum entropy calculations, no discernible solution was found to manifest itself and several of the top nodes had similar likelihood values. As a consequence of this similarity, the procedure of generating further levels of the phasing tree, by identifying the top nodes for a given level and then permuting more phases, was repeated several times. As before, nodes generated were subjected to constrained entropy maximisation. This extensive branching of the phasing tree from level 2 to level 6 is summarised in Table 6.13.

Level	Permute	No. nodes generated	Nodes kept and LLG values
2	4	130-385	215 (0.150)
	16		223 (0.164)
	22		225 (0.159)
	23		231 (0.150)
	32		239 (0.154)
			241 (0.164)
3	17	386-577	443 (0.34)
	26		445 (0.34)
	34		538 (0.34)
	43		540 (0.34)
	49		
4	40	578-641	598 (1.39)
	44		600 (1.39)
	46		630 (1.39)
	57		632 (1.39)
5	31	642-705	651 (2.32)
	33		668 (2.31)
	38		683 (2.32)
	45		700 (2.31)
6	53	706-769	728 (2.35)
	54		761 (2.36)
	55		
	59		

Table 6.13. Summary of phasing tree.

Finally, a distinct solution was seen to emerge at level 6 with nodes 728 and 761 having noticeably higher LLG values. Figure 6.9. illustrates the centroid map generated for node 761. Immediately apparent is the increase in molecular detail over that present in the centroid map shown for level 2.

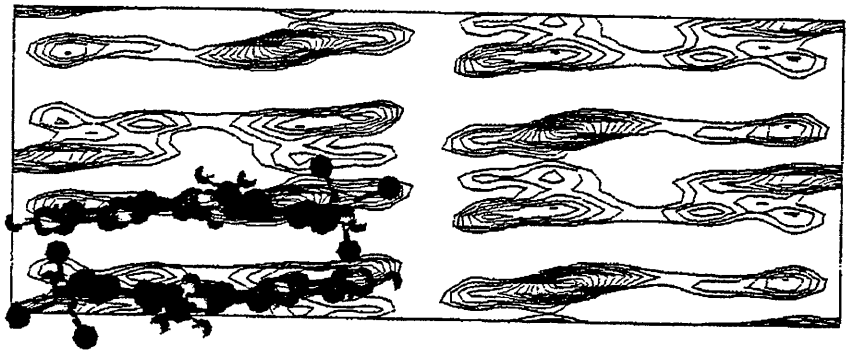


Figure 6.9. Centroid map from node 761.

Examination of this centroid map, which is representative of *h0l* projection, clearly shows two columns of molecules stacking parallel to the shortest unit cell axis. Also, an anti-parallel array of molecular packing can be discerned. Given that in this centroid map molecules are being viewed, essentially edge-on, it is now clear why it was difficult to determine areas of molecular presence.

Dimensions of these molecules of around 16 Å by 2 Å, obtainable from these centroid maps agree favourable with calculated values. This is illustrated in Figure 6.9. by the superimposing of two molecules in the correct orientation within the unit cell. Also, further credence is given to the postulated unit cell contents with the identification of 8 areas of organic presence. However, location of barium positions, which were found to be so dominant in the other projections, is less obvious here. This may be due to dynamical scattering effects or perhaps, due to the crowded nature of this large unit cell.

Overall, it can be concluded that in accordance with the findings of the previous structural review of β-naphthol and acetoacetanilide pigments, Barium Lake Red C exists as the keto-hydrazone tautomer. A pseudo-dimeric relationship exists between one barium atom and two organic moieties, indicated by the location of 4 barium atoms and 8 areas of organic presence. Molecules are seen to pack in a columnar

manner, with an alternate relationship seen to develop between molecules in adjacent columns. Finally, this packing of molecules is seen to be parallel to the shortest unit cell dimension.

6.11. Concluding Remarks.

Throughout all stages of this challenging electron crystallographic analysis, a certain degree of pessimism has prevailed. It was kept in mind that Dorset had stated that study of perbromophthalocyanine represented an extreme case for accepting data to be used in *ab initio* electron crystallographic studies. It was believed that the study undertaken here perhaps lay beyond the criteria required to fulfill a successful structural elucidation. However, after obtaining internally consistent diffraction patterns, *ab initio* phase determination using maximum entropy methods was attempted. The results presented here indicate that, although significant perturbations to the data exist, through such effects as dynamical scattering and the intractable problem of this compounds sensitivity to the electron beam, structural information is obtainable regarding molecules for which all conventional structural methods failed.

7.0. STUDIES OF DYNAMICAL EFFECTS IN SUBSTITUTED PHTHALOCYANINE MOLECULES.

7.1. Introduction to Phthalocyanine.

Phthalocyanine and its halogenated derivatives are extensively used as pigments due to their blue-green shading and excellent properties. In terms of cost, tinctorial strength, brightness of shade and all round fastness properties, they approximate closely to ideal pigments.

The generalised structure of phthalocyanine is given in Figure 7.1. As it can be seen, it is closely related to the porphyrins of chlorophyll and haemoglobin. The phthalocyanine nucleus, an azoporphyrin, can be derived from the parent porphyrin by replacement of the methine groups by nitrogen atoms.

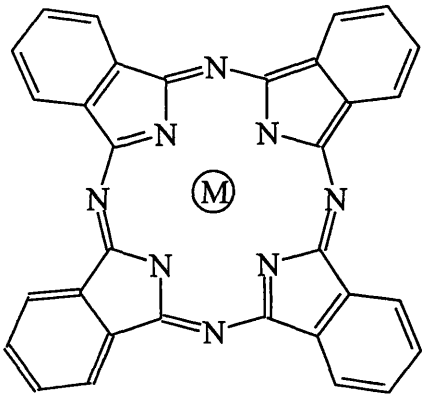


Figure 7.1. Phthalocyanine Molecule.

The phthalocyanine molecule is rigidly coplanar and tetradentate. Usually the metal coordinated is copper, but other elements can be coordinated in this position, including platinum, manganese, aluminium and silicon.

The pigment as shown is blue in colour, with substitution of the aromatic hydrogens by chlorine giving rise to a green pigment. Preparation of the pigment is achieved through the vigorous reaction of phthalic anhydride with urea and a copper salt in a high boiling solvent such as nitrobenzene, with ammonium molybdate as a catalyst.

Phthalocyanine exists in several polymorphic forms, the most stable of which is the β-form. In the β-form, the planar molecules stack in zigzag columns in such a way as to permit distorted coordination of the copper atom through nitrogen atoms. This arrangement is due to the copper atom of one molecule being above the nitrogen atom of another.

The β-form is found to be the most thermodynamically stable polymorph and possesses a green-blue shading, with the other major polymorph, α-form, existing as a red-blue shade. Other polymorphs of phthalocyanine, γ, δ and ε possess distorted α-like structures, but these are of little commercial significance.

Phthalocyanines have been extensively studied by X-ray crystallography, including the first solution of the β-form by J.M. Robertson in the Chemistry Department at Glasgow University.[188] Many other crystallographic studies have been carried out, but of most interest within the context of this research are the commercially important β- and α-forms.

Phthalocyanines have also been extensively studied by electron microscopy. The development of epitaxial specimen preparation techniques[189] and minimal dose techniques,[190] along with improvements in general electron microscope capability have allowed for the examination of the structure of the crystals by high resolution electron microscopy. This work culminated in the direct observation of molecular images of chlorinated copper phthalocyanines by various research groups.[191,192,193,194]

7.2. Previous Electron Crystallographic Studies of Halogenated Phthalocyanines.

Phthalocyanines have been found to be very favourable molecules for study by electron imaging and diffraction techniques. This is especially true for halogenated phthalocyanines, as they exhibit greater resistance to radiation damage within the microscope. In terms of structural analyses by electron crystallographic means, Dorset has extensively studied both copper perchlorophthalocyanine and copper perbromophthalocyanine.[195,196,197,198]

Dorset states that the greatest challenge to *ab initio* electron crystallographic analyses of organic molecules experienced so far has been for the perhalogenated phthalocyanines. Studies where the halogen is chlorine have been the most successful when intensity data has been collected from thin crystals at sufficiently high electron accelerating voltage.[199] It was possible to generate maps with a few identifiable atomic positions via phase determination by symbolic addition, with the complete structure apparent after Fourier refinement. Other routes to structure solution have included use of starting phase sets from the Fourier transform of electron micrographs[200,201] in phase extension via the tangent formula or Sayre equation.[9,202,75]

Structure determination of the perbromo derivative was less favourable even when diffraction patterns were obtained from thin crystals at high accelerating voltage. Direct phase determinations produced results that were of less use than those of the perchlorophthalocyanine. Dynamical scattering was found to affect the accuracy of bond angles and distances by distorting them so that potential maps obtained did not correspond with a chemically correct model.

7.3. Aims of Present Work.

In general, dynamical scattering affects both the intensities and phases of elastically scattered electrons. This can have significant consequences for both images and

diffraction patterns. The purpose of this study is to explore the effects of dynamical scattering on structures determined by direct phasing, achieved through use of maximum entropy methods, of simulated electron diffraction patterns.

Therefore, in common with other studies detailed in this thesis, both *ab initio* phasing and procedures involving image derived phases will be utilised. However, in this instance simulated electron diffraction patterns of copper perchlorophthalocyanine will be used.

7.4. Model Construction and HRTEM Simulation.

Using atomic coordinates listed in Table 7.1, which were obtained from previous studies, an initial model could be constructed using the CERIUS^2 suite of programs.

Atom	x	y
Cu	0.000	0.000
Cl 1	0.081	0.302
Cl 2	0.157	0.202
Cl 3	0.271	0.120
Cl 4	0.405	0.057
N 1	0.000	0.070
N 2	0.093	0.000
N 3	0.117	0.090
C 1	0.136	0.042
C 2	0.203	0.025
C 3	0.265	0.055
C 4	0.327	0.025
C 5	0.055	0.106
C 6	0.036	0.158
C 7	0.074	0.202
C 8	0.036	0.246

Table 7.1. Zonal Atomic Coordinates for Copper Perchlorophthalocyanine.

The plane view of the molecule tilted 26° finds the [001] orientation of the monoclinic unit cell. Dimensions of the *cmm* projection were determined to be d_{100}=17.56 Å and b=26.08 Å. Figure 7.2. illustrates the model structure used throughout this study.

The *CERIUS*^{2} HRTEM (High Resolution Transmission Electron Microscopy) module enables simulation of dynamic electron diffraction patterns and allows transformation of these patterns into real space images, using a multislice simulation technique.[203], [30].

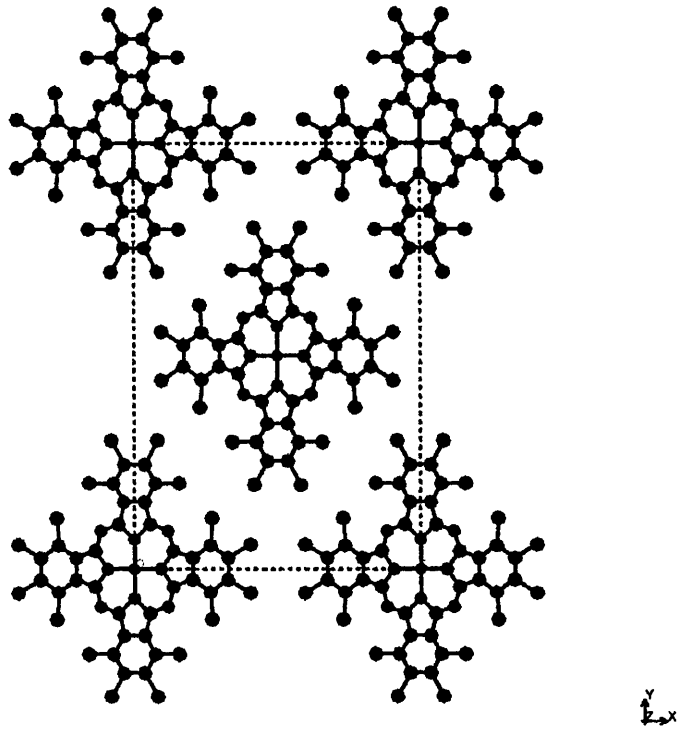


Figure 7.2. Model Structure of Copper Perchlorophthalocyanine.

Dynamical and inelastic scattering are found to affect the reliability of measured intensities in a non-linear manner which is dependent upon sample thickness. As a result of this, simulated dynamical electron diffraction patterns were generated over a range of specimen thickness from 20-200 Å.

The HRTEM simulation within *CERIUS*^{2} consists of four steps. First, the projected potential is calculated. This is regarded as the atomic electrostatic potential of the model cell averaged along the direction of the incident electron beam. Next, the phase transfer function is approximated. Thirdly, the electron beam is propagated through the crystal, calculating the diffracted beam. Finally, this diffracted beam is propagated through the microscope to produce a real space image. Many options exist that can provide flexibility with regard to how the simulation is carried out. These include

altering microscope parameters such as electron energy, aperture size or crystal thickness.

Figure 7.3. (*a-g*) shows the *CERIUS*^{2} generated projected potential map, dynamical diffraction pattern and multislice images for copper perchlorophthalocyanine at thickness of 20, 40, 60, 80, 100, 150 and 200Å.

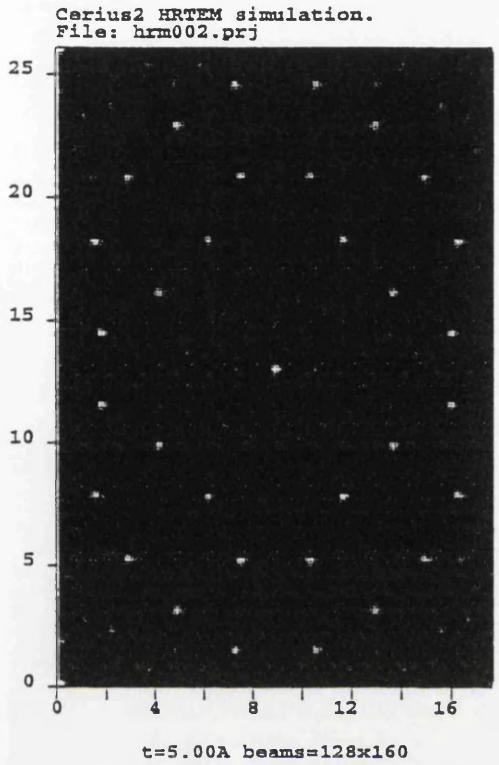
As expected the spread of the diffraction pattern is found to decrease with increasing crystal thickness. This corresponds to a shortening of the reciprocal lattice spikes. Even at 200Å thickness, the data set is at a higher resolution than would be acheived experimentally due to radiation damage.

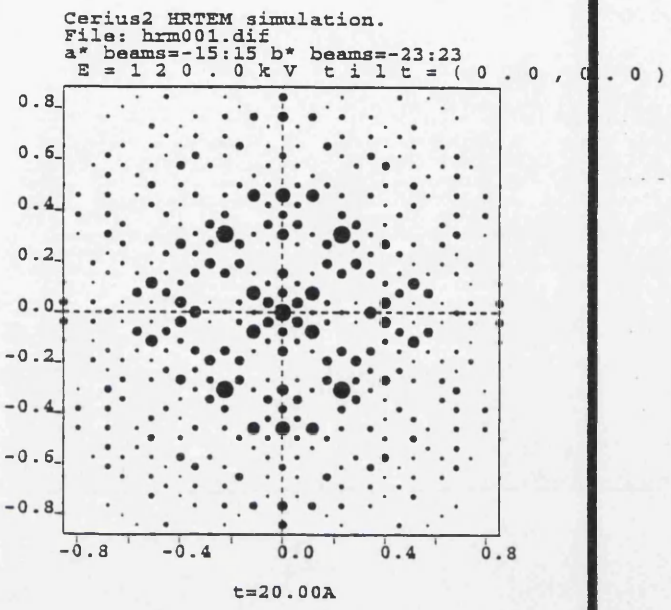
For each appropriate thickness examined, the generated multislice image retains the acknowledged molecular quatrefoil of phthalocyanine. However, as specimen thickness increases, a loss of contrast is seen with respect to the central copper atom. Multislice image calculation is seen to concentrate contrast into the aromatic carbon ring centres, again increasing with increasing crystal thickness.

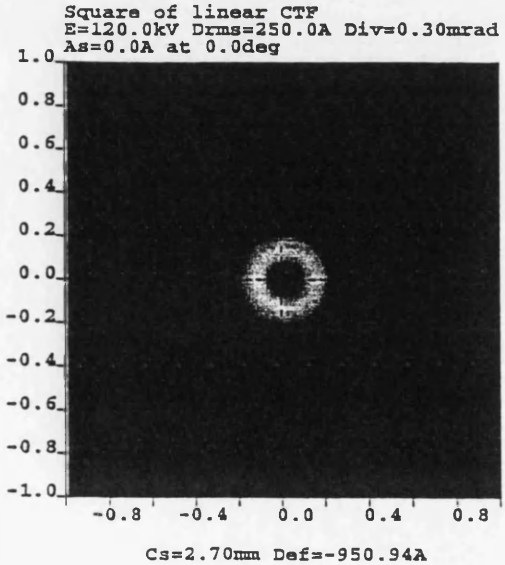
7.5. Data Processing and Manipulation.

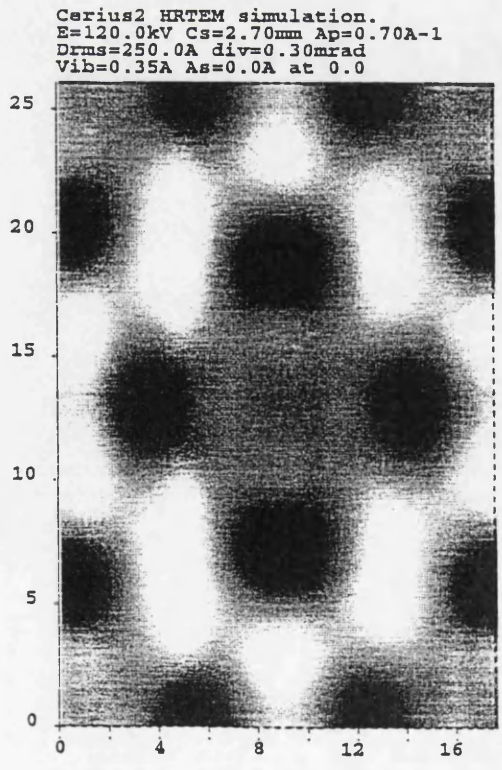
As the intention of this study was to carry out *ab initio* phasing using maximum entropy methods, use was made of the previously described software for quantification of electron diffraction patterns.

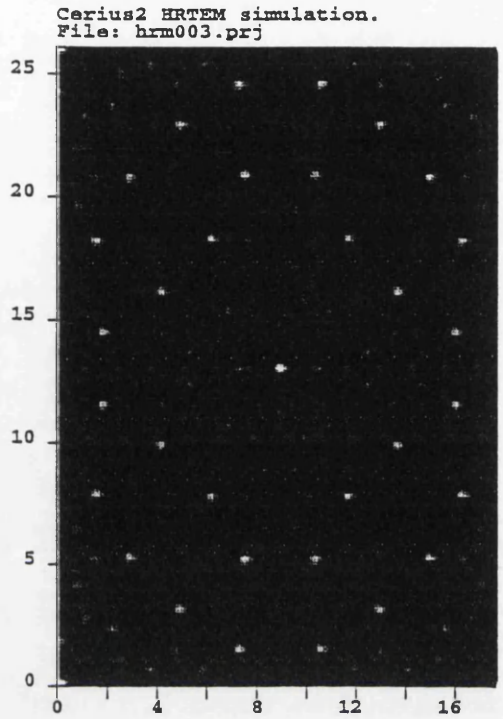
To simulate the acquisition of intensity information, as carried out with experimental data, the simulated diffraction patterns were printed and then digitised using a scanner, ready to be imported into ELD.[68] As described, ELD allows for quantitative evaluation of electron diffraction patterns producing listings of indexed reflections and their extracted intensities.This procedure was repeated for diffraction patterns of each investigated thickness.

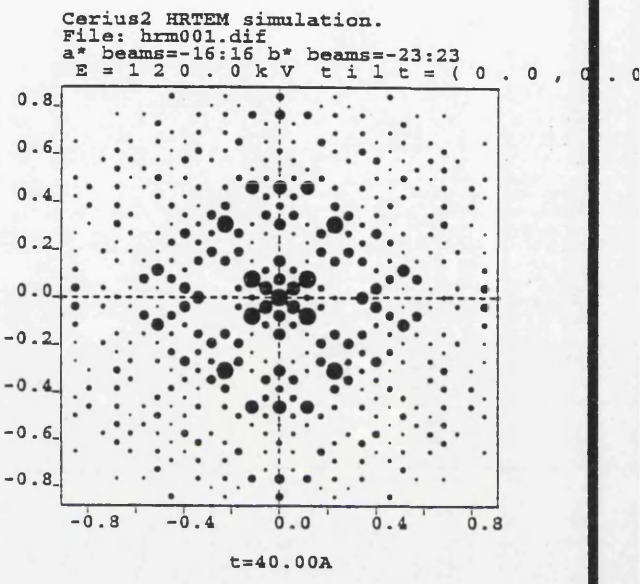

Cerius2 HRTEM simulation.
File: hrm002.prj
25
20
15
10
5
0
0
4
8
12
16
t=5.00A beams=128x160

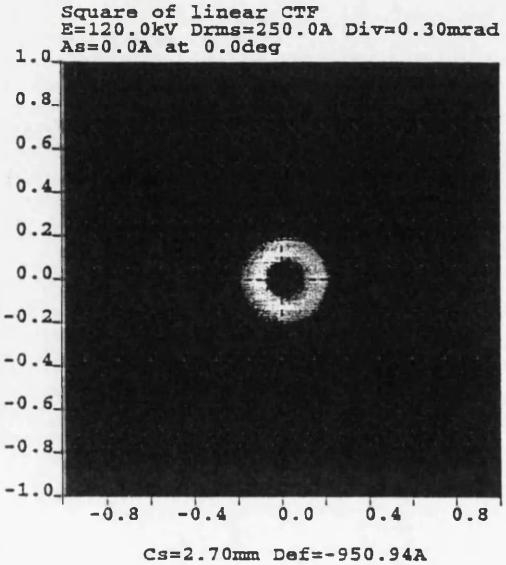

Cerius2 HRTEM simulation.
File: hrm001.dif
a* beams=-15:15 b* beams=-23:23
E = 1 2 0 . 0 k V t i l t = (0 . 0 , 0 . 0)
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-0.8
-0.4
0.0
0.4
0.8
t=20.00A

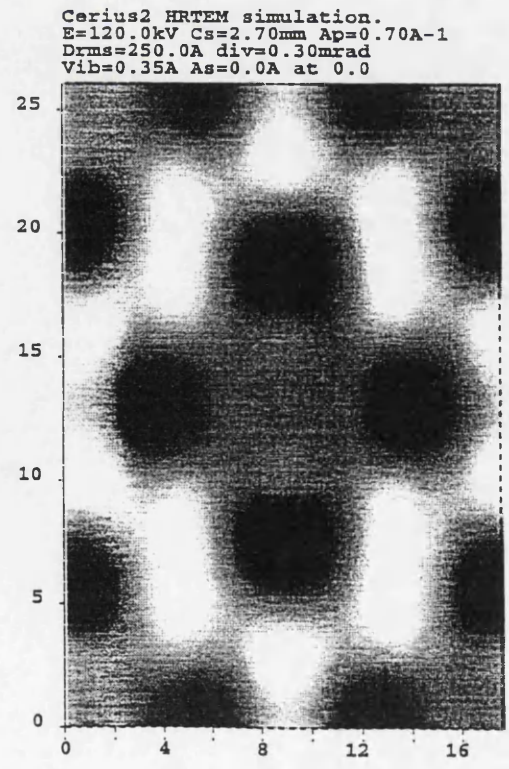

Square of linear CTF
E=120.0kV Drms=250.0A Div=0.30mrad
As=0.0A at 0.0deg
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-0.8
-0.4
0.0
0.4
0.8
Cs=2.70mm Def=-950.94A

