Contributions to
CRYSTALLOGRAPHY
Sivaraj Ramaseshan's 60th Birthday Volume

Edited by
C N R RAO

Indian Academy of Sciences
Bangalore 560 080
CONTRIBUTIONS TO CRYSTALLOGRAPHY
Dedicated to S Ramaseshan
on his
60th Birthday
S. Ramaseshan (born 10 October 1923)
DEDICATION

Professor S Ramaseshan attained the age of 60 on October 10, 1983. Ramaseshan is known to the scientific community in India not only for his scientific contributions, but also for his dedicated service to this Academy. He has served the Academy in various capacities for the past several years. His contribution to the Academy, especially as Editor of Publications and as the Chief Editor of *Pramāṇa* (a journal of physics), have been invaluable. There is little doubt that much of the credit for strengthening and enlarging the publications activity of the Academy goes to Ramaseshan. He is now the President of the Academy.

Sivaraj Ramaseshan was born in Calcutta and had his early education at Nagpur, obtaining his B.Sc(Hons) degree in 1943. He joined the Indian Institute of Science soon after to carry out research work with Professor Sir C V Raman and obtained the Associateship of the Institute in 1948 and the D.Sc., degree from the University of Nagpur in 1951. He was a member of the faculty of the Physics Department of the Indian Institute of Science between 1948 and 1962 and then moved to the Indian Institute of Technology, Madras, as head of the Physics Department. He returned to Bangalore in 1966 to found a new division devoted to materials science in the National Aeronautical Laboratory. The Materials Science Division of this laboratory soon became well known as an outstanding research centre. He returned to the Indian Institute of Science as Joint Director in 1979 and became Director in 1981.

Ramaseshan has contributed to a variety of areas in Crystallography, Physics of Solids, Optics and Materials Science. His sharp intellect and clarity of understanding can be readily seen from his varied research publications. During the early years of his research career, Ramaseshan carried out studies on the crystallography of diamond and magneto-optic rotation. He evolved criteria which determine cleavage planes in diamond and based on surface energies of crystallographic planes, he could explain many characteristic features of the curved morphology of Indian diamonds. He then developed a model to calculate the actual Faraday rotation from the measured, apparent rotation. He studied the effect of stress on optical activity where he showed that the piezo-rotation tensor was an actual tensor of the fourth rank. His interest in Faraday rotation and birefringence led him to study the concept of Poincare representation of polarized light. Ramaseshan later investigated optical and magneto-optical activity in x-rays and neutrons and the Borrmann effect in optics.

One of the major contributions of Ramaseshan has been the application of anomalous scattering of x-rays and neutrons to crystal structure analysis. This
technique has been used by several workers and has become more meaningful with the availability of synchrotron radiation. Ramaseshan and his students have studied crystal structures of several important organic and inorganic systems and a number of leading crystallographers in the country were trained in his laboratory. After moving to the National Aeronautical Laboratory, Ramaseshan and his co-workers investigated a variety of fundamental and applied problems. A new high pressure laboratory was set up at the National Aeronautical Laboratory where several interesting phase transitions were investigated. A compressible ion theory of ionic crystals was developed to explain structural stabilities and phase transitions of a number of systems. Under his guidance, the materials science division contributed much to the development of new techniques, devices and materials.

Professor Ramaseshan was elected a Fellow of the Indian Academy of Sciences in 1955 at the age of 32. He was elected Fellow of the Indian National Science Academy in 1972. He has received the Shanti Swarup Bhatnagar award of the CSIR in Physical Sciences and the Vasvik award in Materials Science. He was a Jawaharlal Nehru Fellow during 1977–1979.

Sivaraj has endeared himself to the scientific community by his fine disposition and sense of humour. He has always been most considerate and helpful to his colleagues. Sivaraj has had the invaluable support of his charming wife, Kausalya, and their three fine daughters who undoubtedly have provided much happiness and encouragement.

On the occasion of his attaining the age of 60, it is a matter of great pleasure to convey the best wishes of the fellows of the Academy and other members of the scientific community, for many years of creative scientific activity and happy life.

To mark this occasion, we have brought out this volume containing papers on crystallography published in the special number (Vol. 92 Numbers 4 and 5) of the Proceedings of the Indian Academy of Sciences (Chemical Sciences).

Bangalore
October 1983

C N R Rao
Vice-President and
Editor of Publications
Indian Academy of Sciences
He still looks young and is so unassuming and easy to talk with that the young visitors in my home who met him recently found it almost impossible to believe that he was nearly sixty, the Director of a great institute, one of the most important scientists in India.

In Oxford, we first began to know something of his qualities when he spent a year in the crystallographic laboratory in 1963–64. It might have been a difficult year for my research group, I was a great deal abroad. In fact, it was a marvellous experience for them, being guided instead by Sivaraj Ramaseshan, with his great learning and his critical understanding. Complicated structures were solved under his hands, mistakes put right, new approaches to structure analysis developed. It was not surprising to find that the whole group had become his devoted friends and admirers—and had developed great respect for Indian science.

We were surprised to learn that soon after his return to India he had become a Director of research not in an academic Institute but in the Aeronautics Research Laboratory, to find him beginning a wholly new line of work, seated in a desk in one corner of a large empty room with one assistant seated in the opposite corner. That the research laboratory grew rapidly, that its work was connected with real life, that it introduced many new productive lines in related industries is now well known, past history.

He changed course again, needing a time to think, he began to work on the past, on the life of C. V. Raman, discovering new and fascinating aspects of Raman’s researches.

And now in his last few working years, his very varied experiences have enabled Sivaraj Ramaseshan to make still more fundamental contributions himself to the development of science in India. In his most recent work as Director of the Indian Institute of Science, he has not only created a happy and fruitful present but has laid the foundation for a future of great promise.

I think of Sivaraj Ramaseshan as one whose influence is extraordinarily creative, who has the power to change the lives of people and nations. His own life was greatly affected by his meeting when young, with Gandhi. Something of Gandhi’s insight, of his care for people as individuals, influence all his actions.

I should add, he belongs to the world, to his friends in many countries, to international science, to everyone’s hopes for the future.

We all greet him on his birthday with great affection and admiration and wish him every happiness, surrounded by his family, and we look forward to many new beginnings in his life’s work.

Dorothy Hodgkin

Crab Mill
Ilmington
CONTENTS

Crystallography

The phase problem of x-ray crystallography

Herbert Hauptman 1

The efficient generation of phase sets by the random approach

K Henderson and M M Woolfson 33

An assessment of criteria of fit in Patterson search

C E Nordman 39

The probability of large structure amplitudes: The space group \( P\bar{1} \)

A J C Wilson 45

Maximum entropy—A new approach to the crystallographic phase problem

Ramesh Narayan, Rajaram Nityananda and G V Vani 51

Extinction correction and synchrotron radiation

P Suortti 69

Mathematical methods for the analysis of charge densities from x-ray diffraction

T N Guru Row 89

Determination of one-dimensional crystal structures using the double Patterson function

Peter Main 95

Precision of determination of \( d \)-spacings using a Guinier camera

Svend Erik Rasmussen 107

The phase problem in neutron diffraction

S K Sikka 113

Neutron diffraction using an electron linear accelerator

B T M Willis 125

Electro- and magnetogyration

I S Zheludev and O G Vlokh 131

A theoretical consideration of the intermediate phase in \( \alpha-\beta \) transition of quartz

N Kato 139

Crystal and molecular structure of pyrrolidine dithio carbamate cobalt(III) chelate

Gopinath Kartha and Kotra V Krishnamurthy 147

(Continued)
Electron density distribution in potassium bis-(carbonato) cuprate(II)  
E N Maslen, N Spadaccini and K J Watson 153

Crystallographic study of push-pull ethylenes  
D Adhikesavalu, Nirupa U Kamath and K Venkatesan 159

Structural evidence for the radical scavenging mechanism of some aminothiol group of radio-protectants  
Kalyani Vijayan, B M Vedavathi and A Mani 167

Structure and interactions of aspirin-like drugs  
M Vijayan 173

Structural changes in the monomeric despentapeptide (B30-B26) insulin crystal  
Ru Chang Bi, Zbigniew Dauter, Eleanor Dodson, Guy Dodson, Federico Gordiano, Rod Hubbard and Colin Reynolds 183

X-ray structural studies on metal complexes with nucleotides and pyridoxal-amino acid Schiff bases. Models for platinum binding to DNA and pyridoxal catalyzed reactions  
M Damodara Poojary, S P Sudhakara Rao and H Manohar 195

Macromolecular crystallography research at Trombay  
K K Kannan and R Chidambaram 207

On the polymorphism of the smectic A phases of highly polar compounds  
N V Madhusudana 219

XANES and EXAFS of copper compounds: Studies of copper carboxylates with metal-metal bonds and of the complex formed by Pseudomonas aeruginosa  
P R Sarode, G Sankar and C N R Rao 237
The phase problem of x-ray crystallography

HERBERT HAUPTMAN
Medical Foundation of Buffalo, 73 High Street, Buffalo, New York 14203, USA

Abstract. Three major themes are developed:
(i) The fundamental principle of direct methods: The structure seminvariants link the observed magnitudes $|E|$ with the desired phases $\phi$ of the normalized structure factors $E$. Specifically, for fixed enantiomorph, the observed magnitudes $|E|$ determine, in general, unique values for all the structure seminvariants; the latter, in turn, as certain well defined linear combinations of the phases, lead to unique values for the individual phases $\phi$.
(ii) The neighbourhood principle: For fixed enantiomorph, the value of any structure seminvariant $T$ is primarily determined, in favourale cases, by the values of one or more small sets of magnitudes $|E|$, the neighbourhoods of $T$, and is relatively insensitive to the great bulk of remaining magnitudes. The conditional probability distribution of $T$, given the magnitudes in any of its neighbourhoods, yields an estimate for $T$ which is particularly good in the favourable case that the variance of the distribution happens to be small.
(iii) The extension concept: By embedding the structure seminvariant $T$ and its symmetry related variants in suitable structure invariants $Q$, one obtains the extensions $Q$ of the seminvariant $T$. In this way the probabilistic theory of the structure seminvariants is reduced to that of the structure invariants, which is well developed.

Finally, the traditional techniques of direct methods are combined with isomorphous replacement and anomalous dispersion.

Keywords. Direct methods; invariant; seminvariant; distribution; neighbourhood; extension.

1. Introduction

1.1 The phase problem

The intensities of a sufficient number of x-ray diffraction maxima determine a crystal structure. The available intensities usually exceed the number of parameters needed to describe the structure. From these intensities a set of numbers $|E_H|$ can be derived, one corresponding to each intensity. However the elucidation of the crystal structure requires also a knowledge of the complex numbers $E_H = |E_H|\exp(i\phi_H)$, the normalized structure factors, of which only the magnitudes $|E_H|$ can be determined from experiment. Thus a “phase” $\phi_H$, unobtainable from the diffraction experiment, must be assigned to each $|E_H|$, and the problem of determining the phases when only the magnitudes $|E_H|$ are known is called “the phase problem”. Owing to the known atomicity of crystal structures and the redundancy of observed magnitudes $|E_H|$, the phase problem is solvable in principle.

1.2 The structure invariants

The values of the individual phases are determined by the crystal structure and the choice of origin. However, there always exist certain linear combinations of the phases whose values are determined by the structure alone and are independent of the choice of origin. These linear combinations of the phases are called the structure invariants.
1.3 The structure seminvariants

For all space groups other than $P1$ the origin may not be chosen arbitrarily if one is to exploit fully the space group symmetries. Those linear combinations of the phases whose values are uniquely determined by the crystal structure and are independent of the choice of permissible origin are known as the structure seminvariants. Thus the collection of structure invariants is a subset of the collection of structure seminvariants. In the space group $P1$ the two classes coincide.

1.4 The fundamental principle of direct methods

For fixed enantiomorph, the observed magnitudes $|E|$ determine, in general, unique values for all the structure seminvariants. The latter in turn, as certain well-defined linear combinations of the phases, lead unambiguously to unique values for the individual phases. Thus the structure seminvariants serve to link the known magnitudes $|E|$ with the desired phases $\phi$ (the fundamental principle of direct methods). By the term “direct methods” is meant that class of methods which exploits relationships among the structure factors in order to go directly from the observed magnitudes $|E|$ to the needed phases $\phi$.

1.5 The neighbourhood principle

For fixed enantiomorph, the value of any structure seminvariant $T$ is primarily determined, in favourable cases, by the values of one or more small sets of observed magnitudes $|E|$, the neighbourhoods of $T$, and is relatively insensitive to the values of the great bulk of remaining magnitudes (the neighbourhood principle). The conditional probability distribution of $T$, given the magnitudes in any of its neighbourhoods, yields an estimate for $T$ that is particularly good in the favourable case that the variance of the distribution happens to be small.

1.6 The extension concept

By embedding the structure seminvariant $T$ and its symmetry related variants in suitable structure invariants $Q$ one obtains the extensions $Q$ of the seminvariant $T$. Owing to the space group-dependent relationships among the phases the value of $T$ is simply related to the values of its extensions. In this way the probabilistic theory of the structure seminvariants is reduced to that of the structure invariants, which is well developed. In particular, the neighbourhoods of $T$ are defined in terms of the neighbourhoods of its extensions.

1.7 Combining direct methods with isomorphous replacement and anomalous dispersion

Most “small” crystal structures, i.e. those with fewer than some 100 independent, nonhydrogen atoms per unit cell, are rather routinely solvable nowadays by traditional direct methods. For the solution of macromolecular structures, on the other hand, the method of isomorphous replacement is universally used, and anomalous dispersion often plays an important supplementary role. One naturally anticipates therefore that integrating the traditional techniques of direct methods with isomorphous replacement and anomalous dispersion will strengthen our ability to solve complex structures. This goal has recently been achieved, and the initial applications strongly suggest that the expected improvement is in fact realized.
2. The phase problem

If one denotes by $\phi_H$ the phase of the structure factor $F_H$, then the relationship between the $F_H$'s and the electron density function $\rho(r)$ is given by

$$F_H = |F_H| \exp(i\phi_H),$$

and

$$\rho(r) = \frac{1}{V} \sum_H F_H \exp(-2\pi i H \cdot r) = \frac{1}{V} \sum_H |F_H| \exp(i \phi_H - 2\pi i H \cdot r),$$

where $V$ represents the unit cell or its volume and $H$ is a reciprocal lattice vector. Only the magnitudes $|F_H|$ of a finite number of structure factors are obtainable from experiment; the values of the phases $\phi_H$, which are also needed if one is to determine $\rho(r)$ from (3), cannot be determined experimentally. It follows that the phase problem, to determine the values of the phases $\phi_H$ of the structure factors $F_H$ when only the magnitudes $|F_H|$ are given, is, in principle, unsolvable when formulated in these terms.

Next, suppose that the real crystal, with continuous electron density $\rho(r)$, is replaced by an idealized one, the unit cell of which consists of $N$ discrete, non-vibrating, point atoms. Then the structure factor $F_H$ is replaced by the normalized structure factor $E_H$ and (1) to (3) are replaced by

$$E_H = |E_H| \exp(i\phi_H),$$

$$E_H = \frac{1}{\sigma_H^{1/2}} \sum_{j=1}^{N} f_j \exp(2\pi i H \cdot r_j),$$

$$\langle E_H \exp(-2\pi i H \cdot r) \rangle_H = \frac{1}{\sigma_H^{1/2}} \left\langle \sum_{j=1}^{N} f_j \exp[2\pi i H \cdot (r_j - r)] \right\rangle_H$$

$$= \frac{f_j}{\sigma_H^{1/2}} \text{ if } r = r_j$$

$$= 0 \text{ if } r \neq r_j,$$

where $f_j$ is the zero-angle atomic scattering factor, $r_j$ is the position vector of the atom labelled $j$, and

$$\sigma_n = \sum_{j=1}^{N} f_j^n, \quad n = 1, 2, 3, \ldots$$

In the x-ray diffraction case the $f_j$ are equal to the atomic numbers $Z_j$ and are therefore all positive; in the neutron diffraction case some of the $f_j$ may be negative.

In practice the magnitudes $|E_H|$ of the normalized structure factors $E_H$ are obtainable (at least approximately) from the observed magnitudes $|F_H|$ while the phases $\phi_H$, as defined by (4) and (5), cannot be determined experimentally. Since one now requires only
the $3N$ components of the $N$ position vectors $r_j$, rather than the much more complicated electron density function $\rho(r)$, it turns out that, in general, the known magnitudes are more than sufficient. This is most readily seen by replacing (5) by

$$|E_H| = \frac{1}{\sigma_h^{1/2}} \left| \sum_{j=1}^{N} f_j \exp (2\pi i \mathbf{H} \cdot r_j) \right|, \quad (8)$$

a system of equations in which the only unknowns are the $3N$ components of the position vectors $r_j$. Since the number of equations (8), equal to the number of reciprocal lattice vectors $\mathbf{H}$ for which the magnitudes $|E_H|$ are observed, usually exceeds the number of unknowns, $3N$, by far, the system (8) is redundant. Thus the phase problem is, in principle, solvable when reformulated in terms of fixed, point atoms.

3. The structure invariants

Equation (6) implies that the normalized structure factors $E_H$ determine the crystal structure. However (5) does not imply that, conversely, the crystal structure determines the values of the normalized structure factors $E_H$ since the position vectors $r_j$ depend not only on the structure but on the choice of origin as well. It turns out nevertheless that the magnitudes $|E_H|$ of the normalized structure factors are in fact uniquely determined by the crystal structure and are independent of the choice of origin but that the values of the phases $\phi_H$ depend also on the choice of origin. Although the values of the individual phases depend on the structure and the choice of origin, there exist certain linear combinations of the phases, the so-called structure invariants, whose values are determined by the structure alone and are independent of the choice of origin.

If the origin of coordinates is shifted to a new point having position vector $r_0$ with respect to the old origin, then, from the definition (equation (5)) of $E_H$, it follows readily that the phase $\phi_H$ of the normalized structure factor $E_H$ with respect to the old origin is replaced by the new phase $\phi'_H$ with respect to the new origin given by

$$\phi'_H = \phi_H - 2\pi \mathbf{H} \cdot r_0. \quad (9)$$

Equation (9) implies that the linear combination of three phases,

$$\psi_3 = \phi_H + \phi_K + \phi_L, \quad (10)$$

is a structure invariant (triplet) provided that

$$\mathbf{H} + \mathbf{K} + \mathbf{L} = 0; \quad (11)$$

the linear combination of four phases,

$$\psi_4 = \phi_H + \phi_K + \phi_L + \phi_M, \quad (12)$$

is a structure invariant (quartet) provided that

$$\mathbf{H} + \mathbf{K} + \mathbf{L} + \mathbf{M} = 0; \quad (13)$$

etc.
4. The structure seminvariants

For all space groups other than \( P1 \) the origin may not be chosen arbitrarily if the simplification permitted by the space group symmetries is to be realized. For example, if a crystal has a centre of symmetry it is natural to place the origin at such a centre while if a two-fold screw axis, but no other symmetry element is present, the origin would normally be situated on this symmetry axis. In such cases the permissible origins are greatly restricted and it is therefore plausible to assume that many linear combinations of the phases will remain unchanged in value when the origin is shifted only in the restricted ways allowed by the space group symmetries. One is thus led to the notion of the structure seminvariant, those linear combinations of the phases whose values are independent of the choice of permissible origin.

4.1 The equivalence concept

For any space group the coordinates of equivalent positions depend upon the choice of origin. Hence the functional form for the geometric structure factor also depends on the choice of origin. Two origins will be said to be equivalent if they give rise to the same functional form for the geometric structure factor. Alternatively, two points are equivalent if they are geometrically situated in the same way with respect to the symmetry elements. Thus, in the space group \( P1 \) all points are equivalent; in \( P1 \) all eight centres of symmetry are equivalent to each other, but no other point is equivalent to any of these eight; in \( P2_1 \) all points on any of the four two-fold screw axes are equivalent to each other, but no other point is equivalent to a point on a two-fold screw axis; in \( P4 \) all points on either of the two four-fold axes are equivalent to each other, all points on either of the two two-fold axes are equivalent to each other, but no point on a two-fold axis is equivalent to any point on a four-fold axis. All points equivalent to a given point are equivalent to each other and are said to form an equivalence class.

4.2 The primary origin

Refer to “International Tables of X-Ray Crystallography”, Vol. 1, in order to define the primary origin for each space group.

4.3 The permissible origins

All points equivalent to the primary origin constitute the permissible origins for each space group.

4.4 The structure seminvariants

The structure seminvariants are those linear combination of the phases whose values are uniquely determined by the crystal structure alone, no matter what the choice of permissible origin. Alternatively, for a given functional form for the geometric structure factor, the values of the structure seminvariants are determined by the structure alone.

The structure invariants and seminvariants have been tabulated for all the space groups (Hauptman and Karle 1953, 1956, 1959; Karle and Hauptman 1961; Lessinger and Wondratschek 1975). A few examples will illustrate the method used to identify the structure seminvariants in the different space groups.
4.5 Space Group $P\overline{1}$

Refer to “International Tables of X-Ray Crystallography”, Vol. 1, to infer that there are eight permissible origins in $P\overline{1}$:

\[(0 0 0), (\frac{1}{2} 0 0), (0 \frac{1}{2} 0), (0 0 \frac{1}{2}), (\frac{1}{2} \frac{1}{2} 0), (\frac{1}{2} 0 \frac{1}{2}), (0 \frac{1}{2} \frac{1}{2}), (\frac{1}{2} \frac{1}{2} \frac{1}{2}). \quad (14)\]

4.5a. The one-phase structure seminvariants $\phi_H$. From (9) it follows that $\phi'_H = \phi_H$ for all permissible origins $r_0$ if and only if $H \cdot r_0$ is an integer, i.e., in view of (14), if and only if the components of $H$ are even. Hence the single phase $\phi_H$ is a structure seminvariant in $P\overline{1}$ if and only if the three components of $H$ are even integers.

4.5b. The two-phase structure seminvariants $\phi_H + \phi_K$. Follow the argument in §4.5a to infer that the linear combination of two phases $\phi_H + \phi_K$ in $P\overline{1}$ is a structure seminvariant if and only if the three components of $H + K$ are even integers.

4.6 Space groups $P2$ and $P2_1$; Unique axis b.

The permissible origins are

\[(0 y 0), (0 y \frac{1}{2}), (\frac{1}{2} y 0), (\frac{1}{2} y \frac{1}{2}), \quad (15)\]

where $y$ is arbitrary. Then from (9) infer that the single phase $\phi_{hkl}$ is a structure seminvariant if and only if $h$ and $l$ are both even and $k = 0$; the linear combination of two phases $\phi_{hkl} + \phi_{h'k'l'}$, is a structure seminvariant if and only if $h_1 + h_2$ and $l_1 + l_2$ are both even and $k_1 + k_2 = 0$; etc.

4.7 Space group $P4$

The permissible origins are

\[(0 0 z), (\frac{1}{2} \frac{1}{2} z), \quad (16)\]

where $z$ is arbitrary. It follows from (9) that the single phase $\phi_{hkl}$ is a structure seminvariant if and only if $h + k$ is even and $l = 0$; the linear combination of two phases $\phi_{h_1k_1l_1} + \phi_{h_2k_2l_2}$, is a structure seminvariant if and only if $(h_1 + k_1) + (h_2 + k_2)$ is even and $l_1 + l_2 = 0$; etc.

4.8 Origin and enantiomorph specification

The theory of the structure seminvariants leads in a natural way to space group dependent recipes for origin and enantiomorph specification.

In general the theory identifies an appropriate set of phases whose values are to be specified in order to fix the origin uniquely. For example, in space group $P1$ the values of any three phases

\[\phi_{h_1k_1l_1}, \phi_{h_2k_2l_2}, \phi_{h_3k_3l_3}, \quad (17)\]
for which the determinant $\Delta$ satisfies

$$\Delta = \begin{vmatrix} h_1 & k_1 & l_1 \\ h_2 & k_2 & l_2 \\ h_3 & k_3 & l_3 \end{vmatrix} = \pm 1,$$

(18)

may be specified arbitrarily, thus fixing the origin uniquely. Once this is done then the value of any other phase is uniquely determined by the structure alone. For enantiomorph specification it is sufficient to specify arbitrarily the sign of any enantiomorph sensitive structure invariant, i.e. one whose value is different from 0 or $\pi$. (See Hauptman 1972, pages 28–52, for further details.)

In the space group $P1$ one again specifies arbitrarily the value (0 or $\pi$) of three phases (17), but now the condition is that the determinant $\Delta$ [defined by (18)] be odd.

Similar recipes for all the space groups are now known and are to be found in the literature cited.

5. The fundamental principle of direct methods

It is known that the values of a sufficiently extensive set of cosine seminvariants (the cosines of the structure seminvariants) lead unambiguously to the values of the individual phases (Hauptman 1972). Magnitudes $|E|$ are capable of yielding estimates of the cosine seminvariants only or, equivalently, the magnitudes of the structure seminvariants; the signs of the structure seminvariants are ambiguous because the two enantiomorphous structures permitted by the observed magnitudes $|E|$ correspond to two values of each structure seminvariant differing only in sign. However, once the enantiomorph has been selected by specifying arbitrarily the sign of a particular enantiomorph sensitive structure seminvariant (i.e. one different from 0 or $\pi$), then the magnitudes $|E|$ determine both signs and magnitudes of the structure seminvariants consistent with the chosen enantiomorph. Thus, for fixed enantiomorph, the observed magnitudes $|E|$ determine unique values for the structure seminvariants; the latter, in turn, lead to unique values of the individual phases. In short, the structure seminvariants serve to link the observed magnitudes $|E|$ with the desired phases $\phi$ (the fundamental principle of direct methods). It is this property of the structure seminvariants which accounts for their importance and which justifies the stress placed on them here.

6. The Neighbourhood principle: Estimating the structure invariants

It has long been known that, for fixed enantiomorph, the value of any structure seminvariant $\psi$ is, in general, uniquely determined by the magnitudes $|E|$ of the normalized structure factors. Recently it has become clear that, for fixed enantiomorph, there corresponds to $\psi$ one or more small sets of magnitudes $|E|$, the neighbourhoods of $\psi$, on which, in favourable cases, the value of $\psi$ most sensitively depends; that is to say that, in favourable cases, $\psi$ is primarily determined by the values of $|E|$ in any of its neighbourhoods and is relatively independent of the values of the great bulk of remaining magnitudes. The conditional probability distribution of $\psi$, assuming as known the magnitudes $|E|$ in any of its neighbourhoods, yields an estimate for $\psi$ which
is particularly good in the favourable case that the variance of the distribution happens to be small [the neighbourhood principle (Hauptman, 1975a, b)].

6.1 The three-phase structure invariant (triplet) in \( P_1 \)

6.1a. The first neighbourhood of the three-phase structure invariant. If \( H, K, \) and \( L \) are three reciprocal lattice vectors satisfying (11), then the first neighbourhood of the three-phase structure invariant \( \psi_3 \) [equation (10)] is defined to consist of the three magnitudes

\[
|E_h|, |E_k|, |E_l|.
\]

(19)

6.1b. The probabilistic background. Suppose that a crystal structure consisting of \( N \) atoms (not necessarily identical) per unit cell in \( P_1 \) is fixed. Denote reciprocal space by \( W \) and by \( W \times W \times W \) the three-fold Cartesian product consisting of all ordered triples \( (H, K, L) \) of reciprocal lattice vectors. Assume also that \( R_1, R_2, \) and \( R_3 \) are fixed positive numbers. Suppose finally that the primitive random variable (vector) is the ordered triple \( (H, K, L) \) of reciprocal lattice vectors which is assumed to be uniformly distributed over the subset of \( W \times W \times W \) defined by

\[
|E_h| = R_1, \quad |E_k| = R_2, \quad |E_l| = R_3,
\]

(20)

and (11). Then the three-phase structure invariant \( \psi_3 \) [equation (10)], as a function of the primitive random variable \( (H, K, L) \), is itself a random variable.

6.1c. The condition probability distribution of the three-phase structure invariant \( \psi_3 \), assuming as known the three magnitudes in its first neighbourhood. Denote by

\[
P_{1|3} = P(\psi | R_1, R_2, R_3)
\]

(21)

the conditional probability distribution of \( \psi_3 \), given (20). Then (Cochran 1955; Hauptman 1976b)

\[
P_{1|3} \approx \frac{1}{2\pi I_0(A)} \exp(A \cos \Psi),
\]

(22)

where \( I_0 \) is the modified Bessel function,

\[
A = \frac{2\sigma^3}{\sigma_{\xi}^2} R_1 R_2 R_3,
\]

(23)

and \( \sigma_{\xi} \) is defined by (7). Since \( A > 0, P_{1|3} \) has a unique maximum at \( \Psi = 0 \), and it is clear that the larger the value of \( A \) the smaller is the variance of the distribution. See figure 1, where \( A = 2.316 \), figure 2, where \( A = 0.731 \). Hence in the favourable case that \( A \) is large, say, for example, \( A > 3 \), the distribution leads to a reliable estimate of the structure invariant \( \psi_3 \), zero in this case:

\[
\psi_3 \approx 0 \text{ if } A \text{ is large}.
\]

(24)

Furthermore, the larger the value of \( A \), the more likely is the probabilistic statement (24). It is remarkable how useful this relationship has proved to be in the applications; and yet (24) is severely limited because it is capable of yielding only the zero estimate for \( \psi_3 \), and only those estimates are reliable for which \( A \) is large, the favourable cases. There is nothing in the older formulations of this relationship which suggests, for
example, how one may obtain a reliable estimate for $\psi_3$ in the neighbourhood of $\pi$. A major goal in the present formulation is to make clear how the results described here are to be extended to structure invariants in general and to yield better estimates which may lie anywhere in the interval ($-\pi, \pi$).

It should perhaps be mentioned in passing that another consequence of the distribution (22) is the so-called tangent formula which is universally used by direct methods practitioners:

$$\tan \phi_h = \frac{\langle |E_k E_{h-k}| \sin (\phi_k + \phi_{h-k}) \rangle_k}{\langle |E_k E_{h-k}| \cos (\phi_k + \phi_{h-k}) \rangle_k},$$

in which $h$ is a fixed reciprocal lattice vector, the averages are taken over the same set of vectors $k$ in reciprocal space, usually restricted to those vectors $k$ for which $|E_k|$ and $|E_{h-k}|$ are both large, and the sign of $\sin \phi_h (\cos \phi_h)$ is the same as the sign of the

Figure 1. The distribution $P_{1/3}$, equation (22) for $A = 2.316$.

Figure 2. The distribution $P_{1/3}$, equation (22) for $A = 0.731$. 
Herbert Hauptman

numerator (denominator) on the right hand side. The tangent formula is usually used to refine and extend a basis set of phases, presumed to be known.

6.2 The four-phase structure invariant (quartet) in $P_1$.

6.2a. The first neighbourhood of the four-phase structure invariant. If $H, K, L,$ and $M$ are four reciprocal lattice vectors satisfying (13), then, in analogy with §6.1a, the first neighbourhood of the four-phase structure invariant $\psi_4$ [equation (12)] is defined to consist of the four magnitudes


(25)

6.2b. The probabilistic background. One employs the same probabilistic background as in §6.1b with the difference that now one specifies four positive numbers $R_1, R_2, R_3,$ and $R_4$ and assumes that the primitive random variable is the ordered quadruple $(H, K, L, M)$ of reciprocal lattice vectors which is uniformly distributed over the subset of the four-fold Cartesian product $W \times W \times W \times W$ defined by

$$|E_H| = R_1, \quad |E_K| = R_2, \quad |E_L| = R_3, \quad |E_M| = R_4.$$  

(26)

and (13). Then the four-phase structure invariant $\psi_4$ [equation (12)], as a function of the primitive random variable $(H, K, L, M)$, is itself a random variable.

6.2c. The conditional probability distribution of the four-phase structure invariant $\psi_4$, assuming as known the four magnitudes in its first neighbourhood. Denote by

$$P_{1|4} = P(\Psi | R_1, R_2, R_3, R_4),$$  

(27)

the conditional probability distribution of $\psi_4$, given (26). Then (Hauptman, 1975a, b),

$$P_{1|4} \approx \frac{1}{2\pi I_0(B)} \exp (B \cos \Psi),$$  

(28)

where

$$B = \frac{2\sigma_4}{\sigma_2} R_1 R_2 R_3 R_4,$$  

(29)

and $\sigma_n$ is defined by (7). Thus $P_{1|4}$ is identical with $P_{1|3}$, but $B$ replaces $A$. Hence similar remarks apply to $P_{1|4}$. In particular, (28) always has a unique maximum at $\Psi = 0$ so that the most probable value of $\psi_4$, given the four magnitudes (26) in its first neighbourhood, is zero, and the larger the value of $B$ the more likely it is that $\psi_4 \approx 0$. Since $B$ values, of order $1/N$, tend to be less than $A$ values, of order $1/\sqrt{N}$, at least for large values of $N$, the estimate (zero) of $\psi_4$ is in general less reliable than the estimate (zero) of $\psi_3$. Hence the goal of obtaining a reliable non-zero estimate for a structure invariant is not realized by (28). The decisive step in this direction is made next.

6.2d. The second neighbourhood of the four-phase structure invariant $\psi_4$. If one assumes as known not only the four magnitudes (25), but the additional three magnitudes $|E_{H+K}|, |E_{K+L}|, |E_{L+H}|$, then, in favourable cases, one obtains a more reliable estimate for $\psi_4$ and, furthermore, the estimate may lie anywhere in the interval $0$ to $\pi$. The idea to use these three additional magnitudes is suggested by the following heuristic argument (Schenk and de Jong 1973; Schenk 1973, 1974; Hauptman 1976a).
Assume first that the four magnitudes (25) are all large:

$$|E_H|, |E_K|, |E_L|, |E_M|$$

are all large, (30) where $H, K, L, M$ satisfy (13). If it should happen that $|E_{H+K}|$ is also large then (24) implies

$$\phi_H + \phi_K + \phi_{-H-K} \approx 0,$$

and, in view of (13),

$$\phi_L + \phi_M + \phi_{H+K} \approx 0.$$  

Then, by the addition of (31) and (32),

$$\phi_H + \phi_K + \phi_L + \phi_M \approx 0.$$  

If $|E_{K+L}|$ is large or if $|E_{L+H}|$ is large then, by symmetry, the relationship (33) again holds. If it should happen that

$$|E_{H+K}|, |E_{K+L}|, |E_{L+H}|$$

are all large, (34) then one would again expect (33) to hold, but now with very high probability. However, the distribution (28) implies that the value of $\psi_4$ must occasionally be in the neighbourhood of $\pi$, even if $B$ is quite large. How is one to identify the small fraction of quartets which are in fact near $\pi$? In view of the foregoing argument it is plausible to suppose that

$$\phi_H + \phi_K + \phi_L + \phi_M \approx \pi$$

precisely in the case that

$$|E_{H+K}|, |E_{K+L}|, |E_{L+H}|$$

are all small. (36) Although the argument presented here is heuristic only and proves nothing, it does suggest the question: What is the conditional probability distribution of the quartet $\psi_4$, assuming as known the seven magnitudes $|E_H|, |E_K|, |E_L|, |E_M|, |E_{H+K}|, |E_{K+L}|, |E_{L+H}|$?

In view of this discussion one defines the second neighbourhood of $\psi_4$ to consist of the seven magnitudes

$$|E_H|, |E_K|, |E_L|, |E_M|, |E_{H+K}|, |E_{K+L}|, |E_{L+H}|.$$  

(37)

6.2e. The conditional probability distribution of the four-phase structure invariant $\psi_4$, assuming as known the seven magnitudes in its second neighbourhood. Assume that the seven positive numbers $R_1, R_2, R_3, R_4, R_{12}, R_{23}, R_{31}$ are fixed. Suppose next that the ordered quadruple of reciprocal lattice vectors $(H, K, L, M)$ is a random variable which is uniformly distributed over the subset of the fourfold Cartesian product $W \times W \times W \times W$ defined by

$$|E_H| = R_1, |E_K| = R_2, |E_L| = R_3, |E_M| = R_4;$$  

$$|E_{H+K}| = R_{12}, |E_{K+L}| = R_{23}, |E_{L+H}| = R_{31};$$  

and (13). Then $\psi_4$, as a function of the primitive random variable $(H, K, L, M)$, is itself a random variable. Denote by

$$P_{\psi_4} = P(\Psi | R_1, R_2, R_3, R_4, R_{12}, R_{23}, R_{31})$$  

(40)
the conditional probability distribution of the quartet $\psi_4$, given the seven magnitudes (38) and (39). Then (Hauptman, 1975a, b; 1976b)

$$P_{\psi|\psi} \approx \frac{1}{L} \exp \left(-2B' \cos \psi\right) I_0\left(\frac{2\sigma_3}{\sigma_2^{3/2}} R_{12} X_{12}\right) I_0\left(\frac{2\sigma_3}{\sigma_2^{3/2}} R_{23} X_{23}\right) \times I_0\left(\frac{2\sigma_3}{\sigma_2^{3/2}} R_{31} X_{31}\right),$$

where

$$B' = \frac{1}{\sigma_2} (3\sigma_3^2 - \sigma_2 \sigma_4) R_1 R_2 R_3 R_4,$$

$$(41)$$

$$X_{12} = \left[R_1^2 R_2^2 + R_3^2 R_4^2 + 2R_1 R_2 R_3 R_4 \cos \psi\right]^{1/2},$$

$$(42)$$

$$X_{23} = \left[R_2^2 R_3^2 + R_1^2 R_4^2 + 2R_1 R_2 R_3 R_4 \cos \psi\right]^{1/2},$$

$$(43)$$

$$X_{31} = \left[R_3^2 R_1^2 + R_2^2 R_4^2 + 2R_1 R_2 R_3 R_4 \cos \psi\right]^{1/2},$$

$$(44)$$

$$(45)$$

$\sigma_n$ is defined by (7), and $L$ is a normalizing parameter, independent of $\psi$, which is not needed for the present purpose.

Figures 3–5 show the distribution (41) (solid line) for typical values of the seven parameters (38) and (39). For comparison the distribution (28) (broken line) is also shown. Since the magnitudes $|E|$ have been obtained from a real structure, comparison with the true value of the quartet is also possible. As already emphasized, the distribution (28) always has a unique maximum at $\psi = 0$. The distribution (41), on the other hand, may have a maximum at $\psi = 0$, or $\pi$, or any value between these extremes, as shown by figures 3–5. Roughly speaking, the maximum of (41) occurs at 0 or $\pi$ according as the three parameters $R_{12}, R_{23}, R_{31}$ are all large or all small, respectively, thus confirming the heuristic argument given earlier. These figures also clearly show the improvement which may result when, in addition to the four magnitudes (38), the three

![Figure 3](image-url)
The phase problem of x-ray crystallography

Figure 4. The distribution (41) (— — —) and (28) (— — —) for the values of the seven parameters (38) and (39) shown. The mode of (41) is 105°, of (28) always 0.

Figure 5. The distribution (41) (— — —) and (28) (— — —) for the values of the seven parameters (38) and (39) shown. The mode of (41) is 180°, of (28) always 0.

magnitudes (39) are also assumed to be known. Finally, in the special case that

\[ R_{12} \approx R_{23} \approx R_{31} \approx 0 \]  

the distribution (41) reduces to

\[ P_{1|\psi} \approx \frac{1}{L} \exp(-2B' \cos \Psi), \]

which has a unique maximum at \( \Psi = \pi \) (fig. 5).

6.3 Space group \( P\bar{1} \)

In complete analogy with \( P1 \), conditional probability distributions for the triplet \( \psi \), in
Herbert Hauptman

Given three magnitudes, and for the quartet $\psi_4$ in $P\bar{1}$, given four or seven magnitudes, are readily derived. The major difference is that now the distributions to be described are discrete, because the invariants take on only the two values 0, $\pi$. Using the definitions and terminology of §§6.1 and 6.2, the present account is greatly abbreviated.