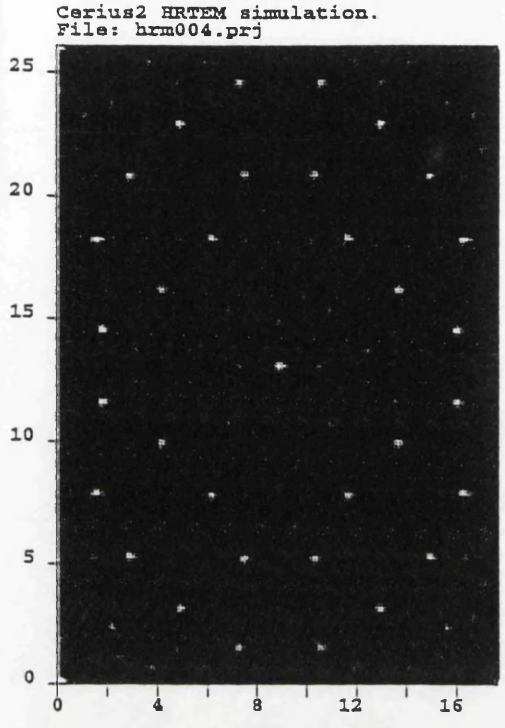

Cerius2 HRTEM simulation.
E=120.0kV Cs=2.70mm Ap=0.70A-1
Drms=250.0A div=0.30mrad
Vib=0.35A As=0.0A at 0.0
25
20
15
10
5
0
0
4
8
12
16
t=20.00A def=-950.9A

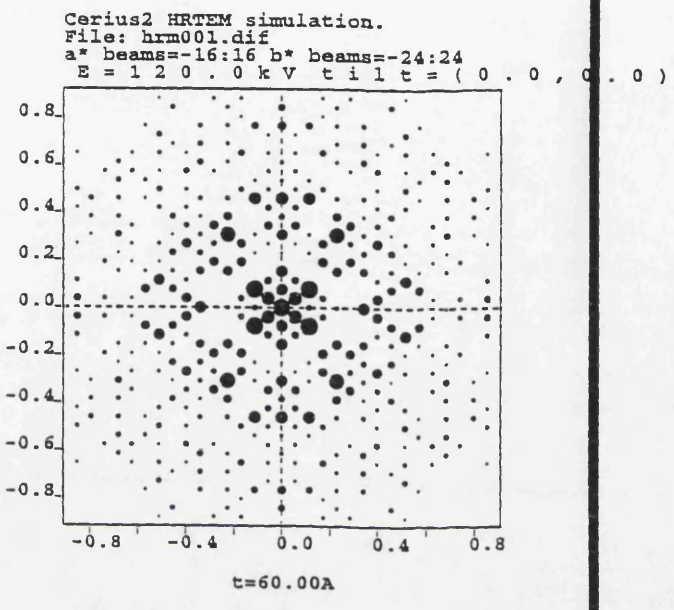

Cerius2 HRTEM simulation.
File: hrm003.prj
25
20
15
10
5
0
0
4
8
12
16
t=5.00A beams=128x160

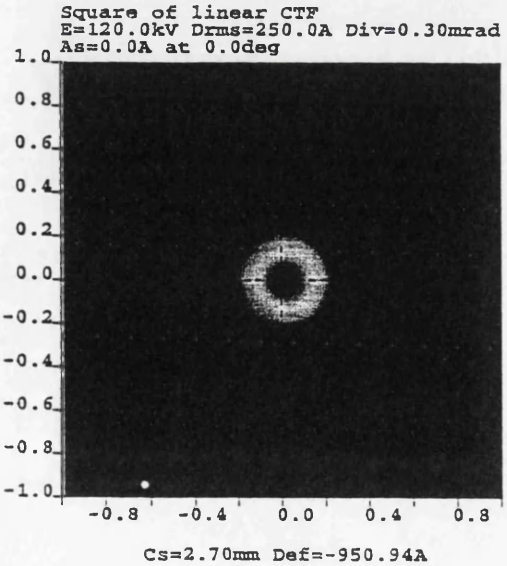

Cerius2 HRTEM simulation.
File: hrm001.dif
a* beams=-16:16 b* beams=-23:23
E = 1 2 0 . 0 k V t i l t = (0 . 0 , 0 . 0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-0.8
-0.4
0.0
0.4
0.8
t=40.00A

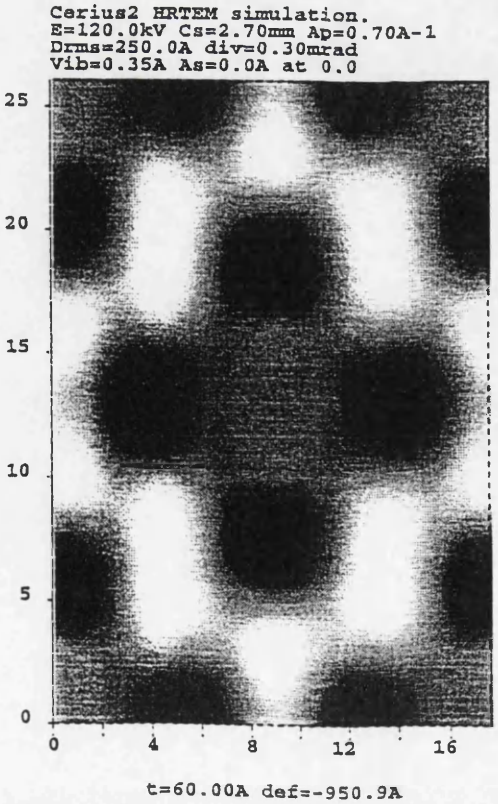

Square of linear CTF
E=120.0kV Drms=250.0A Div=0.30mrad
As=0.0A at 0.0deg
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-0.8
-0.4
0.0
0.4
0.8
Cs=2.70mm Def=-950.94A

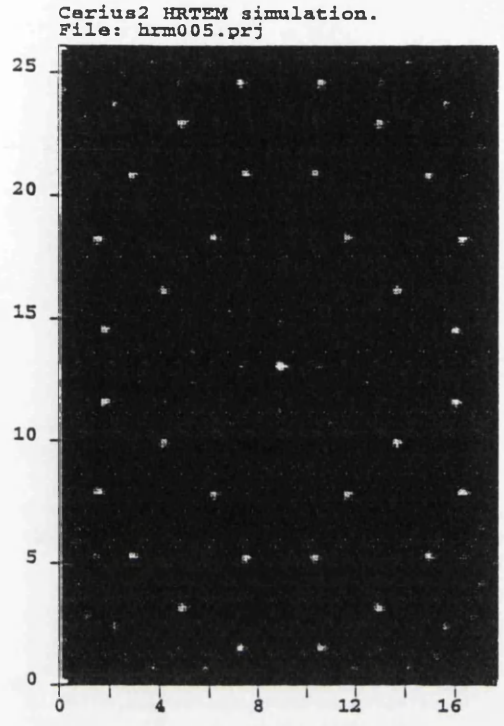

Cerius2 HRTEM simulation.
E=120.0kV Cs=2.70mm Ap=0.70A-1
Drms=250.0A div=0.30mrad
Vib=0.35A As=0.0A at 0.0
25
20
15
10
5
0
0
4
8
12
16
t=40.00A def=-950.9A

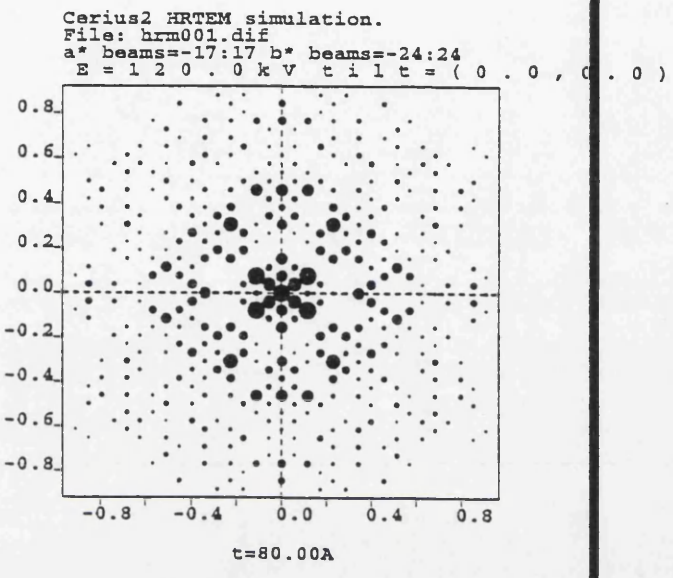

Cerius2 HRTEM simulation.
File: hrm004.prj
25
20
15
10
5
0
0
4
8
12
16
t=5.00A beams=128x160

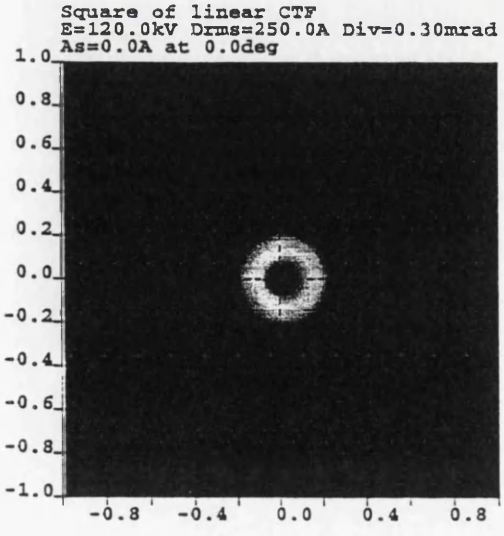

Cerius2 HRTEM simulation.
File: hrm001.dif
a* beams=-16:16 b* beams=-24:24
E = 1 2 0 . 0 k V t i l t = (0 . 0 , 0 . 0)
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-0.8
-0.4
0.0
0.4
0.8
t=60.00A

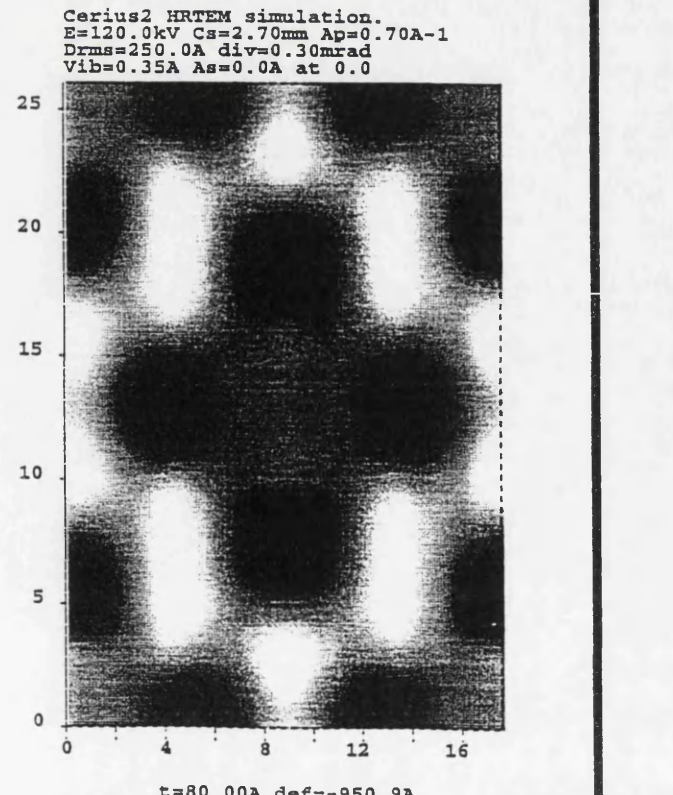

Square of linear CTF
E=120.0kV Drms=250.0A Div=0.30mrad
As=0.0A at 0.0deg
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-0.8
-0.4
0.0
0.4
0.8
Cs=2.70mm Def=-950.94A

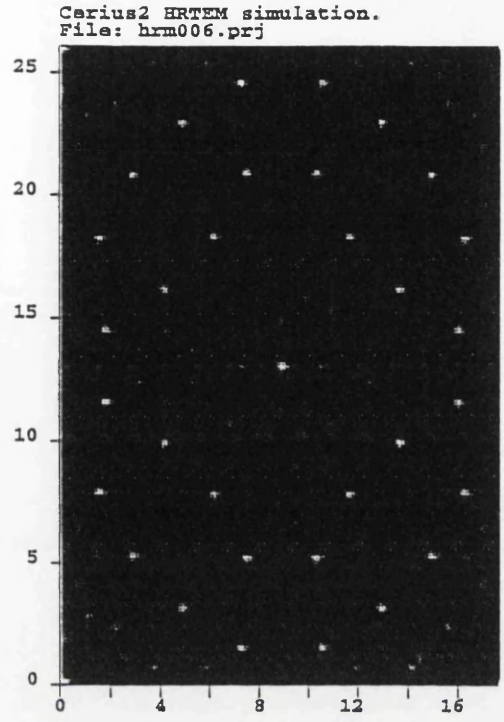

Cerius2 HRTEM simulation.
E=120.0kV Cs=2.70mm Ap=0.70A-1
Drms=250.0A div=0.30mrad
Vib=0.35A As=0.0A at 0.0
25
20
15
10
5
0
0
4
8
12
16
t=60.00A def=-950.9A

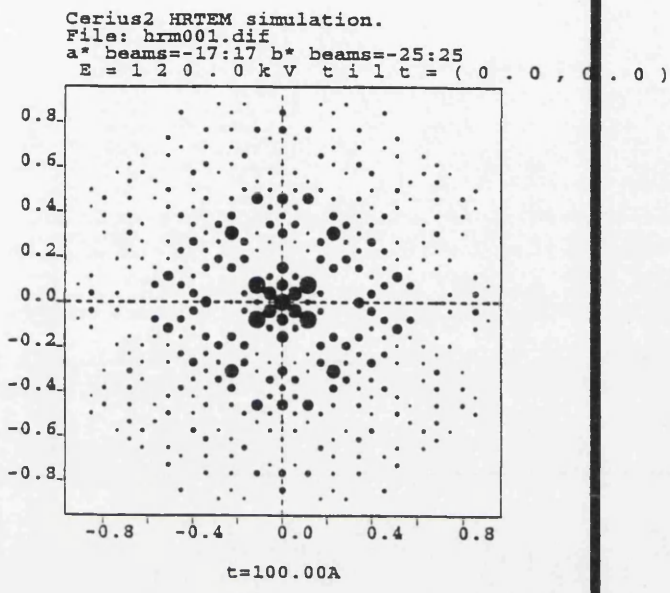

Cerius2 HRTEM simulation.
File: hrm005.prj
25
20
15
10
5
0
0
4
8
12
16
t=5.00A beams=128x160

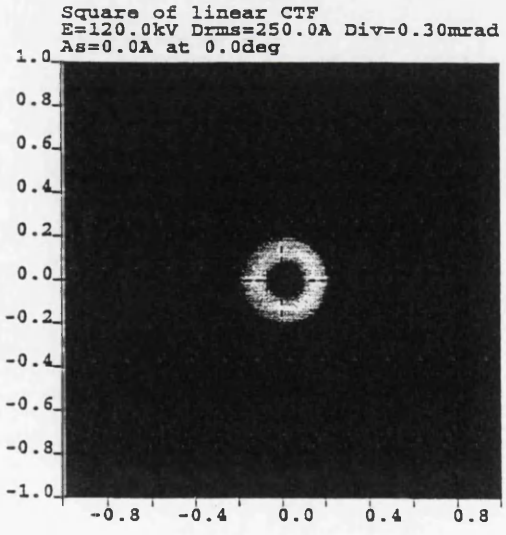

Cerius2 HRTEM simulation.
File: hrm001.dif
a* beams=-17:17 b* beams=-24:24
E = 1 2 0 . 0 k V t i l t = (0 . 0 , 0 . 0)
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-0.8
-0.4
0.0
0.4
0.8
t=80.00A

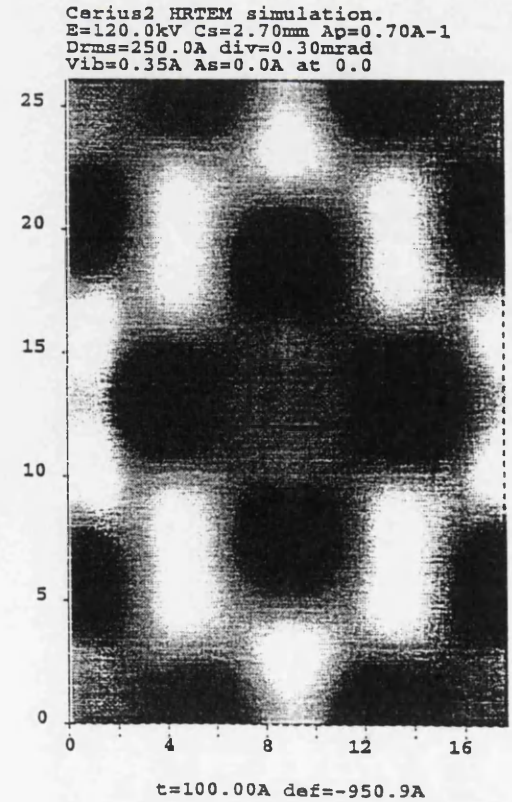

Square of linear CTF
E=120.0kV Drms=250.0A Div=0.30mrad
As=0.0A at 0.0deg
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-0.8
-0.4
0.0
0.4
0.8
Cs=2.70mm Def=-950.94A

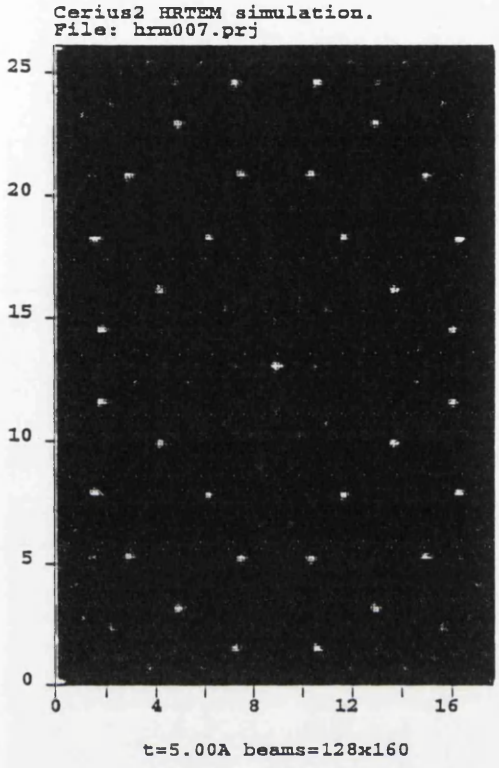

Cerius2 HRTEM simulation.
E=120.0kV Cs=2.70mm Ap=0.70A-1
Drms=250.0A div=0.30mrad
Vib=0.35A As=0.0A at 0.0
25
20
15
10
5
0
0
4
8
12
16
t=80.00A def=-950.9A

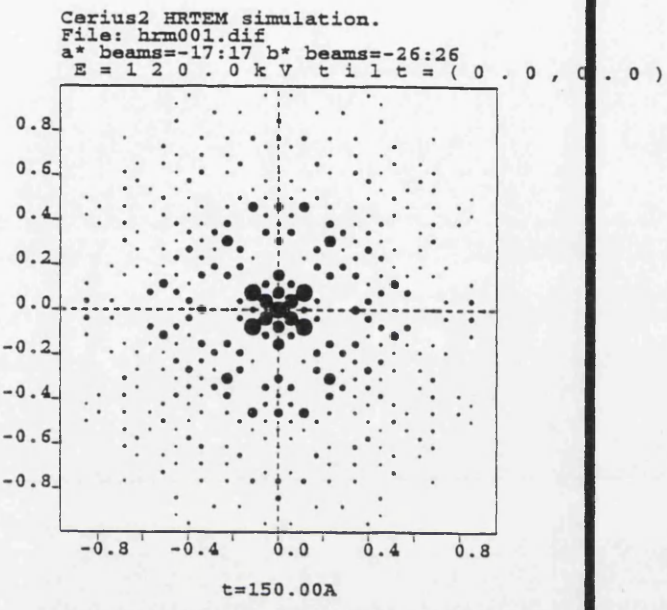

Cerius2 HRTEM simulation.
File: hrm006.prj
25
20
15
10
5
0
0
4
8
12
16
t=5.00A beams=128x160

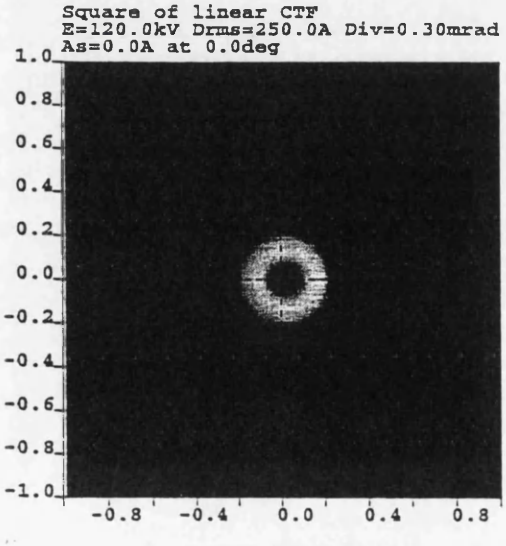

Cerius2 HRTEM simulation.
File: hrm001.dif
a* beams=-17:17 b* beams=-25:25
E = 1 2 0 . 0 k V t i l t = (0 . 0 , 0 . 0)
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-0.8
-0.4
0.0
0.4
0.8
t=100.00A

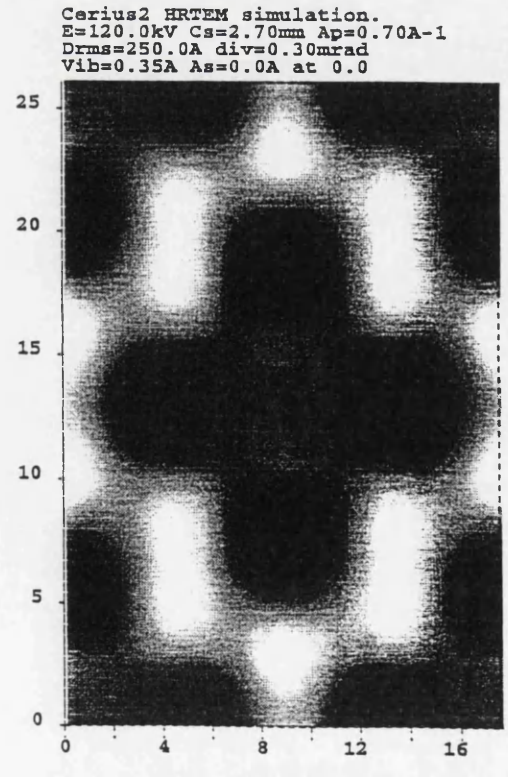

Square of linear CTF
E=120.0kV Drms=250.0A Div=0.30mrad
As=0.0A at 0.0deg
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-0.8
-0.4
0.0
0.4
0.8
Cs=2.70mm Def=-950.94A

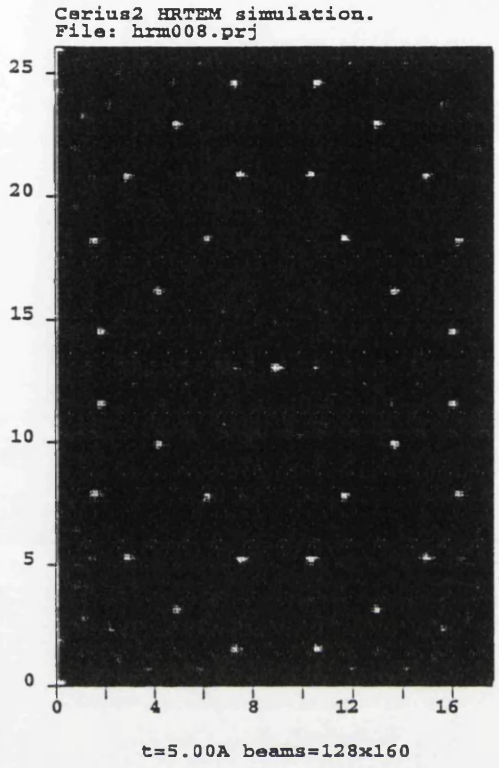

Cerius2 HRTEM simulation.
E=120.0kV Cs=2.70mm Ap=0.70A-1
Drms=250.0A div=0.30mrad
Vib=0.35A As=0.0A at 0.0
25
20
15
10
5
0
0
4
8
12
16
t=100.00A def=-950.9A

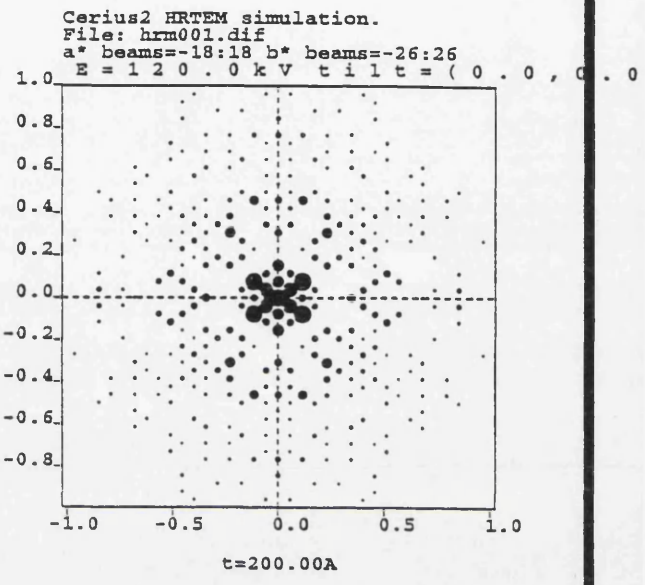

Cerius2 HRTEM simulation.
File: hrm007.prj
25
20
15
10
5
0
0
4
8
12
16
t=5.00A beams=128x160

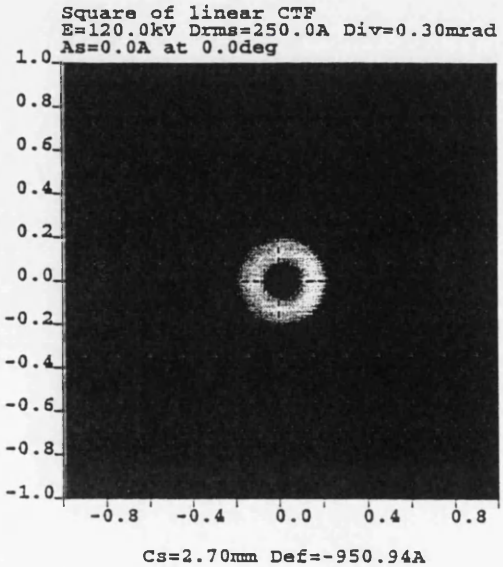

Cerius2 HRTEM simulation.
File: hrm001.dif
a* beams=-17:17 b* beams=-26:26
E = 1 2 0 . 0 k V t i l t = (0 . 0 , 0 . 0)
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-0.8
-0.4
0.0
0.4
0.8
t=150.00A

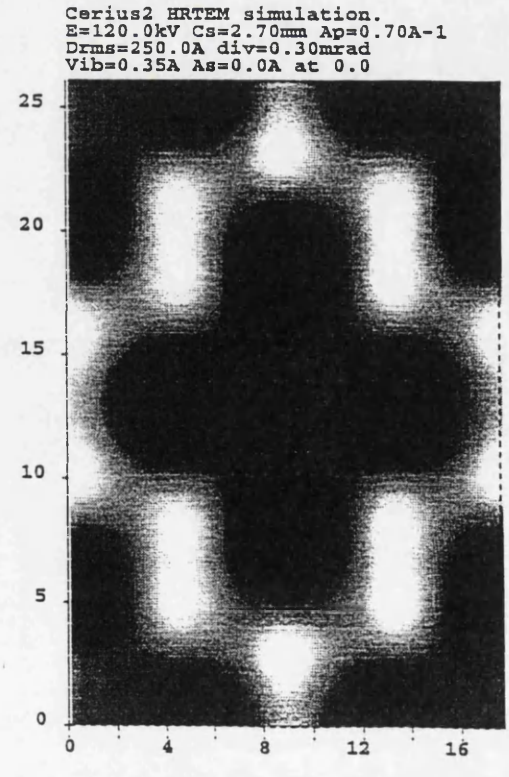

Square of linear CTF
E=120.0kV Drms=250.0A Div=0.30mrad
As=0.0A at 0.0deg
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-0.8
-0.4
0.0
0.4
0.8
Cs=2.70mm Def=-950.94A


Cerius2 HRTEM simulation.
E=120.0kV Cs=2.70mm Ap=0.70A-1
Drms=250.0A div=0.30mrad
Vib=0.35A As=0.0A at 0.0
25
20
15
10
5
0
0
4
8
12
16
t=150.00A def=-950.9A


Cerius2 HRTEM simulation.
File: hrm008.prj
25
20
15
10
5
0
0
4
8
12
16
t=5.00A beams=128x160


Cerius2 HRTEM simulation.
File: hrm001.dif
a* beams=-18:18 b* beams=-26:26
E = 1 2 0 . 0 k V t i l t = (0 . 0 , 0 . 0
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-0.5
0.0
0.5
1.0
t=200.00A

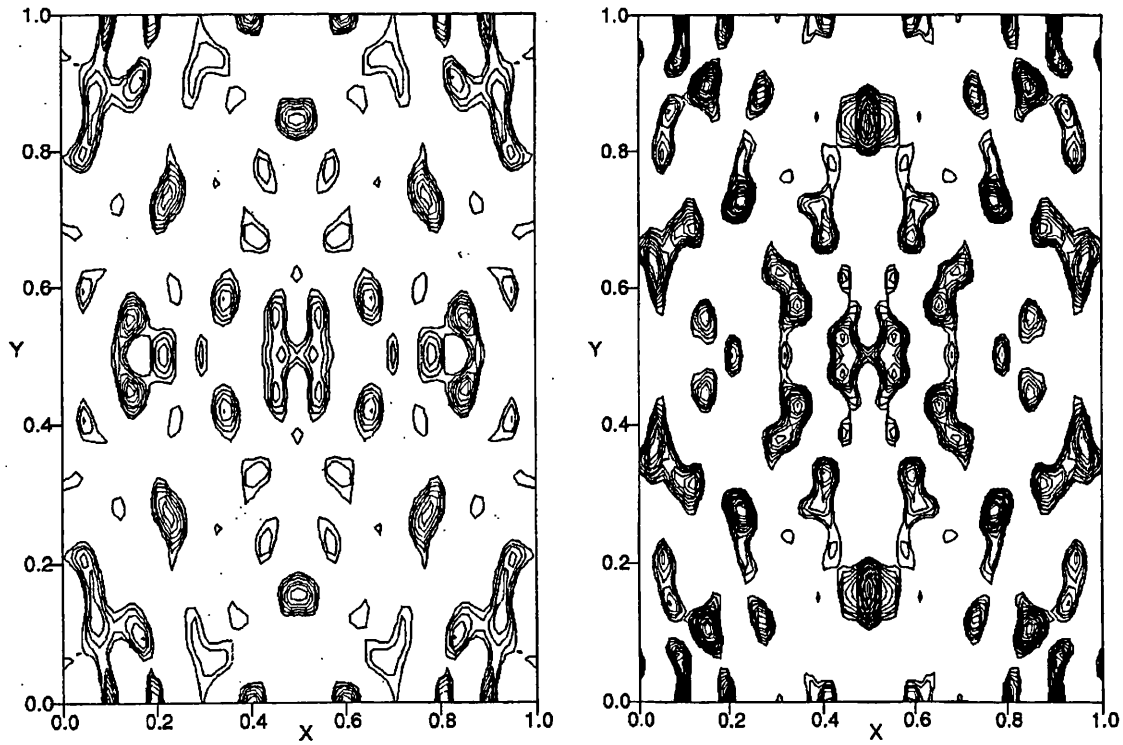

Square of linear CTF
E=120.0kV Drms=250.0A Div=0.30mrad
As=0.0A at 0.0deg
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-0.8
-0.4
0.0
0.4
0.8
Cs=2.70mm Def=-950.94A


Cerius2 HRTEM simulation.
E=120.0kV Cs=2.70mm Ap=0.70A-1
Drms=250.0A div=0.30mrad
Vib=0.35A As=0.0A at 0.0
25
20
15
10
5
0
0
4
8
12
16
t=200.00A def=-950.9A

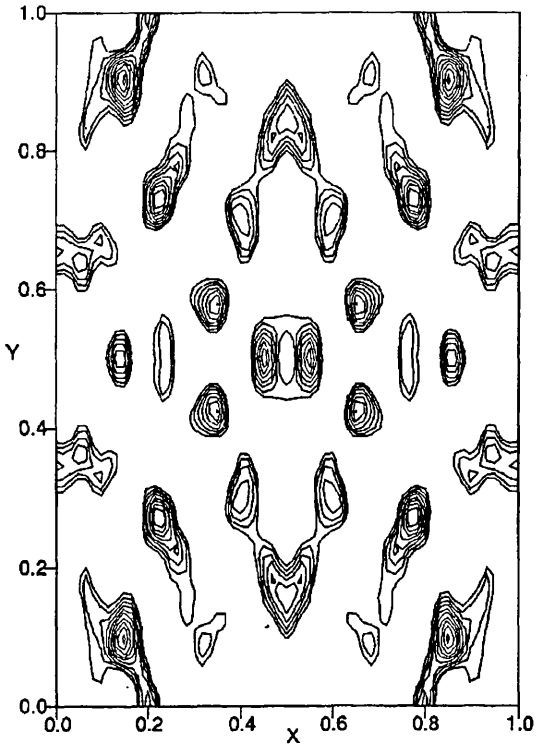
7.6. Patterson Map Interpretation.

As a prelude to the electron crystallographic structure investigation by maximum entropy methods, Patterson maps were generated. Structural information contained within the simulated diffraction pattern is portrayed by the Patterson function. Figure 7.4. (*a-g*) shows the Patterson maps generated for each individual data set.

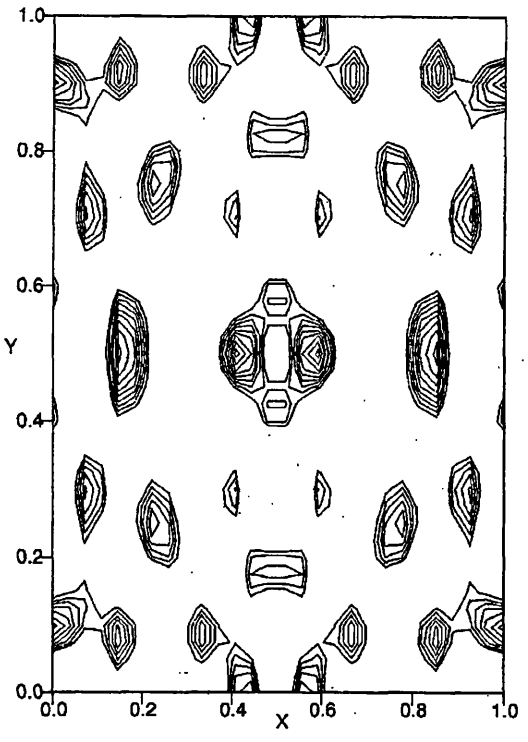
Examination of these Patterson maps showed that the data did indeed show promise for use within further structural calculations. Information regarding atomicity was seen to exist throughout the range of thickness examined. More importantly, the correct molecular orientation in the unit cell could be identified. Therefore, even when dynamical scattering is present, Patterson maps are found to still represent the underlying crystal structure.



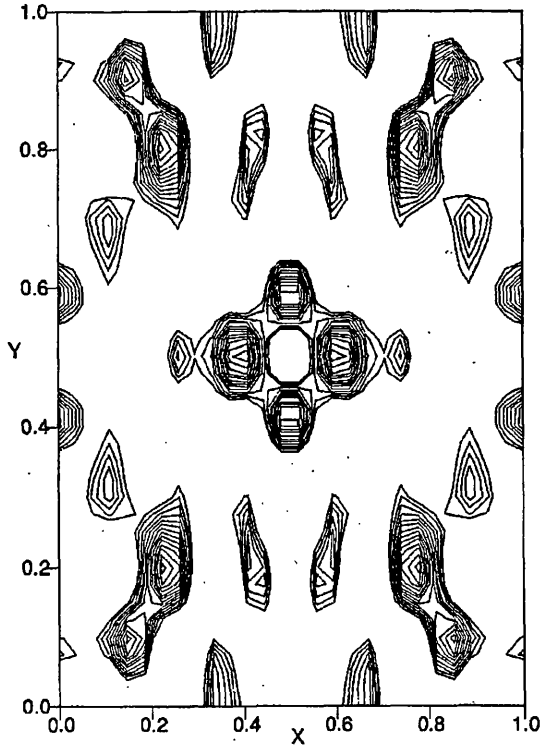
(a) 20 Å (b) 40 Å



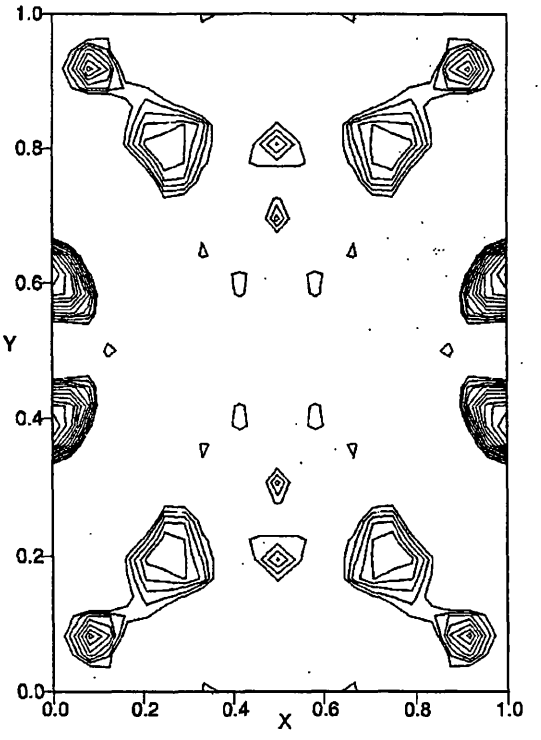
(c) 60 Å



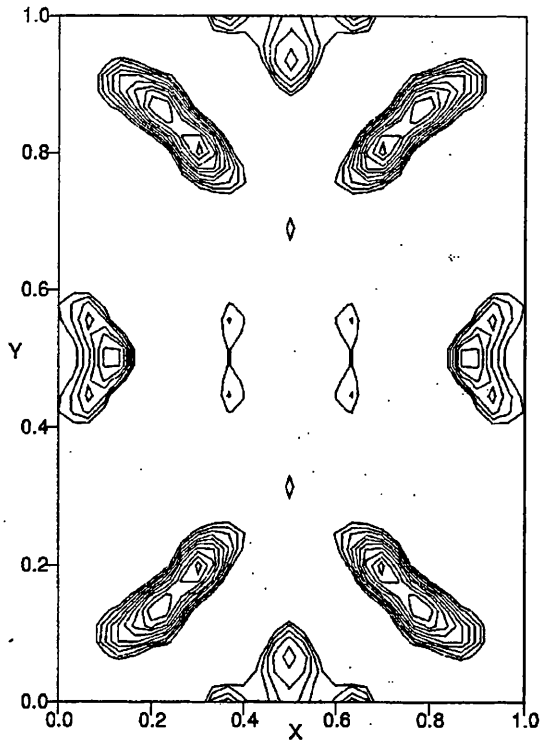
(d) 80 Å



(e) 100 Å



(f) 150 Å



(g) 200 Å

Figure 7.4. contd. Patterson Maps obtained for data from a range of sample thickness 20-200 Å.

7.7. Ab Initio Phase Determination using Maximum Entropy Methods.

In accordance with the theoretical description of maximum entropy methods and the experimental procedures employed elsewhere in this thesis, the following procedure was carried out for each of the datasets investigated.

The diffraction data were normalised using MITHRIL[204,205] to give the unitary structure magnitudes ($|U_h|^{obs}$) and their associated standard deviations using electron scattering factors. Reflections remaining after normalisation were used within MICE.

Due to the projection under investigation being centered, only one reflection was required for origin definition, thus forming the basis set {**H**}. The root node of the

phasing tree was generated by carrying out a constrained entropy maximisation. Here, the phases and amplitudes of the origin phases were used as constraints in producing a maximum entropy distribution $q^{ME}(\mathbf{x})$.

Depending on the size of the data set being investigated, either five or seven reflections were chosen that optimally enlarged the second neighbourhood of the basis set $\{\mathbf{H}\}$. These reflections had phases permuted, thus increasing the number of nodes at this branch of the phasing tree. Each node was subjected to constrained entropy maximisation as before. Log likelihood gains were computed for each and analysis using student *t*-test produced a ranking of nodes.

The best nodes from the first level were kept and used to generate a further level of the tree. Again, depending upon the data set, either two, four or five reflections were selected to undergo phase permutation. This allowed for the generation of up to a maximum of 256 new nodes. Entropy maximisation and analysis using log likelihood gain and *t*-test were repeated. Centroid maps were generated for the solution remaining with the highest LLG.

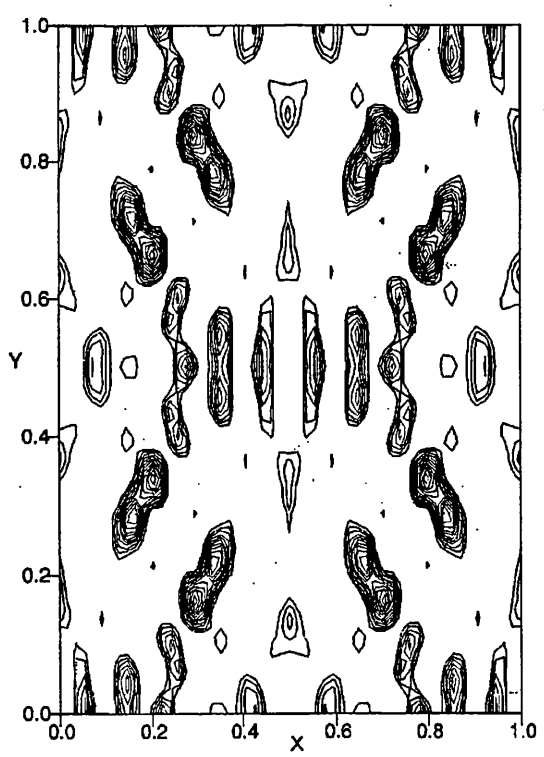
Table 7.2. details the development of the phasing tree for each data set. The number of unique reflection, origin defining reflection, permuted reflections and top nodes, with associated LLG, for each level are given.

			LEVEL 1			LEVEL 2		
Dataset	No. refs.	Origin	Permuted	Nodes	Top LLG	Permuted	Nodes	Top LLG
20	57	(5,3)	7	1-129	0.078(87)	5	130-385	0.26(346)
40	40	(1,3)	7	1-129	0.593(42)	5	130-385	0.94(218)
60	34	(1,3)	7	1-129	0.533(42)	5	130-385	0.94(153)
80	19	(3,3)	5	1-33	0.091(19)	4	34-129	0.22(137)
100	13	(3,3)	7	1-129	3.188(65)	4	130-225	3.73(161)
150	15	(1,3)	5	1-33	0.890(33)	4	34-129	2.36(62)
200	9	(3,5)	5	1-33	0.676(25)	2	34-61	1.75(37)

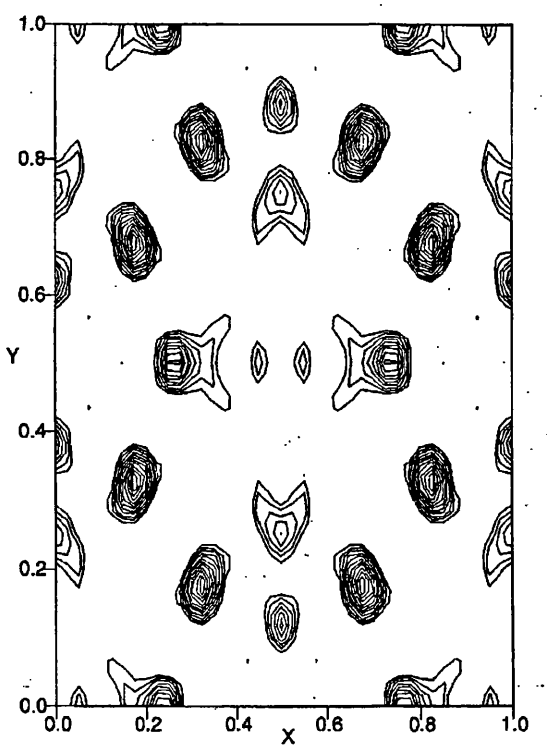
Table 7.2. Details of calculations using MICE.

Figure 7.5. (*a-g*) depicts the best potential maps obtained for each data set from phase determination by maximum entropy means. Upon examination of all maps generated, the orientation of the molecule within the unit cell is apparent. As expected, achieved resolution decreases with increasing crystal thickness. However, in terms of identification of the heavier atom positions, structure solution by *ab initio* means is very successful.

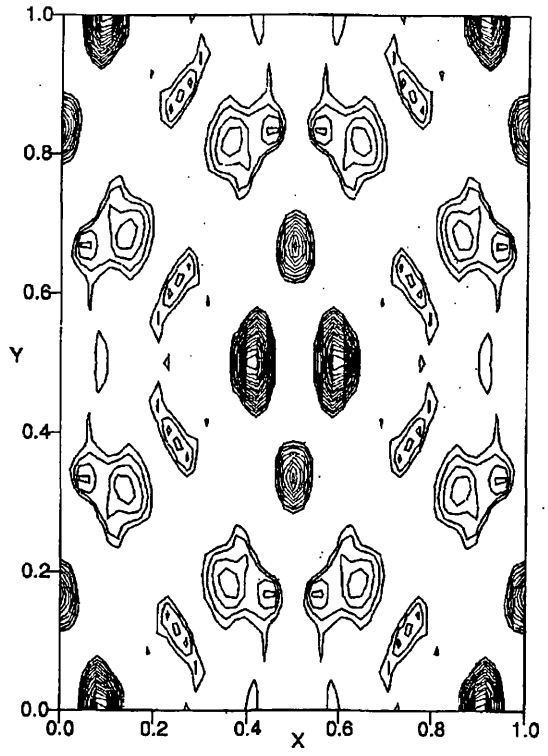
Dominant throughout the whole series of centroid maps are chlorine positions and areas of presence approximating to the aromatic rings of the molecule. At 20 Å thickness, the areas of aromatic presence can be resolved to bond level. As thickness increases, a smearing of these aromatic bonds is seen to give concentrations of density approximating to those aromatic areas. Likewise, with the chlorine positions, loss of individual positions is seen in a merging with aromatic areas. An interesting feature to note is that for specimen thickness up to 60 Å no evidence of the copper atoms is found. However, above 80 Å thickness copper atoms are found to be the most prominent feature of the centroid maps.