6.3a. The conditional probability distribution of the triplet $\psi_3$ in $P\bar{1}$, assuming as known the three magnitudes in its first neighbourhood. Make the same hypotheses as in §6.1. Denote by $P_3^+$ ($P_3^-$) the conditional probability that $\psi_3 = 0 (\pi)$, given the three magnitudes (20). Then (Woolfson 1954; Hauptman 1976b)

$$P_3^\pm = \frac{1}{K_3} Z_3^\pm,$$  \hspace{1cm} (48)

$$Z_3^\pm = \exp(\pm A/2),$$  \hspace{1cm} (49)

and

$$K_3 = 2 \cosh(A/2),$$  \hspace{1cm} (50)

where $A$ is given by (23) and upper (lower) signs in (48) and (49) go together. Equations (48) to (50) should be compared with the $P1$ analogue, (22). It follows that

$$P_3^+ > \frac{1}{2},$$  \hspace{1cm} (51)

so that the estimate $\psi_3 = \pi$, given only the three magnitudes (20), is impossible.

6.3b. The conditional probability distribution of the quartet $\psi_4$ in $P\bar{1}$, assuming as known the four magnitudes in its first neighbourhood. Denote by

$$P_4^+ or P_4^-,$$  \hspace{1cm} (52)

the conditional probability that the quartet $\psi_4$ be equal to 0 or $\pi$ respectively, given the four magnitudes (26). Then (Green and Hauptman 1976; Hauptman et al 1976; Hauptman 1976b)

$$P_4^\pm = \frac{1}{K_4} Z_4^\pm,$$  \hspace{1cm} (53)

where

$$Z_4^\pm = \exp(\pm B/2),$$  \hspace{1cm} (54)

$$K_4 = Z_4^+ + Z_4^-,$$  \hspace{1cm} (55)

and $B$ is defined by (29). The distribution (53)–(55) should be compared with the $P1$ analogue (28). The variance of the distribution (53) is given by

$$\text{Var} = 4 P_4^+ P_4^- (4 Z_4^+ Z_4^-)/(Z_4^+ + Z_4^-)^2.$$  \hspace{1cm} (56)

Clearly

$$P_4^+ > \frac{1}{2},$$  \hspace{1cm} (57)

so that (53) can give only the zero estimate for the quartet. Finally, the smaller the value of the variance, the more reliable is the estimate, zero in this case.

6.3c. The conditional probability distribution of the quartet $\psi_4$ in $P\bar{1}$, assuming as known the seven magnitudes in its second neighbourhood. Denote by

$$P_7^+ or P_7^-$$  \hspace{1cm} (58)
the conditional probability that the quartet \( \psi_4 \) be equal to 0 or \( \pi \) respectively, given the seven magnitudes (38) and (39). Then (Green and Hauptman 1976; Hauptman and Green 1976; Hauptman 1976b; but see also Giacovazzo, 1976a for a different but related result)

\[
P_\gamma^{\pm} = \frac{1}{K_\gamma} Z_\gamma^{\pm},
\]

where

\[
Z_\gamma^{\pm} = \exp \left( \mp B' \right) \cosh \left( \frac{\sigma_3}{\sigma_2^{3/2}} R_{12} X_{12}^{\pm} \right) \cosh \left( \frac{\sigma_3}{\sigma_2^{3/2}} R_{23} X_{23}^{\pm} \right) \times \cosh \left( \frac{\sigma_3}{\sigma_2^{3/2}} R_{31} X_{31}^{\pm} \right),
\]

(59)\(\quad\)

\[
X_{12}^{\pm} = R_1 R_2 \pm R_3 R_4,
\]

(60)\(\quad\)

\[
X_{23}^{\pm} = R_2 R_3 \pm R_1 R_4,
\]

(61)\(\quad\)

\[
X_{31}^{\pm} = R_3 R_1 \pm R_2 R_4,
\]

(62)\(\quad\)

\[
K_\gamma = Z_\gamma^{+} + Z_\gamma^{-},
\]

(63)\(\quad\)

and the variance is given by

\[
\text{Var} = 4 P_\gamma^{+} P_\gamma^{-} = \frac{4 Z_\gamma^{+} Z_\gamma^{-}}{(Z_\gamma^{+} + Z_\gamma^{-})^2}.
\]

(64)\(\quad\)

In contrast to (53), \( P_\gamma^{\pm} \) may lie anywhere in the interval 0 to 1 so that (59)–(64) may give either the 0 or \( \pi \) estimate for the quartet. In complete analogy with \( P_1 \), the estimate is 0 or \( \pi \) according as \( R_{12}, R_{23}, R_{31} \) are all large or all small respectively, the favourable cases. In the extreme case that

\[
R_{12} \approx R_{23} \approx R_{31} \approx 0,
\]

(65)\(\quad\)

(60) reduces to

\[
Z_\gamma^{\pm} = \exp \left( \mp B' \right).
\]

(66)\(\quad\)

Again, the smaller the value of the variance, the more reliable is the estimate, 0 or \( \pi \).

6.4 The discriminant of the five-phase structure invariant (quintet)

6.4a. The first two neighbourhoods. The linear combination of five phases

\[
\psi_5 = \phi_H + \phi_K + \phi_L + \phi_M + \phi_N
\]

is a structure invariant (quintet) if

\[
H + K + L + M + N = 0.
\]

(67)\(\quad\)

(68)\(\quad\)

(69)\(\quad\)

In complete analogy with the earlier work on quartets, the first neighbourhood of \( \psi_5 \) is defined to consist of the five magnitudes

\[
|E_H| = R_1, \ |E_K| = R_2, \ |E_L| = R_3, \ |E_M| = R_4, \ |E_N| = R_5;
\]

(70)\(\quad\)

and the second neighbourhood of the five magnitudes (70), the “main terms”, plus the additional ten magnitudes, the “cross-terms”,

The phase problem of x-ray crystallography
\[ |E_{H+k}| = R_{12}, \quad |E_{H+l}| = R_{13}, \quad |E_{H+m}| = R_{14}, \quad |E_{H+N}| = R_{15}, \]
\[ |E_{K+l}| = R_{23}, \quad |E_{K+m}| = R_{24}, \quad |E_{K+N}| = R_{25}, \]
\[ |E_{L+m}| = R_{34}, \quad |E_{L+N}| = R_{35}, \quad |E_{M+N}| = R_{45}, \]
\[ \text{(71)} \]
i.e. 15 magnitudes |E| in all.

6.4b. The heuristic argument. By means of plausible reasoning, in analogy with §6.2d., one is led to identify those quintets having the extreme values 0 or \( \pi \) as summarized in table 1, the entries of which give the values, large (L) or small (S) of the 15 magnitudes, (70) and (71), in the second neighbourhood of \( \psi_5 \) corresponding to \( \psi_5 = 0 \) or \( \pi \).

6.4c. The probabilistic theory of quintets in \( P1 \) and \( \overline{P1} \) (Schenk 1975; Hauptman 1977; Putten and Schenk 1977; Fortier and Hauptman 1977a, b; Hauptman and Fortier 1977). Using the probabilistic background described earlier and referring to the literature one is led to define the discriminant \( \Delta \) of the quintet \( \psi_5 \) as follows:

\[
\Delta = R_1 R_2 R_3 R_4 R_5 \left\{ \frac{2\sigma^3}{\sigma^2/2} \sum_{15} R_{12}^2 R_{34}^2 \right\} - \frac{2\sigma_3}{\sigma^2/2} (3\sigma_3^2 - \sigma_2 \sigma_4) \sum_{10} R_{12}^2 + \frac{2}{\sigma^2/2} (15\sigma_3^2 - 10\sigma_2 \sigma_3 \sigma_4 + \sigma_2^2 \sigma_5) \}
\[ \text{(72)} \]

where

\[
\sum_{15} R_{12}^2 R_{34}^2 = R_{12}^2 R_{34}^2 + R_{12}^2 R_{35}^2 + R_{12}^2 R_{45}^2 + R_{13}^2 R_{24}^2 \\
+ R_{13}^2 R_{25}^2 + R_{14}^2 R_{23}^2 + R_{14}^2 R_{25}^2 + R_{15}^2 R_{23}^4 + R_{15}^2 R_{24}^2 + R_{15}^2 R_{34}^2 \\
+ R_{12}^2 R_{35}^2 + R_{12}^2 R_{45}^2 + R_{12}^2 R_{34}^2.
\[ \text{(73)} \]

Table 1. The probable values of the structure invariant \( \psi_5 = \phi_H + \phi_K + \phi_L + \phi_M + \phi_N \), given the 15 magnitudes in its second neighbourhood; L means large, S means small.

<table>
<thead>
<tr>
<th>( \psi_5 )</th>
<th>( R_1 )</th>
<th>( R_2 )</th>
<th>( R_3 )</th>
<th>( R_4 )</th>
<th>( R_5 )</th>
<th>( R_{12} )</th>
<th>( R_{13} )</th>
<th>( R_{14} )</th>
<th>( R_{15} )</th>
<th>( R_{23} )</th>
<th>( R_{24} )</th>
<th>( R_{25} )</th>
<th>( R_{34} )</th>
<th>( R_{35} )</th>
<th>( R_{45} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>( \pi )</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>( \pi )</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>( \pi )</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>( \pi )</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>( \pi )</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>( \pi )</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>( \pi )</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>( \pi )</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>( \pi )</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>( \pi )</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>( \pi )</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
</tbody>
</table>

\[ \text{Table 1} \]
As described in the literature the discriminant $\Delta$ is readily computable and strongly correlated with extreme values of the quintet in $P1$ and $P1\dagger$ in the sense that $\psi_5 \approx \pi$ according as $\Delta > 0$ or $\psi_5 \approx \pi$ according as $\Delta \geq 0$, respectively, all in accord with table 1.

6.5 The neighbourhood principle

It has been known for many years that, for a chosen enantiomorph, the values of the structure factor magnitudes $|E|$ uniquely determine, in general, (i.e. provided that no homometric solutions, other than enantiomorphs, exist) the values of the structure invariants. Alternatively, for a fixed functional form for the geometric structure factor and for fixed choice of enantiomorph, the magnitudes $|E|$ uniquely determine the values of the structure seminvariants. A number of formulas explicitly relating the cosine invariants to the magnitudes of all (or large numbers of) structure factors $E$ have in fact been known for some time and these have played an important role in the applications. However their value has been somewhat limited by the requirement that certain kinds of rational dependence among atomic coordinates not be present, and for very complex structures their usefulness is greatly reduced. The results described in the preceding paragraphs however point strongly to a deeper insight. Instead of seeking to express the structure seminvariants in terms of all observed structure factor magnitudes $|E|$, or even in terms of large numbers of $|E|$'s, as the earlier formulas did, it is now suggested that the value of each structure seminvariant $\psi$ is primarily determined, in favourable cases, by the values of one or more small sets of appropriately chosen magnitudes $|E|$, the so-called neighbourhoods of $\psi$, and is relatively insensitive to the vast bulk of the remaining magnitudes (the neighbourhood principle, Hauptman 1975a, b).

Thus in the favourable case that $A$, (23), is large the value of the three-phase structure invariant $\psi_3$ is primarily determined by the three magnitudes (20) and the larger the value of $A$ the more reliable is the estimate for $\psi_3$, zero in this case, as reference to (22) shows. Hence the first neighbourhood of the three-phase structure invariant $\psi_3$ is defined to consist of the three magnitudes (20).

Next, if one is given only the four magnitudes (26), then the most probable value of the quartet $\psi_4$ is zero in view of (28), and in the favourable case that $B$, (29), is large the zero estimate is good. Thus the first neighbourhood of the quartet $\psi_4$ is defined to consist of the four magnitudes (27).

Suppose next that the seven magnitudes (38) and (39) are given. In this case the most probable value of the quartet $\psi_4$ is given by the maximum of the distribution (41) and this estimate may lie anywhere in the range $(0, \pi]$ [or, because of symmetry, in the range $(-\pi, 0)$], depending on the values of the seven parameters (38) and (39). In the favourable case that the variance of the distribution (41) is very small then the estimate for $\psi_4$ is particularly good. The second neighbourhood of the quartet $\psi_4$ is therefore defined to consist of the seven magnitudes (38) and (39).

The variance of the distribution (41) may be greater than or less than that of (28). If $B$, (29), is fixed then, in the favourable cases that the three magnitudes (39) are all large or all small the variance of (41) will be smaller than that of (28) and the estimate for $\psi_4$ given by (41) will then be more reliable than that given by (28). Thus the gain in going
from the first to the second neighbourhood is that the potential for obtaining a distribution with a small variance is increased; (although it should be emphasized that this potential is not necessarily realized in every given case); in short it becomes possible to find a more reliable estimate for \( \psi_4 \), and the estimate is no longer restricted to be zero but may instead lie anywhere in the interval \((0, \pi)\).

The formulation of the neighbourhood principle has stimulated the rapid development of the probabilistic theory of the structure invariants and seminvariants. Thus, with the derivation of conditional probability distributions of the four-phase structure invariants, assuming as known the magnitudes in the third and higher neighbourhoods, the probabilistic theory of quartets is now reasonably complete, at least for the space groups \( P1 \) and \( P\bar{1} \) (Hauptman 1977a, b, c). The theory of quintets and sextets has been initiated (Giacovazzo 1976c; Hauptman 1977; Fortier et al 1977a, b, c; Hauptman et al 1977a, b), and a beginning has been made on the probabilistic theory of the structure seminvariants for selected space groups (Hauptman 1979, 1980; Burla et al 1981; Cascarano et al 1982; Busetta et al 1980; Giacovazzo 1977, 1979, 1980a, b, c; Giacovazzo et al 1979; Giacovazzo and Vickovic 1980). Not only have these developments been heavily dependent on the neighbourhood principle but the results obtained confirm conclusively the validity of the neighbourhood principle for the structure invariants and structure seminvariants in general.

7. The extension concept: Estimating the structure seminvariants

By embedding the structure seminvariant \( T \) and its symmetry related variants in suitable structure invariants \( Q \) one obtains the extensions \( \hat{Q} \) of the seminvariant \( T \). Owing to the space group dependent relations among the phases, \( T \) is related in a known way to its extensions. In particular the neighbourhoods of \( T \) are defined in terms of the neighbourhoods of its extensions. The process will be illustrated in some detail for the three-phase structure seminvariant in \( P\bar{1} \) which serves as the prototype for the structure seminvariants in general, in all space groups, noncentrosymmetric as well as centrosymmetric.

7.1 The three-phase structure seminvariant \( T \) in \( P\bar{1} \)

7.1a. The three-phase structure seminvariant. The linear combination of three phases

\[
T = \phi_H + \phi_K + \phi_L,
\]

is a structure seminvariant in \( P\bar{1} \) if and only if the three components of the reciprocal lattice vector \( \mathbf{H} + \mathbf{K} + \mathbf{L} \) are even. Then the components of the eight reciprocal lattice vectors \( \frac{1}{2}(\pm \mathbf{H} \pm \mathbf{K} \pm \mathbf{L}) \) are all integers.

7.1b. The extensions of \( T \). Following Hauptman (1977b, 1978; but see also Giacovazzo, 1977, for an equivalent procedure) one embeds the three-phase structure seminvariant

\[
T = \phi_H + \phi_K + \phi_L,
\]

and its three symmetry related variants

\[
T_1 = \phi_{-H} + \phi_K + \phi_L,
\]

and its three symmetry related variants

\[
T_1 = \phi_{-H} + \phi_K + \phi_L.
\]
The phase problem of x-ray crystallography

\[ T_2 = \phi_H + \phi_{-K} + \phi_L, \quad (78) \]
\[ T_3 = \phi_H + \phi_K + \phi_{-L}, \quad (79) \]

in the respective quintets

\[ Q_0 = T + \phi_{-\frac{1}{2}(H+K+L)} + \phi_{-\frac{1}{2}(H+K+L)}, \quad (80) \]
\[ Q_1 = T_1 + \phi_{-\frac{1}{2}(-H-K+L)} + \phi_{-\frac{1}{2}(-H-K+L)}, \quad (81) \]
\[ Q_2 = T_2 + \phi_{-\frac{1}{2}(H-K+L)} + \phi_{-\frac{1}{2}(H-K+L)}, \quad (82) \]
\[ Q_3 = T_3 + \phi_{-\frac{1}{2}(H+K-L)} + \phi_{-\frac{1}{2}(H+K-L)}. \quad (83) \]

In view of (76)-(79) and the space group-dependent relationships among the phases, it is readily verified that (80) to (83) are five-phase structure invariants (quintets) and

\[ T = Q_0 = Q_1 = Q_2 = Q_3. \quad (84) \]

Thus the probabilistic theory of the three-phase structure seminvariant \( T \) is reduced to that of the quintets, which is well developed. In particular the neighbourhoods of \( T \) are defined in terms of the neighbourhoods of the quintet.

7.1c. The first two neighbourhoods of the extensions. Clearly only four of the five “main terms” (see §6.4a and Schenk 1975; Hauptman 1977a) of the special quintet \( Q_0 \) are distinct. The first neighbourhood of \( Q_0 \) is accordingly defined to consist of the four magnitudes

\[ |E_H|, |E_K|, |E_L|, |E_{\frac{1}{2}(H+K+L)}|. \quad (85) \]

Referring again to §6.4 it is readily verified that only seven of the ten “cross-terms” of the special quintet \( Q_0 \) are in fact distinct. Thus the second neighbourhood of \( Q_0 \) is defined to consist of the four main terms (85) and the seven cross-terms

\[ |E_{\frac{1}{2}(-H+K+L)}|, |E_{\frac{1}{2}(-H-K+L)}|, |E_{\frac{1}{2}(H+K-L)}|; \]
\[ |E_{H+K}|, |E_{K+L}|, |E_{L+H}|, |E_{H+K+L}|, \quad (86) \]

or eleven magnitudes \( |E| \) in all. (In contrast, the second neighbourhood of the general quintet consists of fifteen magnitudes \( |E| \).)

Not only does §6.4 serve to identify the first two neighbourhoods of the special quintet \( Q_0 \) but table 1 shows in a qualitative way what the relationship between \( Q_0 \) and the magnitudes \( |E| \) in its second neighbourhood must be. Thus table 1 leads, by suitable specialization, to rows 1–8 of table 2. Although table 1 has 16 rows, only eight of these yield entries in table 2 which are internally consistent; the remaining eight rows of table 1, which would lead to contradictory entries in table 2 (arising from the fact that only seven of the cross-terms of the special quintet \( Q_0 \) are distinct), are therefore omitted from table 2. Thus the 16 rows of table 1 yield only rows 1–8 of table 2 for \( Q_0 \).

In a similar way it may be shown that the first neighbourhood of the special quintet \( Q_1 \) consists of the four magnitudes

\[ |E_H|, |E_K|, |E_L|, |E_{\frac{1}{2}(-H+K+L)}|, \quad (87) \]

and the second neighbourhood of the four main terms (87) plus the additional seven cross-terms

\[ |E_{\frac{1}{2}(H+K+L)}|, |E_{\frac{1}{2}(H-K+L)}|, |E_{\frac{1}{2}(H+K-L)}|. \]
Table 2. The probable values of the extensions $Q_{ij}$, $j = 0, 1, 2, 3$, of $T$, given the magnitudes in their second neighbourhoods; $L$ means large; $S$ means small; obtained by suitable specialization from selected rows of Table 1.

<table>
<thead>
<tr>
<th>Row</th>
<th>$Q_j$</th>
<th>$H$</th>
<th>$K$</th>
<th>$L$</th>
<th>$\frac{1}{2}(H+K+L)$</th>
<th>$\frac{1}{2}(H-K+L)$</th>
<th>$L+H$</th>
<th>$H-K$</th>
<th>$K-L$</th>
<th>$L-H$</th>
<th>$H+K+L$</th>
<th>$-H+K+L$</th>
<th>$H-K+L$</th>
<th>$H+K-L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>2</td>
<td>$\pi$</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>S</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>3</td>
<td>$\pi$</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>S</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>4</td>
<td>$\pi$</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>S</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>5</td>
<td>$\pi$</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>L</td>
<td>S</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>6</td>
<td>$\pi$</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>7</td>
<td>$\pi$</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>8</td>
<td>$\pi$</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
<td>S</td>
</tr>
</tbody>
</table>

\[ \begin{array}{ccccccccccccc}
| 9   | 0      | L  | L  | L  | L                | L                | L    | L    | L    | L    | L        | L        | L        | L        |
| 10  | $\pi$  | L  | L  | L  | L                | S                | S    | S    | S    | S    | S        | S        | S        | S        |
| 11  | $\pi$  | L  | L  | L  | S                | L                | L    | L    | S    | S    | S        | S        | S        | S        |
| 12  | $\pi$  | L  | L  | L  | S                | L                | L    | L    | S    | S    | S        | S        | S        | S        |
| 13  | $\pi$  | L  | L  | L  | S                | L                | L    | L    | S    | S    | S        | S        | S        | S        |
| 14  | $\pi$  | L  | L  | L  | S                | L                | L    | L    | S    | S    | S        | S        | S        | S        |
| 15  | $\pi$  | L  | L  | L  | S                | L                | L    | L    | S    | S    | S        | S        | S        | S        |
| 16  | $\pi$  | L  | L  | L  | S                | L                | L    | L    | S    | S    | S        | S        | S        | S        |

\[ \begin{array}{ccccccccccccc}
| 17  | 0      | L  | L  | L  | L                | L                | L    | L    | L    | L    | L        | L        | L        | L        |
| 18  | $\pi$  | 1  | 1  | 1  | S                | S                | L    | L    | S    | S    | S        | S        | S        | S        |
| 19  | $\pi$  | L  | L  | L  | L                | L                | L    | L    | S    | S    | S        | S        | S        | S        |
| 20  | $\pi$  | L  | L  | L  | S                | L                | L    | S    | L    | L    | S        | S        | S        | S        |
| 21  | $\pi$  | L  | L  | L  | S                | L                | L    | S    | L    | L    | S        | S        | S        | S        |
| 22  | $\pi$  | L  | L  | L  | S                | S                | L    | S    | L    | L    | S        | S        | S        | S        |
| 23  | $\pi$  | L  | L  | L  | L                | S                | L    | S    | S    | S    | S        | S        | S        | S        |
| 24  | $\pi$  | L  | L  | L  | S                | L                | S    | S    | S    | S    | S        | S        | S        | S        |

\[ \begin{array}{ccccccccccccc}
| 25  | 0      | L  | L  | L  | L                | L                | L    | L    | L    | L    | L        | L        | L        | L        |
| 26  | $\pi$  | L  | L  | L  | S                | L                | L    | L    | L    | L    | L        | L        | L        | L        |
| 27  | $\pi$  | L  | L  | L  | S                | L                | L    | L    | L    | L    | L        | L        | L        | L        |
| 28  | $\pi$  | L  | L  | L  | S                | L                | L    | L    | L    | L    | L        | L        | L        | L        |
| 29  | $\pi$  | L  | L  | L  | S                | L                | L    | L    | L    | L    | L        | L        | L        | L        |
| 30  | $\pi$  | L  | L  | L  | S                | L                | L    | L    | L    | L    | L        | L        | L        | L        |
| 31  | $\pi$  | L  | L  | L  | S                | L                | L    | L    | L    | L    | L        | L        | L        | L        |
| 32  | $\pi$  | L  | L  | L  | S                | L                | L    | L    | L    | L    | L        | L        | L        | L        |
The phase problem of x-ray crystallography

\[ |E_{-H-K}|, |E_{K+L}|, |E_{L-H}|, |E_{-H+K+L}|, \]  
(88)

i.e. eleven magnitudes \(|E|\) in all. Now the 16 rows of table 1 for \(Q_1\) lead to the eight rows 9–16 of table 2. Particularly noteworthy is the fact that the second neighbourhoods of \(Q_0\) and \(Q_1\) overlap and a main term of one becomes a cross-term of the other.

Table 2 is now completed in the obvious way, rows 17–24 corresponding to \(Q_2\) and rows 25–32 to \(Q_3\). Again the overlap of the second neighbourhoods is to be noted, and the fact that main-terms of one quintet appear as cross-terms in others is particularly noteworthy (because of the implication that, in order to arrive at a proper definition of the first neighbourhood of \(T\), it is necessary to implicate the second neighbourhoods of its extensions).

Certain rows of table 2 are mutually contradictory and others are mutually reinforcing. By combining those rows of table 2 which are mutually reinforcing, (e.g. rows 2 and 10 of table 2 give row 2 of table 3) one obtains, in view of (84), table 3 which leads directly to the first neighbourhood of the structure seminvariant \(T\), as shown next.

7.1.d. The first neighbourhood of \(T\). The first seven entries (under magnitudes \(|E|\)) in rows 2, 8, 11, and 20 of table 3 are identical, thus leading to row 2 of table 4. In a similar way all rows of table 4 are obtained. If one makes the definitions

\[ h = \frac{1}{2}(H + K + L), \quad h_1 = \frac{1}{2}(-H + K + L), \]
\[ h_2 = \frac{1}{2}(H - K + L), \quad h_3 = \frac{1}{2}(H + K - L), \]  
(89)

Table 4 shows that, on the assumption that all three magnitudes

\[ |E_h| = R_1, \quad |E_{h_1}| = R_2, \quad |E_{h_2}| = R_3, \]
(90)

are large, then \(T \approx 0\) if all four magnitudes

\[ |E_{h_3}| = r, \quad |E_{h_3_1}| = r_1, \quad |E_{h_3_2}| = r_2, \quad |E_{h_3_3}| = r_3 \]  
(91)

are large; but if precisely two of the four magnitudes (91) are large and the remaining two are small, then \(T \approx \pi\). Thus the first neighbourhood of \(T\) is defined to consist of the seven magnitudes (90) and (91), i.e., as it turns out, the set-theoretic union of the first neighbourhoods of all of its extensions. For the three-phase structure seminvariant in \(PI\) the “favourable cases” of the neighbourhood principle are defined by the entries of table 4, viz. the three magnitudes (90) are large and either all four magnitudes (91) are large, in which case \(T \approx 0\), or precisely two of the magnitudes (91) are large and the remaining two are small, in which case \(T \approx \pi\).

7.1.e. The conditional probability distribution \(P^\pm\) of the three-phase structure seminvariant \(T\), given the seven magnitudes in its first neighbourhood. Suppose that a crystal structure in \(PI\) is fixed and that the seven non-negative numbers \(R_1, R_2, R_3, r, r_1, r_2, r_3\) are also specified. Denote reciprocal space by \(W\) and by \(W \times W \times W\) the three-fold Cartesian product which consists of all ordered triples \((H, K, L)\) of reciprocal vectors. Suppose finally that \((H, K, L)\) is the primitive random variable which is assumed to be uniformly distributed over the subset of \(W \times W \times W\) defined by (90), (91), and the condition that the components of \(H + K + L\) be even. Then the structure seminvariant \(T\) [equation (75)], as a function of the primitive random variable \((H, K, L)\), is itself a random variable. Denote by \(P^+\) or \(P^-\) the conditional probability, given the seven magnitudes (90) and (91), that \(T\) be 0 or \(\pi\), respectively, or, equivalently, that \(\cos T = +1\).
Table 3. The probable values of the structure seminvariant $T$, given the magnitudes $|E|$ in the second neighbourhoods of its extensions; L means large; S means small, obtained by combining reinforcing rows in Table 2.

<table>
<thead>
<tr>
<th>Derived from rows of table 2</th>
<th>$T$</th>
<th>$\frac{1}{2}(H+K+L)$</th>
<th>$\frac{1}{2}(-H+K+L)$</th>
<th>$\frac{1}{2}(H-K+L)$</th>
<th>$H+K$</th>
<th>$H-K$</th>
<th>$L+H$</th>
<th>$K+L$</th>
<th>$L-H$</th>
<th>$K-L$</th>
<th>$H+K+L$</th>
<th>$H-K+L$</th>
<th>$L+H$</th>
<th>$K+L$</th>
<th>$L-H$</th>
<th>$K-L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 1, 9, 17, 25</td>
<td></td>
<td>L L L L L L L L L L L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 2, 10</td>
<td></td>
<td>L L L L L S S L S L L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 3, 19</td>
<td></td>
<td>L L L L S S L L S L L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 4, 28</td>
<td></td>
<td>L L L L S S L L S L L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 11, 27</td>
<td></td>
<td>L L L L S L L L L S L L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 12, 20</td>
<td></td>
<td>L L L L S L L L L S L L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 18, 26</td>
<td></td>
<td>L L L L S S L L L L S L L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 2, 14</td>
<td></td>
<td>L L L L L L S L S L S S S S S L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 3, 23</td>
<td></td>
<td>L L L L L S L S L S L L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 4, 32</td>
<td></td>
<td>L L L L L S L S L S L L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 6, 10</td>
<td></td>
<td>L L L L L L S L S L S S S S S L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12 11, 31</td>
<td></td>
<td>L L L L L L S L S L S L L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 12, 24</td>
<td></td>
<td>L L L L L L S L L L L S L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14 18, 30</td>
<td></td>
<td>L L L L L L S L L L L S L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 7, 19</td>
<td></td>
<td>L L L L L S L S S S S L L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 16, 20</td>
<td></td>
<td>L L L L L S L S L L S L S L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 22, 26</td>
<td></td>
<td>L L L L L S L L L L S L L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 15, 27</td>
<td></td>
<td>L L L L L L S L L L L S L L L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19 8, 28</td>
<td></td>
<td>L L L L L S L S L L S S S S S L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 6, 14</td>
<td></td>
<td>L L L L L L S S S S S S S S S L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 7, 23</td>
<td></td>
<td>L L L L L L S L S S S S S S S S L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 8, 32</td>
<td></td>
<td>L L L L L L S L S S S S S S S S L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 15, 31</td>
<td></td>
<td>L L L L L L S L S S S S S S S S L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24 16, 24</td>
<td></td>
<td>L L L L L L S L L S S S S S S S L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 22, 30</td>
<td></td>
<td>L L L L L L S L S S S S S S S S L L L L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Herbert Hauptman
Table 4. The probable value of the structure seminvariant $T$, given the seven magnitudes in its first neighbourhood; L means large; S means small; obtained by combining reinforcing rows from the first seven columns (under magnitudes $|E|$) of table 3.

<table>
<thead>
<tr>
<th>Row</th>
<th>Derived from rows of table 3</th>
<th>$T$</th>
<th>$H$</th>
<th>$K$</th>
<th>$L$</th>
<th>$\frac{1}{2}(H + K + L)$</th>
<th>$\frac{1}{2}(-H + K + L)$</th>
<th>$\frac{1}{2}(H - K + L)$</th>
<th>$\frac{1}{2}(H + K - L)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
</tr>
<tr>
<td>2</td>
<td>2, 8, 11, 20</td>
<td>π</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>3</td>
<td>3, 9, 15, 21</td>
<td>π</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>L</td>
<td>S</td>
</tr>
<tr>
<td>4</td>
<td>4, 10, 19, 22</td>
<td>π</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>S</td>
<td>L</td>
</tr>
<tr>
<td>5</td>
<td>5, 12, 18, 23</td>
<td>π</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>L</td>
<td>S</td>
<td>L</td>
</tr>
<tr>
<td>6</td>
<td>6, 13, 16, 24</td>
<td>π</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>L</td>
<td>S</td>
<td>L</td>
</tr>
<tr>
<td>7</td>
<td>7, 14, 17, 25</td>
<td>π</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>S</td>
<td>L</td>
<td>S</td>
<td>L</td>
</tr>
</tbody>
</table>

or $-1$, respectively. The formula for $P^\pm$, the major result of this section, is simply (Hauptman 1980)

$$P^\pm = \frac{1}{K} Z^\mp, \quad K = Z^+ + Z^-,$$

(92)

$$Z^\mp = \exp (\pm V) \sum_{\eta_1, \eta_2, \eta_3 = \pm 1} \exp (W^\pm),$$

(93)

$$V = \left[ \frac{1}{\sigma_2^{a_1/2}} (6 \sigma_3^3 - 6 \sigma_2 \sigma_3 \sigma_4 + \sigma_2^2 \sigma_3) (r^2 + r_1^2 + r_2^2 + r_3^2) - \frac{2}{\sigma_2^{a_1/2}} (2 \sigma_3^3 - 3 \sigma_2 \sigma_3 \sigma_4 + \sigma_2^2 \sigma_3) \right] R_1 R_2 R_3,$$

(94)

$$W^\pm = (\eta_1 r r_1 + \eta_2 \eta_3 r_2 r_3) \left[ \pm \frac{\sigma_3}{\sigma_2^{a_1/2}} R_1 - \frac{1}{\sigma_2} (2 \sigma_3^2 - \sigma_2 \sigma_4) R_2 R_3 \right]$$

$$+ (\eta_2 r r_2 + \eta_3 \eta_1 r_3 r_1) \left[ \pm \frac{\sigma_3}{\sigma_2^{a_1/2}} R_2 - \frac{1}{\sigma_2} (2 \sigma_3^2 - \sigma_2 \sigma_4) R_3 R_1 \right]$$

$$+ (\eta_3 r r_3 + \eta_1 \eta_2 r_1 r_2) \left[ \pm \frac{\sigma_3}{\sigma_2^{a_1/2}} R_3 - \frac{1}{\sigma_2} (2 \sigma_3^2 - \sigma_2 \sigma_4) R_1 R_2 \right],$$

(95)

where the summands of (93) consist of the eight exponentials obtained by permitting each of $\eta_1, \eta_2, \eta_3$ to take on the values $+1$ or $-1$ independently in (95); otherwise upper (lower) signs in (92)--(95) go together.

7.1f. The discriminant of $T$. Just as the discriminant of the quintet, a polynomial in the 15 magnitudes of the second neighbourhood, serves to identify and estimate those quintets having the extreme value 0 or $\pi$, so here there exists a polynomial in the seven magnitudes constituting the first neighbourhood of $T$, the discriminant, extreme values of which serve to estimate reliably the value of $T$. The discriminant $\Delta$ of the three-phase structure seminvariant in $P\Gamma$ is defined by
Herbert Hauptman

\[
\Delta = \left\{ \frac{\sigma_3^3}{\sigma_2^{3/2}} (r^2 r_2^2 + r^2 r_2^2 + r^2 r_3^2 + r^2 r_3^2 + r^2 r_4^2) \right\} - \frac{2\sigma_3^3 - \sigma_2^2 \sigma_3 \sigma_4}{\sigma_2^{3/2}} (r^2 r_1^2 + r^2 r_2^2 + r^2 r_3^2 + r^2 r_4^2 + r^2 r_5^2 + r^2 r_1^2) \\
+ \frac{1}{2\sigma_2^{3/2}} (6\sigma_3^3 - 6\sigma_2 \sigma_3 \sigma_4 + \sigma_2^2 \sigma_5) (r^2 + r_1^2 + r_2^2 + r_3^2) \\
- \frac{2}{\sigma_2^{3/2}} (2\sigma_3^3 - 3\sigma_2 \sigma_3 \sigma_4 + \sigma_2^2 \sigma_5) \right\} R_1 R_2 R_3. 
\] (96)

The applications also show that extreme values of the discriminant \( \Delta \) are well correlated with the values of \( T \) in the sense that \( T \approx 0 \) or \( \pi \) according as \( \Delta \approx 0 \) or \( \Delta \ll 0 \) respectively. Thus, as has been observed many times previously (e.g. Fortier and Hauptman 1977c) the discriminant proves to be a useful and readily computable indicator of the value of the seminvariant.

Inspection of (96) shows that, on the assumption that the three magnitudes (90) are large, \( \Delta \approx 0 \) if all four magnitudes (91) are large; but \( \Delta \ll 0 \) if precisely two magnitudes (91) are large and the remaining two are small. In short \( T \approx 0 \) or \( \pi \) in the respective cases, and the qualitative prediction of table 4 is in fact borne out. Inspection of the distribution, (92)–(95), leads to the same conclusion, but in a less transparent way.

Finally, it is not difficult to show that if one of the magnitudes (91) is outside the observed range, then it is to be replaced by unity, the average value of \( |E|^2 \), in (96) or (92)–(95) (see, for example, Giacovazzo 1975; Heinerman et al 1977).

7.1g. Concluding remarks. In the present section the conditional probability distribution of the three-phase structure seminvariant \( T \), given the seven magnitudes in its first neighbourhood, has been found for space group \( P\bar{1} \). The distribution yields a reliable estimate (0 or \( \pi \)) for \( T \) in the favourable case that the variance of the distribution happens to be small. Because the structure seminvariants are available in much larger numbers than the structure invariants, it is anticipated that the results described here will find important application in the determination of real structures, and some initial calculations confirm this expectation.

Of particular importance is the analysis leading to the definition of the first neighbourhood given in §§7.1b to 7.1d. This work has been written in such a way as to permit easy generalization to the higher order structure seminvariants, and the final results for the four- and five-phase structure seminvariants in \( P\bar{1} \) have been derived. It is expected also that the work described here will serve as a useful guideline for the derivation of analogous distributions in other space groups, noncentrosymmetric as well as centrosymmetric.

8. Direct methods and isomorphous replacement

8.1 Introduction

Direct methods are routinely used nowadays for the solution of small molecular structures, i.e. those containing fewer than some 100 independent non-hydrogen atoms per unit cell. For the solution of macromolecular structures on the other hand, the method of isomorphous replacement is universally used. One naturally anticipates that
the ability to combine the two techniques would yield methods more powerful than either. This fusion of the two techniques is described here, and the initial applications (Hauptman 1982a; Hauptman et al 1982) strongly suggest that the anticipated gain in power is in fact realized.

8.2 The normalized structure factors

If \( f_j \) and \( g_j \) denote atomic structure factors for a corresponding pair of isomorphous structures in \( P1 \), then the respective normalized structure factors \( E_H \) and \( G_H \) are defined by

\[
E_H = E_H \exp(i\phi_H) = \frac{1}{\alpha_{1/2}} \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_j),
\]

\[
G_H = |G_H| \exp(i\psi_H) = \frac{1}{\alpha_{1/2}} \sum_{j=1}^{N} g_j \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_j),
\]

where

\[
\alpha_{mn} = \sum_{j=1}^{N} f_j^{m} g_j^{n},
\]

some of the \( f_j \)’s (or \( g_j \)’s) may be zero (or negative in the neutron diffraction case) and \( \mathbf{r}_j \) is the position vector of the atom labelled \( j \). This formulation includes not only the case of two strictly isomorphous structures but also, for example, the special case that the \( f \) structure is a native protein and the \( g \) structure a heavy-atom isomorphous derivative, as well as the case that one or both sets of data are obtained by neutron diffraction. The \( f_j \) and \( g_j \) are zero-angle atomic scattering factors and are therefore equal to the atomic numbers \( Z_j \) in the x-ray diffraction case but may be negative in the neutron diffraction case.

8.3 The structure invariants

If \( \mathbf{H}, \mathbf{K}, \) and \( \mathbf{L} \) are reciprocal lattice vectors satisfying

\[
\mathbf{H} + \mathbf{K} - \mathbf{L} = 0
\]

then, in sharp contrast to the non-isomorphous case, there exist eight different kinds of three-phase structure invariants:

\[
\zeta_0 = \phi_H + \phi_K + \phi_L,
\]

\[
\zeta_1 = \phi_H + \phi_K + \psi_L, \quad \zeta_1' = \phi_H + \psi_K + \phi_L, \quad \zeta_1'' = \psi_H + \phi_K + \phi_L,
\]

\[
\zeta_2 = \phi_H + \psi_K + \psi_L, \quad \zeta_2' = \psi_H + \phi_K + \psi_L, \quad \zeta_2'' = \psi_H + \psi_K + \phi_L,
\]

\[
\zeta_3 = \psi_H + \psi_K + \psi_L.
\]

8.4 The first neighbourhood of each of the three-phase structure invariants (101)–(104)

Again in contrast to §6.1a, the first neighbourhood of each of the three-phase structure invariants (101)–(104) is now defined to consist of the six magnitudes

\[
|E_H|, |E_K|, |E_L|,
\]

\[
|G_H|, |G_K|, |G_L|.
\]
Thus estimates, for example, of even the "pure" structure invariants $\zeta_0$ [equation (101)] are made to depend on the values of six magnitudes, (105) and (106), rather than on the three magnitudes (105) alone, as in the non-isomorphous case.