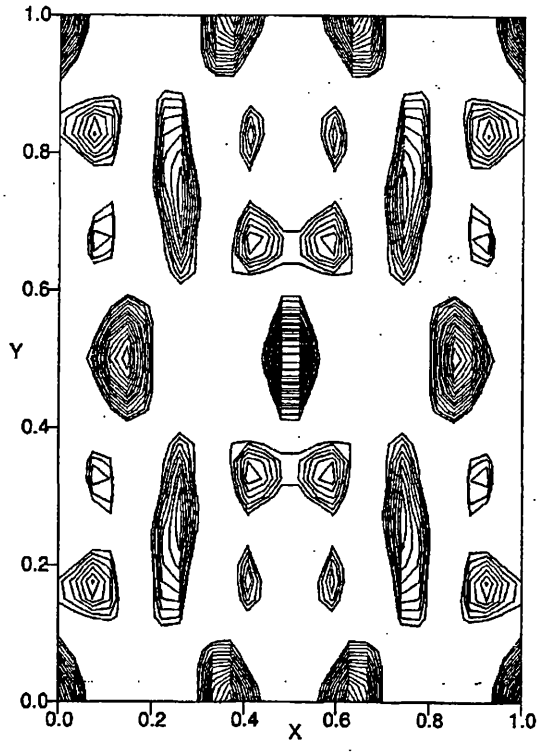
(a) 20 Å



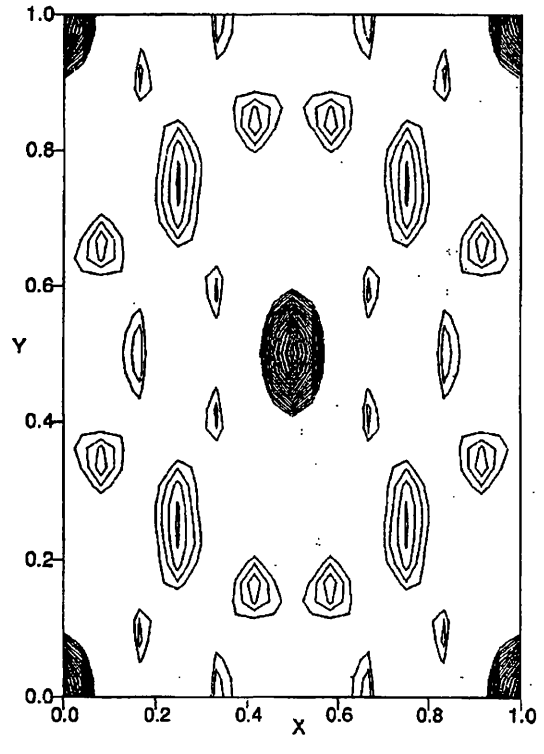
(b) 40 Å



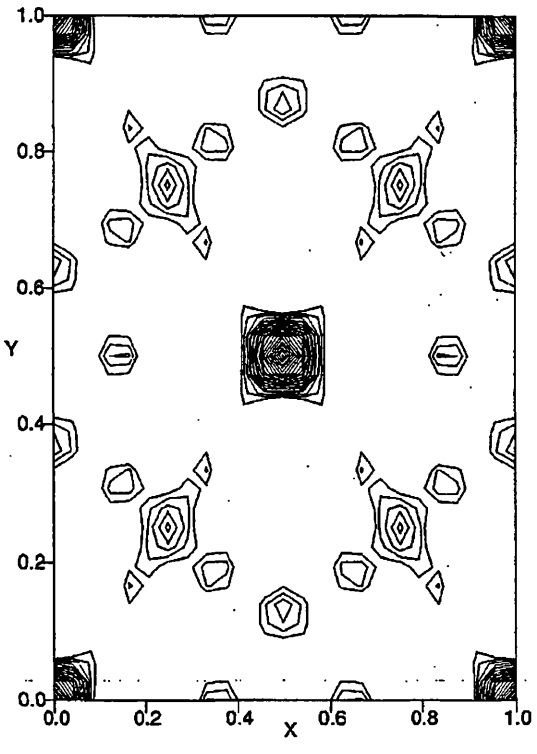
(c) 60 Å



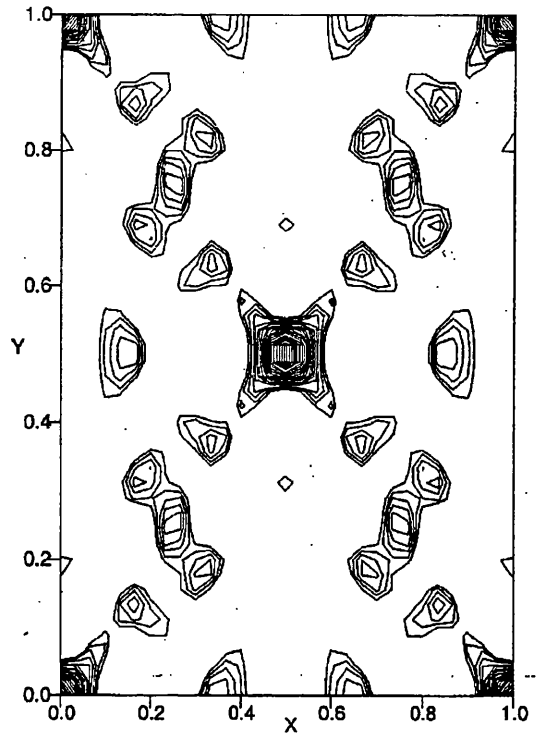
(d) 80 Å



(e) 100 Å



(f) 150 Å



(g) 200 Å

Figure 7.5. (contd.) Centroid maps from *ab initio* phasing.

Overall, results using *ab initio* phasing techniques are very encouraging. It is clear that for electron crystallographic analyses employing maximum entropy methods, strict adherence to the kinematical scattering approximation is not necessarily required. This is exemplified by the results acheived in this study. Under extraneous circumstances of dynamical scattering and the presence of considerable errors, direct phasing by maximum entropy methods is shown to be robust and more importantly successful.

7.8. Use of Image Derived Phases in Maximum Entropy Methods.

Use of phases derived from the Fourier transform of electron micrographs is a well documented route to structure solution by electron crystallography. In these studies, phase information was derived from images, then combined with the previously used

diffraction intensities in maximum entropy procedures and phase searching techniques, to furnish structural information about copper perchlorophthalocyanine.

Using the methodology described in chapter 3 regarding phase retrieval from images, Fourier transforms of previously obtained images of copper perchlorophthalocyanine yielded a large number of phased reflections.

The number of phased reflections assigned to a particular dataset was dependent upon the size of the data set involved. Table 7.3. lists the number of reflections assigned to each dataset.

Dataset	No. Phased Reflections
20 Å	9
40 Å	15
60 Å	7
80 Å	8
100 Å	6
150 Å	5
200 Å	3

Table 7.3. Phased Reflections.

All intensity information was the same as was utilised in the *ab initio* calculations described in the previous section.

Again, diffraction data were normalised using MITHRIL to give unitary structure magnitudes and their associated standard deviations using electron scattering factors. Using the phases assigned for each dataset, the basis set {**H**} was formed. This consists of the relevant phases and amplitudes of the assigned reflections. Thus, the root node of the phasing tree was generated and subsequently subjected to constrained entropy maximisation.

Using a selection algorithm, further reflections were chosen to be added to this basis set. As the phases of these reflections are unknown, all possible combinations of 0° and 180° were used , thus generating new nodes on the phasing tree. Each node was subjected to constrained entropy maximisation as before.

Following convention, analysis of the nodes generated was carried out employing student *t*-test. This generates a ranking of nodes according to their likelihood values. The top nodes found at this level had centroid maps generated.

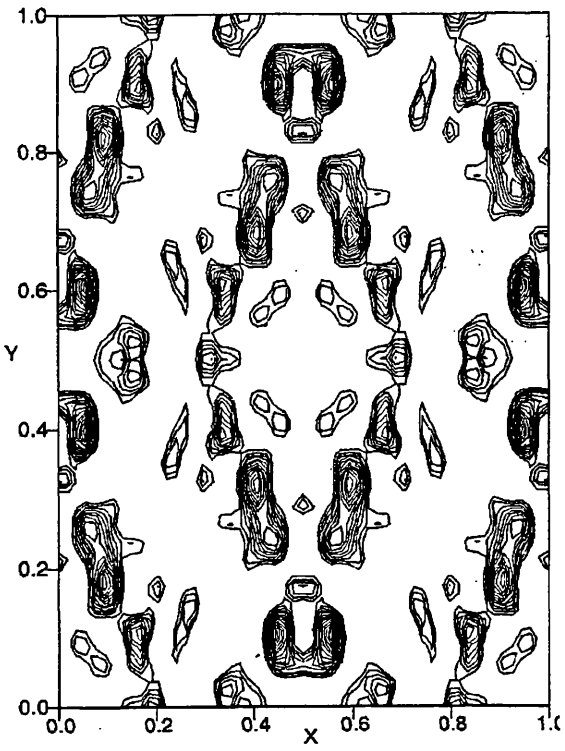
For some of the smaller datasets, this was as far as the size of the dataset would allow the calculations to progress. The larger datasets could allow generation of a further, second level of the phasing tree.

For generation of a second level, the best nodes from the previous level were kept. Further reflections were selected to have their phases permuted. As before, this generates a new family of nodes which undergo entropy maximisation and subsequent analysis and ranking. Centroid maps were generated for the solutions with the highest log likelihood gain. Table 7.4. details the development of the calculations using image derived phases within MICE.

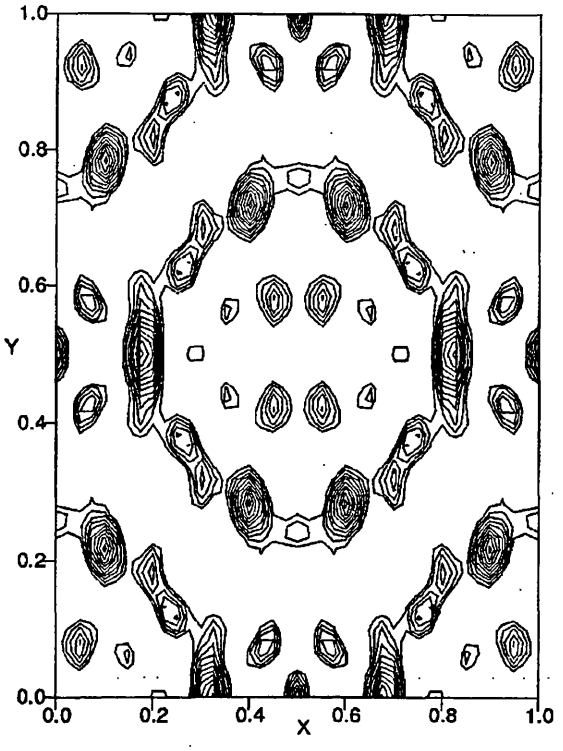
Dataset	No. refs.	Phased	Permuted	Nodes	Top LLG	Permuted	Nodes	Top LLG
20 Å	57	9	7	1-129	0.348(71)	5	130-385	0.95(218)
40 Å	40	15	7	1-129	1.469(98)	5	130-385	5.86(184)
60 Å	34	7	7	1-129	0.400(28)	5	130-385	1.00(213)
80 Å	19	8	5	1-33	5.03(2)	-	-	-
100 Å	13	6	5	1-33	0.108(21)	1	34-45	0.16(36)
150 Å	15	5	5	1-33	2.20(9)	-	-	-
200 Å	9	3	5	1-33	0.755(24)	-	-	-

Table 7.4. Details of calculations using MICE.

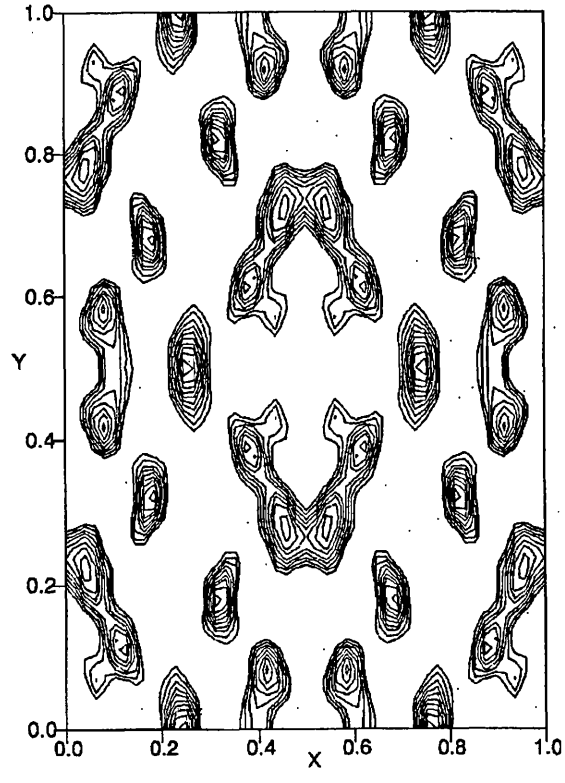
Figure 7.6. (*a-g*) depicts the most successful centroid maps for each dataset of a given specimen thickness of copper perchlorophthalocyanine.



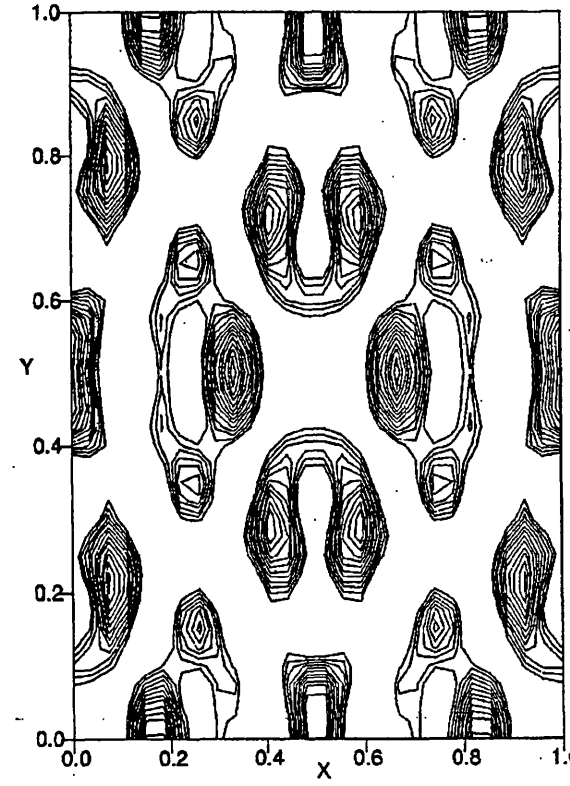
(a) 20 Å



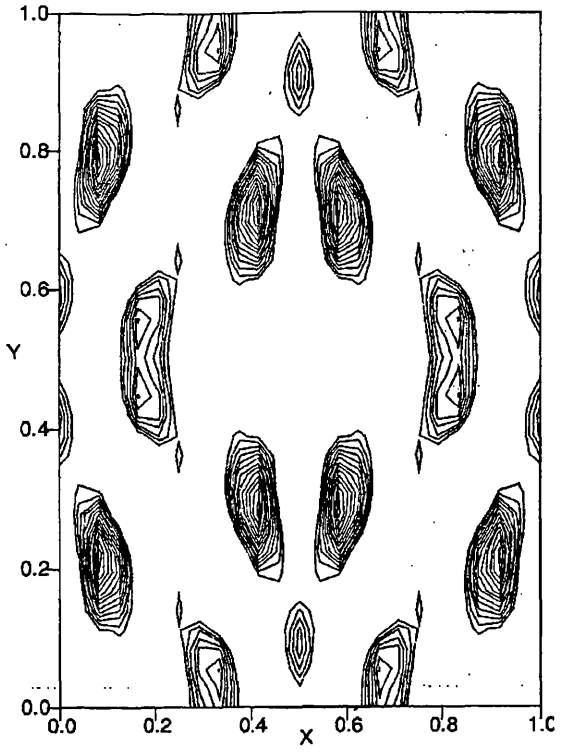
(b) 40 Å



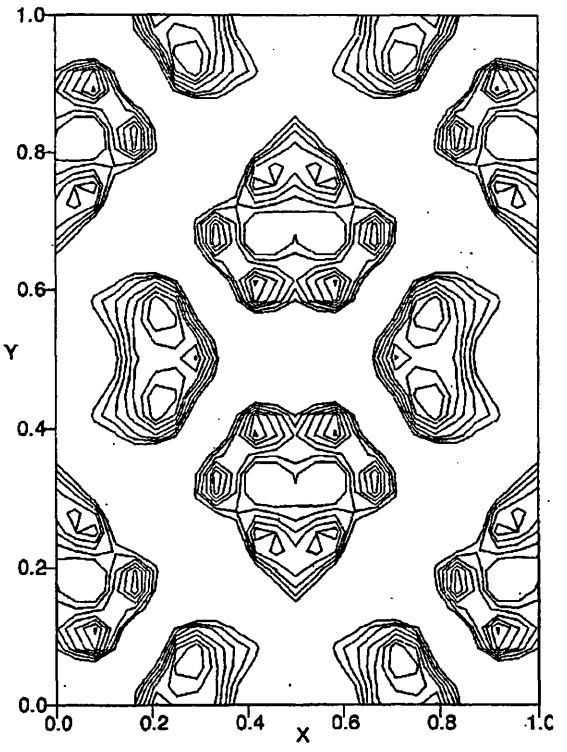
(c) 60 Å



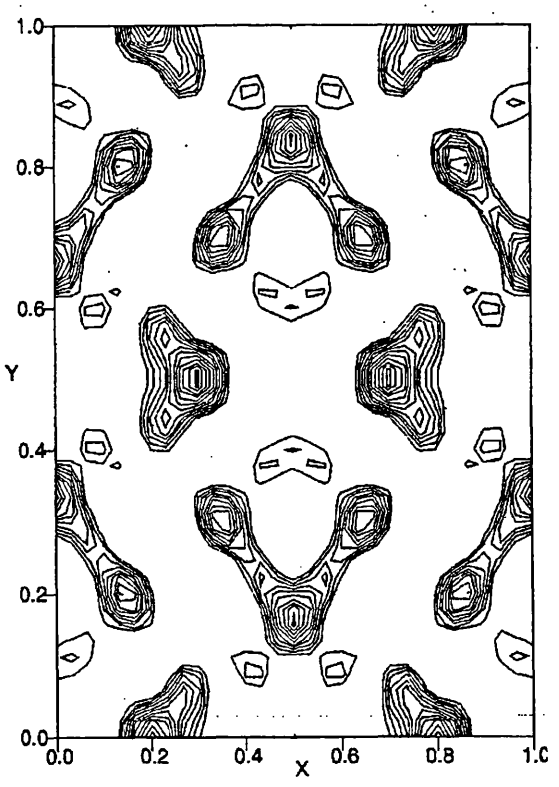
(d) 80 Å



(e) 100 Å



(f) 150 Å



(g) 200 Å

Figure 7.6.(contd.) Centroid Maps from MICE using image derived phases.

As with *ab initio* phasing, the correct molecular orientation is immediately apparent upon inspection of the maps. What is surprising is the superior molecular detail contained in these maps compared to the previously obtained multislice images. For each respective thickness, resolution is greatly enhanced over that of centroid maps obtained from *ab initio* phasing. Therefore, this is another study which gives emphasis to the use of image derived phases in electron crystallography.

An important feature of the maps obtained is that for all datasets, the copper atoms are not present in the centroid maps. This could be due to the fact that the phases obtained are affected by dynamical scattering and that phases are found to have a profound effect of the correctness of any proposed structure. However, with respect to the other atomic species that are present, atomic resolution is greatly enhanced. Inspection of the 20 Å thickness centroid map allows for identification of all chlorine positions, along with aromatic bond presence.

This degree of resolution is lost slightly as thickness increases, with the observation of a merging of the chlorine and aromatic bond positions to give a concentrated area of atomic presence. This merging of atomic areas is retained throughout the series of thickness investigated , even up to and including 200 Å.

Overall, resolution and atomic detail is greatly improved over that of results *from ab initio* phasing, with no appearance of the spurious peaks seen in *ab initio* centroid maps of specimen of greater thickness.

In conclusion, the comments made regarding successful *ab initio* phase determination by maximum entropy methods hold true for this study using image derived phases. The difference here is that resolution and atomic detail is seen to be greatly improved. This is very advantageous to the electron crystallographer in that the methods employed here are found to be stable even when data is of questionable quality and of significant sparseness.

8.0 ELECTRON CRYSTALLOGRAPHIC STUDIES OF AN ALUMINIUM PHTHALOCYANINE POLYMER SYSTEM.

8.1. Introduction.

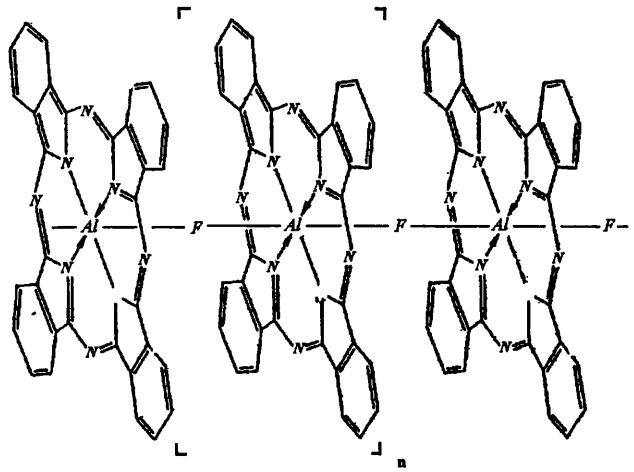


Figure 8.1. Structure of $(AlPcF)_n$ (adapted from ref 206.)

$(AlPcF)_n$, where Pc is phthalocyanine ligand, is a cofacial phthalocyanine polymer system. Several studies have been carried out on this polymer, and ones related to it. A study of $(AlPcF)_n$ by X-ray powder diffraction, allied with other techniques found that the polymer chains were packed parallel to each other, which was confirmed by the spacing data collected.[206]

Much interest was generated in polymers of this type due to the expression of properties such as conductivity. This led to the study of this material by high resolution electron microscopy.[207]

High resolution electron microscopy images of $(AlPcF)_n$ were obtained in which molecular columns could be resolved. This led to the conclusion that the molecules in the crystallites are parallel to each other and orientated vertically. Therefore, the phthalocyanine rings are perpendicular to the Al-F backbone. Descriptions are given

in this work of the specimen preparation techniques necessary, along with methods of collection of such high resolution images.

One of the themes of this thesis is the validation of electron crystallographic methods involving phase determination by maximum entropy means. In accordance with this, use is made of image and diffraction data from the aforementioned electron microscopy study for an electron crystallographic analysis of $(AlPcF)_n$.

As described in the previous theoretical sections, structure elucidation by maximum entropy methods can be attempted through ab initio phase determination or by utilising image derived phases. Both approaches have been used in this work and a detailed description of the experimental procedures involved in structural electron crystallographic analyses of this nature will be given.