8.5 The conditional probability distribution of the three-phase structure invariant

It is assumed that an isomorphous pair of structures in $P1$ with atomic position vectors $r_1, r_2, \ldots, r_N$ is fixed and that the six non-negative numbers $R_1, R_2, R_3, S_1, S_2, S_3$ are specified. Denote reciprocal space by $W$ and by $W \times W \times W$ the three-fold cartesian product consisting of all ordered triples (H, K, L) of reciprocal vectors. Suppose that the primitive random variable is the triple (H, K, L) which is assumed to be uniformly distributed over the subset of $W \times W \times W$ defined by (100) and

$$
\begin{align*}
|E_H| &= R_1, \quad |E_K| = R_2, \quad |E_L| = R_3; \\
|G_H| &= S_1, \quad |G_K| = S_2, \quad |G_L| = S_3,
\end{align*}
$$

where the normalized structure factors $E$ and $G$ are defined in (97)-(99). Then the eight three-phase structure invariants, $\zeta_0, \zeta_1, \zeta_1, \ldots, \zeta_3$ [equations (101)-(104)], as functions of the primitive random variable (H, K, L), are themselves random variables. The conditional probability distribution of $\zeta_0$, for example, assuming as known the six magnitudes $|E_H|, |E_K|, |E_L|, |G_H|, |G_K|, |G_L|$ [equations (107) and (108)] is given by (Hauptman 1982a)

$$
P_0(\Omega_0|R_1, R_2, R_3, S_1, S_2, S_3) \approx \frac{1}{K_0} \exp (A_0 \cos \Omega_0),
$$

where

$$
K_0 = 2\pi I_0(A_0),
$$

and $A_0$ in turn is explicitly expressible in terms of the atomic scattering factors $f_j, g_j, j = 1, 2, \ldots, N$, and the six magnitudes $|E_H|, |E_K|, |E_L|, |G_H|, |G_K|, |G_L|$ [equations (107) and (108)]. Similar formulas hold for $P_1(\Omega_1|R_1, R_2, R_3, S_1, S_2, S_3)$, the conditional probability distribution of $\zeta_1$, assuming as known the six magnitudes $|E_H|, |E_K|, |E_L|, |G_H|, |G_K|, |G_L|$ [equations (107) and (108)], etc. (109) is seen to be identical in form to (22). However the parameter $A_0$ of (109), in sharp contrast to the parameter $A$ of (22), may be positive or negative. Thus (109) has a single maximum in the interval (0, $2\pi$) and this maximum occurs at 0 or $\pi$ according to $A_0 > 0$ or $A_0 < 0$. Again, the larger the value of $|A_0|$ the smaller is the variance of the distribution. Thus, when $|A_0|$ is large one obtains the reliable estimate

$$
\zeta_0 \approx 0 \text{ or } \zeta_0 \approx \pi
$$

(111)

according as

$$
A_0 \gg 0 \text{ or } A_0 \ll 0,
$$

(112)

respectively. Just as (24) has proved to be of great importance in the development of the traditional techniques of direct methods, it is natural to assume that the estimates (111), and the like estimates for the remaining invariants (101)-(104), will play a similar role in the development of techniques combining the traditional direct methods and isomorphous replacement. In particular, one anticipates the existence of a generalized tangent formula which takes into account the estimates (111).
9. Direct methods and anomalous dispersion

9.1 Introduction

It has been known for a long time (Peerdeman and Bijvoet 1956; Ramachandran and Raman 1956; Okaya and Pepinsky 1956) that the presence of one or more anomalous scatterers facilitates the solution of the phase problem; and some recent work (Kroon et al 1977; Henerman et al 1978) employing Bijvoet inequalities and the double Patterson function, leads in a similar way to estimates of the sines of the three-phase structure invariants. This work strongly suggests that the ability to integrate the techniques of direct methods, in particular the recent advances in the mathematical formalism, with anomalous dispersion would lead to improved methods for phase determination. The fusion of these techniques is described here and the anticipated improvement is in fact realized (Hauptman 1982b). Not only do the new formulas lead to improved estimates of the structure invariants but, more important still, because the distributions derived here are unimodal in the whole interval \((0, 2\pi)\), the two-fold ambiguity inherent in all the earlier work is removed. It is believed that this resolution of the two-fold ambiguity results from the ability now to make use of the six individual magnitudes in the first neighbourhood of the structure invariant and the avoidance of explicit dependence on the Bijvoet differences; the explicit use of the Bijvoet differences, as is done in all previous work, leads apparently to a loss of information resulting in a two-fold ambiguity in estimates of the structure invariants. Since, in the presence of anomalous scatterers, the observed intensities are known to determine a unique enantiomorph, and therefore unique values for all the structure invariants, formulas of the kind described here should not be unexpected; nevertheless not even their existence appears to have been anticipated.

9.2 The normalized structure factors

In the presence of anomalous scatterers the normalized structure factor

\[
E_H = |E_H| \exp(i\phi_H),
\]

is defined by

\[
E_H = \frac{1}{\alpha_H^{1/2}} \sum_{j=1}^{N} f_{jH} \exp(2\pi iH \cdot r_j)
\]

\[
\alpha_H = \sum_{j=1}^{N} |f_{jH}|^2.
\]

For a normal scatterer, \(\delta_{jH} = 0\); for an atom which scatters anomalously, \(\delta_{jH} \neq 0\). Owing to the presence of the anomalous scatterers the atomic scattering factors \(f_{jH}\), as functions of \(\sin \theta/\lambda\), do not have the same shape for different atoms, even ap-
proximately. Hence the dependence of the $f_H$ on $|H|$ cannot be ignored, in contrast to the usual practice when anomalous scatterers are not present. For this reason the subscript $H$ is not suppressed in the symbols $f_H$ and $\alpha_H$ [equation (117)].

9.3 The structure invariants

It will be assumed throughout that $H, K, L$ are fixed reciprocal-lattice vectors satisfying

$$H + K + L = 0. \quad (118)$$

Owing to the breakdown of Friedel's law there are now eight distinct three-phase structure invariants:

$$\zeta_0 = \phi_H + \phi_K + \phi_L, \quad (119)$$

$$\zeta_1 = -\phi_H + \phi_K + \phi_L, \quad (120)$$

$$\zeta_2 = \phi_H - \phi_K + \phi_L, \quad (121)$$

$$\zeta_3 = \phi_H + \phi_K - \phi_L, \quad (122)$$

$$\zeta_0 = \phi_H + \phi_K + \phi_L, \quad (123)$$

$$\zeta_1 = -\phi_H + \phi_K + \phi_L, \quad (124)$$

$$\zeta_2 = \phi_H - \phi_K + \phi_L, \quad (125)$$

$$\zeta_3 = \phi_H + \phi_K - \phi_L. \quad (126)$$

9.4 The first neighbourhood of each of the three-phase structure invariants (119)–(126)

In analogy with §8.4, the first neighbourhood of each of the three-phase structure invariants (119)–(126) is defined to consist of the six magnitudes

$$|E_H|, |E_K|, |E_L|, \quad (127)$$

$$|E_H|, |E_K|, |E_L|, \quad (128)$$

which, because of the breakdown of Friedel's Law, are not, in general, equal in pairs.

9.5 The conditional probability distribution of the three-phase structure invariant, assuming as known the six magnitudes in its first neighbourhood.

Fix the reciprocal-lattice vectors $H, K,$ and $L$, subject to (118). Suppose that the six non-negative numbers $R_1, R_2, R_3, R_7, R_7,$ and $R_3$ are also specified. Define the $N$-fold Cartesian product, $V$, to consist of all ordered $N$-tuples $(r_1, r_2, \ldots, r_N)$, where $r_1, r_2, \ldots, r_N$ are atomic position vectors. Suppose that the primitive random variable is the $N$-tuple $(r_1, r_2, \ldots, r_N)$ which is assumed to be uniformly distributed over the subset of $V$ defined by

$$|E_H| = R_1, |E_K| = R_2, |E_L| = R_3, \quad (129)$$

$$|E_H| = R_7, |E_K| = R_7, |E_L| = R_7, \quad (130)$$

where the normalized structure factors $E$ are defined by (114). Then the eight structure invariants

$$\zeta_j, \zeta_j, j = 0, 1, 2, 3, \quad (131)$$
The phase problem of x-ray crystallography

(119)–(126), as functions of the primitive random variables \((r_1, r_2, \ldots, r_N)\), are themselves random variables. Denote by

\[
P_j(\Omega_j) = P_j(\Omega_j | R_1, R_2, R_3, R_T, R_T, R_T),
\]

the conditional probability distribution of the three-phase structure invariant \(\zeta_j\), \(j = 0, 1, 2, 3, 0, 1, 2, 3\), given the six magnitudes (129) and (130) in its first neighbourhood. Then the major result is that the distributions \(P_j(\Omega_j)\) have the pure exponential form (Hauptman 1982b):

\[
P_j(\Omega_j) = \frac{1}{K_j} \exp \{ A_j \cos (\Omega_j - \omega_j) \}, \ j = 0, 1, 2, 3, 0, 1, 2, 3,
\]

where the parameters \(K_j, A_j\) and \(\omega_j\) are expressible in terms of the complex scattering factors \(f_H, f_K, f_L\), presumed to be known, and the observed magnitudes \(|E_H|, |E_K|, |E_L|, |E_H|, |E_K|, |E_L|\). No prior knowledge of the positions of the anomalous scatterers is needed, nor is it required that the anomalous scatterers be identical. Since the \(K_j\)'s and \(A_j\)'s are positive, the maximum of (133) occurs at \(\Omega_j = \omega_j\). Hence when the variance of the distribution is small, \(i.e.\) when \(A_j\) is large, one obtains the reliable estimate

\[
\zeta_j = \omega_j, \ j = 0, 1, 2, 3, 0, 1, 2, 3
\]

for the structure invariant \(\zeta_j\), [(119)–(126)]. Once again a generalized tangent formula, based on the estimates (134) for the invariants \(\zeta_j\), is readily formulated and is expected to yield estimates for the individual phases in the case that anomalous scatterers are present.

10. Concluding remarks

The basic principles of the direct methods of x-ray crystallography have been described. The fundamental role of the structure seminvariants, that they serve to link the observed magnitudes with the desired phases of the structure factors, has been emphasized. The neighbourhood principle, which affirms the intimate relationship which exists between a structure seminvariant \(T\) and one or more small sets of magnitudes \(|E|\), the neighbourhoods of \(T\), has been formulated. Reduction of the probabilistic theory of the structure seminvariants \(T\) to that of the structure invariants has been accomplished by embedding \(T\) and related seminvariants in suitable structure invariants, the extensions of \(T\). Estimates of the structure invariants and seminvariants \(T\) are obtained by means of appropriate conditional probability distributions, assuming as known the magnitudes \(|E|\) in the neighbourhoods of \(T\). Finally, the traditional techniques of direct methods have been combined with isomorphous replacement and anomalous dispersion, thus opening up the possibility that the existing machinery of direct methods may be carried over without essential change to facilitate the solution of macromolecular structures \(via\) isomorphous replacement and anomalous dispersion.

References

Herbert Hauptman

Giacovazzo C 1976a Acta Crystallogr. A32 74
Giacovazzo C 1976b Acta Crystallogr. A32 91
Giacovazzo C 1980c Acta Crystallogr. A36 704
Hauptman H 1972 crystal structure determination: The role of the cosine seminvariants (New York: Plenum Press)
Hauptman H 1975a Acta Crystallogr. A31 671
Hauptman H 1977a Acta Crystallogr. A33 553
Hauptman H 1977b Acta Crystallogr. A33 556
Hauptman H 1977d Acta Crystallogr. A33 568
Hauptman H 1978 Acta Crystallogr. A34 525
Hauptman H 1982a Acta Crystallogr. A38 289
Hauptman H 1982b Acta Crystallogr. A38 632
Karle J and Hauptman H 1958 Acta Crystallogr. 11 264
Okaya Y and Pepinsky R 1956 Phys. Rev. 103 1645
Putten N van der and Schenk H 1977 Acta Crystallogr. A33 856
Putten N van der, Schenk H and Hauptman H 1980b Acta Crystallogr. A36 897
The phase problem of x-ray crystallography

Ramachandran G N and Raman S 1956 *Curr. Sci.* **25** 348
Schenk H 1973 *Acta Crystallogr.* **A29** 480
Schenk H 1974 *Acta Crystallogr.* **A30** 477
Schenk H 1975 *Acta Crystallogr.* **A31** S14
Schenk H and De Jong J G H 1973 *Acta Crystallogr.* **A29** 31
Timkovich L and Dickerson R E 1976 *J. Biol. Chem.* **251** 4033
Woolfson M 1954 *Acta Crystallogr.* **7** 61
The efficient generation of phase sets by the random approach

K HENDERSON and M M WOOLFSON
Department of Physics, University of York, Heslington, York YO1 5DD, UK

Abstract. It is shown that a large number of plausible sets of phases satisfying triple-phase relationships may be found very efficiently and cheaply. Sets of phases satisfying the original set of phase relationships ($\phi$s) may be combined with 'solutions' ($\psi$s) for the set of relationships with any translationally-derived constant angles removed. An application to a known structure demonstrates the validity of the technique and a theoretical interpretation is given for a case where a true solution results from a $\phi$ and $\psi$ combination.

Keywords. Direct methods; phase sets; random approach

1. Introduction

An individual triple-phase relationship, with phases expressed in cycles, may be written as

$$\phi_p \pm \phi_q \pm \phi_r + b \approx n,$$

(1)

where the constant angle $b$ may arise from translational symmetry, $\approx$ means 'is probably close to' and $n$ is some integer. A complete set of such relationships may be written in matrix notation as

$$\kappa(A\phi + b) \approx \kappa n,$$

(2)

where the elements of the diagonal matrix $\kappa$ are the usual $\kappa$ values, $2N^{-\frac{1}{2}} |E_pE_qE_r|$ for an equal-atom structure.

In the YZARC method, developed from ideas given by Baggio et al (1978), initial random values are assigned to about 100 phases which are subsequently refined by an iterative process to satisfy equation (2). Some 50–100 sets of refined phases are then extended by the tangent formula and thereafter treated and evaluated in the same way as sets of phases generated by MULTAN 80 (Main et al 1980).

The YZARC method has been extremely successful in solving structures with up to 100 atoms or so in the asymmetric unit. However it transpires that, in general, the larger the structure the greater is the number of phase sets which need to be examined in order for a correct one to appear. For a small protein it may be necessary for many thousands of phase sets to be generated in order to obtain just one which may reveal the structure. This would be very expensive in computer time if done in a straightforward way but, as will be demonstrated, an economical process is available.

2. $\psi$ sets

Consider a set of equations

$$\kappa A\psi = \kappa m$$

(3)
where $A$ is the same matrix as in (2) and $m$ is a vector whose elements are close to integers. The $\text{YZARC}$ process may be used to find $\psi$'s which satisfy (3).

A $\phi$ solution of (2), $\phi_i$, and a $\psi$ solution of (3), $\psi_j$, are combined to give

$$\kappa\{A(\phi_i \pm \psi_j) + b\} \approx \kappa(n_i \pm m_j).$$

(4)

Since the elements of $n_i \pm m_j$ will be close to integers then $\phi_i \pm \psi_j$ will be near solutions of (2). If there are $P$ $\phi$ solutions and $Q$ $\psi$ solutions then there will be $P(2Q + 1)$ available solutions or near-solutions of (2). The advantage of this is apparent when we consider a numerical case. With $P = Q = 50$ then $P(2Q + 1) = 5050$ and this number of possible phase sets is obtained with little more computer time than is required to generate 100 sets of phases with $\text{YZARC}$.

A question which arises is the following: 'For each $\phi$ solution is there a $\psi$ solution which can be combined with it to give an almost-correct set of phases?'

If $\phi_i$ is the set of true phases then for a particular solution of (2), $\phi_i$, a correction vector would be

$$\Delta \phi_i = \phi_i - \phi_i.$$

If $\Delta \phi_i$ is substituted into (3) and refined and if the refined elements are close to their initial values ($\leq 40^\circ$ different, say) then we may assume that there is an obtainable $\psi$ which will, combined with $\phi_i$, give a substantially correct solution.

Tests with several structures indicate that for about 20% of $\phi$ solutions there is a potentially-available $\psi$ which will lead to a structure solution.

3. Figures of merit for small sets of phases

In the process which is to be described it is necessary to associate a figure of merit (FOM), a measure of plausibility, to sets of 100 or less phases. It is well-known that the FOMS used in MULTAN are very effective but these are applied to phase sets which define the structure well, in the sense that a Fourier synthesis would show the correct structure, or most of it, quite clearly.

We have found a group of FOMS, suitable for a small data set, which can be expressed in terms of the quantities

$$\alpha_h = \int |\sum \kappa_{h,h'} \exp \{i(\phi_h - \phi_{h'} - \phi_{h-h'})\}|.$$

These are:

(i) $\text{ABSFOM} = \sum_h |\alpha_h|/\sum_h \alpha_{h,e}$

(6)

(ii) $R_1 = \sum_h |\alpha_{h,e} - k\alpha_h|/\sum_h \alpha_{h,e}$

(7)

(iii) $R_2 = \{\sum_h (S\alpha_{h,e} - \alpha_h)^2/\sum_h \alpha_{h,e}^2\}^{1/2}$

(8)

and (iv) $\text{MR} = \min_h (\alpha_h/\alpha_{h,e})$

(9)
Efficient generation of phase sets

where an estimate of the theoretically expected value of $\alpha_h$ is

$$\alpha_{h,e} = \sum_{h'} k_{h,h'} I_1 (k_{h,h'})/I_0 (k_{h,h'})$$

(10)

$$k = \sum_{h} \alpha_{h,e}/\sum_{h} \alpha_h,$$

(11)

and $S = \sum_{h} \alpha_{h,e} \alpha_h/\sum_{h} \alpha_{h,e}^2$

(12)

which makes $R_2$ a minimum.

These foms are not good as absolute figures of merit but they do seem to work quite well as ranking figures. In a large number of experiments it has been found that combining a $\psi$ with a $\phi$ usually lowers $\text{ABSFOM}$. For a few combinations there is a small increase in $\text{ABSFOM}$.

Since a promising $\phi \pm \psi$ should have a reasonably large $\text{ABSFOM} (\geq 1$, say) then only those $\phi$ solutions for which $\text{ABSFOM} \geq 0.95$ have been used for combining with $\psi$s. This rule of thumb may need to be modified in the light of further experience.

4. A trial

The idea of combining $\phi$ and $\psi$ sets was tested with the known structure of cortisone, $C_{21}H_{28}O_5; P2_1 2_1 2_1; a = 10.040, b = 23.649, c = 7.784 \text{ A}; Z = 4$ (figure 1).

For a normal $\text{YZARC}$ run 100 sets of 100 phases were refined to satisfy equation (2), expanded to 204 phases and then analysed by the usual $\text{MULTAN}$ procedure. Four correct solutions were found—sets 46, 57, 59 and 98; these revealed all the non-hydrogen atoms of the structure.

For each of the 100 $\phi$s (each containing 100 phases) from the $\text{YZARC}$ run $\text{FOMS}$ were calculated. There were then derived 51 $\psi$ sets by running $\text{YZARC}$ with (3). Each $\phi$ for which $\text{ABSFOM} \geq 0.95$ was combined with each $\psi$ solution and a $\phi \pm \psi$ combination was investigated further if its $\text{ABSFOM} \geq 0.95$ and if its $R_1$ was smaller than that for $\phi$ alone.

Each of the 384 combinations satisfying these criteria was extended to 204 phases by the tangent formula and then subjected to the usual $\text{MULTAN}$ figures of merit. Of the 384 combinations 26 gave the entire structure.

For each of the original $\phi \pm \psi$ combinations (with 100 phases) the $\text{FOMS}$ (6) to (9) were calculated and a combined figure of merit, $\text{CFOM}$, produced from them (as is done in

![Figure 1. A molecule of cortisone.](image-url)
Table 1. The 26 combinations giving a solution for cortisone

<table>
<thead>
<tr>
<th>Set</th>
<th>φ</th>
<th>ψ</th>
<th>CFOM</th>
<th>Set</th>
<th>φ</th>
<th>ψ</th>
<th>CFOM</th>
</tr>
</thead>
<tbody>
<tr>
<td>71</td>
<td>14</td>
<td>21</td>
<td>3.30</td>
<td>378*</td>
<td>98</td>
<td>-3</td>
<td>2.67</td>
</tr>
<tr>
<td>290*</td>
<td>59</td>
<td>-32</td>
<td>3.12</td>
<td>212*</td>
<td>46</td>
<td>3</td>
<td>2.58</td>
</tr>
<tr>
<td>315</td>
<td>71</td>
<td>31</td>
<td>2.94</td>
<td>365</td>
<td>92</td>
<td>-50√</td>
<td>2.55</td>
</tr>
<tr>
<td>215*</td>
<td>46</td>
<td>-19</td>
<td>2.92</td>
<td>188</td>
<td>43</td>
<td>-50√</td>
<td>2.55</td>
</tr>
<tr>
<td>220*</td>
<td>46</td>
<td>24</td>
<td>2.90</td>
<td>189</td>
<td>43</td>
<td>-50√</td>
<td>2.55</td>
</tr>
<tr>
<td>219*</td>
<td>46</td>
<td>-24</td>
<td>2.81</td>
<td>211*</td>
<td>46</td>
<td>-3</td>
<td>2.54</td>
</tr>
<tr>
<td>223*</td>
<td>46</td>
<td>-32</td>
<td>2.80</td>
<td>153</td>
<td>31</td>
<td>-50√</td>
<td>2.54</td>
</tr>
<tr>
<td>216*</td>
<td>46</td>
<td>19</td>
<td>2.78</td>
<td>366</td>
<td>92</td>
<td>-50√</td>
<td>2.49</td>
</tr>
<tr>
<td>270*</td>
<td>57</td>
<td>19</td>
<td>2.77</td>
<td>278*</td>
<td>57</td>
<td>51</td>
<td>2.46</td>
</tr>
<tr>
<td>269*</td>
<td>57</td>
<td>-19</td>
<td>2.77</td>
<td>152</td>
<td>31</td>
<td>-50√</td>
<td>2.46</td>
</tr>
<tr>
<td>273*</td>
<td>57</td>
<td>-24</td>
<td>2.73</td>
<td>264*</td>
<td>57</td>
<td>-3</td>
<td>2.45</td>
</tr>
<tr>
<td>274*</td>
<td>57</td>
<td>24</td>
<td>2.70</td>
<td>279*</td>
<td>57</td>
<td>32</td>
<td>2.45</td>
</tr>
<tr>
<td>381*</td>
<td>98</td>
<td>-32</td>
<td>2.68</td>
<td>265*</td>
<td>57</td>
<td>3</td>
<td>2.34</td>
</tr>
</tbody>
</table>

* ψ^ trivial with all elements nearly equal zero.
+ These ϕs are identical.
√ ψ^50 has all elements close to 0 or π.

MULTAN). Table 1 shows the 26 solutions in ranking order of CFOM.

Set 71 had the highest CFOM of all the 384 combinations produced and led to an E map of exceptionally high quality. It is also worth noting that set 265, the correct solution with the lowest CFOM was 161st in ranking order for the 384 combinations so that all the correct solutions were comfortably contained in the top half of the ranking order.

5. An interpretation

Set 71, that with the highest CFOM, corresponded to the combination φ_{14} + ψ_{21}. Figures of merit were:

\[
\begin{align*}
\text{for } &\phi_{14} \quad \text{ABSFOM} = 1.000 & R_1 = 0.220 \\
&\text{R}_2 = 0.231 & \text{MR} = 0.059
\end{align*}
\]

\[
\begin{align*}
\text{for } &\phi_{14} + ψ_{21} \quad \text{ABSFOM} = 1.060 & R_1 = 0.201 \\
&\text{R}_2 = 0.198 & \text{MR} = 0.290.
\end{align*}
\]

It can be seen that all the figures of merit are better for the combination than for φ_{14} alone. This type of behaviour is rare and in our limited experience seems nearly always to indicate that a correct solution has been found by combination.

In figure 2(a) there is shown a peak-interpreted E map for φ_{14} and in figure 2(b) the corresponding map for φ_{14} + ψ_{21}. It is clear that φ_{14} is showing a partially correct fragment of the molecule but displaced. In figure 2(b) the peaks are ringed corresponding to part of the configuration shown in figure 2(a). The coordinates of two corresponding points are shown and it is clear that the displacement of the fragment in figure 2(a) is approximately (0.44, 0, 0).

Let us assume that for each phase

\[
\phi_{ih} + ψ_{ih} = \phi_{ih},
\]

(13)
Efficient generation of phase sets

![Diagram of molecule](image)

Figure 2. a. The fragment from $\phi_{1+}$. b. The molecule from $\phi_{1+} + \psi_{21}$.

where $\phi_{1h}$ is the true phase. Then since

$$E_h = |E_h| \exp (i\phi_{1h}),$$

we may write

$$E_h = |E_h| \exp (i\phi_{1h}) \exp (i\psi_{21}).$$

A 'true' $E$ map is the Fourier transform of the set of $E_h$. Hence it is the convolution of the $E$ map obtained using the phases $\phi_i$ with the map obtained using phases $\psi_j$ and unit amplitudes. This latter map we may call the error map (Silva and Viterbo 1980).

The error map calculated from $\psi_{21}$ has as its largest peaks:

<table>
<thead>
<tr>
<th>Height</th>
<th>$x$</th>
<th>$y$</th>
<th>$z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3931</td>
<td>0.4326</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>2108</td>
<td>0.4312</td>
<td>0.0941</td>
<td>0.8673</td>
</tr>
<tr>
<td>1927</td>
<td>0.4358</td>
<td>0.0901</td>
<td>0.1737</td>
</tr>
<tr>
<td>1875</td>
<td>0.4330</td>
<td>0.0036</td>
<td>0.6953</td>
</tr>
<tr>
<td>1594</td>
<td>0.0346</td>
<td>0.0276</td>
<td>0.3380</td>
</tr>
</tbody>
</table>

We can recognise the highest peak as the shift in the fragments displayed in the two parts of figure 2. This way of thinking about errors in sets of phases generated by direct methods can be most productive and may lead to even more efficient processes for generating large numbers of plausible phase sets.
6. Conclusion

The tests carried out with cortisone show that the idea of combining $\phi$ and $\psi$ solutions is basically sound. Some questions to which answers are needed are:

(i) Can we recognise and eliminate duplicate $\phi$ and $\psi$ sets?
(ii) Is the number of different $\psi$ sets fairly small and if so how do we find all of them?
(iii) Can we find even better figures of merit for small sets of phases?

For really complicated structures (even small proteins) many thousands or tens of thousands of sets of phases must be investigated. While it seems that we can easily generate many sets we must find good and fast ways to evaluate their plausibility.

Acknowledgements

One of us (KH) is grateful to the Wellcome Trust for support during the period in which this work was being done. The generous support of the Science and Engineering Research Council for this work is also acknowledged.

References

An assessment of criteria of fit in Patterson search

C E NORDMAN
Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA

Abstract. The problem of reliably detecting a known set of \( n \) vectors of weight \( w \), embedded in a heavily overlapped Patterson function \( P \), is investigated by a Monte-Carlo simulation based on searches of computer-generated random number sequences. Several formulations of the criterion of fit were compared. All were found to improve when the criterion was based on a subset of \( m \) "worst fitting" vectors as judged by a low value of \( (R/w) \). The best criteria were \( \Sigma \) \( w \) \( w \), with \( m \approx (0.4-0.5)n \), \( \Sigma \) \( P \) \( w \), with \( m \approx 0.3n \), and \( \Sigma \) \( P \) \( w \), with \( m \approx 0.7n \). In each case the detectability of the embedded vectors \( w \) increases with increasing \( \sigma(w) \) in relation to \( \sigma(N) \), the standard deviation of the overlaid noise. A related simulation of a Patterson search for non-crystallographic symmetry shows that for a given size of the non-crystallographically symmetric region, the detectability increases with the order (2-fold, 6-fold, 12-fold) of the symmetry.

Keywords. Patterson search; signal detection; Monte-Carlo simulation.

1. Introduction

Crystallographic structure-solving techniques based on interpretation of the Patterson function typically involve computer-implemented systematic sampling or exhaustive search of the Patterson function. The manner in which the Patterson function is sampled reflects the nature of the available \emph{a priori} information concerning the unknown crystal structure.

This information may be the knowledge of the structure of some (rigid) fragment present in the molecule. In this case the Patterson function is sampled at points corresponding to the set of vectors within this fragment, or between two properly oriented copies of the fragment. Alternatively, the available information may be the knowledge that the molecule—perhaps a multisubunit protein, or a virus—possesses local non-crystallographic symmetry of a particular kind. In this case the Patterson function may be sampled with a movable "symmetry grid," an array of points in spherical polar coordinates, which possesses the exact point group symmetry of the molecules (Nordman 1980a).

A recent review (Nordman 1980b) describes these methods in greater detail and gives examples of their use in small-molecule crystallography.

In either type of search a criterion of fit, or "image-seeking function," is evaluated at each step in the search. Collectively, these values constitute a "map," generally in three dimensions, the coordinates being angular or translational depending on the nature of the search. Promising orientations or translational positions of the search object are indicated by maxima, or minima, in the map.

In small-molecule problems, where the rigid fragment constitutes most or all of the molecule, the exact formulation of the criterion of fit is not very crucial. Criteria used in such cases include maximizing the sum of the sampled values of the Patterson function (Braun \emph{et al} 1969), or minimizing the sum of the squares of the difference between the
fragment Patterson and the crystal Patterson at all points where the former, unacceptably, exceeds the latter (Huber 1965).

The problem of making the search as discriminating as possible was first considered by Schilling (1970). He showed that it is advantageous to sort the sampled Patterson values $P_i$ and the weights $w_i$ of the known sampling vectors, in order of increasing values of the ratio $P_i/w_i$. The lower the value of this ratio, the worse is the fit, that is, the more poorly is the fragment vector peak $w_i$ accommodated by the Patterson value $P_i$. By including in the criterion of fit only those $(w_i, P_i)$ values which are most discriminating, as indicated by low $P_i/w_i$, Schilling showed that more reliable search results were obtained. The “minimum average” is defined as

$$\text{MIN}(m, n) = \sum_{i=1}^{m} P_i / \sum_{i=1}^{m} w_i \quad m < n,$$

where $n$ is the total number of fragment vectors used, and $m$ is the size of the subset having the $m$ lowest values of $P_i/w_i$. This criterion of fit has found wide use, typically with $m/n = 0.1 - 0.3$.

Another reasonable criterion of fit is the sum of the products of the search vector weights $w_i$ and the sampled Patterson values $P_i$. This function

$$SP = \sum_{i=1}^{n} w_i P_i$$

(2)

tends to be high at the correct solution. Recognizing that $w_i$ represents points in a “model” Patterson, $P_m$, and replacing the sum with an integral, it is seen that (2) is related to $\int P_m(r) P(r) \, dr$. This integral, evaluated as a sum of products of Patterson coefficients, is the criterion of fit used in reciprocal-space search methods, for example, the widely used rotation function (Rossmann and Blow 1962).

In order to assess the potential value of Patterson-space search techniques in macromolecular crystallography, it is of interest to examine different criteria of fit in the hope of finding the one which is most promising in the unfavourable case of a very heavily overlapped Patterson function. It has recently been shown (Nordman and Hsu 1982) that a Monte-Carlo calculation which simulates a Patterson search can be formulated as a one-dimensional problem of detecting a sequence of known numbers embedded in a longer sequence of random numbers. On the basis of relatively limited statistics it was concluded that the criterion (1), with $n = 300$, $m = 50 - 100$, was more successful in finding the “correct” solution than criterion (2).

In this communication a slightly modified Monte-Carlo calculation is used to examine several criteria of fit in the light of much more extensive statistical material. Also, a related formulation is employed in a Monte Carlo simulation of a search for local non-crystallographic symmetry.

2. Simulated structure search

In an actual Patterson search $n$ vectors of weight $w_i$ scan the Patterson function, returning, at each point in the search a set of values $P_i, i = 1, \ldots, n$. The arrays $P_i$ and $w_i$ are sorted in ascending order of $(P_i/w_i)$ and criteria of fit based on the first $m$ entries in the sorted arrays, where $m \leq n$.

A rapidly computable simulation of this is as follows. Let the “Patterson” to be
Criteria of fit in Patterson search

searched be represented by \( P_i = S_i + N_i \) where \( S_i \) and \( N_i \) are sequences of random numbers with positive means \( \langle S \rangle \) and \( \langle N \rangle \) and standard deviations \( \sigma(S) \) and \( \sigma(N) \). These sequences are of equal length \( l \), here taken as 500.

The \( n \) search vectors \( w_j \) in this simulation are a positionally significant subset of the sequence \( S_i \). Without loss of generality the \( w_j \) can be taken as a contiguous sequence \( w_j = S_{j+k0} \) where the \((k_0 + 1)\)th entry in the \( S_i \) sequence is the first in \( w_j \). The \( w_j \)'s are treated as the “known” search vector weights; \( n \) was taken as 300 here. The search is carried out by translating the \( w_j \) sequence along the \( P_i \) sequence, allowing the \( w_j \)'s to sample \( P_{j+k} \) for successive values of \( k \), from zero to \( l-n \). At each of the 201 steps in the search the data are sorted, and several criteria of fit evaluated. A given criterion is judged successful if it assumes a higher value when \( k = k_0 \) than for any of the 200 other values of \( k \). The percentage of successes scored by a given criterion in a large number of independent searches allows us to compare different criteria with one another.

Five different criteria of fit were evaluated. These included the minimum average, \( \min \), as defined in eq. (1), a weighted minimum average

\[
\text{wmin}(m, n) = \frac{\sum_{i=1}^{m} (w_i P_i)}{\sum_{i=1}^{m} w_i^2} \quad m \leq n
\]

and the quantity

\[
\text{msp}(m, n) = \frac{\sum_{i=1}^{m} (w_i P_i)}{\sum_{i=1}^{m} w_i} \quad m \leq n
\]

which is a generalization of \( \text{SP} \) (equation (2)). The criteria \( \sum_{i=1}^{\infty} (1 + w_i \langle w \rangle) P_i / \sum_{i=1}^{\infty} w_i \) and \( \sum_{i=1}^{\infty} (P_i/w_i) \), \( m \leq n \), were also computed. They were less successful than the others, and are not further discussed.

Figure 1 shows the results for the three best criteria of fit. In these calculations the noise, \( N_i \), was taken as Gaussian with a mean of 300 and a standard deviation \( \sigma(N) = 100 \). The sequence \( S_i \), including the vector weights \( w \), was also Gaussian with a mean \( = 3\sigma(S) \), and \( \sigma(S) \) ranging from 15 to 50. For each choice of \( \sigma(S) \), at least 100 runs were calculated.

The results are summarized in figure 1. It should be noted that the ‘noise level’ of the ordinate is \( 1/201 \) or \( 0.5\% \); this would be the statistical chance of a ‘successful’ search with a vanishingly small \( \sigma(S) \). At the upper end, \( \sigma(S) \approx 0.5\sigma(N) \) essentially insures success.

All curves reach maxima at values of \( m < n = 300 \). The position of the maximum appears to be independent of \( \sigma(S)/\sigma(N) \) for any one criterion of fit.

With \( m = 300 \), the \( \min \) criterion (equation (1)) reduces to \( \Sigma_{i=1}^{300} P_i \) divided by a constant. Since the calculation is designed so that \( \langle w \rangle \approx \langle S \rangle \), the \( \min \) criterion is meaningless at \( m = 300 \), and the corresponding points are not shown.

The optimal choice of \( m \) for \( \min \) can be estimated to be approximately \( 0.3n \), in agreement with past experience. The \( \text{wmin} \) criterion tends to have its maximum at \((0.4-0.5)n\), and in essentially every run tends to give slightly higher success rates than \( \min \). The maxima of \( \text{msp} \) lie at approximately \( 0.7n \), and tend to be lower than those of either \( \text{wmin} \) or \( \min \). It should be noted that at \( m = 300 \) \( \text{msp} \) is identical to \( \text{sp} \) (equation (2)). This criterion is distinctly inferior to any of the others.

Additional runs were done with \( N_i \) unchanged, but \( S_i \) uniformly distributed between 0-01 and variable upper limits, up to 60. When compared to the Gaussian runs with equivalent \( \sigma(S) \) no clear difference in success rates could be discerned.
Figure 1. Percentage of successful structure searches as a function of $m \leq n = 300$ for three criteria of fit as defined in the text: $\text{MIN}$ (left), $\text{WMIN}$ (centre) and $\text{MSP}$ (right). For each criterion the percentage is shown for $\sigma(S)/\sigma(N) = 0.15$ (bottom graph), 0.2, 0.25, 0.3, 0.4 and 0.5 (top graph).

Finally the effect of varying the mean of the noise $N_i$ was explored, keeping $\sigma(N)$ constant at 100. Lowering $\langle N \rangle$ to 250 caused some deterioration; raising it to 1000 or above also appeared to cause some decline. One may tentatively conclude that the optimal choice of the constant term in the Fourier synthesis of the Patterson is 3 to 5 times $\sigma(P)$; this choice also achieves a reasonable compromise between numerical accuracy and packing density in the computer memory.

3. Simulated symmetry search

In each simulated symmetry search a set of 50 independent random number sequences

$$F_i = S_i + N_i, \quad i = 1, \ldots, 960$$

was searched. The noise sequences $N_i$ were taken as Gaussian with $\langle N \rangle = 100$ and $\sigma(N) = 20$. The sequences $S_i$ were also Gaussian with $\langle S \rangle = 0$ and $\sigma(S)$ variably chosen from 0.05 $\sigma(N)$ to 0.45 $\sigma(N)$.

One of the 50 $S_i$ sequences was modified, so as to give it $n_s$-fold symmetry, by requiring that $S_{i+j} = S_i$, where $j = 960/n_i$. Thus, this one $S_i$ sequence consists of $n_s$ copies of a random number sequence of length $j$.

For each of the 50 $F_i$-sequences the quantity

$$\sum_{i=1}^j n_i^{-1} \sum_{s=0}^{n_i-1} [F_{i+j s} - \langle F_{i+j s} \rangle_s]^2$$

was evaluated. This quantity is expected to have its lowest value for the "symmetric" sequence. If this is indeed found, the search is taken as successful. The "noise level" for this search, as formulated here, is 1/50 or 2%.

Symmetry searches were carried out for $n_i = 12, 6$ and 2. For each choice of $\sigma(S)$ 50 to 100 searches were done. The results, shown in figure 2 bear out the expected sharp increase in the success rate with increasing strength of the symmetric component, $\sigma(S)$,
4. Conclusions

The results unambiguously demonstrate that the discriminating power of a Patterson-space structure search improves when the "minimum average" principle is applied in calculating the criterion of fit. We conclude that in searches where the detectability of the correct solution is at all in doubt, the benefits of calculating minimum averages are well worth the additional computing time required for the sort, on $P_i/w_i$, which is carried out at every step in the search.

Somewhat more tentatively, the results suggest that the criterion

$$\frac{\sum_{i=1}^{m} w_i P_i}{\sum_{i=1}^{m} w_i^2},$$

with $m \approx (0.4-0.5)n$, is superior to the presently used

$$\frac{\sum_{i=1}^{m} P_i}{\sum_{i=1}^{m} w_i}.$$

Both are consistently superior to the criterion

$$\sum_{i=1}^{n} (w_i P_i).$$

It should be emphasized that it has not been shown or suggested that any of these criteria is the best one that can be formulated. What the best one is, is still an open question.
References

Huber R 1965 *Acta Crystallogr.* **19** 353
Nordman C E 1980a *Acta Crystallogr.* **A36** 747
Schilling J W 1970 *Crystallographic computing* (ed.) F R Ahmed (Copenhagen: Munksgaard) p. 115
The probability of large structure amplitudes: The space group $P\bar{T}$

A J C WILSON
Crystallographic Data Centre, University Chemical Laboratory, Cambridge CB2 1EW, England

Abstract. Until recently, expressions derived for the probability distribution of the magnitudes of structure amplitudes have been valid only for the range from zero up to two or three times the average magnitude. Recent progress in obtaining results valid for larger structure amplitudes (up to unitary structure factors approaching unity) is reviewed; the methods employed include Fourier-series representations, steepest-descent approximations, and computer simulations. Except for the last, the results are valid only for the space group $P\bar{T}$, though those for the space groups belonging to the point group $2/m$ and possibly some other centro-symmetric groups of low symmetry are expected to behave similarly.

Keywords. Crystallographic statistics; large structure amplitudes; probability distribution; statistics.

1. Introduction

The calculation of the probability of a structure amplitude having a particular value is a special case of the random-walk problem. If correlations can be neglected (Wilson 1981), for the space group $P1$ the problem is exactly that of the random walk in two dimensions, and for $P\bar{T}$ it is that of the projection on a line of a random walk in two dimensions. Most attention has been given to walks that do not get far from the starting point—in crystallographic terms structure amplitudes that do not exceed two or three times their average magnitude—but for several years I have tried unsuccessfully (Wilson 1980b) to obtain an estimate of the probability of structure amplitudes with magnitudes approaching the maximum possible. This maximum possible is, of course, the sum of the atomic scattering factors of all the atoms in the unit cell:

$$|F|_{\text{max}} \equiv \Phi = f_1 + f_2 + f_3 + \ldots + f_N.$$  (1)

In direct methods of structure determination the so-called unitary scattering factors and unitary structure amplitudes are often used; these are the ordinary ones divided by $\Phi$, and are thus convenient also in discussing the probability of large structure amplitudes. The unitary atomic scattering factor of the $j$th atom is

$$u_j = f_j/\Phi,$$  (2)

and the unitary structure amplitude is

$$U = |F|/\Phi.$$  (3)

'Large' unitary structure amplitudes are thus somewhat less than unity. Derivations of the mathematical expressions for $p(U)dU$, the probability that a structure amplitude will have a unitary value between $U$ and $U + dU$, are complicated and it is more useful to try to review the present state of knowledge than to work through any derivation. So far I have been able to locate useful analytic results only for structures in the space group $P\bar{T}$.
it seems likely that other low-symmetry centrosymmetric groups will behave very similarly.

2. Exact expression

For the pure mathematician there is no problem. Kluyver (1906) was able to give \( p(U) \) as an integral:

\[
p(U)dU = (2\pi)^{-1} \int_{-\infty}^{\infty} \left[ \prod_{j=1}^{1/2} J_0(2u_jx) \right] \cos(Ux)dxdU,
\]

where \( J_0 \) is the zero-order Bessel function. Kluyver himself recognized, however, that this expression is of little value for actual computation in any range of \( U \).

3. Series representations

For small \( U \), the exact expression (4) reduces to the central-limit Gaussian expression (Wilson 1949), which can be made the basis of a series expansion in terms of Hermite polynomials (see, for example, Karle and Hauptman 1953; Rogers and Wilson 1953; Bertaut 1955; Foster and Hargrées 1963; Shmueli and Wilson 1981, 1982). The expansion, in either the Edgeworth or the Gram-Charlier form, satisfies the Cramér criterion for convergence, but in practice there may be unacceptable negative probabilities in the region of present interest.

Barakat (1973) pointed out that \( p(U) \) vanishes outside the range \(-1\) to \(+1\), so that it can be represented by an ordinary Fourier series. His actual application was to a three-dimensional random walk (with polymer chains in mind), but Weiss and Kiefer (1983) have modified it for the two-dimensional case. Their expression is

\[
p(U)dU = \frac{1}{2} \left[ 1 + 2 \sum_{m=1}^{\infty} C_m \cos \pi mU \right] dU,
\]

where the Fourier coefficients \( C_m \) are given by

\[
C_m = \prod_{j=1}^{1/2} J_0(2\pi mu_j).
\]

The Fourier series has the great advantage that the coefficients are readily obtained for any required number of terms, whereas the calculation of the moments required for the Hermite coefficients rapidly becomes more tedious as the order of the polynomial increases. For most space groups only five-term Hermite expansions are available at present, whereas there is no definite limit to the number that can be used in the Fourier series. In most cases tried twenty terms seem to be ample, there is no significant change in the values of \( p(U) \) when the number is increased to forty (Shmueli 1983, private communication). A series similar to equation (5), but with Bessel functions instead of cosines, can be developed for \( P1 \) (Weiss and Kiefer 1983, private communication).

4. Cramér’s limit theorem

The central-limit theorem is not the only limit theorem, and results due to Cramér
Structure amplitudes

(1938) can be applied in the equiatomic centrosymmetric case (Wilson 1983). The result is

\[ P(U) dU = \left\{ \frac{N}{4\pi\sigma^2} \right\}^{\frac{1}{4}} \left\{ I_0(h) \right\}^{\frac{1}{4}} N \exp\left( -\frac{1}{2}NhU \right) dU, \]  

(7)

where \( h \) is the solution of

\[ I_1(h)/I_0(h) = U, \]  

(8)

the \( I \)'s are modified Bessel functions, and

\[ \sigma^2 = 1 - U/h - U^2. \]  

(9)

For small \( U \) equation (7) reduces to the familiar Gaussian, and for large \( U \) to

\[ p(U) dU = \left\{ \frac{N}{4\pi} \right\}^{\frac{1}{4}} \left\{ \frac{1}{\pi} \right\}^{\frac{1}{4}} N \exp\left( -\frac{1}{2}NhU \right) dU. \]  

(10)

5. Steepest descents

Weiss and Kiefer (1983) use a more general approach, the method of steepest descents (Daniels 1954), and obtain equivalent expressions for the unequal-atom case:

\[ p(U) dU = \left\{ \frac{N}{4\pi\sigma^2} \right\}^{\frac{1}{4}} \left\{ \prod_{j=1}^{N} I_0(NU_jh) \right\} \exp\left( -\frac{1}{2}NhU \right) dU, \]  

(11)

where \( h \) is the solution of the equation

\[ \sum_{j=1}^{N} 2u_j I_1(NU_jh)/I_0(NU_jh) = U, \]  

(12)

and

\[ \sigma^2 = 2N \sum_{j=1}^{N} u_j^2 \left\{ 1 - I_1(NU_jh)/NU_jhI_0(NU_jh) - \left[ I_1(NU_jh)/I_0(NU_jh) \right]^2 \right\} \]  

\[ = 2N \sum_{j=1}^{N} u_j^2 - U/h - 2N \sum_{j=1}^{N} \left[ u_j I_1(NU_jh)/I_0(NU_jh) \right]^2. \]  

(13)

Equations (11)-(14) reduce in an obvious fashion to (7)-(9) in the equal-atom case, in which all \( u_j = N^{-1} \).

6. Resultants near maximum possible

An unequal-atom analogue of (10) is readily obtained from (11), but Weiss and Kiefer (1983) adopt a different approach that leads to a better result. In effect, they expand the Bessel functions in the exact expression (4) into asymptotic series, multiply out, and integrate term by term. The result is a series in ascending powers of \((1 - U)\), of which the first term is

\[ p(U) dU = \left[ (2\pi)^{\frac{1}{4}} N \right]^{\frac{1}{4}} \left( \prod_{j=1}^{N} (2u_j)^{\frac{1}{2}} \right)^{\frac{1}{4}} (1 - U)^{N-1} dU. \]  

(15)
Weiss and Kiefer give two more terms explicitly and recurrence relations for higher terms.

Wilson (unpublished) has developed a calculation that suggests that \( p(U) \) will behave like a power of \((1 - U)\) for large \( U \) for any centrosymmetric space group.

7. Simulation

Shmueli (1982) has described a method of computer simulation that gives an 'experimental' distribution for any postulated structure, or rather for any postulated composition; the effects of correlation of the real atomic positions are not taken into account. None of the theoretical distributions described above take such correlations into account, so a comparison of the simulated distributions with the theoretical is comparing like with like, whereas a comparison with distributions actually observed for real crystals involves a further complication, the full effect of which is unknown (Wilson 1981). Briefly, the simulation method consists in calculating a suitable number of structure amplitudes (3000 is typical), with the usual atomic scattering factors but with \( h_x, ky \) and \( l_z \) or their combinations replaced by computer-generated 'random' numbers. The resulting \( |F| \)'s are sorted by size into suitable groups to give a histogram; examples are given by Shmueli (1982, 1983). The method is applicable to any space group.

The simulation method seems to be the easiest and most flexible method yet devised for obtaining ideal distributions for small and moderate structure amplitudes, but it has one aesthetic and one practical disadvantage. Aesthetically, numerical calculations tied to a specific atomic composition and a specific symmetry are less pleasing than closed or series expressions in which the effect of changing atomic composition or symmetry can be 'seen'. Practically, because of the statistical nature of the simulation, there are random fluctuations in the heights of the bars of the histogram. Wilson (1983, unpublished) has estimated the extent of these; for 3000 simulations and 30 histogram divisions the residual \( R \) between the actual histogram and the ideal distributions should be about 0.07. To obtain reliable estimates of the probability of really large (greater than four times the average, say) structure amplitudes would require a very large number of simulations.

8. Accuracy

The Barakat Fourier series (5) seems to reproduce the exact ideal distribution to several decimal places, the limiting factor being probably the approximations used in the computer calculation of the Bessel functions; at present its use is limited to the space group \( Pt \). Weiss and Kiefer (1983 and private communications) and Shmueli (1982 and private communications) have made various numerical comparisons of the Hermite-Gaussian and steepest-descents expressions with the Fourier results for various total numbers of atoms (10 to 30) and various ratios of atomic scattering factors. As would be expected, the non-Fourier analytic expressions work best for equal-atom structures, and are poor if there is one pair of atoms much heavier than the rest, when the distribution becomes bimodal. The expression (15) is very good for large \( U \). For smaller \( U \), the steepest-descents results are usually, but not always, better than the Gaussian-Hermite. Simulation results agree with the Fourier within the expected \( R \).
The preceding remarks refer to ideal distributions and different methods of approximating them. The distributions observed experimentally for real crystals will show statistical fluctuations of the same type as in those obtained by simulation. The size of the fluctuations from this source (there are several other sources) is inversely proportional to the square root of the number of reflexions observed, but 3000 is fairly typical. Only differences between an observed and a calculated histogram appreciably exceeding an $R$ of 0.07, or between an observed and a simulated histogram appreciably exceeding 0.1, could, with reasonable certainty, be attributed to non-statistical factors (Wilson 1980a). It is hoped to include a discussion of tests of the statistical significance of apparent differences between two distributions, either one observed and one calculated, or two calculated, in a forthcoming paper (Shmueli et al 1983).

Acknowledgements

It is with great pleasure that I have accepted the invitation to contribute to this volume in honour of Professor S. Ramaseshan.

The work was begun before the author’s retirement from the chair of crystallography at the University of Birmingham, and a preliminary version was presented at a meeting of the British Crystallographic Association in March 1983. The revision for publication was greatly aided by conversations with Professor U Shmueli and Dr G H Weiss at the University of Tel Aviv in April 1983.

References

Bertaut E F 1955 Acta Crystallogr. 8 823
Cramér H 1938 Sur un nouveau théorème-limite de la théorie des probabilités No. 736 in the series Actuahtés scientifiques et industrielles (Paris: Hermann)
Foster F and Hargreaves A 1963 Acta Crystallogr. 16 1124
Karle J and Hauptman H 1953 Acta Crystallogr. 6 131
Kluyver J C 1906 Kon. Akad. Wet. Amsterdam 8 341
Rogers D and Wilson A J C 1953 Acta Crystallogr. 6 439
Shmueli U 1982 pp 53–82 in Crystallographic statistics (eds) S Ramaseshan, M F Richardson and A J C Wilson (Bangalore: Indian Academy of Sciences)
Shmueli U 1983 submitted for publication
Wilson A J C 1949 Acta Crystallogr. 2 318
Wilson A J C 1980b Technometrics 22 629
Maximum entropy—A new approach to the crystallographic phase problem

RAMESH NARAYAN*, RAJARAM NITYANANDA and G V VANI
Raman Research Institute, Bangalore 560080, India
*Present address: Theoretical Astrophysics, California Institute of Technology, Pasadena, California 91125, USA

Abstract. We have recently drawn the attention of crystallographers to the maximum entropy method (mem) of image reconstruction which has found application in radio astronomy and geophysics. In this article, we review the theory of the mem as applied to the crystallographic phase problem and discuss points of similarity, and differences, with conventional techniques. We then present a number of illustrative calculations in two and three dimensions on simulated and real structures. These examples show that the mem can be used for phase refinement from partially known phases, say in macromolecular crystallography, and also for solving crystal structures ab initio using only measured Fourier amplitudes. With improved algorithms the method promises to become competitive with current techniques in crystal structure analysis.

Keywords. Crystallography; crystal structure analysis; phase problem; maximum entropy method; direct methods.

1. Introduction

In conventional x-ray crystal structure analysis one measures the intensities of Bragg reflections of the crystal under study. The intensity \( I_{\mathbf{H}_i} \) corresponding to the reciprocal vector \( \mathbf{H}_i \) is related to the complex Fourier structure factor \( F_{\mathbf{H}_i} \) by
\[
I_{\mathbf{H}_i} = F_{\mathbf{H}_i}^* F_{\mathbf{H}_i} = |F_{\mathbf{H}_i}|^2,
\]
where the structure factors \( F_{\mathbf{H}_i} \) and the electron density \( \rho(\mathbf{r}) \) in the crystal unit cell are related by a Fourier transform:
\[
F_{\mathbf{H}_i} = |F_{\mathbf{H}_i}| \exp(\mathbf{i}\varphi_{\mathbf{H}_i}) = \int \rho(\mathbf{r}) \exp(2\pi\mathbf{i}\mathbf{H}_i \cdot \mathbf{r}) \, dV_r,
\]
\[
\rho(\mathbf{r}) = \frac{1}{V_r} \sum_{\mathbf{H}_i} F_{\mathbf{H}_i} \exp(-2\pi\mathbf{i}\mathbf{H}_i \cdot \mathbf{r})
\]
where \( V_r \) is the volume of the crystal unit cell. Thus, one experimentally measures only the structure amplitudes \( |F_{\mathbf{H}_i}| \) whereas according to (2) and (3) one requires both amplitudes as well as phases \( \varphi_{\mathbf{H}_i} \) to compute the electron density. This is the famous phase problem of crystallography—one cannot solve the structure unless one has some estimates of the phases. The magnitude of the problem has been brought out strikingly by Ramachandran and Srinivasan (1970) who have given examples of how true phases, coupled even with random amplitudes, can bring out the true crystal structure; on the other hand, true amplitudes with random phases give no information at all.

Obviously one needs to put in some a priori information in addition to the measured amplitudes if one is to obtain a solution. The most powerful input turns out to be the...
prior knowledge that the crystal consists of atoms so that the (positive) electron density
is strongly clumped in well separated peaks. Further, one generally knows in advance
the number and types of atoms present and also has limited stereochemical
information. Although a rigorous proof is lacking, all these extra constraints in practice
seem to ensure a unique solution to the phase problem, apart from trivial ambiguities
related to the choice of origin or enantiomorph. This solution is however not easy to
determine!

For small structures involving a few atoms, Patterson techniques work quite well.
Here one computes the Fourier transform of the intensities \( I_{H_j} = |F_{H_j}|^2 \). This map can
be shown to be the self-convolution of the electron density (e.g. Ramachandran and
Srinivasan 1970) and so each peak in it corresponds to an interatomic vector in the
original structure. If the peaks are well resolved, which can happen in small structures,
one “inverts” this information on interatomic bonds to solve the structure.

Even larger structures can be solved using the Patterson technique provided a few of
the atoms are significantly heavier (i.e. have more electrons) than the rest; one can
obtain the heavy atom positions from the Patterson map and solve the rest of the structure
by using the heavy atom derived phases for the structure factors. By far the most
challenging aspect of the phase problem in crystallography has been the solution of
large structures of nearly equal atoms. Here some very powerful techniques have been
developed over the years which go collectively under the name of direct methods (see
for example Ladd and Palmer 1980). In a pioneering paper Sayre (1952) mathematically
expressed the equality of atoms by the statement that the squared structure \( \rho^2(r) \) is very
similar to \( \rho(r) \) except for a “normalisation”. He showed that this immediately leads to
relations among the phases of certain sets of reflections—the famous triplet phase
relation. Later work has extended this approach to higher order relations and
probability distributions have been worked out for the values of sums of specific groups
of phases (the so-called structure invariants and seminvariants). Coupled with
multisolution (Karle and Karle 1963, 1966; Germain et al 1971) methods (where one
selects a small set of primary reflections, systematically assigns various phase values to
these and in each case solves the phases of the other reflections using triplet and other
relations), it can be said that the direct methods, exemplified by the computer program
MULTAN and its descendants, are the most powerful techniques available today to solve
crystal structures. However, despite the enviable success that these techniques have had,
it is probably fair to say that the last word has by no means been said on the phase
problem. The search for newer approaches must go on as crystallographers attempt to
solve progressively larger structures.

A technique called the maximum entropy method (MEM), which has been studied for
some time in other fields, has in the last year or two attracted some interest in
crystallography. Two of us (RN and RN) studied the MEM in the context of its application
in radio astronomical image processing and drew attention to its potential importance
in crystallography (Narayan and Nityananda 1981, 1982). In this article we briefly
review some of the work already reported in the field and then describe our more recent
results on simulated and real crystal structures in two and three dimensions. On the
basis of our present studies we believe that the MEM has a promising role to play in
crystallography and merits further investigation. Readers interested in the actual
results of this method applied to crystallography could proceed straightforward to §§5
and 6, although we hope that this will generate enough interest in the MEM to return to
the more theoretical sections.
2. Maximum entropy method

Let us suppose, as is the case in many geophysical and radio astronomical applications, that $F_{hj}$ is available both in amplitude and phase at some of the $H_j$ but is unknown at the rest of the points. Setting the unknown $F_{hj}$ to zero leads to a map with large termination ripple as well as loss of resolution. In the MEM, the unknown $F_{hj}$ are chosen so as to maximise the “entropy” $S$ given by

$$S = \int f[\rho(r)] dV_r, \quad (4)$$

where $f$ is some suitable real function of $\rho(r)$. There is an unresolved debate going on in the literature as to the most suitable form for $f$. One school of thought including the originator of the MEM (Burg 1967; Ables 1972) favours the entropy

$$S_1 = \int \ln[\rho(r)] dV_r, \quad (5)$$

while others (Frieden 1972; Gull and Daneill 1978) prefer

$$S_2 = -\int \rho(r) \ln[\rho(r)] dV_r. \quad (6)$$

The arguments in favour of one or the other form are usually based on information theory or combinatorial probabilistic arguments, verging sometimes on the metaphysical. It has been claimed that the MEM produces as featureless a map as possible consistent with the data (Ponsonby 1973) and that it is maximally non-committal with regard to the unmeasured data (Ables 1972). In an interesting paper Komesaroff and Lerche (1978) (see also Komesaroff et al 1981) showed that there is an intimate connection in one-dimensional problems between the entropy $S_1$ and the positivity constraint on $\rho(r)$; however, this does not generalize very well to higher dimensions.

In contrast to the above approaches which seek to attribute a deep foundation to the MEM, there has been another movement initiated by Hogbom (1978) and Subrahmanya (1978, 1980) who suggested that the successes of the method are primarily due to the a priori information built in, in particular the penalty against negative values. We have extended this view-point (Nityananda and Narayan 1982; Narayan and Nityananda 1983) and carried out a detailed study of the MEM focussing primarily on the maximisation conditions that the reconstruction satisfies. By a simple geometrical argument it was shown that any “entropy” function $f$ which satisfies $f'' < 0$ and $f''' > 0$ implicitly makes the a priori assumption that the map consists of isolated sharp features separated by flat extended regions. The forms (5) and (6) as well as several others, e.g. $\rho^{1/2}, -1/\rho$, etc., satisfy these conditions, and hence all these are suitable “entropies” to generate satisfactory maps. In our view, the MEM is just a variational way of incorporating our a priori information that the true map consists of peaks on a flat background. The relevance of the MEM to crystallography then becomes obvious.

Regardless of the controversy over the foundations of the method, it is well known that the MEM produces excellent reconstructions. Figure 1 shows a two-dimensional simulation of a radio astronomical aperture synthesis reconstruction where the MEM map is clearly far superior to the naive approach of setting unmeasured $F_{hj}$ to zero. Results such as this, coupled with our conviction that the a priori assumptions built into
Ramesh Narayan, Rajaram Nityananda and G V Vani

Figure 1. (a) Simulated "true" map $p(x, y)$ of a radio astronomical source in the sky. (b) Map reconstructed using only "measured" Fourier data $F(h, k)$ corresponding to reciprocal vectors with $|h|, |k| \leq 3$. All the other $F(h, k)$ have been set equal to zero. (c) Reconstruction from the same data using the $S_2$ form of the MEM where the unmeasured Fourier coefficients are optimized to maximise the integral of $-p \ln p$. This reconstruction was obtained by setting the non-linearity parameter $R$ (defined in §3) to the value 100. Full contours correspond to positive values, the successive levels being at 0.01, 0.03, 0.05, 0.1, 0.2, 0.3, 0.5, 0.7 and 1. Negative contours, in dashed lines, are at $-0.01$, $-0.03$ and $-0.05$. (From Nityananda and Narayan 1982).

the MEM are perfectly suited to crystallography, originally led us to propose that the method could help solve the phase problem (Narayan and Nityananda 1981, 1982).

3. Maximum entropy for the phase problem

For the phase problem we write (4) in the form

$$S = \int f \left[ \frac{1}{V_r} \sum_{H_j} |F_{H_j}| \exp \{i(\varphi_{H_j} - 2\pi H_j \cdot r)\} \right] dV_r,$$

where, because $p(r)$ is real, we have

$$\varphi_{H_j} = -\varphi_{-H_j}.$$
Maximum entropy approach to crystallography

Let us define the Fourier coefficients $G_{H_j}$ as follows

$$G_{H_j} = |G_{H_j}| \exp(i\theta_{H_j}) = \int f'(\rho(r)) \exp(2\pi i H_j \cdot r) \, dV_r. \quad (9)$$

Then, maximising the entropy $S$ in (7) with respect to the unknown phases we obtain

(Narayan and Nityananda 1981, 1982)

$$\frac{\partial S}{\partial \varphi_{H_j}} = \frac{2}{V_r} |F_{H_j}| |G_{H_j}| \sin(\varphi_{H_j} - \theta_{H_j}) = 0. \quad (10)$$

Equations (10) (there are as many equations as the number of unknown phases) have to be satisfied in order to maximise the “entropy”. Since $|G_{H_j}|$ and $\theta_{H_j}$ are implicitly functions of all the $\varphi_{H_j}$, these are a highly coupled system of nonlinear equations which will have to be solved numerically by an iterative procedure. However, we note that $\partial S/\partial \varphi_{H_j}$ are the components of the gradient of $S$ with respect to the unknown $\varphi_{H_j}$ and the calculation of $|G_{H_j}|$ and $\theta_{H_j}$ involves only Fourier transforms which can be very efficiently carried out by means of the fast Fourier transform algorithm. Thus gradient-type algorithms to maximise $S$ (see § 4) are easily developed and these are not likely to be very expensive in terms of computing time.

A crucial issue is the question of the uniqueness of the solution. Of course, for every solution there is a whole family related to it by an origin shift and/or enantiomorph reversal. However, apart from this trivial ambiguity, it is possible that there are several distinct maxima representing physically different solutions. To see this let us write the conditions for the vanishing of the gradient in (10) as

$$\theta_{H_j} = \varphi_{H_j} + n_{H_j} \pi, \quad (11)$$

where the $n_{H_j}$ are integers which have either of the two values 0 or 1. One can attempt to solve (10) for each choice of the $n_{H_j}$ (though it is possible that for certain choices there will be no solution at all). In addition, for each choice of the $n_{H_j}$ there may be more than one distinct solution. Presumably, the majority of such solutions correspond to saddle points in the entropy and only a small fraction constitute true maxima. A function of $n$ phase-like variables can be easily seen to have at least one maximum, one minimum, and $n$ saddle points, using the arguments illustrated for $n = 2$ in the context of solid state physics by Ziman (1972). In any case, it is clear that the phase problem introduces a rich new dimension to maximum entropy. The existence of several local maxima (about which we are convinced on the basis of our computer simulations) means that any numerical technique is liable to be trapped in the wrong maximum. The saving grace is that these subsidiary maxima generally display elements of the true structure and are therefore still useful for the elucidation of the structure.

Even if there are multiple MEM solutions to the phase problem, it is crucial that the true structure itself should be close to a solution—only then can we hope to get something out of the method. At a general level, since the electron density distribution in a crystal is always peaky with a flat baseline, it satisfies the a priori input built into the MEM (§2) and hence is likely to be a solution. However, one can more rigorously establish this for equal atom structures. Figure 2a schematically shows the electron density $\rho(x)$ (in one dimension) of a “structure” of equal symmetric well-resolved atoms. Let us for convenience focus on the entropy $S_1$ where $f(\rho) \equiv \ln(\rho)$. Then

$$f'(\rho) = 1/\rho, \quad (12)$$
and its shape as a function of $x$ for the true structure is shown in figure 2b. It is clear that

$$f'[\rho_{\text{true}}(x)] = K + \rho_{\text{true}}(x) \ast S(x),$$

where $K$ is a constant, $\ast$ represents a convolution and $S(x)$ is a symmetric "shape" function. Taking the Fourier transform of (13) we have for any non-origin $H_j$

$$G_{H_j} = F_{H_j} S_{H_j},$$

where $S_{H_j}$, the $H_j$th Fourier coefficient of $S(x)$, is a real number (positive or negative). Thus we automatically have

$$\theta_{H_j} = \varphi_{H_j} \quad \text{or} \quad \varphi_{H_j} + \pi,$$

depending on the sign of $S_{H_j}$. This is identical to the conditions written in (10) or (11) and shows that equal atom structures are automatically solutions of the MEM. As a further bonus, since the shape function $S(x)$ is usually known, one has full information on $S_{H_j}$ and hence one can, in principle, determine all the $n_{H_j}$ in (11). This greatly reduces the number of solutions to the problem. It is clear that the above considerations work for all forms of the entropy and are dimension-independent.

In the case of non-equal atom (i.e., heavy atom) structures, there is a rather subtle point involved. It will be noticed that both the entropies $S_1$ and $S_2$ ((5) and (6)) are defined only for positive values of the argument. Hence, at any intermediate stage of refinement, when one has a lot of spurious features in the map, including negative excursions of the electron density, it is necessary to add a suitable constant $C$ to $\rho(r)$. Any value of $C$ larger than the greatest negative value of $\rho(r)$ is enough to make the problem well-defined. However, it is clear that the reconstructed solutions now have a non-trivial dependence on the value of $C$. Rather than look at $C$ itself, in our work we have found it useful to employ another parameter $R$ given by

$$R = f''(\rho_{\text{min}} + C)/f''(\rho_{\text{max}} + C).$$

We have borrowed this concept from our earlier work in radio astronomy (Nityananda and Narayan 1982) where we showed that $R$ is a measure of non-linearity. For the
Maximum entropy approach to crystallography

...present purposes it is sufficient to note that large values of $C$ imply small values of $R$ (however, $R$ is always $> 1$) while the smallest permitted $C$ gives $R \to \infty$. It is also clear that as $R \to 1$, $f''$ is constant and $f'$ a linear function of $\rho$. The phases $\theta_H$ of $G_H$ in (9) are then just $\varphi_H$ or $\varphi_H + \pi$ and all maps satisfy the MEM conditions (11). Now, in our experience, low values of $R$ lead to a stable centrosymmetric solution even when the true structure is non-centric. When $R$ is increased, the symmetry is broken and one of the two enantiomorphs grows at the expense of the other. This is most clearly seen in the case of a structure with a single heavy atom, where the enantiomorphs are related by a centre of symmetry at the heavy atom location (see figure 3a). As $R$ is increased in magnitude, the ghost peaks in the "wrong" enantiomorph become progressively weaker (figure 3b). This suggests that MEM refinement on non-centric structures, particularly if they have heavy atoms, should be carried out at high values of $R$, i.e. low values of $C$. The role of the constant $C$ vis a vis centric and non-centric structures is discussed in terms of a simple illustrative example in the Appendix.

Although the MEM is a variational method, having its origins in information theory, it turns out surprisingly to have certain features in common with more conventional techniques. If we look at the MEM conditions (10) or equivalently (11), we see that the phases $\theta_H$ and $\varphi_H$ are related. Since $\theta_H$ by (9), is a function of all the $\varphi_H$, through $\rho(r)$, we thus see that the phase relations (11) implicitly involve all the unknown phases. In conventional crystallography Sayre (1952) introduced the concept of triplet phase relations, where the phases of three suitably selected reflections are tied together. This

Figure 3. (a) The true structure consists of 12 atoms shown by solid lines, with a heavy atom located at the strongest peak. When the true phases are refined by the MEM using the entropy form $S_1$ at $R = 9$, the final map has reasonably strong ghost peaks as shown by the dashed lines. These peaks correspond to a spurious enantiomorph molecule sharing the same heavy atom as the true structure. (b) When the refinement is carried out at $R = 100$, the ghost peaks are very much weaker showing that the MEM breaks the symmetry more effectively. On the other hand, if one used $R \ll 9$, one would obtain a completely centrosymmetric map even though the structure amplitudes correspond to a non-centric structure.
has since been extended to quartets, quintets etc., which are believed to be more powerful, though also more difficult to implement (Ladd and Palmer 1980). Since the MEM involves very high order phase relations which are relatively easy to implement through Fourier transforms and the gradient method, there is a case for believing that the method may be an improvement over current direct methods. One can look upon the “non-linearity” parameter \( R \) as determining the extent to which different phases are coupled by the MEM condition, since they become decoupled in the \( R \to 1 \) limit. Since \( R \) can be controlled through the constant \( C \), this may be considered an added advantage.

There has been interest in recent times in a technique called the maximum determinant method (Lajzerowicz and Lajzerowicz 1966; Tsoucaris 1970, 1980) which is based on the Karle-Hauptmann (1950) determinant and involves high order phase relations. We have proved (Narayan and Nityananda 1982) that the maximum determinant method is closely related to maximising the entropy \( S_1 \) (Britten and Collins 1982 have also independently reached the same conclusion); this is therefore another contact between conventional crystallography and the MEM.

### 4. Numerical algorithm

We have carried out a series of calculations on model “structures” in two and three dimensions as well as on a few real structures in three dimensions. We have employed the gradient approach to maximisation of \( S \), using the expression (10) for the components of the gradient. The alternative fixed-point approach which seeks to directly achieve the conditions (11) could be faster when properly implemented. However, fixed point schemes are often notoriously unstable; therefore this approach seemed unsafe to us at this early stage of investigation when speed of convergence is of secondary importance compared to the need to understand the properties of the method. The gradient method on the other hand is very stable and guarantees that \( S \) increases with each iteration.

All our studies have been in the lowest symmetry \( \text{viz } p1 \) (in 2D) and \( P1 \) (in 3D). This was dictated by two considerations. Firstly, conventional methods work very well for higher symmetry crystals, particularly when they are centric. It seemed that the MEM could hope to make the greatest contribution in low symmetry non-centric structures. Secondly, the gradient algorithm requires continuous variation of phases whereas centric structures have only two values for each phase \( \text{viz } 0 \) or \( \pi \).

In our work with the gradient method, we often noticed that a single peak in the map grew at the expense of the others, thus making it virtually impossible to interpret the map. We have been able to suppress this tendency by maximising a modified form of “entropy” \( \text{viz } S' \)

\[
S' = \int \left( f[C_1 + \rho(r)] + f[C_2 - \rho(r)] \right) dV, \tag{17}
\]

where \( C_1 \) is a constant \((> - \rho_{\text{min}})\) which is adjusted to give the required value of \( R \) (equation (16)) and \( C_2 \) is a second constant set to have a value slightly greater than the theoretical height of the heaviest atom (see figure 4). The constant \( C_2 \) ensures that any peak which attains the expected height of an atom does not grow any further, thus giving a chance for the other atoms to develop. The constant \( C_1 \) is regularly updated at each cycle in order to maintain the value of the nonlinearity parameter \( R \). In the early
stages of ab initio structure determination, we usually start with totally random phases for the reflections. The corresponding \( \rho(\mathbf{r}) \) would have no distinct peaks and would have relatively large negative excursions. At this stage \( C_1 \) will have a large value. As the refinement proceeds, peaks will start developing and at the same time the ripple at the bottom would get ironed out, resulting in smaller values for \( C_1 \).

Each cycle of gradient refinement involves the following calculations. Using the current phases \( \varphi_H \), we compute \( \rho(\mathbf{r}) \) by means of a Fourier transform, compute the function \( f \left[ C_1 + \rho(\mathbf{r}) \right] - f \left[ C_2 - \rho(\mathbf{r}) \right] \) and calculate \( G_H \), its Fourier transform. Equations (10) then give the components of the gradient of \( S' \). Now we shift the \( \varphi_H \) by small amounts proportional to the gradient components and compute the new gradient at the shifted phases. We then calculate and apply the shift necessary in order to minimise the magnitude of the gradient at the final point (one could also use the more standard scheme of requiring the final gradient to be orthogonal to the starting one). Thus, one cycle involves four Fourier transforms.

5. Two-dimensional simulations*

We have carried out a number of tests of the MEM on model two-dimensional "structures" with typically 10–15 equal "atoms" in the unit cell. The electron density was sampled on a 32 \times 32 grid and correspondingly the structure factors ranged over reciprocal indices going from \(-15\) to \(+16\) (\(-16\) being identified with \(+16\) because of the translation symmetry in reciprocal space introduced by the discrete sampling in real space).

Figure 5 shows the result of a typical refinement from partial phase information. The structure factors of the model structure with 15 equal atoms were calculated and their phases were given random errors, with a rectangular distribution, of rms 70°. The

---

* The results described here were presented at the XIII National Seminar on Crystallography, Nagpur, India, March 1982.
corresponding electron density map is shown in figure 5a. The structure is still clearly seen but there are many spurious low-level features in the map. Figure 5b shows the remarkable improvement that can be achieved by means of MEM refinement. Here we did 20 cycles of gradient refinement with the entropy $S_1$ modified as in (17), using $R = 10$. The final rms phase error is only 21.7°. Actually, most of the improvement takes place in the first few cycles. For instance, after 10 cycles, the rms is as low as 30.8°. From a number of similar studies we find that the structure refines well from initial phase errors of up to 70–75°. There are many situations, particularly in macromolecular crystallography, where one can obtain limited phase information using techniques such as isomorphous replacement and anomalous dispersion. There appears to be good hope that the MEM can greatly help in phase refinement in such cases. Collins (1982) has published some results on phase refinement on real protein data using the MEM. While there is some improvement, his results are not as impressive as those in figure 5. This could be because of experimental errors in the structure amplitudes or, more probably, because the atoms in his structure are not well resolved (the MEM, being peak-seeking, is likely to work best with high resolution data).

The more difficult, and therefore more interesting problem is to solve a crystal structure *ab initio*, starting from pure amplitudes and no phase information. We have tried this in a number of two-dimensional simulations and figures 6–8 show some typical results. In each case we started the refinement with totally random phases and refined for 20 cycles with entropy $S_1$ at $R = 10$. Although the final map usually had a few spurious peaks, invariably the structure could be identified, with at most one or two atoms not seen. As a more objective test of the power of the method, we carried out a few “blind tests” of the following kind. One of us (Narayan) generated a model and produced a list of structure amplitudes along with stereochemical information on the
Figure 6. The true structure is the same as in figure 5. Random phases were given to the true structure amplitudes and refined using the MEM. Eleven out of 12 atoms can be fitted on peaks in the refined map. The atom marked A is absent with no nearby peak, while the peak S is a strong spurious feature. This is a typical result obtained with the MEM in two-dimensional ab initio structure analysis. Note the origin shift.

Figure 7. The true structure consists of 2 molecules, of 8 equal atoms each, with the orientations and relative spacings shown. The MEM map obtained by refining random starting phases reveals most of the structure except the absent atom A. There is one strong spurious peak S.
Figure 8. The true structure consists of 4 molecules, of 5 equal atoms each, in \( \text{cm} \) symmetry. The \( \text{MEM} \) map, refined from random starting phases, without including the information on the mirror, shows all the 10 independent atoms and has no strong spurious peaks.

"molecule". Another (Vani), given a map with random phases, would refine them by means of the \( \text{MEM} \) and try to interpret the resulting map. In all the three cases where this was tried the structure was solved correctly—usually, one or two atoms did not appear in the \( \text{MEM} \) map, but these could be easily located by standard Fourier techniques (Ramachandran and Srinivasan 1970). Because we start from random phases in our calculations, there is invariably an origin shift in the solution and in 50% of the cases there would also be an enantiomorph reversal.

A point to be emphasized is that all our calculations were carried out in \( \pi \) symmetry which is considered the most difficult. Figure 8 shows an interesting case where the true structure has \( \text{cm} \) symmetry but we did not use the information on the mirror during the refinement. Thus, although this structure has only five independent atoms, we solved it as if it was a 10-atom structure in \( \text{c}1 \) symmetry. It is interesting that the mirror symmetry comes to light during the structure solution, though not perfectly.

Another interesting feature is that in all our calculations we directly used the structure amplitudes \( |F_{Hj}| \) without any modification. In the direct methods a crucial prerequisite is "sharpening" of the data to obtain so-called \( E \)'s, corresponding to a point atom structure. In our experience, the \( \text{MEM} \) works very well with \( F \)'s and, in fact, does not refine so well when \( E \)'s are used.

6. Three-dimensional refinement—real structures

Encouraged by the results in two dimensions we have very recently embarked on three-dimensional structure determination using the \( \text{MEM} \). Due to memory limitations in the
mini-computer (PDP 11/34) with which we worked, we used a 16 x 16 x 16 unit cell for the electron density, corresponding to \((h, k, l)\) ranging from \(-7\) to \(+8\). Once again all refinement was carried out in \(P1\) symmetry using the gradient method. Whereas in two dimensions the entropy \(S_1\) appears to be marginally better than \(S_2\), we find the opposite to be true in three dimensions. Hence all the results described here correspond to the entropy

\[
S'_2 = -\int \{ [C_1 + \rho(r)] \ln [C_1 + \rho(r)] + [C_2 - \rho(r)] \ln [C_2 - \rho(r)] \} \, dV.
\]  

The constant \(C_1\) was adjusted to maintain \(R = 10\) while \(C_2\) was taken to be slightly greater than the theoretical height of the strongest atom. The refinement is typically slower than in two dimensions, needing of the order of 50–100 cycles, though once again most of the gain is achieved early on.