8.2. Experimental Methods.

Use has been made of the original micrographs obtained by Fryer and Kenney in their study of $(AlPcF)_n$. This has included both information from images and diffraction patterns. As described previously, phase information can be retrieved from experimental electron microscope images. Using the appropriate software, CRISP[69], an image can be transferred to PC via a frame grabber. The Fourier transform of the image can be obtained and from this, the necessary phase information is derived.

Intensity information is quantified in an analogous manner, in that the diffraction pattern is transferred to PC using a frame grabber. However, in this instance, ELD[68] software is used for quantification of diffraction pattern. The theoretical basis for both of these procedures has been described in the sections detailing retrieval of information from both diffraction patterns and images.

8.3. Ab initio Phase Determination.

Electron diffraction data for crystallites of $(AlPcF)_n$ allowed determination of the tetragonal unit cell parameters which were found to be $a=b=$ 13.37 Å and $c=$ 3.6 Å, with one molecule per unit cell. Diffraction data utilised here is from the *ab* projection and the plane group is determined to be *p4gm*.

The diffraction data were normalised using MITHRIL to give the unitary structure magnitudes ($|U_h|^{obs}$) and their associated standard deviations using electron scattering factors. The overall isotropic temperature factor *B* was found to slightly negative, with a value of -0.8. Negative temperature factor values are common for sparse electron diffraction datasets such as those examined here. After normalisation, 64 unique reflections remained for use within MICE. U-magnitudes and their resolution in Å are tabulated in Table 8.1.

No.	h	k	$\|U_h\|^{obs}$	d (Å)	No.	h	k	$\|U_h\|^{obs}$	d (Å)
1	0	2	0.567	6.68	33	2	9	0.139	1.45
2	2	0	0.567	6.68	34	9	2	0.139	1.45
3	4	0	0.367	3.34	35	8	2	0.136	1.62
4	0	4	0.367	3.34	36	2	8	0.136	1.62
5	2	2	0.301	4.73	37	3	1	0.136	4.23
6	1	2	0.220	5.98	38	1	3	0.136	4.23
7	2	1	0.220	5.98	39	2	5	0.126	2.48
8	3	2	0.207	3.71	40	5	2	0.126	2.48
9	2	3	0.207	3.71	41	4	8	0.126	1.49
10	1	11	0.203	1.21	42	8	4	0.126	1.49
11	11	1	0.203	1.21	43	1	1	0.125	9.45
12	9	7	0.189	1.17	44	4	7	0.124	1.66
13	7	9	0.189	1.17	45	7	4	0.124	1.66
14	6	0	0.185	2.23	46	1	4	0.123	3.24
15	0	6	0.185	2.23	47	4	1	0.123	3.24
16	10	3	0.180	1.28	48	2	4	0.123	2.99
17	3	10	0.180	1.28	49	4	2	0.123	2.99
18	6	1	0.159	2.20	50	10	2	0.120	1.31
19	1	6	0.159	2.20	51	2	10	0.120	1.31
20	5	10	0.156	1.20	52	3	6	0.120	1.99
21	10	5	0.156	1.20	53	6	3	0.120	1.99
22	4	12	0.154	1.06	54	0	10	0.118	1.34
23	12	4	0.154	1.06	55	10	0	0.118	1.34
24	4	4	0.154	2.36	56	5	3	0.117	2.29
25	9	3	0.151	1.41	57	3	5	0.117	2.29
26	3	9	0.151	1.41	58	5	5	0.116	1.89
27	12	0	0.148	1.11	59	4	5	0.110	2.09
28	0	12	0.148	1.11	60	5	4	0.110	2.09
29	9	1	0.144	1.48	61	0	8	0.101	1.67
30	1	9	0.144	1.48	62	8	0	0.101	1.67
31	6	5	0.142	1.71	63	6	2	0.091	2.11
32	5	6	0.142	1.71	64	2	6	0.091	2.11

Table 8.1. Reflection number, *h*,*k*, the unitary structure factor $|U_h|^{obs}$ and *d* in Å

Subsequent maximum entropy calculations were carried out within MICE. Two suitable reflections were chosen to define the origin and were given phase angles of 0°. The reflections chosen were 8 (3,2) and 9 (2,3) with U magnitudes for both of 0.207. These reflections thus define the basis set {**H**}. Node 1 of the phasing tree was generated by carrying out a constrained entropy maximisation. In this, the phases and amplitudes of the origin phases were used as constraints in producing a maximum entropy distribution $q^{ME}(\mathbf{x})$.

Seven strong reflections were selected to have their phases permuted. This generated 2^7 further nodes, giving a total of 129. Each node was subjected to constrained entropy maximisation as outlined previously and log likelihood gain, LLG, was calculated for each. Analysis of nodes was achieved through utilisation of student *t*-tests, again described previously. The first level of the phasing tree is summarised in Table 8.2.

The top 4 nodes were identified from analysis and subsequently centroid maps were produced for each of these. Upon examination, all maps appeared similar and no discernible molecular features could be identified and it was decided to generate a second level of the phasing tree.

Nodes 5, 21, 59 and 115 were selected to have further phases permuted in the generation of the second level of the phasing tree. Five reflections were selected and there were numbers 12, 13, 16, 17 and 25. This led to the generation of a 128 nodes, giving 257 in total. Constrained entropy maximisation and analysis was repeated. Listed in Table 8.3. are the highest nodes for level 2, along with their entropy values.

Node	From node	Entropy	LLG
197	5	-0.501	1.39
201	5	-0.528	1.59
205	5	-0.528	1.59
217	5	-0.470	1.43
237	21	-0.470	1.43

Table 8.3. Top Nodes from Second Level of Phasing Tree.

Two possible paths to solution are seen to emerge. Centroid maps were generated and upon examination, the recognised phthalocyanine molecular motif could be discerned.

Node	Entropy	LLG	Node	Entropy	LLG	Node	Entropy	LLG
1	-0.095	0.00	44	-0.343	0.61	87	-0.334	0.59
2	-0.385	0.49	45	-0.320	-0.04	88	-0.356	0.66
3	-0.352	-0.02	46	-0.365	0.26	89	-0.314	-0.05
4	-0.356	-0.04	47	-0.329	0.25	90	-0.379	0.50
5	-0.305	0.69	48	-0.337	0.34	91	-0.368	-0.04
6	-0.365	0.25	49	-0.294	0.31	92	-0.339	-0.01
7	-0.327	0.30	50	-0.390	0.20	93	-0.296	0.67
8	-0.347	0.22	51	-0.347	0.31	94	-0.366	0.26
9	-0.302	0.28	52	-0.369	0.26	95	-0.331	0.31
10	-0.372	0.25	53	-0.308	0.25	96	-0.324	0.25
11	-0.347	0.23	54	-0.375	0.47	97	-0.299	0.30
12	-0.329	0.28	55	-0.320	0.00	98	-0.425	-0.06
13	-0.309	0.31	56	-0.343	-0.03	99	-0.371	0.65
14	-0.345	0.01	57	-0.282	0.59	100	-0.355	0.59
15	-0.308	0.52	58	-0.374	-0.02	101	-0.327	-0.03
16	-0.308	0.49	59	-0.378	0.76	102	-0.391	0.22
17	-0.293	-0.06	60	-0.349	0.63	103	-0.339	0.23
18	-0.391	0.47	61	-0.309	-0.08	104	-0.353	0.31
19	-0.377	-0.01	62	-0.351	0.22	105	-0.298	0.32
20	-0.371	-0.01	63	-0.317	0.26	106	-0.411	0.29
21	-0.310	0.69	64	-0.321	0.31	107	-0.352	0.28
22	-0.383	0.26	65	-0.287	0.26	108	-0.350	0.24
23	-0.337	0.33	66	-0.409	0.26	109	-0.308	0.34
24	-0.359	0.28	67	-0.365	0.22	110	-0.360	0.49
25	-0.311	0.31	68	-0.340	0.24	111	-0.327	-0.03
26	-0.387	0.23	69	-0.314	0.30	112	-0.327	-0.03
27	-0.377	0.32	70	-0.373	0.00	113	-0.279	0.64
28	-0.348	0.37	71	-0.323	0.55	114	-0.407	-0.06
29	-0.319	0.33	72	-0.361	0.61	115	-0.381	0.70
30	-0.363	-0.01	73	-0.301	-0.05	116	-0.385	0.67
31	-0.321	0.59	74	-0.386	0.53	117	-0.313	-0.07
32	-0.321	0.60	75	-0.346	-0.05	118	-0.377	0.21
33	-0.304	-0.06	76	-0.325	-0.03	119	-0.330	0.24
34	-0.420	0.27	77	-0.290	0.66	120	-0.333	0.27
35	-0.355	0.29	78	-0.352	0.27	121	-0.292	0.25
36	-0.391	0.27	79	-0.323	0.30	122	-0.376	0.22
37	-0.326	0.33	80	-0.316	0.19	123	-0.344	0.30
38	-0.376	0.50	81	-0.287	0.28	124	-0.332	0.22
39	-0.334	-0.01	82	-0.414	0.20	125	-0.294	0.24
40	-0.361	-0.04	83	-0.394	0.29	126	-0.366	0.46
41	-0.290	0.66	84	-0.359	0.31	127	-0.315	-0.03
42	-0.400	-0.03	85	-0.322	0.31	128	-0.313	-0.04
43	-0.366	0.67	86	-0.390	-0.04	129	-0.287	0.62

Table 8.2. First level of the maximum entropy phasing tree.

Figure 8.2. shows a representative centroid map from level 2 of the phasing tree. The molecular motif is obvious, with the phthalocyanine quatrefoil apparent.

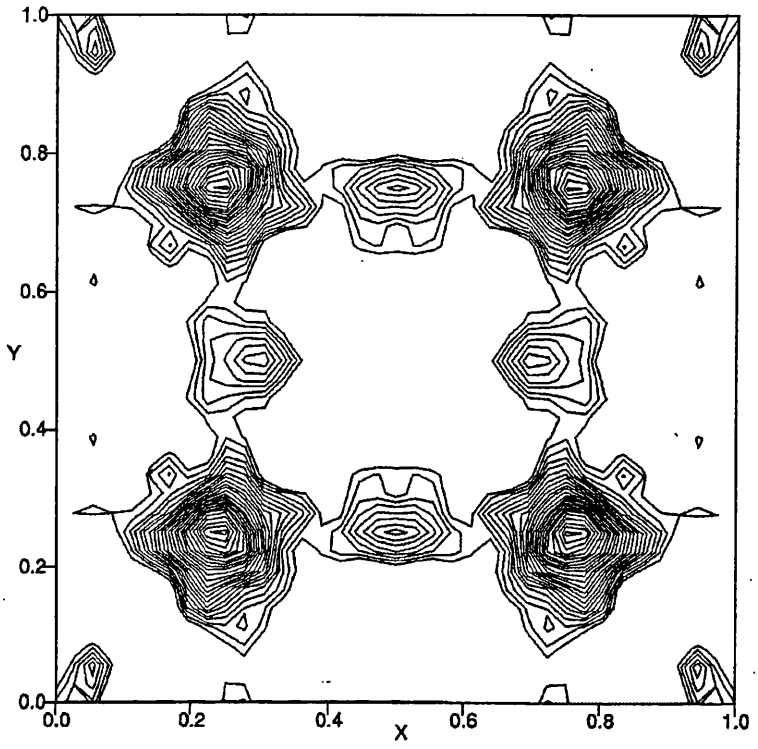


Figure 8.2. Node 205 from level 2 of phasing tree.

As nodes 201 and 205 both have the same likelihood values and give rise to similar centroid maps, it is of value to construct a third level of the phasing tree. The aim of any tree-based phase search is to arrive at a solution deriving from one set of phase choices.

Generation of the third level was achieved by following the same procedure as above. The top five nodes from level two were used, these being nodes 197, 201, 205, 217 and 237. Reflection chosen to have their phases permuted are 32, 33, 34, 50 and 51. This extended the phasing tree to 417 nodes. Again, new nodes generated underwent constrained entropy maximisation. Subsequent analysis and scoring produced a ranking of nodes. Examination of these results shows two nodes with distinctly high likelihoods. Values of entropy and likelihood for these nodes are listed in Table 8.4.

Node	From node	Entropy	LLG
326	217	-0.556	1.65
357	237	-0.559	1.67

Table 8.4. Level three of phasing tree from MICE.

As with previous levels, centroid maps were generated and upon examination node 357 was found to represent best the molecular motif of the aluminium phthalocyanine monomer. Figure 8.3. shows centroid map obtained for node 357.

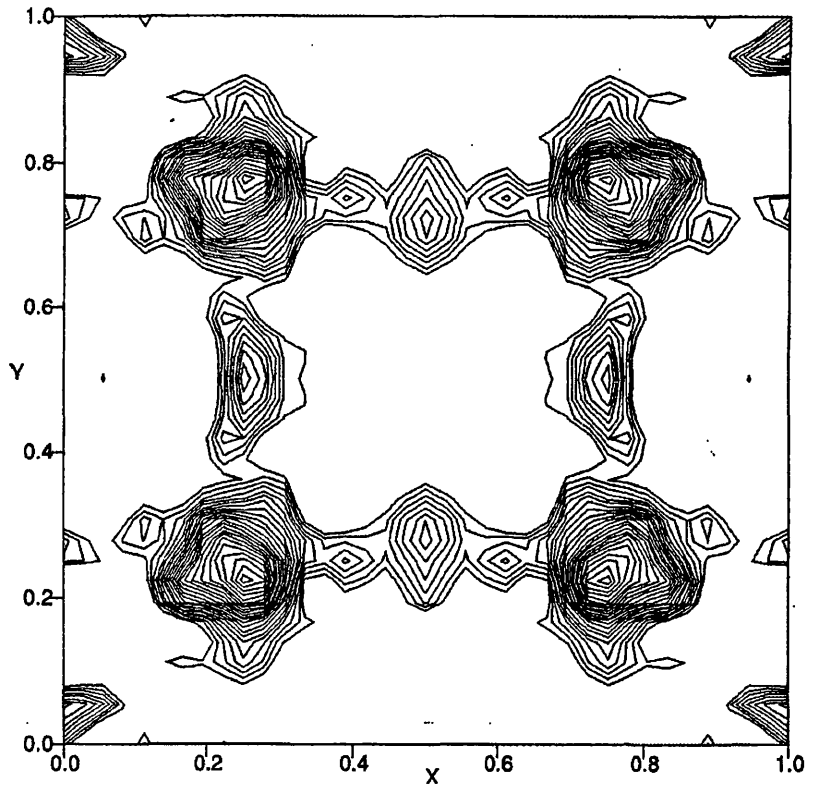


Figure 8.3. Final map obtained from MICE.

It has to be noted, that at no point during this electron crystallographic study was reference made to any model. The procedures involved here were entirely model-free and phase determination was achieved solely by ab initio means. Taking this background into consideration along with the sparseness of the data set involved, phase determination by maximum entropy means is shown to be viable even with poor electron diffraction datasets. As mentioned previous structural studies of

$(AlPcF)_n$ had reached resolution of the molecular columns, it hoped that this study has shown that it is possible to increase that resolution by using electron diffraction data combined with maximum entropy methods. Observation of the molecular quatrefoil of the phthalocyanine polymer has been realised, thus giving credence to structural investigations by these methods from what are regarded as being very sparse data.

8.4. Use of Image Derived Phases in Maximum Entropy Methods.

Previously in this thesis, descriptions have been given of structural analyses that have employed phase information derived from electron microscope images. This phase information is combined with electron diffraction data to achieve enhanced phase extension.

Use was made of previously obtained high resolution electron microscopy images of $(AlPcF)_n$. CRISP software was then used to obtain the Fourier transform of images, which in turn allowed the retrieval of phase information. Combination of the previously studied electron diffraction data with this phase information leads to the collection of structural information through use of maximum entropy methods.

A total of six phases were obtained from Fourier transform studies of images of $(AlPcF)_n$. Table 8.5. lists the relevant reflections along with their *h, k* values and associated *U* magnitudes.

Reflection No.	h	k	$\|U_h\|^{obs}$	Phase
1	0	2	0.567	180
2	2	0	0.567	-180
5	2	2	0.308	-180
6	1	2	0.220	0
7	2	1	0.220	0
43	1	1	0.125	0

Table 8.5. Phased reflections.

All intensity information is the same as was utilised in the previous ab initio calculations and as such, reference is made to Table 8.1. for a complete listing of *h, k* values, unitary structure factors $|U_h|^{obs}$ and resolution, *d* in Å.

The diffraction data were normalised using MITHRIL[3,4] to give unitary structure magnitudes and their standard deviations using electron scattering factors. The phases listed above and their associated *U*-magnitudes were used to form the initial basis set {**H**}. The root node of the phasing tree was generated and subjected to constrained entropy maximisation. Using a selection algorithm, seven further reflections were chosen to be added to this basis set. The reflections chosen were numbered 14, 15, 29, 30, 36, 42 and 58. As their phases were as yet unknown, all possible combinations of 0° and 180° were used. This led to the generation of 128 new nodes (number 2-129) on the phasing tree. As before, each node was subjected to constrained entropy maximisation. The first level of the phasing tree is summarised in Table 8.6.

Analysis of nodes generated was carried out by procedures that have previously been detailed. This employed student *t*-test and nodes were ranked according to likelihood values. Identification of the top eight nodes from this level was followed by generation of centroid maps. Node 94 was found to have the highest likelihood value of 3.381. Examination of all centroid maps found similar structural features. Again, the phthalocyanine molecular motif was obvious and Figure 8.4. shows a representative centroid map of a first level node of the phasing tree.

Node	Entropy	LLG
1	-0.665	0.00
2	-1.139	1.68
3	-1.179	0.59
4	-1.165	1.12
5	-1.181	0.60
6	-1.205	1.59
7	-1.276	0.57
8	-1.122	1.43
9	-1.122	0.71
10	-1.121	2.08
11	-1.229	0.45
12	-1.142	1.76
13	-1.225	0.83
14	-0.996	2.27
15	-1.199	1.20
16	-1.077	1.89
17	-1.256	0.85
18	-1.109	2.75
19	-1.170	1.86
20	-1.205	0.85
21	-1.227	0.03
22	-1.069	2.98
23	-1.193	2.01
24	-1.181	1.25
25	-1.248	0.49
26	-1.134	2.81
27	-1.202	1.79
28	-1.290	0.62
29	-1.261	0.09
30	-0.993	3.01
31	-1.201	2.22
32	-1.203	1.27
33	-1.279	0.55
34	-1.143	1.51
35	-1.185	-0.24
36	-1.181	1.35
37	-1.224	-0.97
38	-1.208	1.20
39	-1.241	0.02
40	-1.090	1.67
41	-1.232	-0.50
42	-1.082	2.16
43	-1.065	0.98
44	-1.084	2.16
45	-1.162	0.41
46	-0.971	2.21
47	-1.051	1.27
48	-1.035	2.02
49	-1.188	0.32
50	-1.120	2.74
51	-1.176	1.21
52	-1.235	0.68
53	-1.272	-1.56
54	-1.075	2.76
55	-1.085	1.93
56	-1.129	1.52
57	-.1246	-0.61
58	-1.112	2.82
59	-1.060	1.93
60	-1.232	0.94
61	-1.303	-1.05
62	-0.969	2.82
63	-1.042	2.12
64	-1.143	1.51
65	-1.314	-0.66
66	-1.138	1.87
67	-1.204	0.75
68	-1.190	1.14
69	-1.247	0.48
70	-1.225	1.87
71	-1.284	0.63
72	-1.084	2.10
73	-1.161	1.41
74	-1.135	2.29
75	-1.115	1.52
76	-1.116	2.05
77	-1.132	1.56
78	-1.020	2.63
79	-1.047	1.98
80	-1.015	2.47
81	-1.109	1.95
82	-1.126	2.81
83	-1.217	1.62
84	-1.232	0.59
85	-1.290	-0.56
86	-1.113	3.39
87	-1.140	2.33
88	-1.090	1.75
89	-1.176	0.91
90	-1.119	3.04
91	-1.081	2.41
92	-1.212	1.10
93	-1.237	0.36
94	-1.001	3.38
95	-1.059	2.77
96	-1.093	1.90
97	-1.226	0.95
98	-1.142	1.41
99	-1.182	0.10
100	-1.197	1.14
101	-1.197	-0.51
102	-1.244	0.89
103	-1.296	-0.25
104	-1.089	1.82
105	-1.123	0.57
106	-1.249	1.56
107	-1.079	1.22
108	-1.175	1.96
109	-1.063	1.01
110	-1.036	2.37
111	-1.023	1.87
112	-1.017	2.47
113	-1.042	1.53
114	-1.133	2.42
115	-1.193	0.89
116	-1.255	0.38
117	-1.147	-0.61
118	-1.181	2.58
119	-1.136	1.83
120	-1.090	1.52
121	-1.099	0.37
122	-1.231	2.55
123	-1.037	2.06
124	-1.277	0.47
125	-1.131	-0.01
126	-1.020	3.07
127	-1.017	2.57
128	-1.085	1.72
129	-1.107	0.59

Table 8.6. First level of maximum entropy-likelihood phasing tree.

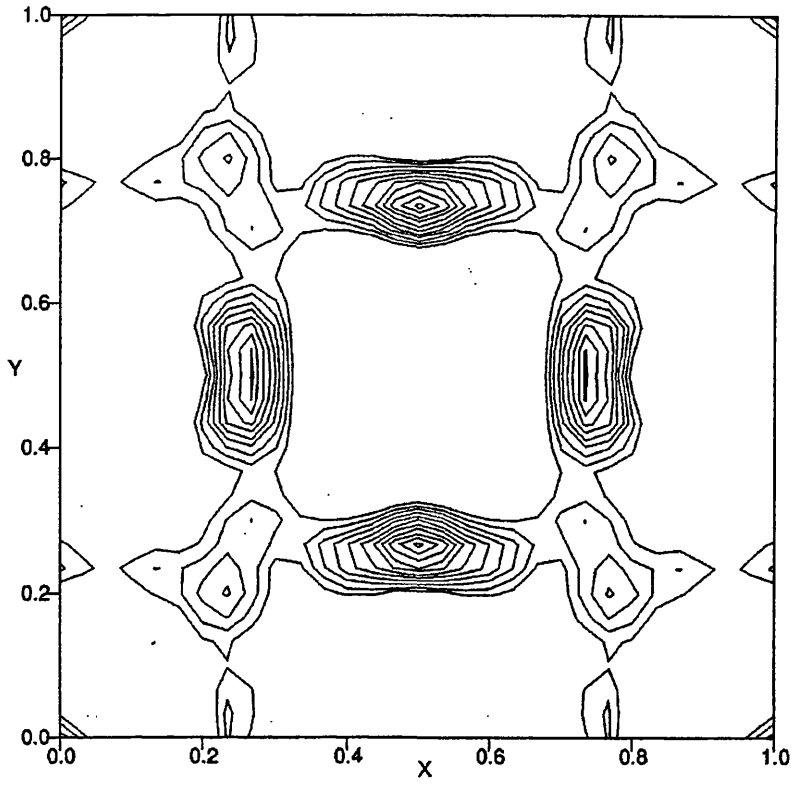


Figure 8.4. Representative centroid map from first level of phasing tree.

From initial inspection of centroid maps the phthalocyanine molecular shape is apparent, but no one solution had been identified as all top 8 nodes had likelihood values that were similar. Therefore, a second level of the phasing tree was generated. For the second level of the phasing tree the top four nodes were used and subsequently a further five reflections had phases permuted. This resulted in the generation of 128 nodes, allowing the second level to be from node 130 through to node 257. Nodes were subjected to entropy maximisation as before, followed by analysis and ranking using likelihood. Table 8.7. lists the top nodes from level 2, as derived from analysis.