To begin with we generated a model structure with 15 equal atoms in \(P1\) symmetry and refined the phases using the MEM, starting from random values. The electron density after 60 cycles of refinement was scanned for peaks and these were listed in decreasing order of magnitude. The structure was very clearly identified from the peaks. Among the first 15 peaks, 10 belonged to atoms in the structure and 5 were spurious. Among the 10 correct peaks, 3 (viz peaks 2, 3, 4 in the order of ranking) were elongated and had a second atom in the tail region. This is a problem we faced in all our three-dimensional calculations—because of the lack of resolution (since we have only 16 grid points along each axis), peaks sometimes merged and two atoms would be represented by a single elongated peak. In such cases, usually one atom is close to the strongest region of the peak and the other weakly present in the tail. Thus, in our simulated 15-atom structure, the first 15 peaks accounted for 13 atoms and in fact the first 8 peaks were all correct and contained 11 atoms. The final 2 atoms in the structure were at peaks 23 and 26. We feel that the performance of the MEM is very encouraging.

Emboldened by this we have tried to “solve” a few real structures using real experimental data. Unfortunately, we could not find in the literature any suitable equal-atom \(P1\) structure whose data would fit within our 16³ grid. As the next best thing, we selected a few structures in \(P1\) symmetry and refined them in \(P1\) symmetry, ignoring the centre of symmetry altogether.

Martuscelli (1969) has reported the crystal structure of octa-trans-2, trans-6-diene-1,8-dioic acid which crystallizes in \(P1\) symmetry with unit cell parameters

\[
a = 12.00 \text{ Å}, 
\]

\[
b = 3.94 \text{ Å}, 
\]

\[
c = 7.72 \text{ Å}, 
\]

\[
\alpha = 121.35^\circ, 
\]

\[
\beta = 139.4^\circ, 
\]

\[
\gamma = 70.05^\circ.
\]

He solved the structure from about 500 observed reflections using model-fitting procedures based on the Patterson maps. Because the cell parameter \(a\) is rather large, our calculations on this structure were hampered by the extremely poor resolution (0.75 Å) in this direction. Also, we could include only reflections with \(h\) in the range \(-7\) to \(+8\) which reduced the number of structure amplitudes we could use to 326. In spite of these limitations, the structure refined quite easily from random phases. After 100 cycles of refinement, the first 12 peaks in the map included all the 12 atoms of the structure. Peaks 7 and 9 were spurious while peaks 1 and 2 included 2 atoms each. This is very gratifying since we had very limited data and severe resolution problems.

Our next attempt on a much larger equal atom structure was not so spectacular though still rather encouraging. Karle (1969) has reported the structure of a thymine-thymine adduct (obtained from irradiated thymine). This is again in \(P1\) symmetry with 38 atoms in the unit cell (2 x 19 independent atoms; \(a = 9.44 \text{ Å}, \)

\[
b = 8.29 \text{ Å}, 
\]

\[
c = 7.72 \text{ Å}, 
\]

\[
\alpha = 121.35^\circ, 
\]

\[
\beta = 139.4^\circ, 
\]

\[
\gamma = 70.05^\circ.
\]

He solved the structure from about 500 observed reflections using model-fitting procedures based on the Patterson maps. Because the cell parameter \(a\) is rather large, our calculations on this structure were hampered by the extremely poor resolution (0.75 Å) in this direction. Also, we could include only reflections with \(h\) in the range \(-7\) to \(+8\) which reduced the number of structure amplitudes we could use to 326. In spite of these limitations, the structure refined quite easily from random phases. After 100 cycles of refinement, the first 12 peaks in the map included all the 12 atoms of the structure. Peaks 7 and 9 were spurious while peaks 1 and 2 included 2 atoms each. This is very gratifying since we had very limited data and severe resolution problems.

Our next attempt on a much larger equal atom structure was not so spectacular though still rather encouraging. Karle (1969) has reported the structure of a thymine-thymine adduct (obtained from irradiated thymine). This is again in \(P1\) symmetry with 38 atoms in the unit cell (2 x 19 independent atoms; \(a = 9.44 \text{ Å}, \)

\[
b = 8.29 \text{ Å}, 
\]

\[
c = 7.72 \text{ Å}, 
\]

\[
\alpha = 121.35^\circ, 
\]

\[
\beta = 139.4^\circ, 
\]

\[
\gamma = 70.05^\circ.
\]
Ramesh Narayan, Rajaram Nityananda and G V Vani

= 7.57 Å, α = 99°0′, β = 91°5′, γ = 89°8′. The data consisted of 1844 independent reflections and the centrosymmetric structure was originally solved by obtaining the structure factor signs using the symbolic addition procedure (Karle and Karle 1966). We attempted to solve the 38 atoms of the structure in P1 symmetry with the mem, using 1370 reflections which fitted within our grid. After 100 cycles of refinement starting from random phases, we found that, of the first 38 peaks, 18 were correct, including 22 atoms (peaks 1, 6, 11 and 14 had two atoms each), while the first 80 peaks had 25 correct peaks with 31 atoms; 7 atoms had no peaks at all in the map. Of the two molecules, one was seen quite clearly while the other was rather weak. The refinement has clearly not been too successful; on the other hand, we believe we could, if necessary, have solved the structure from our map with a little patience if the structure were not already known. Karle (1969) used 6 symbols for which a unique assignment of signs was indicated by the symbolic addition method and the first 19 peaks of the resulting E-map showed the 19 independent atoms in the structure—a truly remarkable performance. However, considering that (i) we had to use less data, (ii) we had poor resolution in the electron density map, (iii) we refined 38 independent peaks, (iv) we worked in a non-centric framework and (v) we have not really optimised our refinement procedure, we feel our results using the mem are quite good. We may point out that a structure with 38 atoms in P1 is not trivial even today and would probably have been considered quite difficult a decade back.

As a final example we attempted a simple heavy atom structure. Dutta and Woolfson (1969) solved the structure of tetraethyl diphosphine disulphide, which crystallizes in P1 symmetry, by direct methods from two projections using two symbols in each. The unit cell has 12 atoms (6 independent atoms if one uses the P1 symmetry), of which 2 sulphur and 2 phosphorus atoms are heavy, while the other 8 are (light) carbon atoms. The unit cell parameters are

\[
a = 8.98 \text{ Å}, \quad b = 6.45 \text{ Å}, \quad c = 6.15 \text{ Å}, \quad \alpha = 113°0′, \quad \beta = 85°2′, \\
\gamma = 102°5′.
\]

Of the 864 reflections observed, we could accommodate only 760 within our grid and the phases of these were refined as usual with the mem. After 70 cycles of refinement the first 12 peaks contained 10 atoms with 2 spurious peaks (8 and 10 in ranking). The first 4 peaks were the 4 heavy atoms and the two atoms that were missing were the weakest in the final solution, having heights only \(\sim 18\%\) of the sulphur atoms. The success of the mem with this structure is interesting as it shows that the method is equally applicable to heavy atom structures (as indeed the direct methods also are, although they were originally developed for equal or near-equal atom structures).

In the various structure determinations described above, there was nothing special about the starting random phases which could have helped in the refinement. We have tried different sets of random phases and usually obtained comparable results. Also, although we refined for \(\sim 100\) cycles in all the cases, most of the refinement is over within the first 20 cycles and the later calculations only help to "sharpen" the peaks and define them better. Finally, we should mention that, as in two dimensions, here again we obtained much better results using the F's rather than E's.

7. Discussion and conclusions

The phase problem of crystallography is probably the most difficult and challenging among the variety of Fourier inversion problems known in several fields. The success of the mem in other fields in different contexts is therefore no guarantee of success when
Maximum entropy approach to crystallography

applied to crystallography. Although we and a few others have been excited about the prospects of the MEM in crystallography, there was until recently only limited reason for optimism. Firstly, we (as well as Britten and Collins 1982) had shown that the MEM is related to the maximum determinant method of crystallography; we had further argued that the MEM phase relations (equations (10)) are higher order phase relations which are likely to be more powerful than the lower order relations (triplets, quartets, etc.) conventionally used and that equal atom structures satisfy the MEM relations. Secondly, Collins (1982) had shown that phase refinement is feasible in macromolecular crystallography, starting from approximately known phases.

We believe that the results presented in this paper greatly improve the case for trying the MEM in crystal structure analysis. We have shown that the method works well in small- and medium-sized real crystals and has the ability to locate a high percentage of the atoms starting from random phases for the structure factors. It is worth stressing once again that all our work has been in P1 symmetry, which is usually considered the most difficult.

A comparison of MEM refinement (as we implement it) with conventional direct methods shows two important differences.

(a) In the direct methods one works with $E$'s and solves for point atoms. On the other hand, we find that the MEM works best with $F$'s. It is possible that this difference arises because the MEM, apart from looking for peaks as do the direct methods, is also concerned about the flatness of the baseline. Since an $E$-map is likely to have a lot of ripple even with correct phases (due to data truncation and experimental errors), this might be a possible explanation.

(b) In the direct methods, the multisolution approach is the rule where a small set of primary reflections are systematically given various phases. One set then produces the structure while the others are usually uninterpretable. In our tests so far we have found that, regardless of the starting set of random phases, the structure usually comes up to about the same level of interpretability. It is possible that we are still working with structures that are too small and that the MEM also may need a multisolution approach with larger structures. On the other hand, there is a distinct possibility that, because of its higher order phase relations, the method is able to attract the solution towards the correct structure from a wider range of starting phases.

Because of the above reasons we feel that the MEM is not just the direct methods in a new disguise but is distinctly different. As such it merits further study in more complicated situations. The ability to work with $F$'s rather than $E$'s might turn out to be a big advantage since the method will be less susceptible to experimental errors. In our view the following further questions now need urgent attention.

(a) Although the gradient method we have employed is useful for the sort of preliminary studies we have made, the future probably belongs to fixed point schemes (e.g. Sayre 1980) which directly seek to satisfy the MEM conditions (10). Not only would the method then easily work for centrosymmetric structures, there is also the real advantage that one can use one's a priori information on the shapes of atoms through the $S_H$ in (14) to determine the $n_H$ in (11). This information cannot be introduced into the gradient algorithm.

(b) One needs sensitive figures of merit to measure the goodness of the refinement, particularly if the MEM is to be combined with the multisolution approach.

(c) Techniques such as isomorphous replacement and anomalous dispersion are employed to give added information in macromolecular phase refinement and these
could be put in variationally into the mem. For instance, in isomorphous replacement, one knows that the electron density distribution in a protein and its derivatives are identical except at the locations of the replaced atoms where the differences are known. So too, in the anomalous dispersion case, where the electron density becomes complex, one has the powerful a priori information that the peaks in the imaginary part occur only at the centres of the peaks in the real part. One technique to introduce these extra constraints is to maximise the integral of \( f(\rho_{\text{derivative}} - \rho_{\text{protein}}) \) and \( f(\rho_{\text{real}} - \rho_{\text{imaginary}}) \), which is equivalent to introducing the a priori input that \( (\rho_{\text{derivative}} - \rho_{\text{protein}}) \) has a flat baseline with sharp well-resolved peaks and that \( (\rho_{\text{real}} - \rho_{\text{imaginary}}) \) has no negative peaks. These and other similar ideas deserve to be pursued, particularly since Collins (1982) has obtained interesting results even with just the bare mem.

Acknowledgements

It is both a privilege and a pleasure to contribute this article to a volume commemorating Prof. S Ramaseshan’s sixtieth birthday. A whole band of friends, colleagues, students, grandstudents, etc have benefited from his initiative, insight and encouragement. We join with them in wishing him many active and fruitful years of scientific work.

Appendix

We present here an illustrative example in one dimension bringing out the crucial role played by the zero Fourier coefficient \( F_0 \). We consider a periodic real function in one dimension which is sampled at 4 points in the unit cell so that the function is completely described by its values \( p_0, p_1, p_2 \) (the origin has been taken at the second point). The Fourier coefficients are also periodic. As the \( \rho \)'s are real, \( F_0 \) and \( F_2 \) are real and \( F^*_{-1} = F_1 = |F_1| \exp(i\varphi_1) \) are the only structure factors with a continuously variable phase; even here, if we eliminate physically identical solutions (related by an origin shift or enantiomorph reversal), the range of \( \varphi_1 \) from 0 to \( \pi/2 \) completely describes all distinct solutions compatible with a given set of amplitudes \( F_0, |F_1|, F_2 \).

Consider the centric “structure” given by

\[
\rho_0 = 2.0, \quad \rho_{-1} = \rho_1 = 0.5, \quad \rho_2 = 0, \quad (A1)
\]

whose Fourier transform gives

\[
F_0 = 1.0, \quad |F_1| = 0.75, \quad \varphi_1 = 0^\circ, \quad F_2 = 0.5. \quad (A2)
\]

We investigate the variation of the entropy

\[
S_1(C, \varphi_1) = \int \ln [C + \rho(j \varphi_1)] \, dx = \sum_{j = -1}^{2} \ln [C + \rho(j \varphi_1)], \quad (A3)
\]

where \( \rho(j \varphi_1) \) are calculated with the true amplitudes and the phase \( \varphi_1 \). Adding \( C \) to all the \( \rho \)’s is equivalent to replacing \( F_0 \) by \( F_0 + C \). Figure 9 shows the variation of \( S_1 \) with \( \varphi_1 \) for different values of \( C \). At high values of \( C \), the maximum value of \( S_1 \) is seen to occur at the correct phase, \( \varphi_1 = 0 \). However, at low values of \( C \) the maximum shifts away from 0 and at \( C = 0.02 \), for instance, occurs at \( \varphi_1 \sim 10^\circ \). Thus, at low values of \( C \), the mem solution is non-centric even though the data correspond to a centric structure.
Consider next the non-centric structure
\[ \rho_{-1} = 0, \rho_0 = 3.0, \rho_1 = 1.0, \rho_2 = 0, \]  \[ F_0 = 1.0, |F_1| = 0.7906, \varphi_1 = 18.43^\circ, F_2 = 0.5. \]  \hspace{1cm} (A4) \hspace{1cm} (A5)

The variation of the entropy \( S_1 \) for this case is shown in figure 10. At low values of \( C \), the maximum value of \( S_1 \) occurs close to the true phase e.g. for \( C = 0.05 \), we obtain \( \varphi_1 \sim 18^\circ \). However, as \( C \) increases, the maximum shifts to lower values of \( \varphi_1 \) until above a critical value of \( C \) the maximum is at \( \varphi_1 = 0 \), i.e. a centric structure. This tendency of a non-centric structure to move towards a centric solution at high \( C \) (or low \( R \)) is also discussed in § 3 with a two-dimensional example.

The above results suggest that the value of \( C \) should be set by the user depending on prior knowledge of the presence or absence of a centre of symmetry. We have found by experience that very low values of \( C \) lead to slow convergence while very high values of \( C \) can cause round-off errors in the calculations. In our work, with these two auxiliary constraints in mind, we constantly update \( C \) at each cycle so as to maintain the non-linearity parameter \( R \) at a value \( \sim 10 \). This has given good results with both centric and non-centric structures.
References


Collins D M 1982 Nature (London) 298 49


Karle I L and Karle J 1963 Acta Crystallogr. 16 969


Karle J and Hauptman H 1950 Acta Crystallogr. 3 181


Sayre D 1952 Acta Crystallogr. 5 60


Tsoucaris G 1970 Acta Crystallogr. 26 492


Extinction correction and synchrotron radiation

P SUORTTI
Department of Physics, University of Helsinki, Siltavuorenpenkeri 20D, SF-00170 Helsinki 17, Finland
Present address: NSLS, Brookhaven, National Laboratory Upton, New York 11973, USA

Abstract. The primary extinction factor $y_p$ is defined as the ratio of the integrated reflection from a coherently diffracting domain to the integrated kinematical reflection from the same domain. When $y_p$ is larger than 0.5 it may be approximated by $y_p = \exp\{-a(\delta)^2\}$, where $a$ is about 0.5 and $\delta$ the average size of the coherent domain when measured in units of the extinction length $\Lambda$, $\delta = D/\Lambda$.

Transfer equations are applied to symmetrical Laue diffraction, and the reflectivity per unit length, $\sigma(\varepsilon)$ is solved from the measured reflecting ratio as a function of the rocking angle $\varepsilon = \theta - \theta_g$.

Measurements with conventional x-ray sources are made on single crystal slabs of Be and Si using AgK$\beta$, MoK$\alpha$, and CuK$\alpha$ radiation. The primary extinction factor $y_p(\varepsilon)$ is solved from a point-by-point comparison of two measurements where the extinction length $\Lambda$ is changed by varying the polarization and/or wavelength of the x-ray beam. The results show that primary and secondary extinction are strongly correlated, and that the customary assumption of independent size and orientation distributions of crystal mosaics is unjustified. The structure factors for Be and Si show close agreement with other recent measurements and calculations.

The limitations of the method are discussed in length, particularly the effects of beam divergences and incoherence of the rays in the crystal. It is concluded that under typical experimental conditions the requirements of the theory are met. Practical limitations arising from the use of characteristic wavelengths and unpolarized radiation prohibit the use of the full potential of the method.

The properties of a synchrotron radiation source are compared with a conventional x-ray source, and it is demonstrated that the experimental limitations can be removed by the use of synchrotron radiation. A diffraction experiment with synchrotron radiation is outlined, as well as generalization of the method to small spherical crystals.

Keywords. Extinction; synchrotron radiation.

1. Introduction

In his paper “The reflection of x-rays from imperfect crystals” Darwin (1922) writes about the theory of extinction which he is going to introduce: “I am afraid I have not succeeded in welding the parts of the argument rigorously together, but in spite of certain gaps in the theory it seems unlikely that there is serious error in the general views to which it leads”. After more than 60 years, our understanding of extinction is still based on Darwin's views, on the concepts of primary and secondary extinction. In the main line of crystallography, correction schemes have been incorporated in procedures for structure determination, but in a sense this has been a step backwards, as the extinction calculations are based on the energy transfer equations, which cover secondary extinction only (Hamilton 1957; Zachariasen 1967; Cooper and Rouse 1970; Coppens and Hamilton 1970; Becker and Coppens 1974a, b, 1975). Moreover, the extinction corrections are derived by least squares fitting to a crystal model which includes theoretical structure factors and extinction parameters. These are highly correlated, and although the extinction correction greatly improves the average quality of the diffraction data it may also introduce artefacts.
Darwin’s paper was purported to provide understanding of the experimental results by Bragg et al (1921, abbreviated hereafter as \( bb \)). Since those days, there has always been a line of direct experimental approach to the problem of extinction. The kinematical limit, where the feedback between the direct and reflected beams can be ignored, has been approached in various ways. These include methods where the active volume of the crystal or the reflectivity per unit volume are reduced (for references, see Mathieson 1979).

Some of the basic ideas of the present work can be found already in Darwin’s paper where he writes: “The methods used in \( bb \) remove the secondary extinction, but are without influence on the primary. In fact, it will appear that no experiments of the present type can possibly remove it; indeed, to do so would require the measurement of the actual sizes of the small blocks of perfect crystal.” . . . “Now, at each instant of the rotation there will be a different amount reflected and therefore a different extinction, and consequently it will not suffice to treat of the mean effect of extinction, without first determining it at every setting of the crystal.” In the following we will see that the degrees of both primary and secondary extinctions are correlated to the reflection profile and each other, and that when the actual sizes of the coherent domains are expressed in units of extinction length the sizes can be determined from two measurements and the effects of primary extinction removed.

The details of the method have been described earlier (Suortti 1982a, b), and the aim of this paper is to outline the method, give some new results, discuss the limitations of the theory and those of an experiment with a conventional x-ray source, and finally describe how the full potential of the method is utilized by the use of synchrotron radiation.

2. Transfer equations

In the following we will assume that the radiation from the x-ray source is incoherent. For a conventional x-ray tube this is obvious, and even with a synchrotron radiation source the possible effects of wave coherence are small. Consider the two waves in figure 1, which illustrates crystal domains with constant but independent lattice phases \( G \). When wave (1) enters domain \((a)\) at an angle \( \theta \), which is close enough to a Bragg angle \( \theta_g = \theta - \epsilon \), two waves of electric displacements \( D_d \) and \( D_g \) are excited, and these have a constant phase difference within \((a)\), i.e. the amplitudes of the waves are coupled. The resulting decrease of \( D_g \) is called primary extinction. On the other hand, the difference between \( D_d \) and \( D_g \) is arbitrary, the intensities from \((a)\) and \((b)\) add up. However, domain \((b)\) is screened by diffraction in \((a)\), in addition to the normal absorption, and this loss due to intensity coupling is called secondary extinction. Also the second diffraction of \( D_g \) in domain \((c)\) decreases the reflected power \( P_g^* \). The flow of energy between \( P_d^* \) and \( P_g^* \) is covered by the energy transfer equations,

\[
\frac{\partial P_d^*}{\partial s_d} = - \mu_d \sigma_g P_g^* + \sigma_g P_g^* \tag{1a}
\]

\[
\frac{\partial P_g^*}{\partial s_g} = - \mu_g \sigma_d P_d^* + \sigma_d P_d^* \tag{1b}
\]

where \((s_d, s_g)\) are oblique coordinates along the direct and diffracted beams, respectively. If we assume a non-polar crystal, \( \sigma_g = \sigma_{-g} \), and ignore the possibility of the Borrmann effect,

\[
\mu_e = \mu_0 + \sigma, \tag{2}
\]
Extinction correction and synchrotron radiation

Figure 1. Diffraction from coherent domains. Within each domain, the lattice phase $G$ is constant, but the differences between $G_a$, $G_b$ and $G_c$ are arbitrary. $D_d$ and $D_g$ are the electric displacements of the direct and reflected beams excited in domain (a). The area of coherent coupling for the ray entering (a) at $A$ and exiting at $B$ is indicated by $s_d$ and $s_g$, and the total area of wave coherence with dots. The shaded areas of (b) and (c) show incoherent coupling of the direct and reflected beams of rays (1) and (2).

where $\sigma = \sigma(e, s_d, s_g)$ is the reflectivity per unit length of the traversed crystalline medium and $\mu_0$ the linear absorption coefficient. The observable reflectivity $\sigma(e)$ is the integral of $\sigma(e, s_d, s_g)$ over the region of interaction for the extended incident beam, and the integrated intensity is found by an angular integration.

3. Primary extinction

The transfer equations for two waves that have amplitude coupling were introduced by Takagi (1962, 1969) and Taupin (1964):

$$\frac{\partial D_d}{\partial s_d} = iK_g D_g \exp\{iG(s_d, s_g)\}, \quad (3a)$$

$$\frac{\partial D_g}{\partial s_g} = i\kappa_g D_d \exp\{-iG(s_d, s_g)\}, \quad (3b)$$

where the coupling coefficient is

$$\kappa_g = r_e \lambda C F_g / V_e. \quad (3c)$$

Here $C$ is the polarization factor ($\cos 2\theta$ when the electric vector is in the plane of diffraction, 1 when perpendicular to that), $\lambda$ the x-ray wavelength, $r_e = e^2/mc^2$ the electron scattering length, $V_e$ the unit-cell volume, and $F_g$ the structure factor. The lattice phase $G = 2\pi g \cdot \mathbf{u}$, where $\mathbf{g}$ is the scattering vector and $\mathbf{u}$ the lattice distortion.

The two waves $D_d$ and $D_g$ stay coherent within the domain where $G$ is constant. As seen in figure 1, the maximum area of coherent coupling is $s_d s_g \sin 2\theta$, and the appropriate average is found by an integration over the exit surface. The “size” of the coherent domain may be defined as the average

$$D = \sqrt{s_d s_g}. \quad (4)$$

Explicit calculation is possible for a given cross-section of a perfect crystal, and detailed results are available for square and circular shapes (Olekhnovich and Olekhnovich 1978, 1980). If the incoming ray is taken as a plane wave, reflectivity curves $R(e)$ can be calculated; an example is illustrated in figure 2. The detailed form of $R(e)$
depends on the crystal shape, wavelength and polarization, but the integral of \( R(\varepsilon) \) is well defined by \( \kappa_p D \), and this is the basis of the extinction correction to be introduced in the following.

The integrated reflectivity due to domains of size \( D \) is (Suortti 1982a)

\[
\sigma_D(\varepsilon) = \int R_D(\varepsilon - \eta)w_D(\varepsilon)\,d\varepsilon
\]

\[
\approx w_D(\varepsilon) \int R_D(\varepsilon - \eta)\,d\varepsilon = Qy_p(D)w_D(\varepsilon).
\] (5)

Here \( \eta \) is the deviation of domains from the average orientation, \( w_D(\varepsilon) \) is the distribution function of domain sizes (which is supposed to be slowly varying in comparison with \( R_D \)), \( Q = (\lambda / \sin 2\theta_B)\kappa_p \) is the integrated kinematical reflectivity per unit length (or integrated reflecting power per unit volume; see James 1962), and \( y_p(D) < 1 \) is the primary extinction factor. The observable reflectivity is

\[
\overline{\sigma}(\varepsilon) = Q \int y_p(D)w_D(\varepsilon)\,dD.
\] (6)

The integrated reflectivity per unit length of the crystallite, \( Q' = Qy_p(D) \), can be expressed in terms of low-order Bessel functions \( J_{2n} \). The argument of the functions is

\[
\delta = \kappa_p H(s_g^2s_g) = D/\Lambda,
\] (7)

where \( \Lambda = (\kappa_p H)^{-1} = V_c/(r_0\lambda CF_g H) \) is the extinction length. Here \( H \) is the long-range order parameter, which for a perfect crystal is the Debye-Waller factor \( \exp(-M) \). The above means that the arguments of the intensity expressions become linear in the size of the coherent domain, when this is measured in the units of \( \Lambda \). The effective size \( \delta \) can be varied by changing \( \lambda, C \) or \( F_g \), and this gives an experimental possibility for
determination of primary extinction.

At small enough values of $\delta$ it follows from the properties of $J_{2n}$ that

$$y_p(\delta) = Q'/Q = Q^{-1} \int R_4(\psi) d\psi \approx \exp\{ - (\alpha \delta)^2 \},$$  

(8)

where the numerical value of $\alpha$ is about 0.5. Model calculations (Olekhnovich and Olekhnovich 1978, 1980) suggest that this approximation is valid when $\delta < 1.5$ or $y_p > 0.5$. It is worth noting that the same functional form can be recovered from Darwin’s (1922) original work,

$$y_p(D) = \frac{\tanh m\theta}{m\theta} \approx \exp\{ - (D/\sqrt{3}\Lambda \sin \theta)^2 \}. 
\quad \text{(8')}
$$

the extra angular factor $(\sin \theta)^{-1}$ arises from the special geometry of a laterally infinite plate.

The primary extinction factor at a given average angle of incidence is weighted by the distribution of domain sizes,

$$y_p(\epsilon)w(\epsilon) = \int y_p(\delta)w(\delta(\epsilon)) d\delta = Q^{-1} \sigma(\epsilon).$$  

(9)

If we assume that $w(\delta(\epsilon))$ is a Gaussian of width $2\Delta_\epsilon$, centred at $\delta_\epsilon$, convolution resulting from substitution of (8) to (9) yields

$$y_p(\epsilon) = \exp\{ - (\alpha' \delta_1)^2 / (1 + \alpha \sqrt{2}\Delta_\epsilon)^2 \} = \exp\{ - (\alpha \delta_\epsilon)^2 \}. 
\quad \text{(10)}
$$

Although idealized, this calculation suggests that the functional form of (8) can be used also for real crystals.

The primary extinction factor can be determined from two measurements of $\sigma(\epsilon)$ where the extinction length is changed. Writing $\delta_1$ and $\delta_2$ for $\delta_\epsilon$ in (10) and assuming that the effective crystal volume is the same in both measurements, one obtains from (10)

$$f(\epsilon) = \frac{y_{p,1}(\epsilon)}{y_{p,2}(\epsilon)} = \frac{Q_2}{Q_1} \frac{\sigma_1(\epsilon)}{\sigma_2(\epsilon)}. 
\quad \text{(11a)}
$$

The correction for primary extinction in measurement 1 is

$$y_{p,1}(\epsilon) = (f(\epsilon))^{\beta}.
\quad \text{(11b)}
$$

where

$$\beta^{-1} = 1 - (\Lambda_1/\Lambda_2)^2. 
\quad \text{(11c)}
$$

The extinction length $\Lambda$ can be varied by changing the polarization factor $C$ or the wavelength $\lambda$. A change in $\lambda$ usually entails a change of $C$, and

$$\beta_C^{-1} = 1 - \left(\lambda_2 C_2/\lambda_1 C_1\right)^2. 
\quad \text{(12a)}
$$

If the polarization of the beam is parallel to the plane of diffraction in one measurement, $C_1 = |\cos 2\theta|$, and perpendicular to this plane in the other, $C_2 = 1$,

$$\beta_C^{-1} = 1 - \sec^2 2\theta. 
\quad \text{(12b)}
$$
4. Secondary extinction

The energy transfer extinction equations must be solved under the boundary conditions. A simple closed-form solution is available only for a parallel-sided crystal, but an iterative solution can be presumably found for any simple polyhedral crystal, when the so-called \( AB \)-extinction formulas by Werner (1974) are used. The quantity determined experimentally is called the reflecting ratio

\[
\frac{P_g^*}{AP_{d,o}} = r^*(\varepsilon),
\]

(13)

where \( A \) is the absorption factor, and \( P_{d,o} \) the incident beam. In the situation illustrated in figure 3, where the crystal is bathed in the incident beam, determination of \( AP_{d,o} \) may require extensive subsidiary measurements. The corrected reflecting ratio, where the decrease of \( P_d \) due to diffraction is counted for, is defined by

\[
r^*(\varepsilon) = y_s(\varepsilon)r(\varepsilon),
\]

(14a)

where \( y_s(\varepsilon) \) is the secondary extinction factor. From conservation of energy, which is demonstrated in the figure,

\[
r(\varepsilon) = \frac{r^*(\varepsilon)}{1 - r^*(\varepsilon)}.
\]

(14b)

The boundary conditions come into the play when the reflectivity \( \sigma(\varepsilon) \) is to be solved from \( r(\varepsilon) \).

Laue diffraction from a parallel-sided crystal plate is illustrated in figure 4, and the powers of the reflected and direct beams can be written as

\[
P_g^* = P_{d,o} \exp(-aT) \sinh(\sqrt{a^2 - b T}),
\]

(15a)

\[
P_d^* = P_{d,o} \exp(-aT) \left\{ (\gamma_g/\sigma) \sqrt{a^2 - b} \cosh(\sqrt{a^2 - b T}) \right. \\
+ \left. \left( \frac{\mu_{\sigma} + \sigma}{2\sigma} \right) (1 - \gamma_g/\gamma_d) \cdot \sinh(\sqrt{a^2 - b T}) \right\}.
\]

(15b)

The geometrical factors are explained in the legend of figure 4, and

\[
a = \frac{1}{2} (\mu_{\sigma} + \sigma)(1/\gamma_g + 1/\gamma_d).
\]

(15c)

\[
b = (\mu_{\sigma}^2 + 2\mu_{\sigma}\sigma)/(\gamma_g\gamma_d).
\]

(15d)

The absorption factor for the transmitted beam is \( A = \exp(-\mu_o T/\gamma_d) \), and the observed

![Figure 3. Crystal bathed in the incident beam. \( P_d^* \) is the measured power of the direct beam and \( P_g^* \) that of the reflected beam, and \( AP_{d,o} \) is the transmitted beam when there is no diffraction. By conservation of energy, \( P_d^* + P_g^* = AP_{d,o} \).](image-url)
reflecting ratio

\[ r^*(\varepsilon) = \exp\left\{ -\frac{1}{2} \mu_o T (1/\gamma_d - 1/\gamma_g) \right\} \times \exp\left\{ -\frac{1}{2} \sigma T (1/\gamma_d + 1/\gamma_g) \right\} \cdot \sinh(\sqrt{a^2 - b T}). \]  

(16)

The reflectivity per unit length, \( r(e) \), can be solved from the experimental \( r^*(\varepsilon) \). The formulas become cumbersome, except in the symmetrical case where \( \gamma_g = \gamma_d = \cos \theta \), and with (14),

\[ r(e) = \sinh(\sigma T / \cos \theta) \approx \sigma(e) T / \cos \theta. \]  

(17)

Small deviations from symmetry can be considered as miscut of the crystal, and corrected for by using series expansions of (15).

5. Experimental results

The method realizes Darwin’s appeal that extinction is to be determined at each setting \( \varepsilon \) of the crystal. This imposes stringent bounds to the experimental situation: on the one hand, the divergences should not smear \( r^*(\varepsilon) \) or \( \sigma(e) \), on the other, \( \gamma_g(D) \) is determined only by integration over \( R_p(e - \eta) \). Therefore, it is advisable to estimate \( \Delta \varepsilon \) in (5) and compare it with typical widths of reflections.

We are interested in the cases where \( \gamma_g(\delta) < 0.99 \), or \( \delta > 0.2 \); for smaller coherent domains the integral over \( R_p(e - \eta) \) is always \( Q \). The total width of \( R_p \) is approximately \( 2\lambda / D \), and at the above limit \( \Delta \varepsilon \approx 10\lambda / \Delta \). Table 1 gives a few typical values for the strongest reflections of various crystals, and the results indicate that \( \Delta \varepsilon \) is of the order of 0.1 mrad for light materials.

The upper limit for the divergence is due to the requirement that \( r^*(\varepsilon) \) should not be appreciably smeared by \( \Delta \varepsilon \). The possible minimum width of a reflection can be estimated as follows: at the kinematical limit \( Q = \sigma_{\text{kin}}(0) \beta \), where \( \beta \) is the integral width of the reflection, and in the present geometry, \( \sigma_{\text{kin}} = r_{\text{kin}}(T / \cos \theta) \). The maximum value of \( r_{\text{kin}}(0) \) is 1, which corresponds to \( r^*(0) = 0.5 \), and so \( \beta_{\text{min}} = Q T / \cos \theta \). For instance, \( \beta_{\text{min}} \) for Be(00-2) measured with MoK\( \alpha \) from a 1.1 mm thick crystal is 0.35 mrad, and
Table 1. Extinction lengths (in units of $\mu$m) and widths of reflectivity curves (in units of mrad) of domains with $D/A = 0.2$ for the strongest reflections of various crystals at a few x-ray wavelengths $\lambda$.

<table>
<thead>
<tr>
<th>Refl.</th>
<th>$\lambda = 0.561$ Å</th>
<th>$\lambda = 0.710$ Å</th>
<th>$\lambda = 1.542$ Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Be(00-2)</td>
<td>30.5 0.02</td>
<td>24.1 0.03</td>
<td>11.1 0.14</td>
</tr>
<tr>
<td>LiF(200)</td>
<td>14.0 0.04</td>
<td>11.0 0.06</td>
<td>5.1 0.30</td>
</tr>
<tr>
<td>Al(111)</td>
<td>12.2 0.05</td>
<td>9.7 0.07</td>
<td>4.5 0.34</td>
</tr>
<tr>
<td>Si(220)</td>
<td>14.9 0.04</td>
<td>11.8 0.06</td>
<td>5.4 0.29</td>
</tr>
<tr>
<td>Fe(110)</td>
<td>4.1 0.14</td>
<td>3.2 0.22</td>
<td>1.5 1.03</td>
</tr>
</tbody>
</table>

for Si(220) of the same thickness 1.5 mrad; these values are an order of magnitude larger than $\Delta \varepsilon$ in table 1.

The above considerations determined the experimental arrangement, which has been explained in detail earlier (Suortti 1982b). In the case of Be, the non-dispersive (1, -1) arrangement of the monochromator and sample crystal could be used, and in the case of Si, $K\alpha_2$ component of radiation could be eliminated by narrow collimation. The measuring geometry is illustrated in figure 5. MoK$\alpha_1$ and AgK$\beta$ were obtained by reflection from a Si(220) flat monochromator, which had a rocking curve width of 0.1 mrad. CuK$\alpha$ was selected by a polarizer-monochromator which was based on anomalous transmission or Borrmann effect. The electric vector of the transmitted beam lies in the diffracting planes and can be rotated by rotating the Ge(220) Borrmann crystal about the axis of the x-ray beam.

Figure 5. Measuring geometry in the plane of diffraction with a Borrmann polarizer $B$ (a and b) and reflecting monochromator $M$ (c). The point focus of the x-ray tube is indicated by $F$, the sample by $C$, and the scintillation counter by $SC$. The forward diffracted beam from the polarizer is used, and the other beam is caught by a beam stop shown in (a). The inserts in (a) and (c) show the non-dispersive setting when the Bragg angle of the polarizer or monochromator is equal to that of the sample crystal.
The measurement on Be was made with (almost) unpolarised MoKα₁ and with polarized CuKα. Si was measured with AgKβ and MoKα₁. The reflections were step-scanned in intervals of 0.05 to 0.1 mrad, and the power of the primary beam was measured by the same scintillation detector. With MoKα₁ and AgKβ an accurately calibrated attenuator was needed. The reflecting ratio \( r(\theta) \), which is corrected for secondary extinction, was calculated at each angle of measurement, and \( y_p(\theta) \) was determined from comparison of two measurements of the same reflection when the extinction length \( \Lambda \) was varied through \( \lambda \) or \( C \). The details of this procedure have been given earlier (Suortti 1982b).

Some of the results are reproduced in table 2 together with earlier experimental and theoretical values, and plots of a few reflections are shown in figure 6. The structure factors of Be are based on quite a few measurements at different locations of the crystal, and the agreement with other recent values is very good, although the average primary and secondary extinction corrections reach 30%, and the peak values of \( y_p^{-1} - 1 \) and \( y_p^{-1} - 1 \) are much larger. The implications of the results have been discussed in detail by Larsen and Hansen (1983), who conclude that the charge density distribution in Be metal can be interpreted by sp³ hybridization, where the orbitals are directed towards the tetrahedral holes. The situation is well described by an LCAO calculation (Dovesi et al 1982), although some of the features are exaggerated. The measurement on Si is a real test case of the method, as the true values of \( F(hkl) \) for MoKα radiation are known very accurately from the Pendellosung measurements by Aldred and Hart (1973). Considering that the experimental situation was quite unfavourable, as the crystal was too thick and so the extinction corrections always large, the results are very satisfactory and substantiate the soundness of the method.

It is worth noting that the degree of primary extinction, \( y_p^{-1}(\theta) - 1 \), follows closely the reflectivity curve when the reflection is narrow, while \( y_p(\theta) \) of a wide reflection is almost constant. The former behaviour is not covered by the customary version of the

### Table 2. Experimental and theoretical structure factors of Be and Si at room temperature.

<table>
<thead>
<tr>
<th></th>
<th>Present</th>
<th>HSL</th>
<th>LH</th>
<th>DPRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Be:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( hkl )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1001</td>
<td>1.90±0.01</td>
<td>1.83</td>
<td>1.85</td>
<td>1.914</td>
</tr>
<tr>
<td>002</td>
<td>3.37±0.02</td>
<td>3.32</td>
<td>3.37</td>
<td>3.397</td>
</tr>
<tr>
<td>101</td>
<td>2.78±0.03</td>
<td>2.83</td>
<td>2.84</td>
<td>2.810</td>
</tr>
<tr>
<td>Si:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( hkl )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>59±60</td>
<td>60±55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>220</td>
<td>69±10</td>
<td>67±84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>44±90</td>
<td>43±96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>222</td>
<td>1.48</td>
<td>1.35 (calc.), 1.50 (ref.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>004</td>
<td>56±70</td>
<td>56±62</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The entries for Be include a measurement with γ-rays (Hansen et al 1983), a conventional measurement from a small crystal with AgKα and MoKα radiations (Larsen and Hansen 1983), and results of an LCAO calculation (Dovesi et al 1982). The theoretical values have been multiplied by the Debye-Waller factor \( \exp(-B \sin^2 \theta / \lambda^2) \), where \( B_{11} = 0.460 \text{ Å}^2 \) (in basal plane) and \( B_{33} = 0.415 \text{ Å}^2 \) (perpendicular to basal plane). The reference values for Si are from a Pendellosung measurement by Aldred and Hart 1973.
Figure 6a. Reflecting ratio $r(\varepsilon)$ and the corresponding correction for primary extinction as a function of the rocking angle $\varepsilon = \theta - \theta_{\parallel}$ of the Be crystal. Open circles (○) give $r(\varepsilon) = r_{2}(\varepsilon)$ with CuKα radiation, and the results with MoKα are brought to the same scale by multiplying by $(Q_{2}/Q_{1})(\cos \theta_{1}/\cos \theta_{2})$; filled circles (●) indicate $r(\varepsilon) = r_{1}(\varepsilon)$ before and crosses (×) after the correction for primary extinction.

Figure 6b. See the legend of figure 6a.

Figure 6c. Reflecting ratio of the measurements with polarized CuKα radiation. The lowest curve is the measured profile $r^{*}(\varepsilon)$ when $C = 1$, the middle one, $r(\varepsilon)$, is corrected for secondary extinction, and the top curve is the kinematical reflection profile calculated from the comparison with the measurement with $C = |\cos \theta_{\parallel}|$. The scale is the same as in figure 6a.
mosaic-crystal model, where the orientation and size distributions are independent. This implies a constant primary extinction over the profile, and the rest of the effect is termed secondary extinction, which is either orientation-determined (type I) or size-determined (type II) according to Zachariasen (1967) (see the next section). The present results show, however, that $y_p(\varepsilon)$ and $y_g(\varepsilon)$ are often strongly correlated, and that interpretations based on the mosaic-crystal model may lead to false conclusions (cf. Olekhnovich et al 1980).

6. Limitations of the method

Questions of validity concern the underlying principles as well as the practical realization of the method. Some of the problems have been discussed in detail in earlier papers (Suortti 1982a, b), and will be only mentioned here.

6.1 Divergences of the beams

The central question is that of the divergences of the beams inside the crystal. In Darwin's words: "... each layer will, on account of diffraction, spread out incident parallel rays into a certain range of angles and so will continually change the angle at which they attack the successive layers". In practice, the incident beam has certain divergences, and figure 7 illustrates how the divergences develop in diffraction by coherent domains. The divergence $\Delta \varepsilon$ is taken to be large enough to cover $R_g(\varepsilon)$ when $y_p < 0.99$. The largest domain reflects only a narrow band of the available rays, and its contribution to the reflectivity is $\sigma_1(\varepsilon, \varepsilon + \Delta \varepsilon) = Q y_{p,1}$; in this case the divergence of the direct beam is retained, and the $g$-beam has less divergence. This is convoluted, however, by the orientation distribution of the domains, and the divergence of the macroscopic $g$-beam roughly equals to that of the $d$-beam. The second domain has a reflectivity curve which fits the divergence $\Delta \varepsilon$, and so $d$- and $g$-beams have equal divergences. For this domain already $\sigma(\varepsilon, \varepsilon + \Delta \varepsilon) \approx Q$. The third, very small domain spreads out the $g$-beam, and the successive reflections make the same happen to the $d$-beam.

From the above discussion it is clear that when the domains are large enough to exhibit primary extinction the divergence of the incident beam can be selected such that it is retained in the successive reflections. In that case the energy transfer equations (1)
Figure 7. Reflection of a beam with divergence $\Delta \varepsilon$ from coherent domains of variable size. In the lower part, the width $x$ and divergence $x'(=\varepsilon)$ of the beam are illustrated as a phase-space diagram together with the reflectivity curves of domains (1) and (2).

apply, and the effects of primary extinction on $\sigma(\varepsilon)$ are covered by $y_p(\varepsilon)$.

If the effects of primary extinction are negligible,

$$\int R_\delta(\varepsilon-\eta) d\varepsilon = Q$$

for all $\delta$. The average kinematical reflectivity for the extended beam becomes

$$\sigma_{\text{kin}}(\varepsilon) = \int \int R_\delta(\varepsilon-\eta) w_\delta(\varepsilon) d\eta d\delta = \int R(\varepsilon-\eta) w(\eta) d\eta,$$

where $R(\varepsilon-\eta)$ is the average reflectivity of the domains oriented in an angle $\eta$, and $w(\eta)$ is the normalized abundance of these domains. In his theory of secondary extinction, Zachariasen (1967) distinguishes two cases on the basis of the wider, dominant distribution,

$$\sigma_{\text{kin}}(\varepsilon) = Q w(\varepsilon), \quad \text{type I}, \quad (20a)$$

$$\sigma_{\text{kin}}(\varepsilon) = R(\varepsilon), \quad \text{type II}. \quad (20b)$$

Figure 8 illustrates the effects of diffraction on the divergences in the two cases. In the type-I crystal the beams are not substantially spread, but in a type-II crystal the divergences of the beams cover the whole reflecting range already after a few reflections inside the crystal. For a type-I crystal the preceding formalism applies, and it can be shown that the correction for secondary extinction would be correct to the first order also for a type-II crystal.

As pointed out by Kato (1980a), the problems arising from the convolution of scan and divergences can be avoided when the incident beam is a spherical wave. A first-order correction for secondary extinction would still be possible, but in a correction for primary extinction a constant $y_p$ should be assumed.
6.2 Incoherence of the beams

The basic idea of the method is that the \( d \)- and \( g \)-beams experience a sufficient number of reflections as to lose mutual coherence. In the terminology introduced by Kato (1980b), a ray entering the crystal leaves the coherent channel. The number of reflections experienced by a narrow ray can be estimated from figure 4. The approximation (8) is valid only when \( \delta < 1.5 \) or when the domains are smaller than \( 3\lambda/2 \). In the symmetrical Laue geometry, the area of interaction for each ray is \( 2T^2 \tan \theta \), and so there are about \( (T/\lambda)^2 \tan \theta \) domains of the maximum size in this area. Simultaneously reflecting is only the fraction \( \Delta \varepsilon/\beta \), and with the typical values quoted earlier, 10 to 100 domains contribute to reflection of each ray.

Although the above result suggests that the requirements of the theory are met in a typical measurement, it is worth studying what happens to a ray that stays in the coherent channel. This means that the \( g \)-beam from the first coherent domain is not re-reflected, and so the contribution to the measured reflecting ratio is \( \delta r^*(e) = \delta r(e) = \sigma(e, s_d, s'_d) \delta T_{\text{eff}} \). This should not be corrected for the effects of secondary extinction through (14), as there are none, but only for primary extinction. Accordingly, there is the possibility of over-correction, which should be eliminated by maximizing the number of domains in the area of interaction, so that the coherent contribution to \( r^*(e) \) would be small. In general, the smaller \( r^*(e) \) the smaller is the correction for secondary extinction and also the possible error. Multiple peaks of reflections are due to diffraction from independent coherent domains, and this kind of data should be treated with special care.

The inverse problem of a crystal subject to secondary extinction only is trivial. In that case \( f(e) \) in (11a) is always unity, and therefore \( r_p(e) = 1 \). However, the reflectivity curves \( \sigma_1(e) \) and \( \sigma_2(e) \) of a type-II crystal should not be compared, if these are measured with different wavelengths, because the width of \( R(e) \) is proportional to \( \lambda \). In that case, only the integrals are comparable.

---

Figure 8. Ewald construction for diffraction from a type-I crystal (a) and a type-II crystal (b). The broken lines in (a) show the effect of variation of the incident beam direction. The inserts illustrate the active volume of the diffraction domain, when \( \Delta \varepsilon \) is the divergence of the incident beam.
6.3 Measuring geometry

The application of the method requires that the incident beam has small divergences, which do not smear the reflection profile. In practice, $\Delta \sigma$ must be only about 0.1 mrad. If the crystal is rocked across the beam only in one direction, the other divergence is allowed to be much larger, and the required power of $P_0 = 10^4$ to $10^5$ c/sec is easily achieved. In the measurements with polarized x-rays the plane of diffraction is rotated by 90° about the axis of the beam, and the solid angle of the incident beam should be only about $10^{-8}$ sr. This requirement cannot be fulfilled with a conventional source, and in the measurement quoted earlier an algorithm was developed for solution of $y_p(e)$. This was based on assumptions of the relation between $r^*(e)$ and $y_p(e)$, which was deduced from measurements with two wavelengths.

Another practical limitation of the method arises from the discrete set of wavelengths that is available. The wavelengths should be optimized in regard of absorption and to make $\beta$ in (11) and (12) sufficiently different from zero. In general, the use of a polarized beam should be favoured, as in this case the change in $\Lambda$ does not introduce any change in the illuminated volume or other geometrical parameters, and more fundamentally: scans in reciprocal space can be made identical. However, $2\theta$ values of the first reflections may become too small for a polarization measurement, and so $\lambda$ must be varied.

7. Characteristics of synchrotron radiation

Synchrotron radiation of x-ray wavelengths (sxr) is produced in storage rings where the electron (or positron) energy is a few GeV. The spectral and angular distributions of the radiation from various elements of the ring differ greatly, and the present discussion is limited to the continuous spectrum from a bending magnet or wiggler. Figure 9 shows a universal spectral curve. Typically $B = 1 - 2 T$, $E = 2$ to 5 GeV, and so $\lambda_c$ order of 1 A. The photon flux is integrated over the vertical direction, where the sxr beam is well collimated, and a typical horizontal aperture is order of 1 mrad. The electron or

![Figure 9. Universal spectral curve for synchrotron radiation from a bending magnet of an electron (positron) storage ring. The numbers correspond to unit of horizontal divergence (mrad), electron beam current (mA), and electron energy (GeV). The flux is integrated over the vertical divergence, and $\lambda_c$ is the characteristic wavelength, which corresponds to the median of the power distribution.](image-url)
The positron current of the present-day rings is about 100 mA, and so the maximum flux to a 0-1% bandwidth ($\Delta \lambda / \lambda = 10^{-3}$) is $10^{12}$ to $10^{13}$ photons in second. The flux on the sample depends on the actual optical solution of the beam line, and a schematic construction with focusing mirrors and a two-crystal monochromator is shown in figure 10. The insert shows a calculated intensity distribution at the sample site, and the flux is about $10^{12}$ photons/sec with $\Delta \lambda / \lambda = 10^{-4}$ (Hastings et al. 1983). The flux needed for the present method is far less, but these estimates ensure that there is sufficient flux at all the wavelengths needed even with very narrow collimation and good energy resolution.

The intensity and divergence distributions of the SXR beam can be illustrated by a phase-space diagram (figure 11). Each electron under radial acceleration emits a cone of radiation, which is linearly polarized in the plane of the electron orbit and elliptically outside; the figure shows the polarization components separated. The photon opening angle depends on the electron energy and the x-ray wavelength. The parallel component has almost a Gaussian distribution with variance

$$\sigma_{\parallel} = 0.57 \gamma^{-1} (\lambda / \lambda_{0})^{0.43}, \quad (21)$$

where $\gamma = E/mc^2$. Typically $\sigma_{\parallel}$ is order of 0.1 mrad. The beam sweeps in the horizontal plane, so the horizontal distribution is uniform, and the divergence is determined by the horizontal aperture. The electrons move in bunches, which have cross-section of order 1 mm$^2$, and the effective source is the convolution of the radiation and electron beam distributions, as illustrated in figure 11.

![Figure 10. Schematic construction of an SXR beam line with 1:1 focusing. M1 and M2 are horizontally focusing grazing incidence mirrors, and the two-crystal monochromator is designed to retain the beam position. The distances $z$ are in mm and typical for an x-ray beam line. The lower part shows an intensity distribution at the sample site, as obtained from a ray-tracing calculation (Hastings et al. 1983).](image-url)
The phase-space diagram propagates down the beam line. In the general case it is convoluted by the focussing functions, but figure 12 shows a simple case which does not involve any curved crystals or mirrors. In the \textit{sxr} source the position and divergences are coupled, and so the intensity distributions at the sample site are non-uniform, in contrast to those from a conventional x-ray source (cf. figure 7). A scan by a crystal is shown in the figure. A small crystal with a very narrow reflectivity curve can be used as a probe, but in our case the sample crystal has reflection profiles that are wide in comparison with the divergence of the \textit{sxr} beam. The polarization of the beam is an important parameter of the experimental situation. In a simple case, as illustrated in figure 12, the degree of polarization could be calculated from the source parameters, but generally it must be measured. The results in the figure show that perfect linear

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure11.png}
\caption{Components of the radiation distributions of one electron (left, $= \perp$ in the orbit plane, $\perp$ perpendicular to the plane). Synchrotron radiation source in phase-space (right, vertical distributions). The dotted line is the $1\sigma$-contour for the electron beam. The solid line is the contour of the parallel component, and the broken line that of the perpendicular component.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure12.png}
\caption{The phase-space diagram at distance $z$ from the source; the figure is obtained from figure 11 by $y \rightarrow y + y'z$. The shaded rectangle shows the effective area of reflection from a small crystal, and the arrow indicates the angular scan. The lower part shows the measured average distributions of the parallel and perpendicular components across the beam.}
\end{figure}
polarization is not found even in the orbit plane unless a polarizing x-ray optical element such as a Borrmann crystal is inserted in the beam line (Materlik and Suortti 1983).

8. Outline of extinction measurement

The principle of the determination of primary extinction is illustrated in figure 13. According to (8), the primary extinction factor \( y_p \) of a coherent domain is approximately a Gaussian function of \( \Lambda^{-1} \). The width of the Gaussian is inversely proportional to the size of the domain, and generally varies with the rocking angle \( \varepsilon \) of the crystal. When a conventional x-ray source is used \( \Delta \varepsilon \) may be appreciable, as in the measurement with a beam polarized in the plane of diffraction, and \( y_p(\varepsilon) \) varies in a non-linear fashion within \( \Delta \varepsilon \). Moreover, \( \Lambda_1 \) and \( \Lambda_2 \) are determined by the available wavelengths, so that the situation cannot be optimized. These restrictions are removed by the use of SXR.

In conclusion, a practical geometrical arrangement, which is planned for the crystallography beam line of the NSLS at Brookhaven National Laboratory is depicted in figure 14. The 4-circle diffractometer is built only for a vertical scan, but an extra scan in the horizontal plane is easily adapted. The beam is monochromatized with two Si crystals in \((n, -n)\) setting, and polarized with a Borrmann crystal placed in front of the incident beam monitor. The estimated flux on 1 mm\(^2\), when there are no focusing elements in the beam line, is \(10^8\) to \(10^9\) c/sec. This high flux makes feasible very detailed studies of the diffraction properties of crystals with a probing beam of size \(10^{-2}\) mm\(^2\) or less.

The method can be extended to crystals which are used customarily in crystallography. Through careful measurements of the intensity of the incident beam, \( r^* \) can be determined for a crystal bathed in the beam, and the secondary extinction factor is found by (14). Kawamura and Kato (1983) have given the relation between \( \sigma, \mu_0 \) and the radius of a spherical crystal in an easily applicable form. Their calculation is for the average \( \sigma \) and \( \mu_0 \), but it is valid also for well collimated beams, if the conditions discussed in § 6.1 are met.

![Figure 13. Primary extinction factor \( y_p \) as a function of the rocking angle \( \varepsilon \) of the crystal and the extinction length \( \Lambda \). The domain size \( \delta = \delta(\varepsilon) \) is largest at \( \varepsilon = 0 \). The divergence of the beam \( \Delta \varepsilon \) makes the measurement cover a corresponding range of \( y_p \)-values (thick lines), shown for two values of \( \Lambda \).](image-url)
Figure 14. A modified 4-circle diffractometer at an SXR beam line: (a) vertical scan, (b) horizontal scan. B is the Borrmann polarizer, the incident power $P_0$ is measured by an ionization chamber. The crystal is set to the reflecting position by the $\chi$, $\phi$, $\psi$- and $\omega$-circles, scanned by $\omega$ and $2\theta$ in the vertical plane, or by $\chi$ and $2\theta'$ in the horizontal plane. The additions to the original construction (the $\psi$- and $2\theta'$-circles) are needed for a horizontal scan; $2\theta'$-circle is indicated by broken lines. D is the detector and M the monitor; the roles are interchanged in the two measurements.

Acknowledgements

The author is grateful to N K Hansen and F K Larsen for communicating their unpublished results, and to J R Schneider and A Freund for loan of the deformed Si crystal.

References

Bragg W L, James R W and Bosanquet C H 1921 Philos. Mag. 42 1
Coppens P and Hamilton W C Acta Crystallogr. A26 71
Darwin C G 1922 Philos. Mag. 43 800
Hamilton W C 1957 Acta Crystallogr. 10 629
James R W 1962 The optical principles of the diffraction of x-rays (London: Bell)
Larsen F K and Hansen N K 1983 Univ. Aarhus unpublished report
Materlik G and Suortti P 1983 J. Appl. Crystallogr. to be published
Mathieson A McL 1979 *Acta Crystallogr.* A35 50
Suortti P 1982a *Acta Crystallogr.* A38 642
Suortti P 1982b *Acta Crystallogr.* A38 648
Takagi S 1962 *Acta Crystallogr.* 15 1311
Werner S A 1974 *J. Appl. Phys.* 45 3246
Zachariasen W H 1967 *Acta Crystallogr.* 23 558
Mathematical methods for the analysis of charge densities from x-ray diffraction*

T N GURU ROW
Physical Chemistry Division, National Chemical Laboratory, Pune 411 008, India

Abstract. Accurate x-ray and neutron studies have demonstrated that atomic deformations due to chemical bonding can be reliably mapped. Several new mathematical techniques which have been devised to obtain quantitative information from these maps are discussed along with their relative merits and demerits. Methods of obtaining theoretical densities modified to resemble the experimental framework are explained. The implications of these modified models are outlined.

Keywords. Electron density modelling; theoretical approaches; x-ray diffraction; charge densities.

1. Introduction

In the last two decades, with the advent of automatic diffractometers and high speed computers, it has become possible to collect and process large amounts of accurate x-ray data. These accurate x-ray intensity measurements combined with neutron diffraction results have demonstrated that the atomic deformations due to chemical bonding can be reliably mapped in molecular crystals. The so-called \( X-N, X-X \) (high order) and other related difference maps have revealed several qualitative features of electron density distributions. Since, x-ray scattering is primarily due to the interaction of electromagnetic radiation with the electrons in the crystal, the charge density is the main information which is readily obtained. Several new mathematical techniques have been invoked to extract quantitative information from such charge density distributions. Direct evaluation of one-electron properties using modified fourier series summations is one such approach (Coppens and Guru Row 1978). Another popular approach is the modelling of density distributions by including additional deformation parameters in the least squares refinement. There are several models in practice and in this article salient features of such models will be discussed.

Recently, developments in theoretical methods have facilitated the calculation of charge densities of sufficient accuracy so that these can be compared with experimental charge distributions. However, these are limited for molecular crystals consisting of only light atoms. These comparisons would reveal the deficiencies in both theory and experiment. The state of art now (Coppens 1981) has reached a stage that it is possible to derive properties like net charges, molecular dipole moments and electrostatic potentials (Stewart 1973, 1979; Coppens et al 1980) and to fit atom-centred spherical harmonic functions (Hirshfeld 1977; Stewart 1976; Hansen and Coppens 1978). The theoretical comparisons, however, are to be made with those from a very large basis set (Stevens et al 1978; Stevens 1980) and not with semi-empirical or an approximate minimal basis \textit{ab initio} calculations.

* NCL communication number 3322
2. Density modelling

The choice of a model is in its efficiency, simplicity and computer adoptability. It should also fit the experimental density within the accuracy of the experiment. Since, the model is to be built into the least squares procedure the adequate ratio of observations to parameters should be maintained to get meaningful outputs. The model should be consistent with theoretical properties expected out of the system. An interesting possibility is to use these conditions as constraints during refinement.

The electron density, in general, can be represented in terms of a finite number of analytical functions as

\[ \rho = \sum_i P_i f_i(r), \]  

where \( P_i \) are the corresponding populations. These population parameters enter the least squares formalism as refinable parameters along with the usual parameters like scale factor, positional and thermal parameters etc. However, the electron density defined in (1) is to be convoluted over the probability functions corresponding to the thermal motion to represent the actual dynamic model density.

\[ \rho_{\text{model (dynamic)}} = \sum_i P_i \int f_i(r - u) p_i(u) \, du \]  

This formalism also filters out the noise from the experiment because \( f_i \)’s have limited flexibility while calculating model deformation density as

\[ \Delta \rho_{\text{model}} = \rho_{\text{model}} - \sum \rho_{\text{isolated, spherical atoms}} \]  

The second term in expression (3) is usually calculated from Hartree-Fock wavefunctions.

2.1 Two-centre models

Earlier models tried to express electron density as a sum over atomic orbital products (Stewart 1969; Jones et al 1972).

\[ \rho_{\text{model}} = \rho_{\text{core}} + \sum_{jk} P_{jk} \phi_j \phi_k. \]  

These models, however, soon became unpopular because the orbital products for an extended basis set became unmanageable and large correlations between orbital products hampered the refinement process. Also, the expression of thermal motion for such a two-centre model becomes ambiguous. However, there is a model due to Helner (1977) which is still a two-centre model, but expresses the density in the bonding regions and at lone pair sites by ellipsoidal ‘charge clouds’. The structure factor expression then becomes

\[ F_{\text{model}} = \sum_i \rho_{\text{core}} \exp(-B_i S_i^2) \exp(2\pi i) S \cdot r_i + \sum_{\text{clouds}} P_j \exp(-B_j S_j^2) \exp(2\pi i) S \cdot r_j. \]  

The shape and size of each ‘charge cloud’ is expressed in terms of an anisotropic thermal parameter.
Analysis of charge densities

2.2 One-centre models

The simplest of these would contain only spherical deformation functions, the spherical atom approximation is maintained and only the occupancy of the valence shell is allowed to refine (Stewart 1970; Coppens et al 1971)

\[ \rho_{\text{model}} = \sum_i [\rho_{\text{core}} + P_i \rho_{\text{valence}}(r)]. \]  

(6)

This is also called the L-shell projection method. But, this model does not allow for changes in shape of density distribution and thus contradicts chemical bonding theory. A more flexible formalism, which has become the most popular of available models, is the radial or kappa refinement (Coppens et al 1979). Here,

\[ \rho_{\text{model}} = \sum_i [\rho_{\text{core}} + P_i \rho_{\text{valence}}(\kappa_i r)]. \]  

(7)

and it allows for the refinement of \( P_i \), the population of \( \kappa_i \), the radial expansion and contraction of the valence shell. The functions \( \rho_{\text{core}} \) and \( \rho_{\text{valence}} \) are the Hartree-Fock densities of the free atoms normalized to one electron. Obviously, this model is limited to spherical or monopolar deformations. A more general fit to account for the asphericity of atoms when they become a part of a molecule could be obtained through a multipolar expansion of the valence density. General multipole models (Stewart 1976; Hirshfeld 1977; Hansen and Coppens 1978) aim at such a fuller description of densities

\[ \rho_{\text{model}} = \sum_i \rho_{i, \text{pseudoatom}} \]  

(8)

Each pseudoatom is given as a finite multipole expansion of the form

\[ \rho_i(r) = \sum_{i,m} R_i(r) A_{i,m}(\theta, \phi), \]  

(9)

where \( r, \theta, \phi \), are the polar coordinates about nucleus \( i \), \( R_i(r) \) is the radial distribution function and \( A_{i,m}(\theta, \phi) \) is the angular part.

In the formalism due to Stewart (1976), the pseudoatom density is expressed as a spherical core density together with a multipolar valence density as

\[ \rho_i(r) = \rho_{\text{core}} + \sum_{i,m} P_{im} R_i(r) Y_{lm}(\theta, \phi), \]  

(10)

\( R_i(r) \) is the radial part and is expressed in terms of Slater type functions,

\[ R_i(r) = N r^n \exp(-\zeta r), \]  

(11)

with \( N \) as the normalization constant. The angular functions are real spherical harmonics expressed in terms of the associated Legendre polynomials

\[ y_{lm+}(\theta, \phi) = p_{l}^{m+}(\cos \theta) \cos \phi, \]

\[ y_{lm-}(\theta, \phi) = p_{l}^{m-}(\cos \theta) \sin \phi, \]  

(12)

with

\[ p_{l}^{m}(Z) = \frac{(1-Z^2)^{m/2}}{2^l l!} \frac{d^{m+l}}{dZ^{m+l}}(Z^2 - 1) \]  

(13)

For \( l = 4 \), (i.e. functions up to the hexadecapole level) there will be 25 population and \( l \) radial parameters per atom which enter the least squares formalism.
In another approach (Hirshfeld 1977), the density is expressed as deviations from the overall atom density as

$$\rho_i(r) = \rho_{\text{atom}}(r) + \delta\rho_i(r), \quad (14)$$

where

$$\delta\rho_i(r) = \sum_{i,k} P_{ik} N_i r^n \exp(-\alpha r) \cos^n \theta_k, \quad (15)$$

$\theta_k$ is the angle between $r$ and a specified vector $\mathbf{K}$ chosen to minimize overlap between deformation functions. This model yields 35 refining parameters ($l = 4$) along with one radial parameter per atom. Electroneutrality is maintained through a constraint.

Hansen and Coppens (1978) in their formalism of a multipole model combine the kappa model with spherical harmonic angular and Slater type radial functions for all $l > 1$. This yields 26 more refining parameters. The density is defined as

$$\rho_i(r) = \rho_{\text{core}}(r) + \rho_{\text{valence}}(r) + \sum_{l,m} P_{lm} N_l r^n \exp(-\zeta r) Y_{lm\pm}(r/r), \quad (16)$$

$P_c$, $P_v$, and $P_{lm}$ are population coefficients such that

$$P_c + P_v + P_{lm} = \text{total number of electrons} \quad (17)$$

$\rho_{\text{core}}$ and $\rho_{\text{valence}}$ are Hartree-Fock densities normalized to one electron. The valence function is allowed for $\kappa$ refinement. $\zeta$ is another refining radial parameter. The angular functions are derived in terms of local atomic coordinate systems which allows for non-crystallographic constraints during refinement. Normalization is such that

$$\int_0^{2\pi} \int_0^\pi |y_{lm\pm}| \sin \theta \, d\theta \, d\phi = 2. \quad (18)$$

### 2.3 Merits and demerits

Among the one-centre models, the kappa refinement formalism has become most popular. Coppens once remarked that it is a 'poor man’s charge density' approach. This methodology leads to reasonable estimates of several properties including net atomic charges and dipole moments. The simplicity of this formulation even in case of molecules containing heavier atoms is the main reason for its popularity. The computer program RADIOL (Coppens et al 1979) based on this model has been extensively used.

The multipole models, however, are more difficult to handle but produce similar results. In general, this aspherical atom formalism leads to a significant reduction in the error function thus providing a better physically meaningful fit to the experimental data. The various multipole models proposed lead to similar results; which has been demonstrated recently (Baert et al 1982). However, the molecular dipole moments derived from Hirshfeld approach are generally high. This is attributed to the presence of diffuse monopolar functions in the multipolar expansion. A serious drawback comes from the experiment. The limited number of unique observations, the accuracy and the resolution places a limitation on the density features particularly near atomic centres. Correlations between functions on neighbouring atoms lead to problems particularly in multipole expansion. In fact it is not just the correlation, but the radial extent of the function itself (for $l > 1$) which may have appreciable value at neighbouring sites. The program MOLLY (Hansen and Coppens 1978) which builds the multipolar expansion into a regular least-squares program is widely used.
3. Theoretical methods

The advances in theoretical chemistry and physics have led to very sophisticated model development of molecular systems. The quality of theoretical methods, for molecular crystals, is good enough for comparison with experimental densities. Thermal motion is the most important factor to be accounted for before such comparisons are attempted. There are two ways in which this can be done: (i) to obtain static theoretical density and apply thermal smearing; compare with experimental density. (ii) to deconvolute thermal motion from experimental density; compare with static theoretical density. It is not the purpose of this article to review the various methods of static density calculation. It is enough to state that the availability of experimental deformation densities could be used as a sensitive test of convergence on molecular crystals. Earlier calculations involving semi-empirical, minimal basis set and SCF type formalisms failed to produce reasonable deformation density in the bonding regions. These also resulted in large densities at lone-pair regions. As bond density is a phenomenon akin to chemical theory and is consistently observed in experimental maps, it can be stated that the experimental densities clearly discriminate against less advanced calculations (Coppens and Stevens 1977). In a recent study (Stevens 1980) theoretical calculations of various degrees of complexities by \textit{ab initio} self-consistent field method using extended basis set of Gaussian orbitals have been performed. To facilitate detailed comparison of these theoretical densities with experimental deformation density, a second order difference function

$$\Delta(\Delta \rho) = \Delta \rho_{\text{exp}} - \Delta \rho_{\text{theory, dynamic}}$$

is defined. The thermally-averaged theoretical electron density $\langle \rho \rangle$ can be obtained, within the Born-Oppenheimer approximation, as

$$\langle \rho \rangle = \int \phi^2(R_1, R_2, \ldots R_N) \rho(R_1, R_2, \ldots R_N) d(R_1, R_2, \ldots R_N)$$

where $\phi^2$ is the nuclear distribution function and $R_1, R_2, \ldots R_N$ are the nuclear position vectors. But, for molecular crystals, since the internal modes of vibration are much smaller than the external modes of vibration the dynamic theoretical density can be calculated to a good approximation by convolution with rigid body thermal motions (Stevens et al 1978). These values can be readily computed from experiment (Schomaker and Trueblood 1968). There is one more feature in molecular crystals which in theoretical density estimates cannot be fully accounted for. This is the intermolecular interactions.

The alternative approach of comparing experimental densities with static theoretical densities is after deconvolution. However, it is not yet clear as to which of the two approaches is to be preferred. In general, the comparisons are to be made only with a large basis set calculations and the low level calculations are not up to the resolution reached by experimental densities. For example, in case of oxalic acid (Stevens 1980) \textit{ab initio} SCF calculations were performed using the program \textit{hondo} (Dupuis et al 1976) with an extended basis set of Gaussian orbitals. An extended basis set including \textit{d}-functions, (11, 5, 1/6, 1) primitive Gaussians contracted to a $\langle 4, 3, 1/4, 1 \rangle$ set was used. Another calculation using 4-31G basis with $\langle 8, 4, 0/4, 0 \rangle$ contracted to $\langle 3, 2, 0/2, 0 \rangle$ set was also performed. The double difference maps calculated based on expression (19) clearly indicated that the extended basis set generally is in excellent
agreement. The 4-31 G set showed positive differences in the bonding regions, while in lone pair regions the differences were negative indicating the inadequacies of the theoretical density.

4. Conclusions

The various models with their adoptability to a least-squares formalism have helped to extract maximum information out of the experiment. However, caution is advised not to extend the refinable parameters to the extent of sacrificing the existing overdeterminacy of the data set. Systematic errors and the sharp fourier cut-off still limits the accuracy. The deformation maps from the various models, particularly from the multipolar expansion contain less noise than the experimental $X-N$ or $X-X$ (high order) maps. It is generally advised to carefully examine the residual maps before attempting to evaluate one-electron properties. The fact that the experimental density is comparable only with the density from a very large basis set is welcome since at this level it is possible to look for very small effects quite accurately.

Acknowledgements

The author gratefully recalls his association with Prof. P Coppens and his group which inspired this article.

References

Coppens P 1981 Coordination Chemistry (IUPAC) 21 53
Coppens P, Pautler D and Griffin J F 1971 J. Am. Chem. Soc. 93 1051
Coppens P and Stevens E D 1977 Israel J. Chem. 16 175
Hirshfeld F L 1977 Israel J. Chem. 16 226
Stewart R F 1979 Chem. Phys. Lett. 65 335
Determination of one-dimensional crystal structures using the double Patterson function

PETER MAIN
Physics Department, University of York, Heslington, York Y01 5DD, UK

Abstract. The properties and previous use of the double Patterson function in x-ray crystallography are briefly reviewed followed by an account of a new iterative technique, based on the double Patterson, which is being developed by the author. The technique starts with an approximation to the double Patterson which does not require phases, then improves the approximation by making it conform to the known projections and known magnitudes of the Fourier coefficients. The latter are 3-phase structure invariants and estimating their phases is an important step in the direct determination of structure factor phases. Tests carried out on one-dimensional centro-symmetric structures show that the technique is successful. At best, it determines correct phases for all 3-phase invariants. At worst, it fails to improve on the estimate of all phases being zero. It consistently does very much better than the $B_{3,0}$ formula which is also based on the double Patterson. Further development is necessary to apply the technique to non-centrosymmetric structures and to real structures in three dimensions.

Keywords. X-ray crystallography; double Patterson; crystal structure determination; direct method; image reconstruction.

1. Introduction

A project to investigate the feasibility of using the double Patterson function to solve crystal structures has recently been started by the author. Some of the interesting early results are presented in this paper, but as very little use has been made of the double Patterson, it will be helpful to start with a brief review of the properties and previous use of the function.

The Patterson function itself was introduced by Patterson (1935)

$$P(u) = \int \rho(x)\rho(x-u)dV,$$

$$= \frac{1}{V} \sum_{h} |F(h)|^2 \exp(-2\pi i h \cdot u). \quad (1)$$

The importance of this function stems from the fact that its Fourier coefficients are the intensities $|F(h)|^2$ and it requires no knowledge of phases for its calculation. It has an appreciable value only when $u$ is an interatomic vector and so gives information on vector distances between atoms in the structure.

A generalisation of this function was given by Sayre (1953) who pointed out the existence of the double Patterson function, defined as:

$$D(u, v) = \int \rho(x)\rho(x+u)\rho(x+v)dV$$

$$= \frac{1}{V^2} \sum_{h} \sum_{k} F(h)F(k)F(-h-k)\exp[-2\pi i (h \cdot u + k \cdot v)] \quad (2)$$

95
This is a six-dimensional function which has an appreciable value only when \( u \) and \( v \) are interatomic vectors referred to the same atom as origin, i.e. if there are atoms at vector displacements \( u \) and \( v \) from any other atom in the structure. It is therefore a vector map of the structure, but unlike the Patterson function, requires a knowledge of the phases for its calculation. The Fourier coefficients are the 3-phase structure invariants \( F(h)F(k)F(-h-k) \) which clearly depend upon phases as well as magnitudes. This is precisely the interest in the double Patterson function. Use of the 3-phase structure invariants forms the basis of most direct methods of phase determination which are now widely used in crystal structure analysis. It follows that if an approximation to the double Patterson can somehow be obtained, phases for the 3-phase structure invariants may be estimated, thus making direct phase determination very much easier.

Sayre’s generalisation went further than equation (2) indicates. He pointed out the existence of the set of density functions

\[
D(n, u_1, u_2, \ldots, u_n) = \int_V \rho(x)\rho(x+u_1)\rho(x+u_2) \ldots \rho(x+u_n) dV
\]

\[
= \frac{1}{V^n} \sum_{h_1} \ldots \sum_{h_n} \left( \prod_{j=1}^n F(h_j) \right) F(-h_1 - h_2 - \ldots - h_n)
\]

\[
\exp[-2\pi i(h_1 \cdot u_1 + h_2 \cdot u_2 + \ldots + h_n \cdot u_n)]
\]

(3)

which can have appreciable values only when \( u_1, u_2, \ldots, u_n \) are all inter-atomic vectors referred to the same atom as origin. These functions were written down explicitly by Vaughan (1958). Note that \( D_0(\equiv F(0)) \) gives the contents of the unit cell, \( D_1 \) corresponds to the Patterson function and \( D_2 \) is the double Patterson. The Fourier coefficients of all these functions \( D_n \) are structure invariants of order \( n+1 \). This paper considers only the \( D_2 \) function which uses the 3-phase invariants.

2. Elementary properties

Since there is no single source giving properties of the double Patterson, some of the more important and useful of them are collected together here.

2.1 Number and content of peaks

From the definition of the double Patterson function in (2), it is possible to deduce that the total number of peaks in the function is \( N^3 \), where \( N \) is the number of atoms in the unit cell. These peaks are distributed as follows:

(i) \( N \) peaks superimposed at the origin.

(ii) \( N(N-1) \) peaks in each of the principal sections \( D(u, O), D(O, v) \) and \( D(u, u) \), excluding the origin.

(iii) \( N(N-1)(N-2) \) peaks elsewhere in the function.

The definition also shows that the content of each peak is proportional to the product of atomic numbers of the three atoms giving rise to the peak.

From (iii) above, it is clear that if \( N \leq 2 \) there will be no peaks in general positions in the double Patterson. For the set of density functions given in (3), this may be stated more generally as

\[
D_n(u_1, u_2, \ldots, u_n) = 0, u_1, \ldots, u_n \neq 0, \text{ when } n \geq N.
\]

(4)
It is interesting to compare this condition with that for Karle-Hauptman determinants (Karle and Hauptman 1950) whereby a determinant of order \( m \) is identically zero if \( m > N \), i.e.

\[
\Delta_m = 0, \text{ when } m > N.
\]  

The lowest order determinant which fulfils this criterion is \( N + 1 \). Such a determinant contains structure invariants up to order \( N + 1 \) in its expansion. This is the same as the order of structure invariants in the lowest order \( D_n \) function which fulfils the criterion (4).

2.2 Symmetry

The complete symmetry of the double Patterson function has been described by Vaisberg (1974), so only the symmetry used later in this paper will be given here. The double Patterson has the same lattice, in the appropriate three-dimensional subspaces, as that of the crystal. It also possesses a particular minimum symmetry which is easily worked out. If there are two atoms at vector displacements \( u \) and \( v \) from a third, the two atoms are related by a vector displacement of \( u - v \). Taking each atom in turn as origin gives rise to three related peaks in the double Patterson. The resulting symmetry can be represented by

\[
D(u, v) = D(-v, u - v) = D(v - u, -u),
\]  

which is a 3-fold axis. In addition, the identities of \( u \) and \( v \) are arbitrary and can therefore be interchanged. This gives

\[
D(u, v) = D(v, u),
\]  

which is a mirror plane. This means \( D(u, v) \) has a 3-fold axis and a mirror plane as minimum symmetry.

The equivalent symmetry can also be demonstrated in reciprocal space. The \( h, k \) Fourier coefficient of \( D(u, v) \) is given by (2) as \( F(h)F(k)F(-h - k) \). Clearly we have

\[
F(h)F(k)F(-h - k) = F(k)F(-h - k)F(h) = F(-h - k)F(h)F(k),
\]  

and

\[
F(h)F(k)F(-h - k) = F(k)F(h)F(-k - h).
\]  

For a one-dimensional noncentrosymmetric crystal, these results show that the two-dimensional space group of the double Patterson must be \( p3m1 \). Similarly, for a centrosymmetric crystal, the symmetry of the double Patterson will be \( p6mm \).

Since the double Patterson is a vector map of the structure, the (three-dimensional) crystal space group translations do not affect its symmetry. Vaisberg has shown that the number of possible symmetry groups of the double Patterson is 73, corresponding to the 73 symmorphic three-dimensional space groups.

2.3 Principal projections

The projection of the double Patterson function on to a three-dimensional crystal subspace is given by

\[
D(u) = \int_D D(u, v)dv
\]
\[
\int_V \frac{1}{V^2} \sum_h \sum_k F(h)F(k)F(-h-k)\exp[-2\pi i(h \cdot u + k \cdot v)] \, dv
\]

\[
= \frac{F(0)}{V} \sum_h |F(h)|^2 \exp(-2\pi i h \cdot u)
\]

(9)

which is the Patterson function multiplied by \( F(0) \) and therefore known. Because of
the symmetry of the double Patterson, there will be three such identical projections.

2.4 Principal sections

Consider now the three identical principal sections \( D(u, O) \), \( D(O, v) \), \( D(u, u) \). An
expression for \( D(u, O) \) can be obtained directly from (2) as

\[
D(u, O) = \int \rho^2(x)\rho(x + u) \, dV,
\]

\[
= \frac{1}{V^2} \sum_h \sum_k F(h)F(k)F(-h-k) \exp(-2\pi i h \cdot u),
\]

(10)

where the section is seen to be the convolution of the electron density inverted in the
origin with the squared structure.