Node	From node	Entropy	LLG
136	94	-1.134	3.97
149	94	-1.071	3.92
152	94	-1.099	3.95
153	94	-1.089	3.83
213	30	-1.127	4.09

Table 8.7. Top nodes from level two of phasing tree.

Inspection of the phasing tree hierarchy finds that nodes with the highest likelihood values are derived from two nodes, 94 and 30 respectively. Figure 8.5. depicts a centroid map obtained for node 136. Clarity of map has improved and map is found to contain more structural detail.

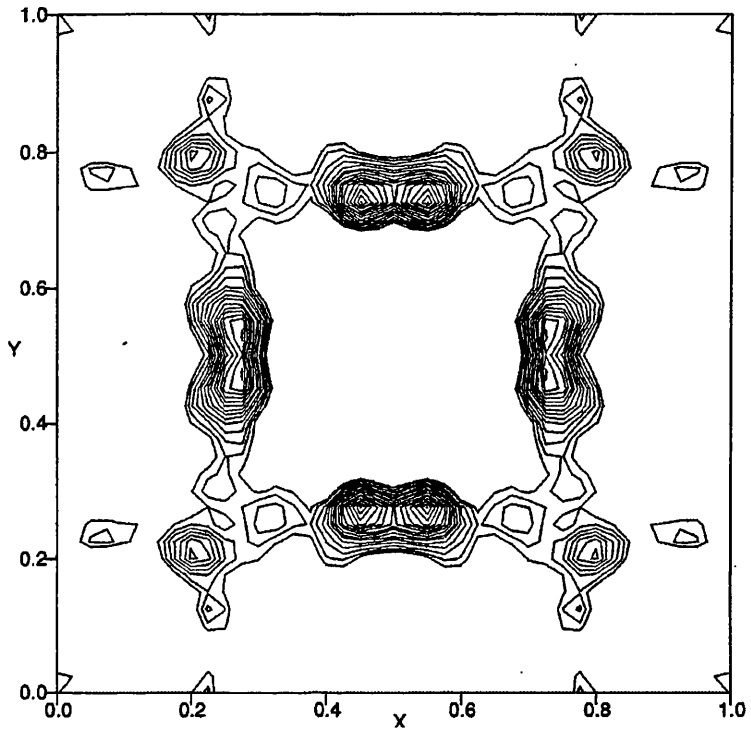


Figure 8.5. Centroid map of node 213.

Again, no unique solution could be determined and maps for all top nodes of this level were found to have little distinction between them, in terms of structural information. Thus, a final third level of the phasing tree was generated using nodes 136, 149, 152, 153 and 213.

This time, the reflections chosen to have their phases permuted were numbered 12, 17, 27, 31 and 51. Thus, level three of the tree ran from node 258 through to 417. All of these nodes were subjected to constrained entropy maximisation, as described previously. Subsequent analysis and scoring produced a list of nodes with the most correct phase choices, or highest likelihood values. Table 8.8. lists these nodes, along with their associated entropy and likelihood values.

Node	From node	Entropy	LLG
299	149	-1.331	6.18
395	213	-1.322	6.28
397	213	-1.325	6.10
399	213	-1.364	6.37
401	213	-1.353	6.19

Table 8.8. Highest nodes from level three of phasing tree.

Following established convention, centroid maps were generated and upon inspection node 399 was found to represent the best structure solution. Figure 8.6. shows the centroid map obtained for this node.

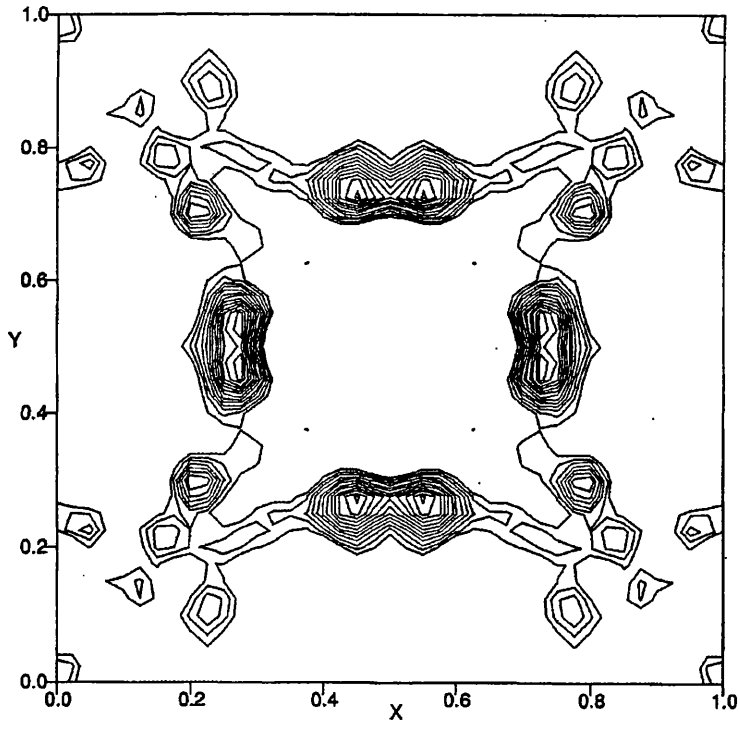


Figure 8.6. Final centroid map from MICE using image derived phases.

However, some spurious peaks are found to exist that are not apparent in structures determined by ab initio means. Further work involving assignation of more phases could perhaps obtain a clearer structure using image derived phases. However, this

does not negate the fact that in this study, the structural motif is apparent and phase determination using maximum entropy methods is still shown to be very robust.

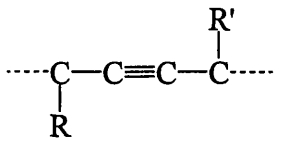
8.5. Conclusions.

Once again, when using sparse and what is regarded as poor quality electron diffraction data, phase determination by maximum entropy means overcomes these problems to provide structural information about previously unknown structures. The power of these methods cannot be underestimated and it is for the benefit of anyone undertaking such structural investigations to have recourse to such methodologies. It has been demonstrated here that phase determination by these procedures is possible and when carried out, structural information is obtainable via routine calculations.

9.0. ELECTRON CRYSTALLOGRAPHY OF POLY[1,6-DI(N-CARBAZOLYL)-2,4-HEXADIYNE].

9.1. Introduction.

Poly[1,6-di(N-carbazolyl)-2,4-hexadiyne] belongs to a class of polymers that are obtained via solid state polymerisation reactions of diacetylenes [208,209] giving rise to large and almost defect free monocrystalline polymers.



where R,R' are

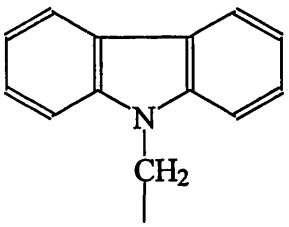


Figure 9.1. Structure of Poly[1,6-di(N-carbazolyl)-2,4-hexadiyne]

The material can be polymerised quantitatively by heat or γ radiation, with large changes in the lattice parameters between monomer and polymer.[210]

X-ray crystal structure analysis was carried out to provide a basis for interpreting the highly anisotropic properties of the polymer, along with its interesting mechanical, thermomechanical and photoconductive properties.[211] Crystals were found to be monoclinic, space group $P2_1/c$ with cell parameters a = 12.9, b = 4.91, c = 17.4 Å and β = 108.3°. The calculated density is 1.301 g cm^{-3} with 2 monomer units per unit cell.

This crystal structure determination showed that the polymer backbone has an alternating double-single-triple-single bonding pattern and is a planar conjugated system. It is to these results that the validation of structure obtained by maximum entropy methods has recourse.

Studies have also been carried out investigating radiation damage and the application of high resolution electron microscopy to polydiacetylene single crystals.[212] These studies used techniques previously described for examination of beam sensitive materials, thus facilitating the collection of both images and diffraction patterns. It is these images and electron diffraction patterns that will form the basis of this electron crystallographic study of poly[1,6-di(N-carbazolyl)-2,4-hexadiyne].

Previously in this thesis, the theoretical methodology employed in a electron crystallographic analysis of this kind has been greatly detailed. A description will now be given of the use of maximum entropy methods in the determination of the structure of poly[1,6-di(N-carbazolyl)-2,4-hexadiyne], both by ab initio phase determination and use of image derived phases.

9.2. Experimental Methods.

Electron diffraction intensities were quantified using the ELD system[68]. The electron diffraction pattern was transferred to PC using a frame grabber. Intensities were evaluated by the ELD software, as previously described in the section detailing intensity retrieval from electron diffraction patterns.

For the phases derived from images, again use was made of a frame grabber and this time, CRISP software[69]. The Fourier transform of the image provides low resolution phase information. Again, the theoretical basis of CRISP and retrieval of phases from images has been described elsewhere in this thesis.

9.3. Ab Initio Phase Determination.

The diffraction data obtained by the above methods were normalised using MITHRIL to give the unitary structure magnitudes ($|U_h|^{obs}$) and their associated standard deviations using electron scattering factors. 64 independent reflections remained after normalisation. The overall isotropic temperature factor *B* was calculated to be -2.279 $Å^2$. It is common to obtain negative temperature factors when using such incomplete datasets, as found with electron diffraction projection data. The plane group of this *ac* projection was determined to be p2gg. The U-magnitudes , their associated standard deviations and their resolution in Å are tabulated in Table 9.1.

No.	h	k	$\|U_h^{obs}\|$	d(Å)	No.	h	k	$\|U_h^{obs}\|$	d(Å)
1	0	2	0.453	6.12	33	4	12	0.108	0.93
2	2	0	0.436	8.26	34	3	6	0.105	1.74
3	0	4	0.294	3.06	35	6	5	0.104	1.60
4	4	0	0.258	4.13	36	0	12	0.104	1.02
5	2	2	0.247	4.31	37	1	6	0.102	1.95
6	2	1	0.216	6.03	38	4	8	0.101	1.31
7	3	2	0.205	3.57	39	4	4	0.100	2.16
8	1	11	0.178	1.09	40	5	2	0.099	2.59
9	7	9	0.170	1.05	41	9	3	0.099	1.51
10	9	7	0.163	1.10	42	12	4	0.093	1.13
11	3	10	0.155	1.12	43	4	2	0.092	3.01
12	0	6	0.148	2.04	44	2	10	0.091	1.16
13	5	10	0.148	1.05	45	1	5	0.090	2.32
14	3	9	0.147	1.23	46	4	5	0.090	1.87
15	11	1	0.144	1.44	47	5	3	0.089	2.25
16	5	6	0.143	1.53	48	0	10	0.088	1.22
17	1	9	0.139	1.32	49	5	5	0.088	1.73
18	10	5	0.130	1.21	50	7	4	0.088	1.64
19	1	3	0.129	3.70	51	5	4	0.086	1.96
20	1	4	0.127	2.85	52	12	0	0.080	1.38
21	4	7	0.127	1.46	53	0	8	0.079	1.53
22	2	9	0.125	1.28	54	2	3	0.078	3.27
23	6	1	0.124	2.52	55	8	2	0.076	1.79
24	6	0	0.120	2.75	56	1	1	0.071	8.63
25	3	5	0.119	2.02	57	8	4	0.068	1.51
26	10	3	0.117	1.39	58	10	0	0.067	1.65
27	2	5	0.115	2.17	59	2	6	0.067	1.85
28	4	1	0.113	3.59	60	1	2	0.065	5.23
29	9	2	0.111	1.62	61	6	2	0.064	2.26
30	2	8	0.110	1.43	62	10	2	0.064	1.48
31	6	3	0.110	2.01	63	8	0	0.061	2.07
32	9	1	0.108	1.74	64	3	1	0.060	4.52

Table 9.1. Reflection number, h, k, the unitary structure factor $|U_h|^{obs}$ and *d* in Å.

Subsequent maximum entropy calculations were carried out using the MICE computer program. Two suitable reflections were selected to define the origin and given phase angles of 0°. The reflections chosen were 6 (2,1) and 20 (1,4) with U magnitudes of 0.216 and 0.127 respectively. At this stage, it has to be emphasised that no recourse was made to image derived phases as this is a strict ab initio phase determination. These reflections thus defined the basis set {H}. Node 1, the root node, of the phasing tree was generated by carrying out a constrained entropy maximisation, in which the phases and amplitudes of the origin phases were used as constraints to produce a maximum entropy distribution $q^{ME}(\mathbf{x})$.

Seven strong reflections were selected and given permuted phases , thus generating 2^7=129 nodes. Each node was subjected to constrained entropy maximisation, as outlined previously and the log likelihood gain, LLG, computed for each. Analysis of these nodes was achieved through utilisation of t-tests, again described in the theoretical sections of this thesis. The phasing tree thus generated is summarised in Table 9.2.

The top 8 nodes were identified from analysis and subsequently centroid maps were produced for each of these. Node 30 was found to have the highest LLG of 3.461 and its centroid map is given in Figure 9.2. The remaining maps were very similar, with distinct areas of atomic presence identifiable.

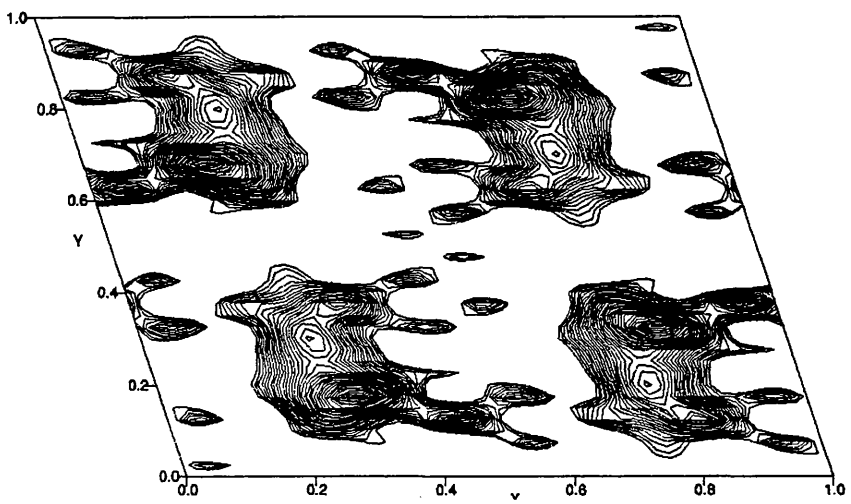


Figure 9.2. Level 1 Centroid map from MICE

Node	Entropy	LLG	Node	Entropy	LLG	Node	Entropy	LLG
1	-0.088	0.00	44	-0.488	1.10	87	-0.573	1.98
2	-0.642	1.00	45	-0.591	2.54	88	-0.482	2.40
3	-0.542	2.40	46	-0.539	2.82	89	-0.552	1.12
4	-0.538	1.46	47	-0.582	1.77	90	-0.549	2.08
5	-0.457	2.33	48	-0.494	2.16	91	-0.452	2.31
6	-0.584	2.71	49	-0.538	0.79	92	-0.562	1.46
7	-0.554	2.07	50	-0.533	1.37	93	-0.477	2.45
8	-0.516	2.63	51	-0.594	1.99	94	-0.561	2.85
9	-0.462	1.08	52	-0.433	1.08	95	-0.532	1.81
10	-0.539	1.33	53	-0.517	2.72	96	-0.552	2.46
11	-0.613	2.11	54	-0.547	2.87	97	-0.486	1.27
12	-0.470	1.14	55	-0.607	1.43	98	-0.562	2.24
13	-0.525	2.47	56	-0.479	2.36	99	-0.510	2.83
14	-0.542	3.21	57	-0.513	1.00	100	-0.521	1.18
15	-0.600	1.28	58	-0.600	1.11	101	-0.487	2.65
16	-0.434	2.23	59	-0.532	2.49	102	-0.555	2.99
17	-0.511	1.24	60	-0.512	1.49	103	-0.469	1.47
18	-0.538	1.26	61	-0.462	2.49	104	-0.541	2.35
19	-0.606	2.61	62	-0.655	2.74	105	-0.468	0.98
20	-0.510	1.05	63	-0.545	1.90	106	-0.472	1.59
21	-0.595	2.57	64	-0.516	2.50	107	-0.551	3.25
22	-0.494	2.77	65	-0.434	0.98	108	-0.474	1.22
23	-0.566	2.16	66	-0.634	1.76	109	-0.548	2.47
24	-0.502	2.47	67	-0.518	2.82	110	-0.496	2.59
25	-0.550	0.86	68	-0.507	1.70	111	-0.554	1.72
26	-0.578	1.59	69	-0.419	2.25	112	-0.469	2.20
27	-0.485	2.07	70	-0.566	2.19	113	-0.538	1.18
28	-0.574	1.06	71	-0.544	1.80	114	-0.520	2.01
29	-0.495	2.34	72	-0.502	2.63	115	-0.583	2.41
30	-0.576	3.46	73	-0.452	1.28	116	-0.421	1.18
31	-0.526	1.94	74	-0.511	1.77	117	-0.507	2.89
32	-0.533	2.38	75	-0.566	2.49	118	-0.510	2.65
33	-0.491	0.96	76	-0.449	1.43	119	-0.585	1.42
34	-0.558	1.57	77	-0.501	2.63	120	-0.440	2.30
35	-0.542	2.59	78	-0.519	2.61	121	-0.493	1.24
36	-0.528	1.16	79	-0.610	1.08	122	-0.569	1.37
37	-0.495	2.43	80	-0.420	2.11	123	-0.513	2.89
38	-0.571	3.05	81	-0.514	1.65	124	-0.506	1.72
39	-0.501	1.58	82	-0.512	1.83	125	-0.450	2.56
40	-0.563	2.27	83	-0.554	2.86	126	-0.594	2.46
41	-0.492	0.79	84	-0.485	1.32	127	-0.527	1.80
42	-0.467	1.12	85	-0.523	2.42	128	-0.493	2.51
43	-0.557	2.67	86	-0.472	2.22	129	-0.413	1.06

Table 9.2. The first level of the maximum entropy phasing tree.

A second level of the phasing tree was generated by taking the top 8 nodes from level 1 and permuting phases for 5 more reflections. This gave rise to 256 new nodes. Entropy maximisation and analysis using *t*-test was repeated. This time, from examination of results of analysis, it could be seen that nodes with highest likelihood were derived mostly from nodes 30 and 14. Highest nodes for level 2, along with their entropy are given in Table 9.3.

Node	From	Entropy	LLG
130	30	-0.762	4.82
194	14	-0.748	4.60
199	14	-0.743	4.58
247	38	-0.694	4.47
195	14	-0.664	4.46

Table 9.3. Top Nodes from Second Level of Phasing Tree.

Again, centroid maps were generated and this second level gives rise to clearer maps as shown for node 130 in Figure 9.3.

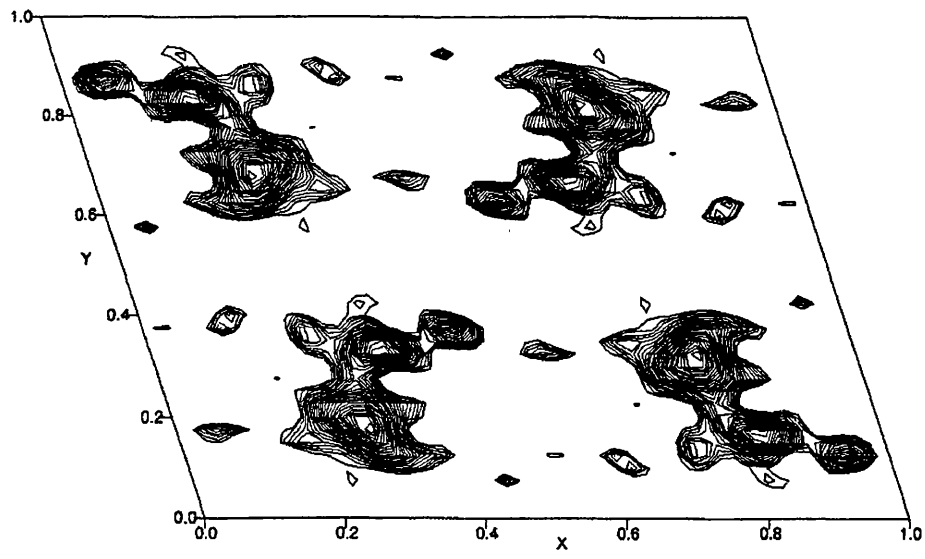


Figure 9.3. Node 130 from level 2 of phasing tree.

As no solution had become apparent from a single node of the phasing tree a third level was generated. This was achieved by following the procedure outlined above. Top five nodes from level two were used, these being 130, 194, 199, 247 and 195. Phases were permuted for 5 reflections, thus giving a third level of nodes from number 386 through to 545. Subsequent analysis and scoring of nodes produced listings of the top nodes. For the third level of the phasing tree the top 3 nodes are listed in Table 9.4. and are all seen to derive from node 195 of the previous level.

Node	From	Entropy	LLG
525	195	-0.766	4.83
543	195	-0.850	4.82
544	195	-0.835	4.88

Table 9.4. Level 3 of phasing tree from MICE.

As all of these solutions were derived from a single node, i.e. a single set of phase choices, no further permutation of phases or generation of subsequent levels of the phasing tree were generated. Therefore, in accordance with above procedure centroid maps were produced for the top node at this level, node 544. The centroid map for node 544 is given in Figure 9.4. As validation of the structure solution achieved through utilisation of maximum entropy methods, comparison can be made with the structure solution resulting from X-ray analysis. It has to be emphasised that the maps generated by the maximum entropy procedure in no way consulted this information. Figure 9.5. shows the same *ac* projection that is being studied here, as retrieved from the Cambridge Structural Database and subsequently manipulated using CERIUS[2]. The correspondence with the determined crystal structure is immediately apparent.

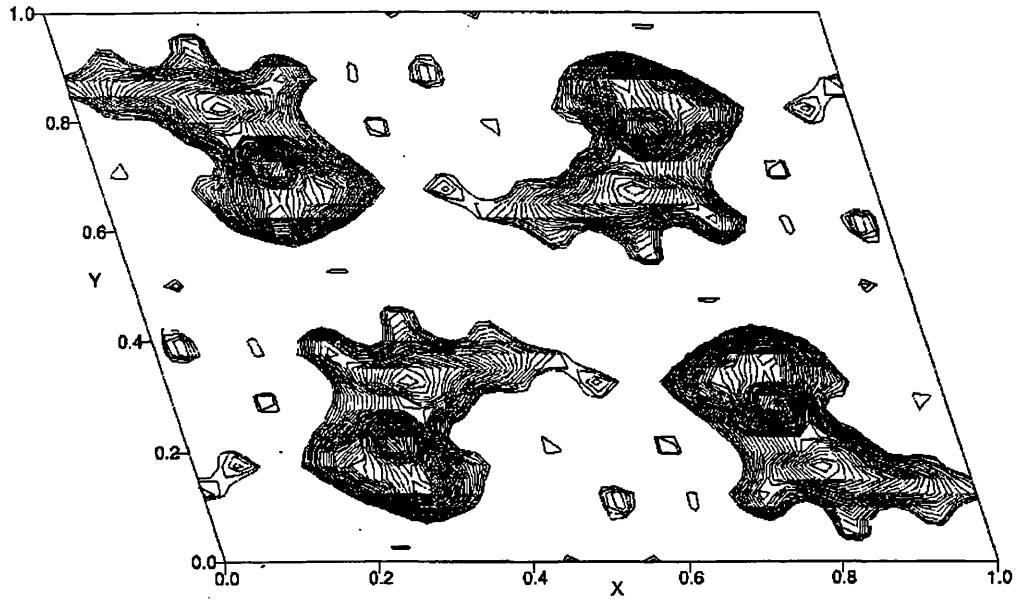


Figure 9.4. Final map obtained from MICE.