An interesting result is obtained if we assume the structure factors obey Sayre's
equation (1952) so that

\[
F(-h) = \frac{1}{V} \sum_k F(k)F(-h-k),
\]

(11)

where \( f \) and \( g \) are the scattering factors of the real and squared atoms respectively.
From (10) and (11) we obtain

\[
D(u, O) = \frac{1}{V} \sum_h \sum_k g \cdot f \cdot |F(h)|^2 \exp(-2\pi i h \cdot u),
\]

(12)

which is a sharpened Patterson function. The three principal sections mentioned above
can therefore be calculated if the structure factors obey Sayre's equation, i.e. if the
atoms are equal and resolved.

3. Vector space methods

A small number of people have used the double Patterson in different ways for crystal
structure analysis. The methods used can be classed as either vector space or reciprocal
space methods and a representative selection of both types are presented in this and the
next section.

As the double Patterson is a six-dimensional function, it is impractical to use it as a
common means of structure determination. It is only recently that computers have
become sufficiently powerful to contemplate calculating it at all. However, certain
three-dimensional sections have interesting properties which have proved to be useful.
3.1 Hoppe section

Hoppe (1957) considered the three-dimensional section

\[ D(u, r) = \int \rho(x + u) [\rho(x) \rho(x + r)] dV \]  \hspace{1cm} (13)

where \( r \) is a fixed vector. If \( r \) is a unique interatomic vector, the function \( \rho(x) \rho(x + r) \) will contain a single peak. The convolution of this with the electron density, as in (13), will therefore give a single image of the structure. If such a section can be calculated, even approximately, the structure will be solved. On the other hand, if \( r \) is an \( n \)-fold interatomic vector \( D(u, r) \) will contain \( n \) translated and superimposed images of the structure and will not be easy to interpret. The situation becomes much simpler if \( r \) is a vector between two heavy atoms. In this case, the peak in \( \rho(x) \rho(x + r) \) corresponding to the superposition of the heavy atoms will be much greater than all other peaks. This results in the corresponding image of the structure in \( D(u, r) \) being stronger than the rest, allowing it to be recognised.

Simonov and Vaisberg (1970) made use of the Hoppe section in the solution of the synthetic silicate \( \text{Na}_2\text{Mn}_2\text{Si}_2\text{O}_7 \) in the space group \( P2_1/n \). Since phases are not known, they used the approximation that the phase of \( F(h)F(k)F(-h-k) \) is zero for all \( h \) and \( k \). It is well-known that this is the most likely phase if only the unit cell contents are known (Cochran 1955). If the magnitudes of the corresponding normalized structure factors are large, the approximation is good for structures of this size. The Mn-Mn vector was easily identified in an ordinary Patterson function, then the appropriate Hoppe section based on this vector was calculated using the approximate formula

\[ D(u, r) = \frac{1}{V^2} \sum_h |F(h)| \left\{ \sum_k |F(k)|F(-h-k)\exp(-2\pi i h \cdot r) \right\} \exp(-2\pi i h \cdot u), \]  \hspace{1cm} (14)

where \( r \) is the known Mn-Mn vector. Most of the structure was recognised in the resulting map.

A more usual method of solving this structure would be to use the minimum function (Buerger 1953) since a unique heavy atom vector is available. Simonov and Vaisberg did, in fact, calculate the minimum function based on the Mn-Mn vector and obtained very similar results to the Hoppe section. Since the Hoppe section requires a lot more computing, it would appear to offer no advantage over the minimum function under these circumstances.

3.2 Symmetry section

Let us assume the crystal space group contains a pair of symmetry elements such that an atom at \( r \) is reproduced at \( C_1r + d_1 \) and \( C_2r + d_2 \), where \( C_1 \) and \( C_2 \) are point group operators and \( d_1 \) and \( d_2 \) are translation vectors. Taking the point \( r \) as origin, the vector distances to the other two atoms are

\[ u = C_1r + d_1 - r, \]  \hspace{1cm} (15)

and

\[ v = C_2r + d_2 - r. \]
That is, a single atom at \( r \) gives rise to a peak at \( (u, v) \) given by (15) in the double Patterson due to this pair of symmetry elements. The double Patterson can therefore be expressed in terms of \( r \) only as

\[
D(r) = \frac{1}{V^2} \sum_h \sum_k F(h)F(k)F(-h-k) \exp\{ -2\pi i [h^T (C_1 - I) + k^T (C_2 - I)] \cdot r + h \cdot d_1 + k \cdot d_2 \}. \tag{16}
\]

Such a three-dimensional section of the double Patterson is called a symmetry section and it contains peaks at atomic sites. Symmetry sections are the equivalent of Harker sections in ordinary Patterson functions and have been studied by Biyushkin (1973).

As with the calculation of the Hoppe section, phases will not be known but the approximation that all phases are zero may be made. Since the peak heights are proportional to \( Z^3 \), heavy atoms should be identified easily even with this approximation to the phases. Biyushkin and Belov (1965) calculated such approximations to symmetry sections in order to solve \( \text{CoCl(NO}_2\text{)}_2 \cdot (\text{NH}_3) \cdot \text{CH}_2\text{C(NH}_2\text{)}_2 \) in the space group \( P2_1/c \). They calculated two different two-dimensional sections through the four-dimensional double Patterson of a projection of the structure. The symmetry sections gave the coordinates of the Co and Cl atoms and one light atom, from which the complete structure was obtained using weighted Fourier syntheses.

4. Reciprocal space methods

In both examples of vector space methods, the double Patterson function or its sections were calculated under the assumption that the phase of \( F(h)F(k)F(-h-k) \) is zero for all \( h \) and \( k \). This is precisely the assumption made in direct methods of phase determination which are now widely used in crystal structure analysis. It would seem therefore that double Patterson methods as described in the previous section add no new information to the analysis and so cannot be made more powerful than direct methods. In order to make progress, better estimates for the phases of the double Patterson coefficients must be obtained. This is the purpose of the \( B_{3,0} \) formula of Karle and Hauptman (1958) which is based on an approximation to the double Patterson function as shown by Vaughan (1958) and will be derived here.

Consider the product of three Patterson functions defined by

\[
Q(u, v) = P(u)P(v)P(u-v). \tag{17}
\]

If \( u, v \) and \( u - v \) are all interatomic vectors, \( Q(u, v) \) will have a large value. This should be compared with the double Patterson function \( D(u, v) \), which has a large value when \( u \) and \( v \) are interatomic vectors referred to the same atom as origin. Thus \( Q(u, v) \), which can be calculated without knowledge of phases, contains all the peaks present in \( D(u, v) \). Unfortunately, it also contains many more peaks not present in the double Patterson and it will always be centrosymmetric. However, using \( Q(u, v) \) as an approximation to \( D(u, v) \) gives an estimate of the Fourier coefficients of the double Patterson.

The Fourier transform of the product of Pattersons involves the convolution of their Fourier coefficients. Leaving out the origin peak and the principal sections of \( Q(u, v) \), the \( h, k \) Fourier coefficient is proportional to

\[
\langle (|E(l)|^2 - 1)(|E(h+l)|^2 - 1)(|E(h+k+l)|^2 - 1) \rangle_l. \tag{18}
\]
To put this on the correct scale, we divide by the average value of \((|E(l)|^2 - 1)^3\) and multiply by the total double Patterson density represented by (18). The latter is given by the total density less that contained in the three missing sections, giving

\[
(\sigma_1^3 - 3\sigma_1\sigma_2 + 2\sigma_3) \frac{\langle(|E(l)|^2 - 1)(|E(h + l)|^2 - 1)(|E(h + k + l)|^2 - 1)\rangle_1}{\langle(|E(l)|^2 - 1)^3\rangle_1}, \tag{19}
\]

where \(\sigma_n = \sum_{j=1}^{N} Z_j^n\).

The three principal sections are restored by assuming they are Patterson functions as in (12), giving the quantity

\[
(\sigma_1\sigma_2 - \sigma_3) \frac{(|E(h)|^2 - 1) + (|E(k)|^2 - 1) + (|E(-h - k)|^2 - 1)}{\langle(|E(l)|^2 - 1)\rangle_1} + \sigma_3, \tag{20}
\]

where the origin peak has also been added. Since the function \(Q(u, v)\) is centrosymmetric, the Fourier coefficient represented by (19) and (20) is always real and is given by

\[
\sigma_2^{3/2} |E(h)E(k)E(-h - k)| \cos(\phi(h) + \phi(k) + \phi(-h - k)). \tag{21}
\]

Remembering that \(\langle |E(l)|^2 - 1 \rangle_1 = 1\), a combination of (19), (20) and (21) yields the formula

\[
|E(h)E(k)E(-h - k)| \cos(\phi(h) + \phi(k) + \phi(-h - k)) = \frac{(\sigma_1^3 - 3\sigma_1\sigma_2 + 2\sigma_3)}{\sigma_2^{3/2}} \frac{\langle(|E(l)|^2 - 1)(|E(h + l)|^2 - 1)(|E(h + k + l)|^2 - 1)\rangle_1}{\langle(|E(l)|^2 - 1)^3\rangle_1} + \frac{\sigma_1\sigma_2 - \sigma_3}{\sigma_2^{3/2}} (|E(h)|^2 + |E(k)|^2 + |E(-h - k)|^2 - 3) + \frac{\sigma_3}{\sigma_2^{3/2}}. \tag{22}
\]

The \(B_{3,0}\) formula for unequal atoms is given by Hauptman (1964) as

\[
|E(h)E(k)E(-h - k)| \cos(\phi(h) + \phi(k) + \phi(-h - k)) = \frac{(\sigma_1^3 - 3\sigma_1\sigma_2 + 2\sigma_3)}{\sigma_2^{3/2}} \frac{\langle(|E(l)|^2 - 1)(|E(h + l)|^2 - 1)(|E(h + k + l)|^2 - 1)\rangle_1}{\langle(|E(l)|^2 - 1)^3\rangle_1} + \frac{\sigma_3}{\sigma_2^{3/2}} (|E(h)|^2 + |E(k)|^2 + |E(-h - k)|^2 - 2), \tag{23}
\]

which differs from (22) in the magnitude of the second term. Since this is of the order of \(1/N\) of the first term, there is little practical difference between the two formulae.

The reason why this formula does not reliably indicate the value of \(\cos(\phi(h) + \phi(k) + \phi(-h - k))\) is that \(Q(u, v)\) is too poor an approximation to \(D(u, v)\). For all but the most trivial structures, most of the peaks in \(Q(u, v)\) do not belong to \(D(u, v)\) at all and the proportion of wrong peaks increases with the size of the structure. This is seen in figure 1 which shows the double Patterson of a four-atom one-dimensional centrosymmetric structure compared with the corresponding \(Q(u, v)\) function. Even for this simple structure, a large proportion of the peaks off the principal sections are spurious.
The successful applications of $B_{3,0}$ have mainly been with fairly small structures and almost never in a situation where the structure could not be solved by other means.

5. A new iterative technique

In the previous section it was shown that the $B_{3,0}$ formula was based on a very poor approximation to the double Patterson function. One way of improving the performance of $B_{3,0}$ will be to obtain a better approximation than the $Q(u, v)$ function. Fortunately this should be possible, since not all the available information on the double Patterson has been used to set up $Q(u, v)$. An obvious omission is that the known projections of the double Patterson have not been used and, indeed, $Q(u, v)$ will not give the correct projections at all. It is proposed, therefore, to use $D(u, v)$ as a first approximation to $D(u, v)$ and then to modify it by making it compatible with the three known principal projections described in § 2.3.

Reconstructing images from projections is now a well-established process and many different techniques for doing it have been developed. The ones worth considering for the present application are the algebraic reconstruction technique (ART) (Gordon et al. 1970) and the simultaneous iterative reconstruction technique (SIRT) (Gilbert 1972) in either their additive or multiplicative forms. Since the number of known projections of the double Patterson function is very small (only three), the problem of reconstructing the function from them is very underdetermined. However, Minerbo and Sanderson (1977) have investigated the formation of an image from only two or three projections and they report acceptable results for simple images when the multiplicative version of ART is used. They also show this corresponds to the maximum entropy solution in the case of three projections. A multiplicative algorithm was therefore chosen to reconstruct the double Patterson. Since it is easier to maintain the symmetry of the double Patterson using SIRT, the multiplicative version of SIRT (MSIRT) was used.

This is an iterative technique which operates on the double Patterson function evaluated on a grid of points $D(u_i, v_j)$. It updates the value of each grid point according to

$$D^{p+1}(u_i, v_j) = D^p(u_i, v_j)s_is_js_{i+j},$$
where
\[ s_i = \sigma_1 P(u_i) / \sum_k D^p(u_i, v_k), \]  \tag{24}

and \( p \) is the number of the iteration. The effect is to modify the double Patterson so its three principal projections match the known projections \( \sigma_1 P(u) \).

A small amount of additional information may be used if it is assumed that the principal sections of \( D(u, v) \) are also known as described in § 2.4. In this case the sections are left out of the calculation of scale factors in (24) and they are also left unmodified by the \textsc{msirt} algorithm. The scale factors are now calculated as

\[ s_i = \left( \frac{\sigma_1^3 - 3\sigma_1\sigma_2 + 2\sigma_3}{\sigma_1^2 - \sigma_2} \right) \frac{P(u_i)}{\sum_k D^p(u_i, v_k) - 2D^p(u_i, 0)}. \]  \tag{25}

Perfect reconstruction of the double Patterson function will not be possible in a single application of \textsc{msirt} since the problem is severely under-determined. At most, only a small improvement over \( Q(u, v) \) can be expected. However, still more information is available since the magnitudes of the Fourier coefficients of \( D(u, v) \) are all known. The simplest way to make use of this is to transform the approximate \( D(u, v) \) into reciprocal space, correct the Fourier coefficients to match the known magnitudes and then to transform them back again. Using the projections in vector space and the magnitudes in reciprocal space should bring a lot more information to bear on the double Patterson than is used in the original \textsc{bsrt} formula.

All of this can be brought together in the following iterative procedure to obtain a good estimate of the double Patterson function:

(a) use \( Q(u, v) \) defined in (17) as a first approximation to \( D(u, v) \);
(b) scale the density to match the known total density (total density off principal sections of \( D(u, v) \) is \( \sigma_1^3 - 3\sigma_1\sigma_2 + 2\sigma_3 \) and set all negative regions of \( D(u, v) \) to zero;
(c) make the approximate \( D(u, v) \) compatible with the known projections by applying \textsc{msirt};
(d) for equal-atom structures, the principal sections of \( D(u, v) \) can be set equal to their known values;
(e) transform to reciprocal space;
(f) alter the calculated magnitudes \( |E_{\text{calc}}(h, k)| \) such that if \( |E_{\text{calc}}(h, k)| > E_{\text{obs}}(h, k) \) then \( |E_{\text{calc}}(h, k)| = E_{\text{obs}}(h, k) \). Other magnitudes and all phases are left unchanged.
(g) other relationships among structure factors may be used here such as Sayre's equation to refine phases—see the next section for a description of this;
(h) transform back to vector space and repeat from (b) until the double Patterson has converged or until reasonable estimates of the phases of the Fourier coefficients have been obtained.

6. Results

The iterative procedure described in the last section requires a huge amount of computing, even for small structures. However, if it forms the basis of a method of solving structures that could not otherwise be solved, it will be worth developing. To see
how the procedure behaves, it was applied to one-dimensional centrosymmetric model structures. The electron density was made discrete, i.e. it existed on a grid of evenly spaced points, and it consisted of equal point atoms. This bears little resemblance to real structures, but it reduced the amount of computing to easily manageable proportions.

As might be expected, the structures which were easy to solve were those in which the electron density was predominantly zero with only a small number of atoms. The $B_{3,0}$ formula correctly calculated the phases of the strongest Fourier coefficients and it was unnecessary to use any better approximation to the double Patterson. Likewise, those structures which contained a large number of atoms with only a few points of zero electron density were also easily solved. In this case, the signs of the Fourier coefficients were predominantly negative and were correctly calculated by $B_{3,0}$.

Between these two extremes, a range of structures could be set up of various degrees of difficulty. Three examples are set out in table 1. The electron density is evaluated at 32 points in the unit cell and only the asymmetric unit is given in the table. A measure of the progress of the technique is given by the weighted rms phase error of the double Patterson Fourier coefficients $E(h)E(k)E(-h-k)$, the weights being the magnitudes of the coefficients. The very weakest coefficients whose phases are not of interest were left out of the average. At iteration 0, the rms error is calculated using an estimate of zero for all the phases, which is the normal estimate in direct methods. Iteration 1 represents the results of applying $B_{3,0}$, that is, phases calculated directly from the $Q(u,v)$ function. Thereafter each iteration consists of correcting the magnitudes in reciprocal space and modifying the double Patterson in vector space as described in the previous section.

In the first example $B_{3,0}$ is seen to provide very little improvement in the phases, but all phases are determined accurately after a few iterations of improving the double Patterson. It was found that making use of Sayre's equation at step (g) of the algorithm speeded up convergence enormously, although it was not necessary for the eventual success of the method. Sayre's equation in (11) can be modified to give

$$\frac{q}{f} |F(h)|^2 = \frac{1}{V} \sum_k F(h)F(k)F(-h-k),$$

Table 1. Three examples of one-dimensional structures referred to in the text. The phase errors are those of the double Patterson Fourier coefficients (3-phase invariants) and are expressed in degrees. Note that random phases will produce an rms error of $127^\circ$.

| (a) asymmetric unit of $\rho(x)$ | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| asymmetric unit of $P(u)$ | 12 | 4 | 2 | 4 | 5 | 2 | 6 | 4 | 4 | 8 | 7 | 0 | 2 | 6 | 5 | 4 | 6 |
| iteration number | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 0 | 1 | 2 | 5 | 4 | 6 | 7 |
| wtd. rms phase error | 109 | 102 | 77 | 35 | 19 | 10 | 6 | 0 |
| (b) asymmetric unit of $\rho(x)$ | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| asymmetric unit of $P(u)$ | 12 | 4 | 1 | 2 | 7 | 6 | 5 | 2 | 4 | 6 | 7 | 2 | 2 | 6 | 7 | 4 | 2 |
| iteration number | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 0 | 1 | 2 | 5 | 4 | 6 | 7 |
| wtd. rms phase error | 113 | 122 | 52 | 36 | 33 | 34 | 35 | 37 |
| (c) asymmetric unit of $\rho(x)$ | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| asymmetric unit of $P(u)$ | 12 | 2 | 3 | 4 | 5 | 6 | 5 | 2 | 4 | 8 | 3 | 4 | 3 | 6 | 6 | 4 | 2 |
| iteration number | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 0 | 1 | 2 | 5 | 4 | 6 | 7 | 0 | 20 |
| wtd. rms phase error | 113 | 128 | 124 | 126 | 125 | 121 | 121.1 | 113 |
in which the right side represents a projection in reciprocal space of the double Patterson coefficients. The left hand side is independent of phase and so is always known. That is, Sayre’s equation gives known projections in reciprocal space, corresponding to the known sections in vector space. A simple, though unconventional, way of applying Sayre’s equation is to apply \( \text{sirt} \) to modify the double Patterson coefficients to give the correct projections and this is what was done.

The second example in table 1 is a more difficult structure than the first. It can be seen that the \( B_{3,0} \) results give worse estimates of phases than assuming they are all zero. The double Patterson iteration, on the other hand, rapidly improves the phases and converges to an rms error of about 35°. Even though this is not a complete success, the improvement in phase estimates is significant. As a measure of this improvement, an attempt was made to solve the structure using conventional symbolic addition with estimates of zero for the 3-phase invariants and then with estimates given by the double Patterson iteration technique. In both cases, sixteen plausible sets of signs were obtained for the ten largest \( E \)'s and the quantity \( \sum_k E(h)E(k)E(-h-k) \) was calculated for each set as a simple figure of merit. In the conventional calculation, the correct set of signs gave the worst figure of merit. With the double Patterson coefficients as the 3-phase invariants, the correct set of signs gave the best figure of merit. Clearly, the structure is solved much more easily after double Patterson iteration.

The third example is of a more difficult structure still. Here the \( B_{3,0} \) formula calculates random phases and the double Patterson iteration cannot improve on the original estimates of all phases equal to zero.

7. Discussion

In all structures tested, the double Patterson iteration technique did significantly better than the \( B_{3,0} \) formula in estimating the signs of the 3-phase invariants. At worst, it was as good as assuming all signs were positive as in the example in table 1(c) and at best was capable of estimating all signs correctly, even starting from an rms error of over 100° as in table 1(a).

There are two features of the structures tested which determine their difficulty. One is the number of zeros in the Patterson function. The structure in table 1(a) produces a single zero in the Patterson whereas the example in table 1(b) has no zeros. Since there is a restriction that the double Patterson density cannot be negative, a zero in the Patterson function automatically leads to correct values at all points in the double Patterson contributing to that point in the projection. Therefore, the more zeros there are in the Patterson function, the easier the structure.

The second feature for consideration is how flat the Patterson function is. Structures whose Patterson functions are fairly flat are more difficult to solve than those with prominent peaks or deep troughs. The example in table 1(c) has a much flatter Patterson than that in table 1(b) and it has already been shown to be a more difficult structure. The variances of the Patterson density are 2.7 and 4.3 respectively.

An examination of the structures that failed to solve, or were difficult, revealed another interesting fact. A large proportion of the signs of the \( \Sigma_1 \)-type coefficients, i.e. those of the form \( E(h)E(h)E(-2h) \), were wrong. The \( \Sigma_1 \) formula can often give wrong results, but it is disturbing that under perfect conditions, without experimental error, the approximation to the double Patterson used here also estimates these signs
wrongly. It ought to be more reliable than the $\Sigma_1$ formula itself. This has led the author to abandon the $\Sigma_1$ formula as an indicator of phases prior to phase determination by the tangent formula of Karle and Hauptman (1956).

It is clear from these tests that information about 3-phase invariants can be obtained by objective procedures involving the double Patterson function. It is interesting that no use is made of probability relationships as in conventional direct methods; neither is Sayre’s equation necessary, although it is useful to accelerate the convergence of the procedure. Unfortunately, it is impossible to predict how the technique will behave with real crystal structures in three dimensions. More elaborate testing will have to be done to determine this.

A problem not yet solved is how to define the enantiomorph for non-centrosymmetric structures. It is possible that the determination of $\cos(\phi(h) + \phi(k) + \phi(-h - k))$ may be sufficient, but it would be more satisfactory if the actual phases could be obtained.

The larger the structure, the more under-determined is the image reconstruction from projections in vector space. This may well be the limiting factor in the whole procedure, although where the limit lies is completely unknown as yet.

The computing time will be large for real structures, but the fastest computers should now be able to handle a realistic calculation. This will limit the method to those crystal structures that cannot otherwise be solved, or require several man-years work as in macromolecular crystallography. Even if the method proves to be of no practical use, the results presented in this paper are very interesting in themselves.

References

Biyushkin V N and Belov N V 1965 Sov. Phys. Crystallogr. 9 655
Cochran W 1955 Acta Crystallogr. 8 473
Hauptman H 1964 Acta Crystallogr. 17 1421
Hoppe W 1957 Acta Crystallogr. 10 751
Karle J and Hauptman H 1950 Acta Crystallogr. 3 181
Karle J and Hauptman H 1956 Acta Crystallogr. 9 635
Karle J and Hauptman H 1958 Acta Crystallogr. 11 264
Minerbo G N and Sanderson J G 1977 Los Alamos Informal Report LA-6747-MS, New Mexico: Los Alamos Scientific Laboratory, USA
Patterson A L 1935 Z. Krist. 90 517
Sayre D 1952 Acta Crystallogr. 5 60
Sayre D 1953 Acta Crystallogr. 6 430
Vaughan P A 1958 Acta Crystallogr. 11 111
Precision of determination of d-spacings using a Guinier camera

SVEND ERIK RASMUSSEN
Chemistry Department, Aarhus University, DK-8000 Aarhus C, Denmark

Abstract. The precision of determination of lattice spacings using a commercial triple Guinier camera and a precision comparator is evaluated. Determination of lattice constants on germanium, tungsten and various samples of mullite are given. It is found that d-spacings can be routinely determined with uncertainties of the order of 1 part in 10000 in the analytically most useful Bragg angle range from 0 to 45° in theta.

Keywords. Precision; d-spacings; Guinier camera.

1. Introduction

For a perfect camera the uncertainty in d-spacing measurements is given by the expression: \( \Delta d/d = \cot \theta \Delta \theta \), assuming that the wavelength of the x-ray beam is either known accurately or is considered a reference value. For cubic crystals the relative uncertainty of the lattice constant \( \Delta a/a \) likewise varies as \( \cot \theta \Delta \theta \). It has become standard practice to determine lattice constants for cubic and other crystals using extrapolation methods as the uncertainty of a d-spacing measurement should vanish at \( \theta = 90° \). Details on determination of lattice constants using powder measurements, and references are given by Klug and Alexander (1974).

The standard Guinier camera employing transmission geometry normally limits the angular range to \( \theta_{\text{max}} = 45° \). Extrapolation methods are of little use in this case. Precision determinations require very accurate angle measurements. Since a good Guinier camera gives very sharp lines when a suitable sample is employed, it should be possible to obtain accurate d-spacing when using Guinier techniques. Most powder lines which are suitable for identification purposes are located in the “Guinier range”, and it is of practical importance, especially for analytical purposes, to examine the accuracy obtainable with the Guinier camera, especially as it is of much lower cost than a powder diffractometer and as it has a better resolving power than a standard diffractometer.

Using a Guinier camera the Bragg angle \( \theta \) is determined by measuring the distance \( l \) between the line from the primary beam and the powder line from a given d-spacing. For a camera of diameter 114-592 mm the nominal \( \theta \)-value is found as \( \theta(\text{degrees}) = l(\text{mm})/4 \) and \( \Delta \theta(\text{degrees}) \) should accordingly be given as \( \Delta l(\text{mm})/4 \). Film shrinkage and other sources of error, such as lack of reproducibility of sample to film distance, require calibration of the film, e.g. by using a reference sample with accurately known lattice constants for accurate determinations of d-spacings.

We have been using a comparator yielding nominal uncertainty of 0.2 \( \mu \text{m} \) in measuring \( l \)-values. This uncertainty is far below other uncertainties connected with d-spacing measurements, and we assume therefore that the reproducibility of our d-spacing measurements reflect the sum of other errors connected with the methods applied such as imperfection in camera construction and deviations from linearity in film shrinkage and the like.
2. Experimental

We use an Enraf-Nonius FR 552 Guinier camera with a Philips fine-focus (0.4 × 8 mm) x-ray tube. The camera has three compartments allowing simultaneous exposure of three samples. Pure CuKα₁ radiation is used in all three compartments.

The comparator was made by Zeiss, Jena, and was modified according to Tomkins and Fred (1951) and Bennet and Koehler (1959). Line positions can be read with a nominal uncertainty of 0.2 μm and peak intensities can be determined over a range from 1 to 100. Our samples have been exposed at 22–24°C. No strict temperature control has been kept. Thermal expansion coefficients of many simple inorganic compounds are of the order of magnitude of 10⁻⁵/°C. An indeterminacy of 1–2° would allow Δd/d to be approximately 10⁻⁴ for most “hard” substances.

In the past, we have used one compartment in our triple Guinier camera for the reference sample and corrected θ-values in the two other compartments using the calibration values obtained from our standard. However, when the comparator became available to us, we found that this procedure is not valid when the precision of measuring l is better than 0.1 mm. Therefore we now mix our reference material with the sample when high accuracy is required. Normally, we use semiconductor grade silicon, and when silicon is not well suited we use germanium. Unfortunately, germanium powder is quickly contaminated by GeO₂ and only freshly powdered germanium is a suitable reference material. Although we can normally reproduce readings to the nominal uncertainty of 0.2 μm when repeating measurements of a given line within a short time scale of a few minutes, the reproducibility of measuring lines on a given film at different days is about 0.005–0.01 mm. Corrections from different exposures can vary as much as 0.1 mm.

For comparison we have also used a standard Bragg-Brentano type powder diffractometer equipped with a diffracted beam graphite monochromator and a strip chart recorder. We use Co-radiation with the diffractometer and both the 𝛼₁ and the 𝛼₂ components are registered. The diffractometer allows angles to be measured up to 158°.

3. Discussion

The corrections Δθ can be reproduced within an uncertainty of about 0.01 mm by a polynomiun of second degree. We have tried two expressions:

\[ Δθ = a + bθ + cθ^2 \]  and \[ Δθ = \frac{a \cdot θ + b}{1 + cθ} \]

and determined a, b and c by ordinary least-squares methods using a microcomputer.

In principle, Δθ should vanish at θ = 0. Experience shows, however, that it is advantageous to include a constant term in the expressions. In most cases there is little difference between the applicability of the two expressions. Some care must, however, be exercised when using the hyperbolic expression as “c” is normally negative and a singularity sometimes occurs.

The validity of our approach has been checked by measurements on high purity samples of germanium and of tungsten using semi-conductor grade silicon as reference sample. Lattice constants were calculated both by using \((tgθ)^2\) as weights and by using unit weights. No significant difference was found, although the weights differ with a
factor of 10 when the \((tg\theta)^2\) weighting scheme is used.

For comparison the lattice constants were determined also from powder diffractometer measurements using an instrument allowing measurements up to 158° in 2-theta. Peak positions were read from strip chart recordings. Also in this case we used silicon as an internal standard.

The lattice constants were also in this case calculated using both \((tg\theta)^2\) weighted and unweighted data. The range of weights in this case was from 1–120 (W) and 1–28 (Ge). The \((tg\theta)^2\) weighting scheme has some similarity with an extrapolation procedure. The lattice constants for Ge and W determined by both methods are given in table 1. There are no significant differences between the lattice constants obtained by the two methods.

As a further check, we tried to distinguish between two samples of mullite of slightly different compositions. Mullite is a crystalline phase composed of \(\text{Al}_2\text{O}_3\) and \(\text{SiO}_2\). It exists in the interval 60–67 mole% \(\text{Al}_2\text{O}_3\). The literature gives conflicting information on the melting of mullite. Some authors claim that the melting is congruent, others that it melts incongruently. References on such observations and a phase diagram of \(\text{Al}_2\text{O}_3\), \(\text{SiO}_2\) are given by Aramaki and Roy (1959).

Neuhaus and Richartz (1958) have grown single crystals of mullite of composition \(2\text{Al}_2\text{O}_3\), \(\text{SiO}_2\) by the Czochralski method. This supports the assumption of congruent melting. We have also at Aarhus succeeded in growing transparent mullite crystals of 2:1 composition by Czochralski growth using an iridium crucible to contain the melt.

The quality of these crystals was assessed by neutron and x-ray diffraction. We were not, however, able to grow single crystals from melts of composition \(3\text{Al}_2\text{O}_3\), \(2\text{SiO}_2\). These experiments gave invariably a polycrystalline mass. We examined the differences in lattice constants between 2:1 and 3:2 mullites by powder diffraction using both the Guinier camera and the powder diffractometer. In both types of measurements silicon was added as internal standard.

Precision measurements of Guinier diagrams of the two samples showed small but significant differences. All d-spacings of the 2:1 sample were larger than the corresponding ones from the 3:2 sample. The differences varied from 0.0176 Å for the largest d-spacing measured (541 Å) to 0.0011 Å for the smallest spacing (1.24 Å). Expressed in theta-angles the difference varied from 0.04 to 0.1 degrees.

The powder diagrams were indexed preliminarily using lattice constants obtained

Table 1. Lattice constants of Ge and W by Guinier and diffractometer methods using Si as reference material. \(\text{CuK}_{\alpha_{1}} = 1.540598\) Å, \(\text{CoK}_{\alpha_{1}} = 1.78890\) Å. Lattice constant, \(a\), of Si = 5.43083 Å.

<table>
<thead>
<tr>
<th>Method</th>
<th>No. of lines</th>
<th>Radiation</th>
<th>Weighting scheme</th>
<th>Sample</th>
<th>(a) (Å)</th>
<th>(\sigma_a) Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guinier</td>
<td>5</td>
<td>(\text{CuK}<em>{\alpha</em>{1}})</td>
<td>((tg\theta)^2)</td>
<td>Ge</td>
<td>5.65844</td>
<td>(6.39 \times 10^{-4})</td>
</tr>
<tr>
<td>Guinier</td>
<td>5</td>
<td>(\text{CuK}<em>{\alpha</em>{1}})</td>
<td>unit weights</td>
<td>Ge</td>
<td>5.65859</td>
<td>(1.03 \times 10^{-3})</td>
</tr>
<tr>
<td>Diffract.</td>
<td>7</td>
<td>(\text{CuK}<em>{\alpha</em>{1}})</td>
<td>((tg\theta)^2)</td>
<td>Ge</td>
<td>5.65700</td>
<td>(7.31 \times 10^{-4})</td>
</tr>
<tr>
<td>Diffract.</td>
<td>7</td>
<td>(\text{CuK}<em>{\alpha</em>{1}})</td>
<td>unit weights</td>
<td>Ge</td>
<td>5.65661</td>
<td>(1.31 \times 10^{-3})</td>
</tr>
<tr>
<td>Guinier</td>
<td>3</td>
<td>(\text{CuK}<em>{\alpha</em>{1}})</td>
<td>((tg\theta)^2)</td>
<td>W</td>
<td>3.1651</td>
<td>(3.57 \times 10^{-3})</td>
</tr>
<tr>
<td>Guinier</td>
<td>3</td>
<td>(\text{CuK}<em>{\alpha</em>{1}})</td>
<td>unit weights</td>
<td>W</td>
<td>3.1652</td>
<td>(4.62 \times 10^{-3})</td>
</tr>
<tr>
<td>Diffract.</td>
<td>6</td>
<td>(\text{CoK}<em>{\alpha</em>{1}})</td>
<td>((tg\theta)^2)</td>
<td>W</td>
<td>3.16459</td>
<td>(1.58 \times 10^{-4})</td>
</tr>
<tr>
<td>Diffract.</td>
<td>6</td>
<td>(\text{CoK}<em>{\alpha</em>{1}})</td>
<td>unit weights</td>
<td>W</td>
<td>3.16454</td>
<td>(2.23 \times 10^{-4})</td>
</tr>
</tbody>
</table>
### Table 2. Lattice constants of two mullite samples. One of composition $2\text{Al}_2\text{O}_3$, $1\text{SiO}_2$ (2:1 mullite) and another of composition $3\text{Al}_2\text{O}_3$, $2\text{SiO}_2$ (3:2 mullite). Silicon was used as reference sample. Values of constants employed are given in table 1.

<table>
<thead>
<tr>
<th>Method</th>
<th>No. of lines</th>
<th>Radiation</th>
<th>Weighting scheme</th>
<th>Sample</th>
<th>$a$ (Å)</th>
<th>$b$ (Å)</th>
<th>$c$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guinier</td>
<td>35</td>
<td>CuK$_{x_1}$</td>
<td>$(tg\theta)^2$</td>
<td>2:1 mullite</td>
<td>7.5927</td>
<td>7.6840</td>
<td>2.8882</td>
</tr>
<tr>
<td>Guinier</td>
<td>35</td>
<td>CuK$_{x_1}$</td>
<td>unit weights</td>
<td>2:1 mullite</td>
<td>7.5924</td>
<td>7.6841</td>
<td>2.8882</td>
</tr>
<tr>
<td>Guinier</td>
<td>37</td>
<td>CuK$_{x_1}$</td>
<td>$(tg\theta)^2$</td>
<td>3:2 mullite</td>
<td>7.5748</td>
<td>7.6836</td>
<td>2.8857</td>
</tr>
<tr>
<td>Guinier</td>
<td>37</td>
<td>CuK$_{x_1}$</td>
<td>unit weights</td>
<td>3:2 mullite</td>
<td>7.5750</td>
<td>7.6838</td>
<td>2.8857</td>
</tr>
<tr>
<td>Diffract.</td>
<td>19</td>
<td>CoK$_{x_1}$</td>
<td>unit weights</td>
<td>2:1 mullite</td>
<td>7.5852</td>
<td>7.6831</td>
<td>2.8878</td>
</tr>
<tr>
<td>Diffract.</td>
<td>16</td>
<td>CoK$_{x_1}$</td>
<td>unit weights</td>
<td>3:2 mullite</td>
<td>7.5745</td>
<td>7.6839</td>
<td>2.8855</td>
</tr>
</tbody>
</table>

from measurements on a Picker four-circle diffractometer on a single crystal. Thirty seven reflections could be indexed unambiguously and these were used in a least-squares analysis for determinations of lattice constants. Standard deviations of the Guinier data are smaller than those from the diffractometer data. The data are given in table 2.

### 4. Conclusion

It is possible to determine $d$-spacings of high accuracy, $\Delta d/d \approx 10^{-4}$ in the analytically most important angular range using a commercially available Guinier camera, provided that the powder lines are measured with a precision corresponding to an uncertainty of $\theta$ of about 0.002–0.005°. Although our comparator is hand-operated, the measuring process is not unduly time-consuming. Automated equipment does exist and Edmonds and Henslee (1978) report on the use of microcomputer-controlled film densitometry applied to Guinier films. Their results are comparable with ours. Computing facilities need not be very sophisticated for handling computations like evaluating measurements, indexing lines and making least-squares determinations of lattice constants. We employ a microcomputer with 48 K bytes memory and two floppy disk drives, but also a smaller computer could do most of the computations. Search/match procedures are often helped by the high precision of the data and the cost of a Guinier camera + comparator + microcomputer is lower than that of a powder diffractometer of similar capacity.

### Acknowledgements

I am indebted to the Carlsberg Foundation for providing the cost of the microcomputer and to Mrs B Lundtoft for assistance in data collection.
References

The phase problem in neutron diffraction

S K SIKKA
Neutron Physics Division, Bhabha Atomic Research Centre, Trombay, Bombay 400 085, India

Abstract. The structure of any crystal can now be determined ab initio from the neutron Bragg intensity data alone, without recourse to the x-ray structural analysis. This has been made possible by the reduction in data collection times due to the availability of increased neutron fluxes at the samples and the extensive development of the phase determining procedures for neutron diffraction in the last 15 years. In this review, we describe the applications of direct methods, anomalous dispersion techniques and difference Patterson methods and discuss why these methods are applicable in neutron diffraction. Their limitations are also discussed. Some newer methods like resonance-modulated diffraction and use of ‘Renninger effect’ to measure the structure invariants are also touched upon.

Keywords. Neutron diffraction; Patterson method; anomalous dispersion techniques; direct methods.

1. Introduction

Neutron diffraction has distinct advantages in structural analysis of crystals: (i) to locate hydrogen atoms, thereby providing in hydrogen-bonded crystals a very precise stereochemical information about hydrogen bonds and (ii) to differentiate the neighbouring elements in the periodic table, say, nitrogen from carbon and oxygen, in crystals of organic and biological molecules. Lately the determination of deformation charge densities in organic molecules by combined x-ray and neutron studies has also become possible. For a recent review of the above applications of neutron diffraction see Chidambaram (1981).

Till late sixties, the neutron diffraction of a given crystal was studied only after analysis of its structure by x-rays. Once the x-ray non-hydrogen skeleton of the structure was available, the phase problem really did not exist in neutron diffraction. This is because, from neutron scattering length considerations (see table 1) a typical organic crystal containing C, H, N, Cl and O atoms etc may be assigned a formula, C_p H_q. Then the average contribution of non-hydrogen atoms to neutron intensities

\[ \sigma_x = \frac{P(0.66)^2}{P(0.66)^2 + Q(0.372)^2} \]  

for \( P = Q \) case is 0.7. For this \( \sigma_x \), 92% of phases evaluated from non-hydrogen atom positions only, according to the theory of Ramachandran and Parthasarathy (1965), should be within 90° of the true phases, a criterion which usually leads to the final structure. In an actual case, carbon monoxide derivative of myoglobin (Norvell et al 1975), the true and the calculated phases with omission of hydrogen atoms differed on the average by 31° up to 1.8 Å Bragg resolution.