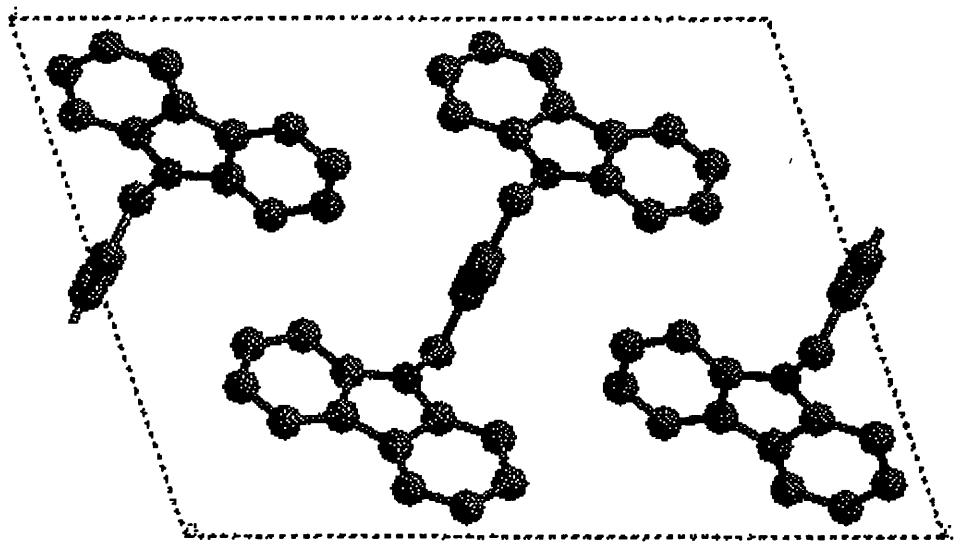


Figure 9.5. *ac* projection from solved X-ray structure.

9.4. Use of Image Derived Phases in Maximum Entropy Methods.

In chapter 3, the methodology of electron crystallographic structure determination was described and examples were given of structure determinations that had employed phase information derived from electron microscope images. In an analogous manner, phase information was derived from images of single crystals of poly[1,6-di(N-carbazolyl)-2,4-hexadiyne]. This phase information was combined with the previously used diffraction intensities in entropy maximisation procedures and tree directed phase search techniques to furnish structural information.

A total of 9 phases were available from Fourier transform studies of images. Table 9.5. lists the relevant reflections, along with their *h, k* values and associated *U* magnitudes.

No.	h	k	$\|U_h^{obs}\|$	Phase
1	0	2	0.453	0
2	2	0	0.436	0
5	2	2	0.247	180
19	1	3	0.129	0
37	1	6	0.102	180
45	1	5	0.091	180
54	2	3	0.078	180
56	1	1	0.072	0
64	3	1	0.060	0

Table 9.5. Phased reflections.

All intensity information is the same as was utilised in the ab initio calculations described previously and reference is made to Table 9.1. for a complete listing of *h, k,* unitary structure factors $|U_h^{obs}|$ and resolution, *d* in Å.

The diffraction data were normalised using MITHRIL[8,9] to give unitary structure magnitudes and their standard deviations using electron scattering factors. The phases listed above and their associated *U*-magnitudes were used to form the initial basis set {**H**}. The root node of the phasing tree was generated and subjected to constrained entropy maximisation. Using a selection algorithm, seven reflections were chosen to

be added to this basis set with serial numbers 10, 11, 12, 13, 24, 28 and 52. As their phases were unknown, all possible combinations of 0° and 180° were used, thus generating 128 new nodes (number 2-129) on the phasing tree. Each node was subjected to constrained entropy maximisation as before. The phasing tree for the first level of the calculation is summarised in Table 9.6.

Node	Entropy	LLG	Node	Entropy	LLG	Node	Entropy	LLG
1	-0.740	0.98	44	-0.738	1.58	87	-0.752	3.33
2	-0.741	1.93	45	-0.682	1.57	88	-0.797	2.47
3	-0.709	1.94	46	-0.833	2.00	89	-0.803	2.63
4	-0.764	1.69	47	-0.848	1.73	90	-0.776	2.34
5	-0.714	1.69	48	-0.762	2.36	91	-0.777	1.96
6	-0.717	2.67	49	-0.772	2.22	92	-0.674	2.55
7	-0.768	2.64	50	-0.713	2.85	93	-0.692	2.25
8	-0.798	2.30	51	-0.739	2.48	94	-0.837	2.81
9	-0.814	2.07	52	-0.763	2.24	95	-0.787	2.54
10	-0.827	1.90	53	-0.781	2.07	96	-0.754	2.99
11	-0.743	1.62	54	-0.744	3.35	97	-0.748	2.70
12	-0.699	2.07	55	-0.761	3.08	98	-0.725	2.46
13	-0.648	1.62	56	-0.810	2.94	99	-0.748	1.72
14	-0.820	2.32	57	-0.824	2.84	100	-0.789	1.61
15	-0.821	1.86	58	-0.787	2.17	101	-0.782	1.26
16	-0.752	2.47	59	-0.780	2.02	102	-0.751	2.93
17	-0.748	2.03	60	-0.705	2.21	103	-0.773	2.43
18	-0.719	2.56	61	-0.709	2.24	104	-0.829	2.26
19	-0.737	2.67	62	-0.823	2.64	105	-0.839	1.91
20	-0.731	2.22	63	-0.817	2.47	106	-0.818	1.69
21	-0.758	2.29	64	-0.749	2.99	107	-0.784	1.49
22	-0.737	3.31	65	-0.753	3.02	108	-0.724	1.68
23	-0.756	3.45	66	-0.736	1.92	109	-0.712	1.56
24	-0.795	2.81	67	-0.723	1.92	110	-0.849	2.05
25	-0.809	2.89	68	-0.735	1.64	111	-0.850	1.78
26	-0.794	2.51	69	-0.728	1.62	112	-0.771	2.44
27	-0.768	2.19	70	-0.733	2.72	113	-0.775	2.33
28	-0.690	2.70	71	-0.760	2.68	114	-0.710	2.81
29	-0.678	2.36	72	-0.798	2.17	115	-0.754	2.28
30	-0.828	2.95	73	-0.809	2.03	116	-0.751	2.04
31	-0.791	2.65	74	-0.804	1.92	117	-0.804	1.66
32	-0.752	3.11	75	-0.754	1.56	118	-0.755	3.34
33	-0.747	2.79	76	-0.688	2.10	119	-0.757	2.96
34	-0.737	2.26	77	-0.668	1.66	120	-0.819	2.65
35	-0.732	1.85	78	-0.842	2.32	121	-0.826	2.52
36	-0.811	1.58	79	-0.808	1.91	122	-0.769	2.12
37	-0.752	1.52	80	-0.755	2.52	123	-0.804	1.80
38	-0.738	2.78	81	-0.750	2.09	124	-0.703	2.10
39	-0.778	2.40	82	-0.718	2.36	125	-0.733	2.08
40	-0.826	2.33	83	-0.753	2.46	126	-0.837	2.56
41	-0.839	2.03	84	-0.709	1.99	127	-0.814	2.40
42	-0.833	1.56	85	-0.772	2.00	128	-0.752	2.91
43	-0.760	1.57	86	-0.749	3.19	129	-0.757	2.96

Table 9.6. First level of maximum entropy-likelihood phasing tree.

Analysis of nodes generated was carried out, employing student *t*-test and nodes were ranked according to likelihood values. The top eight nodes from this initial level of the phasing tree were identified and centroid maps generated. Node 23 had the highest LLG, with a value of 3.45 and the centroid map for this node is given in figure 9.6.

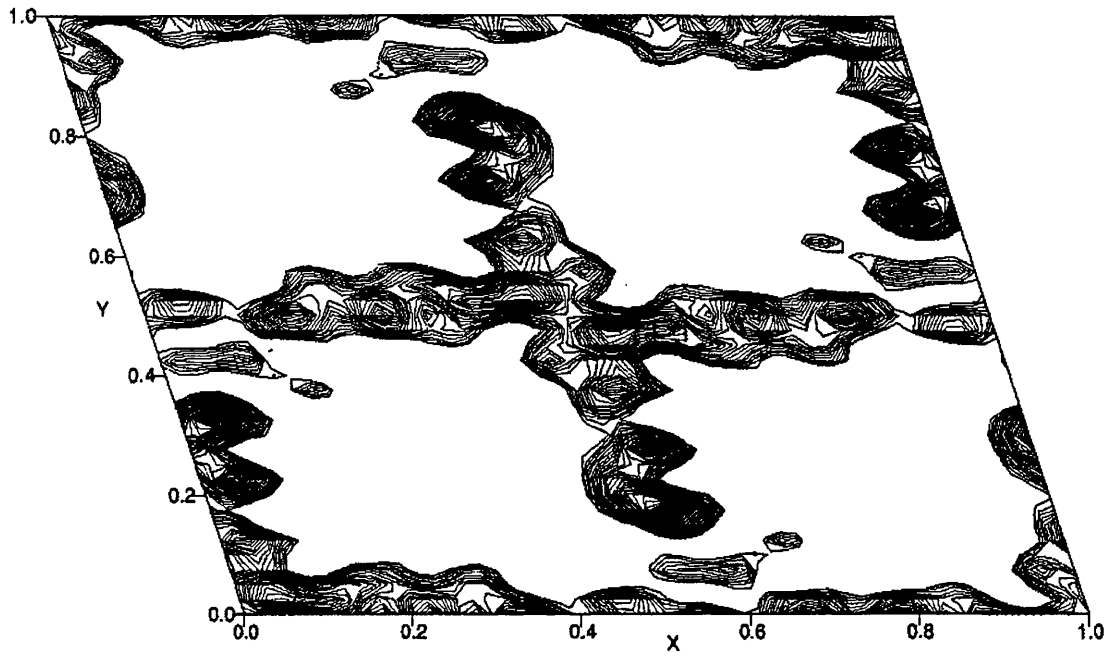


Figure 9.6. Centroid map from first level of phasing tree.

From initial inspection, it is immediately apparent that in comparison with centroid maps obtained by ab initio means areas of electron density are in reverse. Within MICE it was possible to construct Babinet maps. This was carried out and inspection of resulting maps does indeed have areas of density at the correct location, but also maps have less spurious detail as expected. Figure 9.7. gives an example of the babinet maps obtained for node 23 above. Centroid maps generated for the top eight nodes were all found to be similar to that of node 23. These eight nodes were kept and used to generate a second level of the phasing tree.

Further reflections were chosen to have permuted phases. These were reflections 14, 18, 20, 21 and 23. This gave rise to a second level of nodes from number 130 to 321.

Nodes were subjected to constrained entropy maximisation as before, followed by analysis and ranking using likelihood.

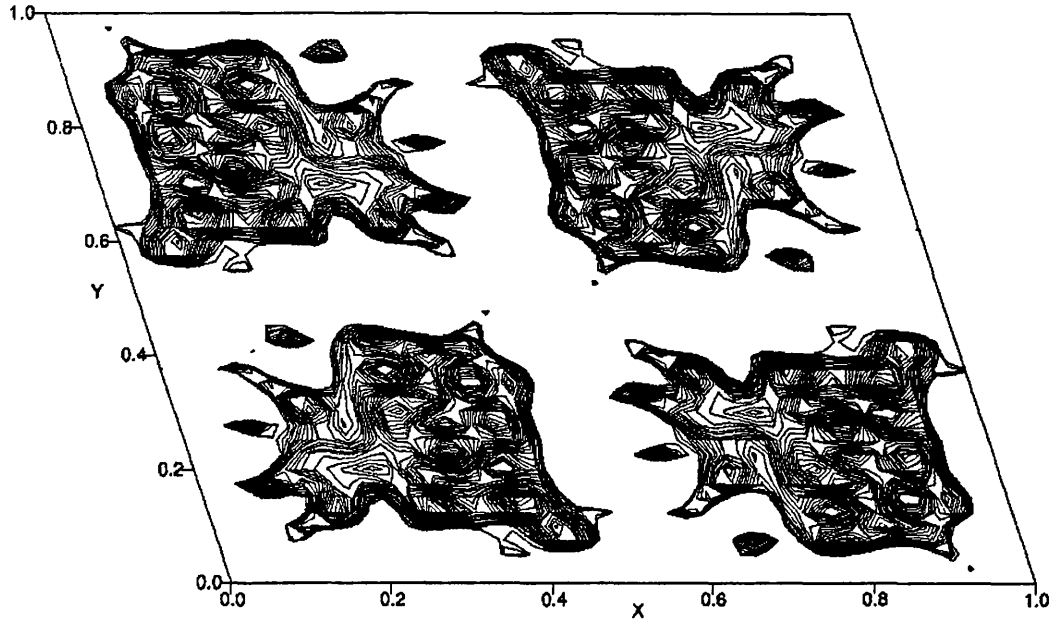


Figure 9.6. Babinet map from first level of phasing tree.

Table 9.7. lists the top nodes from analysis for level 2, with node 214 having the highest likelihood of 4.31. Inspection of the phasing tree hierarchy finds that nodes with the highest likelihood for level two are derived from nodes 54 and 118 alone.

Node	From	Entropy	LLG
166	54	-1.030	4.12
182	54	-0.934	4.16
214	118	-0.953	4.31
216	118	-0.939	4.10

Table 9.7. Top nodes from level two of phasing tree.

Figure 9.8. illustrates the centroid map obtained from node 214. Clarity of the map has improved and highest peaks are easier resolvable.

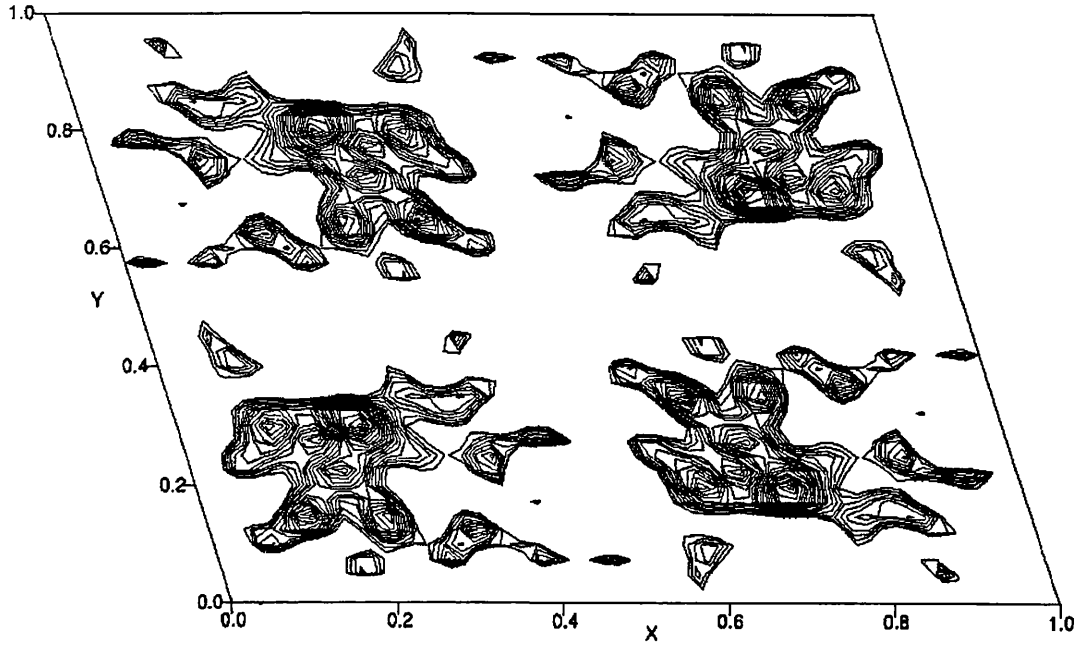


Figure 9.8. Centroid map of node 214.

As no unique solution could be associated with a single node on the phasing tree, a further third level was generated using nodes 166, 182, 214 and 216 from level two.

Reflections selected to have their phases permuted were 15, 26, 32, 38 and 50. Thus, level three of the phasing tree ran from node 322 to 449. All of these nodes had their entropy maximised under constraints and subsequent analysis and scoring produced a list of the nodes with highest likelihood, or more correct phase choices. Table 9.8. lists these nodes, along with their associated entropy and likelihood values.

Node	From node	Entropy	LLG
350	214	-1.019	5.10
402	166	-1.126	5.04
410	166	-1.132	5.25
411	166	-1.113	5.04
412	166	-1.135	5.05

Table 9.8. Highest nodes from level three of phasing tree.

Following with the convention described, centroid maps were generated and upon inspection node 410 represented the best solution to the structure. As with phase determination by ab initio means, comparison is made with the recognised X-ray solution. There is a marked similarity between the two structures, when viewed in the same orientation. This is represented graphically in Figures 9.9. and 9.10. which are the obtained centroid map and solved X-ray structure, respectively.

9.5. Conclusions

It has been shown that there is much to be gained from attempting maximum entropy phase determination by ab initio means. The methodology and procedure is rigorous and is stable even with very limited data as seen here, when basis set contained only 65 reflections. Comparison with X-ray structure is very agreeable.

Furthermore, if phase information can be retrieved from even low resolution images, this is of tremendous aid in the subsequent maximum entropy calculations. These phases can be though of as helping to direct the search through the phasing tree with greater ease, thus arriving at a more probable structure solution.

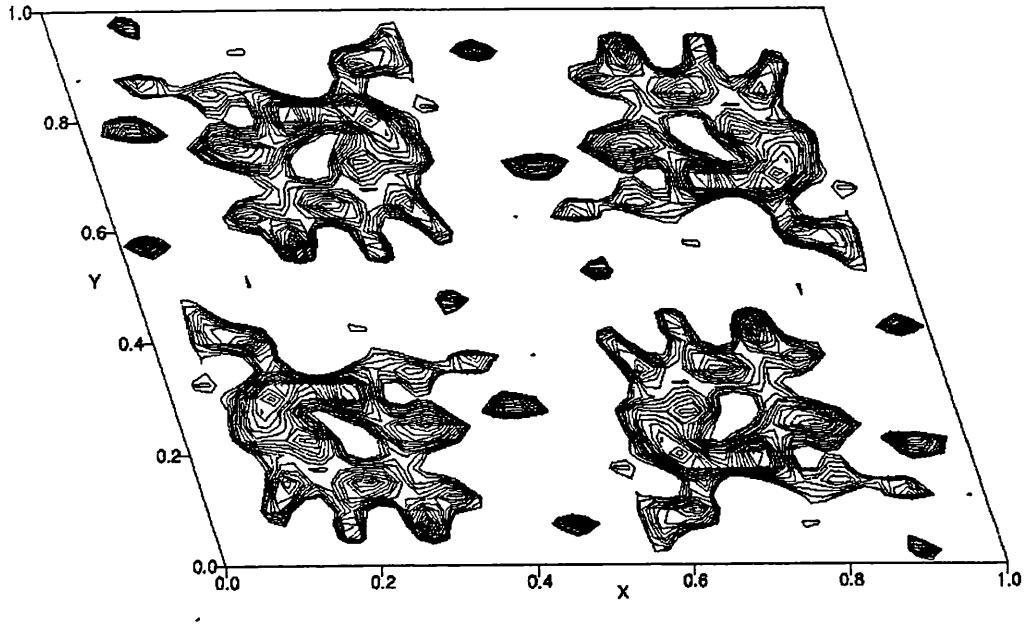


Figure 9.9. Final centroid map from MICE.

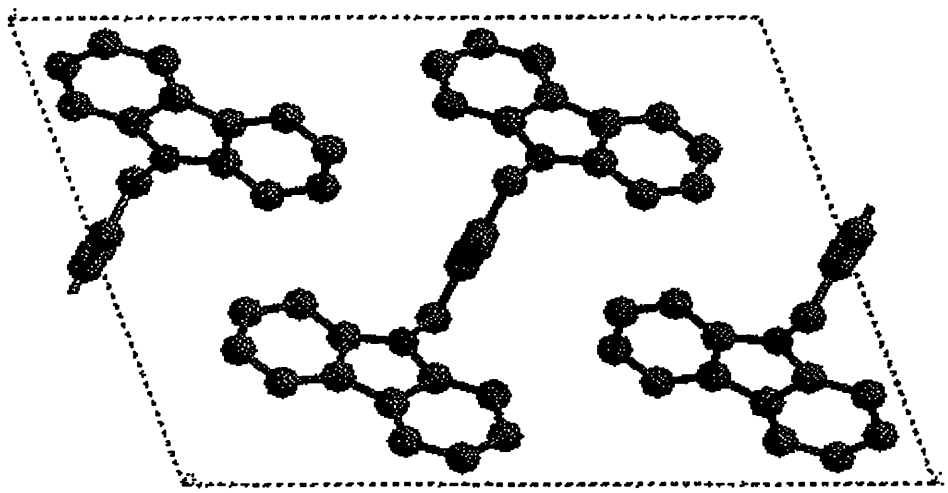


Figure 9.10. Solved X-ray structure for comparison.

10.0. CONCLUDING REMARKS.

The important thing in science is not so much to obtain new facts, as to discover new ways of thinking about them.

Sir W.L. Bragg.

Throughout this thesis it has been illustrated that studies employing electron crystallographic techniques combined with maximum entropy methods can, and do provide structural information about materials that conventional X-ray crystallographic techniques have failed. Therefore, by developing new methodology and techniques in electron crystallography, while having recourse to some of the previously developed theories of X-ray crystallography, it is now possible to think about structure determination of materials in a new way. In conclusion, it can be stated that electron crystallography has established itself as a new and viable route to structure elucidation, which is the aim of all crystallographic studies.

REFERENCES

1. Perkin, W.H., (1896). *Journal of the Chemical Society*, **69**, 596-637.
2. Fytelson, M. (1982) *Kirk-Othmer Encyclopaedia of Chemical Technology*, 3rd edition, J.Wiley and Sons, New York, pp838-870.
3. Zollinger, H.. (1991). *Color Chemistry*. VCH Publishers, New York.
4. Patai, S. (1978). *The Chemistry of Diazonium and Diazo groups*,. J. Wiley and Sons, New York, pp. 511-591
5. Challis, B.C., Ridd, J.H.(1968) *Journal of the Chemical Society*, 5197-5208.
6. March, J. (1992). *Advanced Organic Chemistry*, J.Wiley and Sons, New York, p636.
7. Hughes, E.D., Ingold, C.K., and Ridd, J.H. (1958) *Journal of the Chemical Society*, 58-65, 65-69, 77-82, 82-88.
8. Duckett, G.R.(1987) Ph.D. Thesis, University of Glasgow.
9. Miller, G.G.(1992) Ph.D. Thesis, University of Glasgow.
10. Connell, G. Unpublished work.
11. De Broglie, L. (1924) *Philos. Mag.* **47**, 476.
12. Planck, M. (1890) *Ann. Phys. Chem.* **39**, 161.
13. Einstein, A. (1905) *Ann. Phys.* **17**, 549-560.
14. Davisson C.J. and Germer, L.H. (1927) *Nature* (London) **119,** 558-560.
15. Thomson, G.P and Reid, A. (1927) *Nature (London)* **119**, 890-895.
16. Busch, H. (1926) *Ann.Physik*, **81**, 974-993.
17. Busch, H. (1927) *Arch. Elektrotech*, **18**, 583-594.
18. Ruska, E. (1934) *Z. Physik*, **87**, 580-602.
19. Hirsch, P.B., Howie, A., Nicholson, R.B., Pashley, D.W. and Whelan, M.J. (1971) *Electron Microscopy of thin crystals*, Butterworth, London.
20. Buseck, P., Cowley, J. and Eyring, L. eds.,(1988) *High Resolution Transmission Electron Microscopy and Associated Techniques*, New York, Oxford University Press.
21. Bragg, W.H. and Bragg, W.L. (1913) *Proc. R. Soc. London.* **A88,** 428-438.
22. Vainshtein, B.K. (1994) *Fundamentals of Crystals*, Springer-Verlag, New York., p339.

23. Mott, N.F. (1930) *Proc. Roy. Soc. Lond.* **A127**, 658.
24. *International Tables for Crystallography Vol C* (1992) Kluwer Academic Publishers, Dordrecht
25. Doyle and Turner, (1968) *Acta Crystallogr.* **A24**, 390-397.
26. Jap, B.K. and Glaeser, R.M. (1980) *Acta Crystallogr.* **A36**, 57-67.
27. Bethe, H.A. (1928). *Ann. Phys.* **87**, 55-129.
28. MacGillavry, C.H. (1940) *Physica* **7**, 329-343.
29. Dorset, D.L. (1995) *Structural Electron Crystallography*, Plenum Press, New York.
30. Cowley, J.M. and Moodie, A.F. (1957) *Acta Crystallogr.* **10**, 609-619.
31. Howie, A. and Whelan, M.J. (1961) *) Proc. R. Soc. London, Ser. A* **263**, 217.
32. Sturkey, L. (1957) *Acta Crystallogr.* **10**, 858-859.
33. Fujiwara, K. (1959) *J. Phys. Soc. Japan* **14**, 1513-1524.
34. Sturkey, L. (1962) *Proc. R. Soc. London* **80**, 321.
35. Dorset, D.L. (1983) *Polymer* **24**, 291-294.
36. Fryer, J.R. (1980) *Journal of Microscopy* **120**, 1-14.
37. Fryer, J.R. and Holland, F. (1984) *Proc. R. Soc. London* **A393**, 353-369.
38. Fryer, J.R. and Holland, F. (1983) *Ultramicroscopy* **11**, 67-70.
39. Fryer, J.R., McConnell, C.H., Zemlin, F. and Dorset, D.L. (1992) *Ultramicroscopy* **40**, 163-169.
40. Murata, Y., Fryer, J.R., Baird, T. and Murata, H. (1977) *Acta Crystallogr.* **A33**, 198-200.
41. Fryer, J.R. (1987) *Ultramicroscopy* **23**, 321-328.
42. International Study Group (1986) *J. Microsc.* **141**, 385-391.
43. Salih, S.M. and Cosslett, V.E. (1975) *J. Microsc.* **105**, 269-276.
44. Zemlin, F., Beckmann, E., Reuber, E., Zeitler, E. and Dorset, D.L. (1985) *Science* **229**, 461-462.
45. Siegel, G. (1972) *Naturforschung* **279**, 325-332.
46. Salih, S.M. and Cosslett, V.E. (1974) *Phil. Mag.* **30**, 225-228.
47. Holland, F., Fryer, J.R. and Baird, T. (1983) *Inst. Phys. Conf. Ser* **68**, 19-22.
48. Fryer, J.R. and Smith, D.J. (1982) *Proc. R. Soc. London* **A381**, 225-240.
49. Fryer, J.R. (1978) *Acta Crystallogr.* **A34**, 603-607.

50. Glaeser, R.M. (1985) *Ann. Rev. Phys. Chem.* **36**, 243-275.

51. DeRosier, D.J. and Klug. A. (1985) *Nature* **217**, 130-134.

52. Henderson, R. and Unwin, P.N.T. (1975) *Nature* **257**, 28-32.

53. Nagales, E., Wolf, S.G. and Downing, K.H. (1997) *Journal of Structural Biology* **118**, 119-127.

54. Walz, L., Hirai, T., Murata, K., Heymann, J.B., Mitsuoka, K., Fujiyoshi, Y., Smith, B.L., Agre, R. and Engel, A. (1997) *Nature* **387**, 624-627.

55. Rigamonti, R. (1936) *Gazz. Chim. Ital.* **66**, 174-182.

56. Vainshtein, B.K. (1955) *Zh. Fiz. Khim.* **29**, 327-344.

57. Lobachev, A.N. and Vainshtein, B.K. (1961) *Sov. Phys. Cryst.* **6**, 313-317.

58. Dvoryankin, V.F. and Vainshtein, B.K. (1960) *Sov. Phys. Cryst* **5**, 564-574.

59. Dvoryankin, V.F. and Vainshtein, B.K. (1962) *Sov. Phys. Cryst* **6**, 765-772.

60. Vainshtein, B.K., D'Yakon, I.A. and Ablov, A.V. (1971) *Sov. Phys. Dokl.* **15**, 645-647.

61. D'Yakon, I.A., Kairyak, L.N., Ablov, A.V. and Chapurina, L.F. (1977) *Dokl. Akad. Nauk SSSR* **236**, 103-105.

62. Cowley, J.M. (1953) *Acta Crystallogr.* **6**, 846-853.

63. Cowley, J.M. (1956) *Acta Crystallogr.* **9**, 391-395.

64. Lipson, H. and Cochran, W. (1966) *The Determination of Crystal Structures*, revised and enlarged ed., Cornell University Press, Ithaca, pp377-381.

65. Dorset, D.L. (1996) *Acta Crystallogr.* **B52**, 753-769.

66. Dorset, D.L. (1991) *Ultramicroscopy* **38**, 23-40.

67. Baldwin, J. and Henderson, R. (1984) *Ultramicroscopy* **14**, 319-336.

68. Zou, X.D., Sukharev, Y. and Hovmöller, S. (1993) *Ultramicroscopy* **49**, 147-158.

69. Hovmöller, S. (1992) *Ultramicroscopy* **41**, 121-135.

70. Cowley, J.M., ed. (1992) Electron Diffraction Techniques (IUCr Monograph on Crystallography 3), Vols. I and II, Oxford University Press, Oxford.

71. Henderson, R., Baldwin, J.M., Ceska, T., Zemlin, F., Beckmann, E. and Downing, K.H. (1990) *J. Mol. Biol.* **213**, 899-929.

72. Wenk, H.R., Downing, K.H., Meisheng, H. and O'Keefe, M.A. (1992) *Acta Crystallogr.* **A48**, 700-716.

73. Dong, W., Baird, T., Fryer, J.R., Gilmore, C.J. MacNicol, D.D., Bricogne, G., Smith, D.J., O'Keefe, M.A. and Hovmöller, S. (1992) *Nature* **355**, 605-609.

74. Dorset, D.L., McCourt, M.P., Fryer, J.R., Tivol, W.F. and Turner, J.N. (1994) *Micros. Soc. Am. Bull.* **24**, 398-404.

75. Dorset, D.L., Kopp, S., Fryer, J.R. and Tivol, W.F.(1995) *Ultramicroscopy* **57**, 59-89.

76. Wilson, A.J.C. (1950) *Acta Crystallogr.* **3**, 397-398.

77. Kitaigorodskii, A.I. (1973) *Molecular Crystals and Molecules*, New York, Academic Press.

78. Brisse, F. (1989) *J. Electron Micro. Tech.* **11**, 272-279.

79. Perez, S. and Chanzy, H. (1989) *J. Electron Micro. Tech.* **11**, 280-295.

80. Voigt-Martin, I.G., Yan, D.H., Wortmann, R. and Elich, K. (1995) *Ultramicroscopy*, **57**, 29-43.

81. Voigt-Martin, I.G., Yan, D.H., Yakimansky, A., Schollmeyer, D., Gilmore, C.J. and Bricogne, G. (1995) *Acta Crystallogr.* **A51**, 849-868.

82. Buerger, M.J. (1959) *Vector Space and its Applications in Crystal Structure Investigation*, Wiley, New York.

83. Dorset, D.L. (1976) *Biochim. Biophys. Acta*, **424**, 396-403.

84. Dorset, D.L. (1987) *Biochim. Biophys. Acta*, **898**, 121-128.

85. Woolfson, M.M. (1987) *Acta Crystallogr.* **A43**, 593-612.

86. Harker, D. and Kasper, J.S. (1948) *Acta Crystallogr.* **1**, 70-75.

87. Karle, J. and Hauptman, H.A. (1950) *Acta Crystallogr.* **3**, 181-187.

88. Sayre, D. (1952) *Acta Crystallogr.* **5**, 60-65.

89. Cochran, W. (1952) *Acta Crystallogr.* **5**, 65-67.

90. Zachariasen, W.H. (1952) *Acta Crystallogr.* **5**, 68-73.

91. Hauptman, H.A. and Karle, J. (1953) *The Solution of the Phase Problem. I. The Centrosymmetric Crystal*, ACA Monograph, No. 3. Polycrystal Book Service, New York.

92. Cochran, W. (1952) *Acta Crystallogr.* **8**, 473-478.

93. Wilson, A.J.C. (1942) *Nature* **150**, 151-152.

94. Karle, J. and Haputman, H.A. (1956) *Acta Crystallogr.* **9**, 635-651.

95. Sheldrick, G.M. (1990) *Acta Crystallogr.* **A46**, 467-473

96. White, P.S. and Woolfson, M.M. (1975) *Acta Crystallogr.* **A31**, 53-56.

97. Main, P., Fiske, S.J., Hull, S.E., Lessinger, L., Germain, G., Declercq, J.P. and Woolfson, M.M. (1980) *MULTAN80: A system of computer programs for the automatic solution of crystal structures from x-ray diffraction data.* Universities of York and Louvain.

98. Giacovazzo, C. (1980) *Direct Methods in Crystallography*, Academic Press, London.

99. Giacovazzo, C. [ed.] (1992) *Fundamentals of Crystallography.* Oxford University Press, Oxford.

100. Bricogne, G. (1984) *Acta Crystallogr.* **A40**, 410-445.

101. Cramér, H. (1946) *Mathematical Methods of Statistics*, Princeton University Press, Princeton.

102. Klug, A. (1958) *Acta Crystallogr.* **A46**, 515-543.

103. Bricogne, G. (1991) *Crystallographic Computing 4: Techniques and New Technologies*, edited by N.W. Isaacs and M.R. Taylor, Clarendon Press, Oxford, pp 60-79.

104. Bayes. T. (1763) *Phil. Trans. Roy. Soc.* **53**, 370-418.

105. Daniel, G.J. (1991) in *Maximum Entropy in Action*, edited by B. Buck and V.A. MacAulay Clarendon Press, Oxford, p7.

106. Shannon, C.E. and Weaver, W. (1949) *The Mathematical Theory of Communication*, University of Illinois Press, Urbana.

107. Bricogne, G. (1988) *Acta Crystallogr.* **A44**, 517-545.

108. Bricogne, G. (1991) *Maximum Entropy in Action*, edited by B. Buck and V.A. MacAuley, Clarendon Press, Oxford.

109. Bricogne , G. (1991) *Acta Crystallogr.* **A47**, 803-829.

110. Daniels, M. E. (1954) *Ann. Math. Stat.* **25**, 631-650.

111. Reid, N. (1988) *Stat. Sci.* **3**, 213-238.

112. Dunitz, J.D. (1995) *X-ray Analysis and the Structure of Organic Molecules*, 2nd corr. reprint, VHCA, Basel, pp153-154.

113. Bricogne, G. (1993) *Acta Crystallogr.* **D49**, 37-60.

114. Bricogne, G. and Gilmore, C.J. (1990) *Acta Crystallogr.* **A46**, 284-297.

115. Gilmore, C.J., Bricogne, G and Bannister, C. (1990) *Acta Crystallogr.* **A46**, 297-308.

116. Gilmore, C.J. and Bricogne, G. (1992) *Crystallographic Computing 5: From Chemistry to Biology*, edited by D. Moras, A.D. Podjarny and J.C. Thierry, Oxford University Press, Oxford, pp 298-307.

117. Gilmore, C.J., Shankland, K. and Bricogne, G. (1993) *Proc. Roy. Soc. London Ser. A* **442**, 97-111.

118. Shankland, K., Gilmore, C.J., Bricogne, G. and Hashizume, H. (1993) *Acta Crystallogr.* **A49**, 493-501.

119. Sim, G.A. (1959) *Acta Crystallogr.* **12**, 813-815.

120. Sim, G.A. (1960) *Acta Crystallogr.* **13**, 511-512.

121. Gilmore, C.J. (1996) *Acta Crystallogr.* **A52**, 561-589.

122. Gilmore, C.J., Henderson, K. and Bricogne, G. (1991) *Acta Crystallogr.* **A47**, 830-841.

123. Tremayne, M., Lightfoot, P., Glidewell, C., Mehta, M.A., Bruce, P.G., Harris, K.D.M., Shankland, K., Gilmore, C.J. and Bricogne, G. (1992) *J. Solid State Chem.* **100**, 191-196.

124. Tremayne, M., Lightfoot, P., Harris, K.D.M., Shankland, K., Gilmore, C.J., Bricogne, G. and Bruce, P.G. (1992) *J. Mater. Chem.* **2**, 1301-1302.

125. Lightfoot, P., Tremayne, M., Harris, K.D.M., Glidewell, C., Shankland, K., Gilmore, C.J. and Bruce, P.G. (1993) *Mater. Sci. Forum* **133-136**, 207-212.

126. Xiang, S., Carter, C.W.C.Jr, Bricogne, G. and Gilmore, C.J. (1993) *Acta Crystallogr.* **D49**, 193-212.

127. Doublié, S., Xiang, S., Gilmore, C.J., Bricogne, G. and Carter, C.W.C.Jr (1994) *Acta Crystallogr.* **A50**, 164-182.

128. Henderson, R., Baldwin, J.M., Downing, K.H., Lepault, J. and Zemlin, F. (1986) *Ultramicroscopy* **19**, 147-178.

129. Gilmore, C.J., Shankland, K. and Fryer, J.R. (1993) *Ultramicroscopy* **49**, 132-146.

130. Gilmore, C.J., Nicholson, W.V. and Dorset, D.L. (1996) *Acta Crystallogr.* **A52**, 937-946.

131. Baird, T., Gall, J.H., MacNicol, D.D., Mallinson, P.R. and Mitchie, C.R.(1988) *J.Chem. Soc. Chem. Commun.* 1471-1473.

132. Whitaker, A. (1978) Journal of the Soc. Dyers and Col., 431-435.

133. Whitaker, A. (1988) *Journal of the Soc. Dyers and Col.*, 294-300.

134. Allen, F.H., Davies, J.E., Galloy, J.J., Johnson, O., Kennard, O., Macrae, C.F., Mitchell, E.M., Mitchell, G.F., Smith, J.M. and Watson, D.G. (1991) *J. Chem. Inf. Comp. Sci.* **31**, 187-204.

135. Allen, F.H. and Kennard, O. (1993) *Chem. Des. Autom. News* **8**, 1-30.

136. Allen, F.H. and Kennard, O. (1993) *Chem. Des. Autom. News* **8**, 31-37.

137. *CERIUS*^{2} . Molecular Simulations Inc., San Diego.

138. Sutton, L.E. (1958) Tables of Interatomic Distances and Configuration in Molecules and Ions. Chemical Society Special Publication, **11**, Chemical Society , London.

139. Kennard, O. in *International Tables for Crystallography*, Vol. 3, Kynoch Press, Birmingham, pp 257-276.

140. Allen, F.H., Kennard, O., Watson, D.G., Brammer, L., Orpen, A.G. and Taylor, R. (1987) *Journal of the Chemical Society, Perkin Transactions* **2**, S1-S19.

141. International Tables for Crystallography (1969) Vol I, Kynoch Press, Birmingham. pp530-535.

142. Mighell, A.D., Himes, V.L. and Rodgers, J.R. (1983) *Acta Crystallogr.* **A39**, 737-740.

143. *Crystal Identification File* (1982) National Bureau of Standards

144. Badcock, T.D. (1992) Ph.D. Thesis, Scottish College of Textiles, Galashiels.

145. Latimer, W.M. and Rodebush, W.H. (1920) *J. Amer. Chem. Soc.* **42**, 1419-1433.

146. Taylor, R. and Kennard, O. (1984) *Acc. Chem. Res.* **17**, 320-326.

147. Taylor, R., Kennard, O. and Versichel, W. (1984) *Acta Crystallogr.* **B40**, 280-288.

148. Pimental, G.C. and McClellan, A.L. (1960) *The Hydrogen Bond.* W.H. Freeman, San Francisco.

149. Liebermann, C. (1883) *Chem. Ber.* **16**, 2858.

150. Zincke, T. and Bindewald, H. (1884) *Chem. Ber.* **17**, 3026.

151. Morgan, K.J.(1961) *J. Chem. Soc.*, 2151-2159.

152. Burawoy, A., Salem, A.G. and Thompson, A.R. (1952) *J. Chem. Soc*, 4793-4798.

153. Brown, C.J. and Yadav, H.R. (1984) *Acta Crystallogr.* **C40**, 564-566.

154. Cambridge Structural Database (1995). *CSD User's Manual.* Cambridge Crystallographic Data Centre, Cambridge.

155. Kobelt, D., Paulus, E.F. and Kunstmann, W. (1972) *Acta Crystallogr.* **B28**, 1319-1322.

156. Jarvis, J.A.H. (1961) *Acta Crystallogr.* **14**, 961-964.

157. Kobelt, D., Paulus, E.F. and Kunstmann, W. (1974) *Z. Kristallogr.* **139**, 15-19.

158. Whitaker, A. (1977) *Z. Kristallogr.* **146**, 173-184.

159. Olivieri, A.C., Wilson, R.B., Paul, I.C. and Curtin, D.Y.(1989) *J. Am. Chem. Soc.* **111**, 5525-5532.

160. Salmen, R., Malterud, K.E. and Pedersen, B.F. (1988) *Acta Chem. Scand.* **A42**, 493-499.

161. Whitaker, A. (1977) *Z. Kristallogr.* **147**, 99-112.

162. Grainger, C.T. and McConnell, J.F. (1969) *Acta Crystallogr.* **25**, 1962-1970.

163. Whitaker, A. (1980) *Z. Kristallogr.* **152**, 227-238.

164. Guggenberger, L.J. and Teufer, G. (1975) *Acta Crystallogr.* **B31**, 785-790.

165. Whitaker, A. (1977) *Z. Kristallogr.* **145**, 271-288.

166. Diamantis, A.A., Manikas, M., Salam, M.A. and Tiekink, E.R.T. (1992) *Z. Kristallogr.* **202**, 154-156.

167. Alcock, N.W., Spencer, R.C., Prince, R.H. and Kennard, O. (1968) *J. Chem. Soc. A* , 2383-2388.

168. Sutor, D.J. (1963) *J. Chem. Soc.*, 1105-1110.

169. Paulus, E.F., Rieper, W. and Wagner, D. (1983) *Z. Kristallogr.* **165**, 137-149.

170. Gridunova, G.V., Tafeenko, V.A., Tambieva, O.A., Lisitsyna, E.S. and Medev, S.V. (1991) *Zh. Strukt. Khim.* **32**, 53-55.

171. Paulus, E.F. and Rieper, W. (1985) *Z. Kristallogr.* **171**, 87-100.

172. Whitaker, A. (1984) *Z. Kristallogr.* **167**, 225-233.

173. Whitaker, A. (1986) *Acta Crystallogr.* **C42**, 1566-1569.

174. Whitaker, A. (1985) *Z. Kristallogr.* **171**, 17-22.

175. Whitaker, A. (1987) *Acta Crystallogr.* **C43**, 2141-2144

176. Whitaker, A. and Walker, N.P.C. (1987) *Acta Crystallogr.* **C43**, 2137-2141.

177. Whitaker, A. (1983) *Z. Kristallogr.* **163**, 139-149.

178. Paulus, E.F. (1984) *Z. Kristallogr.* **167**, 65-72

179. Whitaker, A. (1984) *Z. Kristallogr.* **166**, 177-188.

180. Whitaker, A. (1985) *Z. Kristallogr.* **170**, 213-223.

181. Whitaker, A. (1983) *Z. Kristallogr.* **163**, 19-30.

182. Whitaker, A. (1985) *Z. Kristallogr.* **171**, 7-15.

183. Golinsk, B. (1988) *Z. Kristallogr.* **184**, 161--167.

184. British Patent No. 23831

185. British Patent No. 23832

186. United States Patent No. 733280

187. *maXus*. MAC Science Ltd, Yokohama, Japan and Nonius B.V., Delft, The Netherlands.

188. Robertson, J.M. (1936) *Journal of the Chemical Society*,1195-1209.

189. Ashida, M. (1966) *Bull. Chem. Soc. Japan* **39**, 2632-2638.

190. Williams, R.C. and Fisher, H.W. (1970) *J. Mol. Biol.* **52**, 121-123.

191. Uyeda, N., Kobayashi, T., Suito, E., Harada, R. and Watanabe, M. (1972) *J. Appl. Physics* **43**, 5181-5189.

192. Murata, Y., Fryer, J.R. and Baird, T. (1976) *J. Micros.* **108**, 261-275.

193. Fryer, J.R. (1977) *Inst. Phys. Conf. Ser* No. 36, pp 423-242.

194. Fryer, J.R. (1978) *Acta Crystallogr.* **A34** 603-607.

195. Dorset, D.L., Tivol, W.F. and Turner, J.N. (1991) *Ultramicroscopy* **38**, 41-45.

196. Dorset, D.L., Tivol, W.F. and Turner, J.N. (1992) *Acta Crystallogr.* **A48**, 562-568.

197. Dorset, D.L., McCourt, M.P., Fryer, J.R., Tivol, W.F. and Turner, J.N. (1994) *Microsc. Soc. Am. Bull.* **24**, 398-404.

198. Dorset, D.L. (1997) *Acta Crystallogr.* **A53**, 356-365.

199. Tivol, W.F., Dorset, D.L., McCourt, M.P. and Turner, J.N. (1993) *Microsc. Soc. Am. Bull.* **23**, 91-98.

200. Uyeda, N., Kobayashi, T., Ishizuka, K. and Fujiyoshi, Y. (1978-1979) *Chem. Scr.* **14**, 47-61.

201. O'Keefe, M.A., Fryer, J.R. and Smith, D.J. (1983) *Acta Crystallogr.* **A39**, 838-847.

202. Fan, H.F., Xiang, S.B., Li, H.F., Pan, Q., Uyeda, N. and Fujiyoshi, Y. (1991) *Ultramicroscopy* **36**, 361-365.

203. Saxton, W.O., O'Keefe, M.A., Cockayne, D.J.H. and Wilkens, M. (1983) *Ultramicroscopy* **12**, 75-86.

204. Gilmore, C.J. (1984) *J. Appl. Cryst.* **17**, 42-46.

205. Gilmore, C.J. and Brown, S.R. (1988) *J. Appl. Cryst.* **22**, 571-572.

206. Linsky, J.P., Paul, T.R., Nohr, R.S. and Kenney, M.E. (1980) *Inorg. Chem.* **19**, 3131-3135.

207. Fryer, J.R. and Kenney, M.E. (1988) *Macromolecules* **21**, 259-262.

208. Wegner, G. (1969) *Z. Naturforsch.* Teil B, **24**, 824-832.

209. Weger, G. (1972) *Makromol. Chem.* **154**, 35-48.

210. Yee, K.C. and Chance, R.R.(1977) *J. Polym. Sci. Polym. Phys. Ed.* **16**, 431-441.

211. Apgar, A.P. and Yee, K.C. (1978) *Acta Crystallogr.* **B34**, 957-959.

212. Read, R.T. and Young, R.J. (1984) *J. Mater. Sci.* **19**, 327-338.