Two reasons can be attributed for not carrying out neutron structure determinations ab initio. First, because of low incident neutron fluxes from nuclear reactors (\( \sim 10^4 \) neutrons/cm²/sec at the sample) in earlier times, the neutron Bragg intensity
Table 1. Neutron scattering amplitudes for some elements.

<table>
<thead>
<tr>
<th>Z</th>
<th>Element</th>
<th>$b(10^{-12} \text{ cm})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H</td>
<td>0.372</td>
</tr>
<tr>
<td></td>
<td>$b^*$</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>$b_-$</td>
<td>0.07</td>
</tr>
<tr>
<td>1</td>
<td>D</td>
<td>0.67</td>
</tr>
<tr>
<td>6</td>
<td>C</td>
<td>0.662</td>
</tr>
<tr>
<td>7</td>
<td>N</td>
<td>0.940</td>
</tr>
<tr>
<td>8</td>
<td>O</td>
<td>0.58</td>
</tr>
<tr>
<td>17</td>
<td>Cl</td>
<td>0.98</td>
</tr>
<tr>
<td>48</td>
<td>$^{113}\text{Cd}$</td>
<td>$0.725 + 4.507i$ (at $\lambda = 0.678 \text{ Å}$)</td>
</tr>
<tr>
<td>62</td>
<td>$^{149}\text{Sm}$</td>
<td>$0.795 + 6.051i$ (at $\lambda = 0.915 \text{ Å}$)</td>
</tr>
<tr>
<td>66</td>
<td>Dy</td>
<td>1.69</td>
</tr>
<tr>
<td>92</td>
<td>U</td>
<td>0.85</td>
</tr>
</tbody>
</table>

* $b_-$ for $I + \frac{1}{2}$ and $b_-$ for $I - \frac{1}{2}$ compound nucleus states.

data collection was very time-consuming compared to x-rays. Second, because of the differences in which x-rays and neutrons interact with the atoms making up the crystals, it was suspected that the customary phasing methods of x-rays like the heavy atom technique, direct methods etc could not be employed in neutron diffraction. However, with the advent of high flux reactors ($\sim 10^6$–$10^7$ neutrons/cm$^2$/sec at the sample) and building of automatic four-circle diffractometers, the neutron Bragg intensity data collection times have been drastically reduced and full three-dimensional data are now routinely recorded (see Sequeira et al 1978). These have motivated many attempts to solve the crystal structures directly from neutron data. These have been quite successful and have shown that the direct methods and anomalous dispersion techniques of x-rays can be applied in neutron diffraction as well. Since this volume is dedicated to Prof. Ramaseshan, it is appropriate here to review these successes, as he was the first to recognise the higher potential of neutron resonance scattering compared to the x-ray anomalous dispersion scattering for solving the phase problem. Before we do so, it is instructive first to examine the causes of the differences in x-ray and neutron phase problems in some detail.

2. Differences between x-ray and neutron phase problems

The differences arise from the fact that the scattering agents in atoms are different for the two radiations. Whereas the x-rays are essentially scattered by atomic electrons, the scattering centres for neutrons are predominantly the nuclei. The nuclear scattering length for neutrons can be written as

$$b = b_0 + b' + ib''.$$  (2)

The term $b_0$ represents the hard sphere or potential scattering contribution and is
independent of neutron energy, \( b' \) and \( b'' \) are the energy dependent resonance scattering contributions. In most cases, \( b'' \) is negligible. It is this energy-dependent \( b' \) term which makes the neutron scattering amplitude an irregular but featureful function of the atomic number.

Expression (2) is similar to

\[
f = f_0 + \Delta f' + i\Delta f''
\]

the x-ray scattering factor of the atom. \( \Delta f' \) and \( \Delta f'' \) are significant only when the wavelength of the radiation is close to the absorption edge of this atom.

Neutron scattering amplitudes for some elements are given in table 1, from which many things are apparent.

2.1 Equal atoms

The neutron scattering factors for different nuclei are of the same order of magnitude and do not display as much variation as in the x-ray case, for which it changes by a factor of 90 from the hydrogen to the uranium atom. This, although conferring the major advantage to neutron diffraction for location of protons, leads to the fact that none of the nuclei present in the crystal will dominate the scattering of neutrons and hence there will be no heavy atoms for neutron diffraction. The variation in \( b' \)'s for different elements is such that at the most, the neutron 'heavy atoms' are of strength equivalent to that of a Cl atom in a crystal containing C, H, O atoms etc. for the x-ray case*†. Further, as the scattering amplitude of the hydrogen nucleus is comparable to other elements, there are more numbers of atoms to be located at the phasing stage in a neutron structure analysis. Consequently, the neutron Patterson map will be more overcrowded and difficult to interpret directly.

2.2 Negative scattering factors

The x-ray scattering factors are all positive while for neutrons, they are of both signs. This fact may be expected to have the following consequences:

(a) The neutron Patterson map will now contain both positive and negative vector peaks and hence some of them will be removed from the map by their overlap, thereby making difficult the deconvolution of the Patterson function with the help of superposition methods like the minimum and sum functions etc. However for small structures, this has not proved a serious drawback. The structures of potassium hydrogen chloromaleate (Ellison and Levy 1965) and BeSO₄·4H₂O (Sikka and Chidambaram 1969) have been solved by the combination of symmetry minimum and minimum functions.

(b) The negativity of \( b \) also appears to rule out the possibility of using direct methods in neutron diffraction as hydrogen nucleus (\( b \) is negative) are present in almost every interesting crystal. This led Karle (1966) to propose the squared structure approach to

*¹¹³Cd,¹⁴⁹Sm,¹⁶⁴Dy etc. can act as heavy atoms for structures, say of 200 atoms per one heavy atom in the unit cell. However, they are of wrong type. They cannot be easily introduced into such structures and also their high absorption cross-sections for neutrons pose problems in data collection.

† In some of the structures e.g. UO₂(NO₃)₂·6H₂O (Taylor and Muller (1965) and H₄AuCl₄·2H₂O (Williams and Peterson 1969) which have been solved by the 'heavy atom' neutron Patterson method, this situation prevailed.
circumvent the difficulty caused by the negative scattering factors. He showed that from the observed neutron normalized structure factors \((E_h)\), it is possible to derive another set of normalized structure factors \((V_h)\) corresponding to a structure in which all the atoms scatter with the square of the original scattering factors but whose positions are the same as in the original structure. The scattering density would then become positive everywhere and the phase problem for this ‘squared’ crystal can be solved using the direct methods. The success of this approach was demonstrated by Ellison and Levy (1967), who used it to solve the structure of glycolic acid.

Sikka (1969, 1970), however, re-examined the consequences of the negativity of \(b\) for hydrogen atoms and found that these are not as serious, as assumed for the straight employment of direct methods in neutron diffraction. The amount of average neutron scattering from hydrogen atoms present in a crystal defined by

\[
\sigma = \frac{\sum_{j=1}^{N} b_j^2}{\sum_{j=1}^{N} b_j^2}
\]

is rarely more than 30% in a crystal. Up to this \(\sigma\), the criterion for the application of direct methods in x-rays i.e. the normalized structure factors for actual and the squared structures \((E_h\) and \(V_h\)) should have the same sign or nearly equal phases, is very well satisfied in neutron diffraction for \(E_h \geq 1\) as shown in figure 1.

### 2.3 Isotopic differences

The neutron scattering amplitudes vary from isotope to isotope. This makes the difference Patterson technique\(^*\) more attractive for neutrons. This map now will

\[P(E_h, V_h)\]

Figure 1. The probability \(P(E_h, V_h)\) that \(E_h\) and \(V_h\) will be alike in sign for different \(E_h\) and \(\sigma\) or \(x = P/Q\) (from Sikka 1969).

\(^*\) Difference Patterson can in principle be utilized with polarised neutrons for crystals containing ordered magnetic atoms, for whom the scattering factors are neutron polarisation-dependent. As is well-known, for spin-up polarisation of neutrons, the scattering by the magnetic atom is proportional to \((b + p)\) and for the spin-down case, it is proportional to \((b - p)\). Here \(p = \text{magnetic atom form factor. For Ni, } p = 0.13 \times 10^{-12} \text{ cm at } \sin \theta/\lambda \approx 0\) compared to \(b = 1.02 \times 10^{-12} \text{ cm.}\)
The phase problem in neutron diffraction

contain only \( RR \) and \( RO \) vectors \((R = \) replaceable atoms, \( O = \) other atoms) with peak heights \((b_{R}^{2} - b_{O}^{2})\) and \([ (b_{R1} - b_{R2}) \times b_{0} ] \) and since \( b_{R1} \) and \( b_{R2} \) can be of opposite sign, \( RO \) vectors can get strengthened. However, one has to know the positions of the \( R \) atoms to recover the structure from the difference Patterson map. Again as these will not be heavy scatterers of neutrons, unlike in x-rays, these may be difficult to locate.

A neat application of the isotopic replacement method is due to Johnson (1967) in 3 endo-phenyl-2-endo norbornanol (60 atoms in the asymmetric unit). For this compound the neutron intensity data were collected from two crystals—one containing hydrogen atoms and the other in which 4 of the hydrogen atoms had been selectively substituted by deuterium atoms. The replaceable atoms were first found by a \((|F_H| - |F_D|)^2\) Patterson (similar to \((\Delta|F|)^2\) synthesis of Rossmann (1961) for anomalous dispersion). A difference Patterson map \((F_D^2 - F_H^2)\) helped to locate additional eight atoms and the remaining atoms were found by the usual Fourier synthesis.

2.4 Complex scattering amplitudes

The neutron scattering amplitudes for some nuclei like \(^{113}\text{Cd},^{149}\text{Sm},^{155,157}\text{Gd},^{135}\text{Xe}\) etc which have high resonant absorption for thermal neutrons are complex and hence the anomalous dispersion method of x-rays becomes applicable in neutron diffraction (Peterson and Smith 1962). However, there are quantitative differences in the values of real and imaginary dispersion terms for the two radiations (Ramaseshan 1966; Dale and Willis 1966) as shown below for \(^{113}\text{Cd}\)

<table>
<thead>
<tr>
<th></th>
<th>(b'/b_0)</th>
<th>(b''/b_0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>x-rays</td>
<td>~0.3</td>
<td>~0.3</td>
</tr>
<tr>
<td>Neutrons(^{113}\text{Cd})</td>
<td>7.4 at (\lambda = 0.55) Å</td>
<td>12.4 at (\lambda = 0.68) Å</td>
</tr>
<tr>
<td></td>
<td>-7.4 at (\lambda = 0.8) Å</td>
<td></td>
</tr>
</tbody>
</table>

The neutron values of these ratios are an order of magnitude higher* than for the x-rays and can be varied significantly by a variation in wavelength (figure 2). This larger

![Figure 2](image-url)  
* These ratios for \(^{149}\text{Sm},^{157}\text{Gd}\) and \(^{135}\text{Xe}\) are even higher
anomalous dispersion effect for neutrons can, therefore, be used to tackle more complex structures than are possible by x-ray anomalous dispersion method. The power of this technique is shown in figure 3, taken from Sikka and Rajagopal (1975), by high average values of the Bijvoet ratios:

\[
\langle \delta \rangle = \frac{\langle \Delta I / I \rangle}{I} = \frac{\left| F(hkl) \right|^2 - \left| F(\bar{h}\bar{k}\bar{l}) \right|^2}{0.5 \left[ \left| F(hkl) \right|^2 + \left| F(\bar{h}\bar{k}\bar{l}) \right|^2 \right]}
\]

(5)

for hypothetical structures containing N atoms and one resonant atom (natural Sm in this case, containing 13% \(^{149}\text{Sm}\) resonant isotope). Note that in the x-ray case, structures have been solved when \(\langle \delta \rangle\) has been as low as 0.1.

3. Applications and status of direct and anomalous scattering methods in neutron diffraction

3.1 Direct methods

The application of direct methods (\(\Sigma_1, \Sigma_2\) tangent formulas, and cosine and sine invariants methods etc) for neutron diffraction both for centrosymmetric and non-centrosymmetric structures is now an accepted fact. So far, about 25 structures have been solved by this technique and these have not revealed any significant differences between the application of direct methods in x-ray and neutrons. The errors in phases (figure 4) determined by this technique are similar for the two radiations \(\langle |\phi_0 - \phi_{\text{TAN}}| \rangle_{av} = 28^\circ\) for the neutron study of \(L_4\)-threonine (Ramanadham et al 1973) compared to the value 22° for x-ray study of alkaloid Panamine (Karle and Karle...
The phase problem in neutron diffraction

119

Figure 4. Frequency distribution of the error $|\phi_0 - \phi_{\text{TAN}}|$ for non-centrosymmetric neutron reflections in L-t-threonine (from Ramanadham et al 1973).

1966). The largest structure solved using neutron data, by routine application of the program MULTAN (Germain et al 1971), up-to-date, is that of Melampodin containing 54 atoms in the asymmetric unit (Bernal and Watkins 1973). Out of the 22 largest maxima on the $E$ map of this structure, only four were spurious. In view of the above, it may now be said that direct methods can be applied in a neutron structure analysis for almost any crystal to which they are applicable in x-rays.

3.2 Anomalous dispersion method

In spite of the higher power of this technique in neutron structure analysis for large crystals like proteins (as discussed in §2.4), the method has not lived up to its early promise. So far, the method has been tested on six small structures

(i) Cd(NO$_3$)$_2$ $\cdot$ 4D$_2$O (MacDonald and Sikka 1969),
(ii) Sm(BO$_3$)$_3$ $\cdot$ 9H$_2$O (Sikka 1969a),
(iii) NaSm (EDTA) $\cdot$ 8H$_2$O (Koetzle and Hamilton 1975),
(iv) Cd(tartrate) $\cdot$ 5H$_2$O (Sikka and Rajagopal 1975),
(v) Cd-Histidine $\cdot$ 2H$_2$O (Bartunik and Fuess 1975),
(vi) agua(L-glutamato)Cd(II) $\cdot$ H$_2$O (Flook et al 1977),

and one protein, $^{113}$Cd myoglobin (Schoenborn 1975). The method is beset with the following difficulties, especially for complex crystals.

3.2a. Phase ambiguity resolution: It is well known that phases determined by the anomalous dispersion method at a single wavelength for a crystal have two fold ambiguity:

$$\phi = \phi_A + \frac{\pi}{2} \pm \theta, \quad (6)$$

$$\cos \theta = \frac{|F(hk\ell)|^2 - |F(\bar{h}\bar{k}\bar{\ell})|^2}{4|F'_A||F''_A|}. \quad (7)$$
In the x-ray case, this ambiguity is resolved or circumvented by (i) the heavy atom method (Ramachandran and Raman 1956), (ii) the double-phase Fourier synthesis method (Ramachandran 1964), (iii) the sine-Patterson method (Okaya et al. 1955) and (iv) the two wavelength method (Raman 1959; Ramaseshan et al. 1957). For neutron diffraction as the anomalous site is a light weight from scattering point of view, method (i) cannot be used. Attempts by Schoenborn to use this method in $^{113}$Cd myoglobin did not yield good results. Methods (ii) and (iii) have successfully been applied in Cd(NO$_3$)$_2$·4D$_2$O (see figure 5), Cd(tartrate)·5H$_2$O and agua(L-glutamato) Cd(II)·H$_2$O (see figure 6). But for large molecules, these methods may not be applicable. This is because, if hydrogen atoms are present in the structure (this will be almost so in every large crystal), then because of their negative scattering length, they will give peaks of opposite polarity in the positive and negative images of the structure which are present in Fourier or Patterson maps of the crystal due to these two methods and thus cause confusion in picking out the correct structure. The limitations of the two wavelength method (successful in Sm(Br$_3$)$_3$·9H$_2$O and Na Sm(EDTA)·8H$_2$O) have been examined by Sikka (1969b). If the phase ambiguity has to be resolved then the two wavelengths chosen for data collection have to be on the opposite side of the resonance. For Cd, this is not practicable ($\lambda_1 < 0.68 < \lambda_2$ and at $\lambda_1$, the incident neutron flux will be low, scattered intensity will also be less ($\alpha/\gamma$) and there will be intense overlapping of reflections) unless a hot source coupled with the reactor is used. The best anomalous scatterer for the two wavelength method seems to be $^{157}$Gd (see figure 7). It has resonance at 1.6 Å. However, no structure containing this isotope has been studied so far.

Another method suggested to circumvent the phase ambiguity is to use the tangent formula of direct methods (Sikka 1973). The basis of this is that in a tangent formula calculation, if the input phases are within 45° of their true values, the tangent formula will determine and refine the phases to a correct set. Now this input set of phases can be obtained in two ways from the anomalous dispersion data.

(i) One can use the two-wavelength method for resolving the phase ambiguity for a limited set of reflections as suggested by Bartunik and Fuess (1975). For Cd Histidine·2H$_2$O, these authors could measure the phases utilizing anomalous dispersion, with $\langle |\phi_0 - \phi_{ANM}| \rangle_{AV}$ of only 2.3°.

(ii) The centroid of the two phases, $\phi_1$, and $\phi_2$ given by formula (6) can also be used. As pointed out by Sikka (1973) and others, the errors in phases by choosing the centroid will be limited to 90° and there will be reasonable number of reflections with errors less than 45°. Weighted tangent-formula calculations for L-aspartatocadmium(II) trihydrate (P$2_1$/c$2_1$) as given in table 2, demonstrate the success of this approach.

3.2b. Location of the anomalous scatterer. In all the above methods for ambiguity resolution of phases, prior knowledge of the position of the anomalous scatterer is implied. Again the task of locating these atoms in neutrons is more difficult than for x-rays (see § 2.1). The position of the anomalous scatterer can, in principle, be found by the two wavelength methods (for which $b_\gamma (\lambda_1) = b_\gamma (\lambda_2)$) of Singh and Ramaseshan (1968) and Sikka (1969c). Singh and Ramaseshan's method is an analytical method which combines data at these two-wavelengths to give $F^2_A$, the contribution due to the

* $b_\gamma = b_0 + b'$, the total real part of scattering from the anomalous atom.
anomalous atoms alone. The position of the anomalous atoms can then be determined by a Patterson synthesis with $F_2^2$ as coefficients. This method has been tested by Koetzle and Hamilton (1975) on NaSm(EDTA)·8H$_2$O. Sikka's method is based on the cosine synthesis of Okaya and Pepinsky (1961). However, when the anomalous scatterer is
122  S K Sikka

Figure 7. Optimum wavelength regions for $^{157}\text{Gd}$, $^{149}\text{Sm}$ and $^{113}\text{Cd}$ for two different absorption factors (from Jauch and Dachs 1981).

Table 2. Tangent formula refinement results on L-asparatocadmium(II) trihydrate for circumventing the phase ambiguity in neutron anomalous dispersion method.

| Set number | Input set | Number of reflections in the input | $\delta$ limit | Weight assigned in first cycle | $\langle |\phi_0 - \phi_{\text{TAN}}| \rangle$ for acentric reflections | Number of centric reflections with wrong phase |
|------------|-----------|-----------------------------------|----------------|-------------------------------|-------------------------------------------------|---------------------------------|
| 1          | $\phi_0$  | 144                               | $-$            | 1                             | 22.4°                                           | 2                               |
| 2          | $\phi_x$  | 88                                | 90°            | $|\cos \delta|$               | 22.5                                           | 3                               |
| 3          | $\phi_x$  | 88                                | 90°            | $|\cos \delta|^2$             | 22.5                                           | 1                               |
| 4          | $\phi_x$  | 17                                | 20°            | $|\cos \delta|$               | 24.7                                           | 4                               |

$\delta$ is the half angle of the phase ambiguity in (6) and (7). $\phi_x$ is the centroid phase, $\phi_0$ is the true value of phase and $\phi_{\text{TAN}}$ is the one calculated by the tangent formula. Out of 144 reflections, 90 were acentric. $\langle |\phi_0 - \phi_{\text{TAN}}| \rangle$ is the average error after 7 cycles of refinement.

$^{113}\text{Cd}$, both these methods require the use of one neutron wavelength of about 0.7 Å, which as pointed above is not very practicable. The method which has shown more promise is the $(\Delta F)^2$ synthesis of Rossmann (1961), used by Sikka and Rajagopal (1975) in Cd(tartrate)·5H$_2$O and by Schoenborn (1975) in Cd myoglobin. For the latter structure, the Harker section at $\lambda = 1.25$ Å is shown in figure 8. The Patterson density is, clearly, much higher at the Cd–Cd vector site.

Recently, a novel way to integrate the techniques of direct methods and anomalous dispersion technique for x-rays has been attempted by Hauptman (1982) which avoids the prior determination of the anomalous dispersion sites. It will be interesting to extend this approach to neutron diffraction also.

Figure 8. Harker section for Cd myoglobin at $\lambda = 1.25$ Å for $(\Delta F)^2$ synthesis (from Schoenborn 1975). The densest peak is the Cd–Cd vector peak.
4. Some new techniques

None of the techniques described below has been used in an actual structure analysis but the physics underlying them is attractive.

4.1 Resonance modulated diffraction

This method has been suggested by Hayter et al. (1974). It is based on the fact that the neutron scattering lengths of some nuclei are spin-dependent (see table 1 for the hydrogen nucleus). If these nuclei in crystals can be polarised, then the neutrons, which are parallel to the direction of nuclear polarisation, see a coherent scattering amplitude of \( b^+ \) and the neutrons, which are anti-parallel, will be scattered with amplitude \( \frac{b^+ + 2ib^-}{2} \) (Schermer and Blume 1968). Now if the intensity data are recorded from unpolarized and polarized crystals, the phase problem can be solved by utilizing the difference Patterson technique.

Since, for the hydrogen nuclei, \( b^+ \) and \( b^- \) are large, this technique according to Hayter (1976) opens up the possibility of using these hydrogens as heavy atoms for phase determination. Experiments so far (using microwaves for polarizing hydrogens at 1.5°K) have only demonstrated the nuclear polarization effects on Bragg intensities and no actual structure determination has been carried out yet.

4.2 Three-beam method of Post (1977) for x-rays

Post (1977, 1979) has demonstrated that, in principle, the phase problem in x-ray diffraction can be resolved if the Renninger effect, in which three Bragg planes simultaneously reflect giving out three diffracted beams, is made use of. For in this case, according to the dynamical theory, the interactions among these beams are coherent and their intensities will contain information about the phases. Post, using a single crystal of \( \text{Al}_2\text{O}_3 \), has experimentally shown that the intensity distributions are indeed dependent upon the products \( F_H F_K F_{H-K} \), \( K = (h_2 k_2 l_2) \) and \( H-K \), the difference), the structure invariants as the theory predicts.

The above method can be employed in neutron diffraction also, as the dynamical theory is common to both x-rays and neutrons (Shull 1973). The four-circle diffractometers in conjunction with position sensitive neutron detectors or the film methods in neutron diffraction (Hohlwein 1975) can be employed to measure these intensity distributions. However, it remains to be seen even for x-rays, whether the method will be useful for imperfect crystals, such as are usually employed for crystal structure analysis.

Acknowledgement

The author is thankful to Dr R Chidambaram for useful discussions.

---

* The phases of these structure invariants (products of structure factors the sum of whose indices are zero) are invariant to the choice of the origin and a knowledge of a few of these is enough to initiate the use of direct methods for the solution of phase problem.
References

Chidambaram R 1981 Int. Quantum Chem. 20 449
Dale D and Willis B T M 1966 AERE R-5195
Hauptman H 1982 Acta Crystallogr. A38 289
Koetzle T F and Hamilton W C 1975 Anomalous scattering (eds) S Ramaseshan and S C Abrahams 489
(Copenhagen: Munksgaard)
Norvell J C, Nunes A C and Schoenborn B P 1975 Science 190 568
Okaya Y, Saito Y and Pepinsky R 1955 Phys. Rev. 28 1857
Ramachandran G N and Raman S 1956 Curr. Sci. 25 348
Ramachandran G N and Parthasarathy S 1965 Science 150 212
Raman S 1959 Z. Kristall. 111 301
Ramanadham M, Sikka S K and Chidambaram R 1973 Pramana 1 247
Ramaseshan S 1966 Curr. Sci. 35 87
Scherrer R I and Blume M 1968 Phys. Rev. 166 554
Schoenborn B P 1975 Anomalous scattering (eds) S Ramaseshan and S C Abrahams (Copenhagen: Munksgaard)
Shull C G 1973 J. Appl. Cryst. 6 257
Sikka S K and Rajagopal H 1975 Anomalous scattering (eds) S Ramaseshan and S C Abrahams (Copenhagen: Munksgaard) 503
Neutron diffraction using an electron linear accelerator

B T M WILLIS
Materials Physics Division, AERE, Harwell, Oxon OX11 ORA, England

Abstract. The Harwell electron linear accelerator HELIOS has been used for studying the crystal structures of polycrystalline materials by neutron diffraction. Atomic positions and atomic vibrational amplitudes can be determined equally well with the pulsed method as with the more conventional method employing a high-flux nuclear reactor as source. The powder pattern can be indexed more easily with pulsed neutrons.

Keywords. Linear accelerator; neutron diffraction.

1. Introduction

HELIOS, the 136 MeV electron linear accelerator at Harwell, has been used since 1981 to study the physics and chemistry of condensed matter. The accelerator, whose main features have been described by Lynn (1980), may be viewed as a precursor of the more intense Spallation Neutron Source (sns) which is under construction nearby and due to operate in 1985.

Pulsed neutron techniques are superior to those employing the continuous neutron beam from a nuclear reactor for the study of high-energy excitations in a solid or for the measurement of the structure factors of liquids or amorphous materials where the observations must be made over a wide range of sin Θ/λ. It is less obvious that there are also advantages in using the pulsed method to examine polycrystalline materials by time-of-flight neutron diffraction. We shall illustrate these advantages by describing the results of some experiments with HELIOS and the Harwell back scattering spectrometer (bss). The main features of the bss have been described by Windsor et al (1977). We shall not consider neutron diffraction from single crystals, although it is in this field that pulsed neutrons are particularly effective in solving the crystallographic phase problem using anomalous dispersion (see Ramaseshan and Narayan 1981).

2. Time-of-flight crystallography

The derivation is given in this section of the basic formulae required for interpreting the time-of-flight diffraction patterns from a polycrystalline sample examined with a pulsed neutron source.

Pulsed neutrons emerge from the target in the ‘condensed matter cell’ of HELIOS by a two-stage process. Bremsstrahlung radiation with a maximum photon energy equal to the energy of the electron beam (136 MeV) arises from interaction of the electrons with the heavy-metal target; the gamma-rays then take part in a photonuclear reaction to produce fast neutrons. These neutrons are subsequently slowed down to the thermal, or epithermal, range by a moderator consisting of a slab of ordinary water whose moderator time matches the electron pulse length. The pulse width of the moderated
Figure 1. Geometry of back-scattering spectrometer, bss. M is the neutron moderator, S the sample and D the detector. D is coupled to a time-of-flight analyser with 2048 time channels.

neutrons varies with their wavelength, but the mean time spread at the moderator of the neutrons in a single pulse is about 5μsec. This time spread, together with the distance \( L = L_0 + L_1 \) (see figure 1) travelled by the neutrons from the moderator to the detector, determine the resolution characteristics of the diffraction pattern.

Let \( t \) be the time taken by neutrons of velocity \( v \) to travel the distance \( L \). Then

\[
t = \frac{L}{v} = \frac{(Lm_n/h)\lambda}{},
\]

where we have used the de Broglie equation

\[
\lambda = \frac{h}{(m_n v)},
\]

relating the wavelength \( \lambda \) and the velocity. \( h \) is Planck’s constant and \( m_n \) the neutron mass.

Putting the Bragg equation

\[
\lambda = 2d \sin \Theta
\]

into equation (1), where \( 2\Theta \) is the scattering angle for the planes of spacing \( d \), gives

\[
t = (505.54 L \sin \Theta)d,
\]

with \( t \) expressed in μsec, \( L \) in metres and \( d \) in Å units. The path length \( L \) in the bss is 14 m and the scattering angle is 170°, so that the time-of-flight of the neutrons diffracted by planes of 1 Å spacing is a little over 7000 μsec. Clearly, with an initial pulse length of 5 μsec very little uncertainty in the flight time is introduced by the finite pulse length.

Equation (4) shows that the time \( t \) is directly proportional to the spacing \( d \). For \( d \) values in the range 0.5 Å < \( d \) < 2.5 Å, the corresponding times are between 3,500 and 17,500 μsec. The neutron counts are recorded in 2048 time channels with a separation of a few μsec between adjacent channels, giving a diffraction pattern such as that shown in figure 2.

Differentiating equation (3) gives

\[
\Delta \lambda/\lambda = \cot \Theta \Delta \Theta,
\]

and so high resolution is achieved by working at a high scattering angle 2Θ in the back scattering region, where \( \cot \Theta \) tends to nothing as \( \Theta \) approaches 90°. If the pattern is recorded for a \( \Theta \) value in the forward scattering region, then the lines in figure 2 would be less sharply defined.

3. Diffraction at low \( \sin \Theta/\lambda \): measurement of \( d \)-spacings

The first stage in solving a crystal structure is the measurement of the unit-cell dimensions. Using a polycrystalline sample the crucial requirement is the accurate
Neutron diffraction using an electron linac

Figure 2. Time-of-flight diffraction pattern of silicon at low Q. The observed spectrum has been ratioed with a vanadium spectrum: vanadium is an incoherent neutron scatterer, whose spectrum gives the wavelength dependence of the incident beam flux.

measurement of the first 20-30 lines of the powder pattern. This point has been emphasised by Shirley (1980): “There can be few fields of endeavour in which careful data preparation is as dramatically rewarded as in computer-based powder indexing. A properly measured pattern is usually solved almost immediately, while mediocre data can run forever”. In this context a “properly measured pattern” means that the d-spacings must be known to a few parts in 10,000.

Figure 2 shows a portion of the low sin \( \Theta/\lambda \) pattern from silicon, and table 1 indicates the kind of accuracy that can be achieved by measuring the corresponding interplanar spacings. In this table \( t_{\text{obs}} \) is the time-of-flight at the maximum of each diffraction peak: the neutron counts are recorded in time channels 8 \( \mu \)sec in width, and \( t_{\text{obs}} \) can be measured to a few tenths of a microsecond by fitting a suitable analytical function (Albinati and Willis 1982) to each peak. The observed \( d \)-spacing in the third column of the table was derived from the relation

\[
t_{\text{obs}} = cd_{\text{obs}},
\]

where \( c \) is an instrumental constant given by (1) and (3) as

\[
c = (2m_n/h) L \sin \Theta.
\]

\( d_{\text{calc}} \) in the last column of table 1 was calculated from

\[
d_{\text{calc}} = a_0 (h^2 + k^2 + l^2)^{-1/2}
\]

with \( a_0 \), the cubic lattice parameter of silicon at 20°C, equal to 5.43070 Å. The measured spacings are within 2 parts in \( 10^4 \) of the calculated values.
The exceptionally good resolution $\Delta d/d$ arises in time-of-flight crystallography because the entire diffraction pattern can be collected in back scattering. To obtain comparable resolution at large $d$ spacings with a reactor source it is necessary to design the diffractometer to give optimum focusing conditions at low values of $\sin \Theta/\lambda$, but this is not normally done as it will spoil the high-$\sin \Theta/\lambda$ portion of the pattern (Hewat 1975).

**4. Diffraction at high $\sin \Theta/\lambda$**

4.1 *Measurement of atomic positions*

To assess the accuracy of the time-of-flight technique in determining atomic positions, a diffraction pattern was recorded from monoclinic zirconia, $\text{ZrO}_2$, which has a distorted fluorite-type structure. There are three atoms at general positions (not fixed by symmetry) in the unit cell.

Data were collected up to $\sin \Theta/\lambda = 1.5 \text{Å}^{-1}$ with the back-scattering spectrometer, using 15 g of sample and 30 hr of counting time. The data were refined by the Rietveld method (see Albinati and Willis 1982) in which the entire profile of the diffraction pattern is matched to the calculated profile. The final results are shown in figure 3 and

![Figure 3. Portion of the time-of-flight diffraction pattern of monoclinic zirconia. Points are the raw data points, and the continuous line the calculated best-fit profile.](image-url)
Neutron diffraction using an electron linac

Table 2. Fractional atomic coordinates of monoclinic zirconia (space group P2\(_1\)/c)

<table>
<thead>
<tr>
<th>Atom</th>
<th>Linac method</th>
<th>Nuclear reactor method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>X (Å)</td>
<td>y (Å)</td>
</tr>
<tr>
<td>Zr</td>
<td>0.2756(4)</td>
<td>0.0393(4)</td>
</tr>
<tr>
<td>O(_{I})</td>
<td>0.0681(5)</td>
<td>0.3308(4)</td>
</tr>
<tr>
<td>O(_{II})</td>
<td>0.4494(4)</td>
<td>0.7570(5)</td>
</tr>
</tbody>
</table>

\((\sin \Theta/\lambda)_{\text{max}} = 1.5 \, \text{Å}^{-1}\)

\((\sin \Theta/\lambda)_{\text{max}} = 0.8 \, \text{Å}^{-1}\)

table 2. Twenty five parameters were varied (including four lattice parameters, three isotropic temperature factors and three coefficients defining the background), but the parameters representing the atomic positions only are given in the table.

The same sample of zirconia has also been analysed by conventional neutron powder diffraction. The high-resolution neutron powder diffractometer at AERE Harwell was used with the PLUTO nuclear reactor as source, and a Rietveld refinement carried out with a computer program adapted to constant-wavelength, angle-dispersive measurements. These results are listed in table 2 alongside those from the time-of-flight technique.

The numbers in parentheses in table 2 are estimated standard deviations (referring to least significant digits). There is satisfactory agreement between the two sets of coordinates. The greater precision for the coordinates derived by the Linac method is associated with the higher accessible value of \(\sin \Theta/\lambda\).

4.2 Measurement of atomic vibrational amplitudes

To examine the accuracy to be expected from determining atomic vibrational amplitudes, a polycrystalline sample of thoria was chosen. Th\(_2\)O\(_3\) crystallises in the cubic fluorite arrangement, so that the atomic positions are fixed by symmetry and the vibrational parameters alone are to be derived from the analysis.

Neutron diffraction data were collected in two separate time ranges:

- 4,000 to 17,500 μsec (with 8 μsec time channels)

and

- 2,500 to 5,000 μsec (with 2 μsec time channels).

The maximum value of \(\sin \Theta/\lambda\) was 0.9 Å\(^{-1}\) in the first case and 1.5 Å\(^{-1}\) in the second. Analysis was performed by the Rietveld method and the results are given in table 3.

The table shows that by extending \(\sin \Theta/\lambda\) from 0.9 Å\(^{-1}\) to 1.5 Å\(^{-1}\) the uncertainty in the measurement of the atomic vibration parameters is reduced by a factor of two. Compared with a reactor-based instrument, we gain by using the 'hot' neutrons from the epithermal spectrum of HELIOS. The theoretical values in table 2 were derived from a force-constant model used in fitting phonon dispersion relations to neutron inelastic scattering spectra (Dolling et al 1965).

5. Conclusions

We have shown that pulsed neutron diffraction, using a back-scattering spectrometer
Table 3. Thermal displacements of thorium dioxide.

<table>
<thead>
<tr>
<th>Maximum sin Θ/λ value of reflections in pattern (Å⁻¹)</th>
<th>Number of reflections</th>
<th>Mean square displacement of thorium (Å²)</th>
<th>Mean square displacement of oxygen (Å²)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Measured</td>
<td>Theory</td>
</tr>
<tr>
<td>0.9</td>
<td>44</td>
<td>2.5 ± 0.6</td>
<td>2.7</td>
</tr>
<tr>
<td>1.5</td>
<td>130</td>
<td>2.8 ± 0.4</td>
<td></td>
</tr>
</tbody>
</table>

and the Harwell electron linear accelerator, can lead to accurate measurements of d-spacings at low values of sin Θ/λ and of vibrational parameters and atomic coordinates at high values of sin Θ/λ. The performance of the pulsed neutron instrument matches or exceeds that of the best conventional instrument on a nuclear-reactor source. This augurs well for the future of the Spallation Neutron Source, which is under construction at the Rutherford Appleton Laboratory of the UK. The neutron flux of the SNS will be over one hundred times that of HELIOS.

Acknowledgements

It is a pleasure to acknowledge the help received from my Harwell colleagues, Dr R N Sinclair and Dr C G Windsor. The computer program used for the Rietveld analysis of time of flight data was provided by Professor R von Dreele of Arizona State University, USA.

References

Albinati A and Willis B T M 1982 J. Appl. Crystallogr. 15 361
Lynn J E 1980 Contemp. Phys. 21 483
Shirley R 1980 NBS Special Publication No. 567 361
Electro- and magnetogyration

I S ZHELUDEV and O G VLOKH*
Institute of Crystallography, Academy of Sciences of the USSR, Moscow 117333, Leninsky pr., 59, USSR
* Lvov State University, Lvov 290000, USSR

Abstract. A general picture of electro- and magnetogyration is presented together with the results obtained in the study of these phenomena in the crystals of linear dielectrics and ferroelectrics.

Keywords. Electrogyration; magnetogyration; crystallography; phase transition.

1. Introduction

Electrogyration and magnetogyration as new physical phenomena were predicted rather recently on the basis of general relations of tensor crystallography (Zheludev 1964, 1980, 1982). After several experiments on the observation and study of electrogyration (Vlokh 1970; Vlokh et al 1972, 1975) these investigations were rewarded with a diploma for a discovery (Zheludev and Vlokh 1980). The first communication on the observation of magnetogyration appeared quite recently (Vlokh 1981, 1982).

2. Electrogyration

Electrogyration, by definition, consists of the effect of electric field on the gyrotropy of crystals. Analytically, it can be represented by the power expansion of gyration tensor $g_{ij}$ (second-rank axial tensor) in electric fields $E$:

$$g_{ij} = g_{ij}^0 + \gamma_{ijk} E_k + \beta_{ijkl} E_k E_l + \ldots,$$

where $g_{ij}^0$ is the gyration tensor in the absence of the field, and $\gamma_{ijk}$ and $\beta_{ijkl}$ are third- and fourth-rank axial tensors describing linear and quadratic electrogyration effects, respectively. Linear electrogyration occurs in all crystal classes except for $m3m, 43m$ and $432$; with tensor $\gamma_{ijk}$ being the same as those describing piezomagnetism (Zheludev 1983).

Phase transition in ferroelectrics with the appearance (or disappearance) of spontaneous polarization $P_s$ provides a natural possibility to observe and study spontaneous electrogyration. The most simple and reliable are the cases where a centrosymmetric optically inactive crystal acquires gyrotropy. This phenomenon was observed for a number of crystals. In particular, the studies of the phase transition in triglycine sulphate (TGS) crystals near 49°C show that at temperatures less than $T_c$ specific rotation of the polarization plane $\rho$ is really proportional to $P_s$ (Vlokh et al 1971).
For some ferroelectrics, appearance of $P_s$ of opposite orientations results in the formation of the domains with enantiomorphism of different signs having therefore optical activity of different signs. This, particularly, enables us to observe both change in sign of optical activity and gyration hysteresis in ferroelectrics (e.g. in $5\text{PbO} \cdot 3\text{GeO}_2$ crystals) during domain repolarization (Iwasaki et al 1972). Of special interest is the possibility to obtain high-induced electrogyration in ferroelectric crystals in the phase transition region. Anomalies of induced electrogyration in the phase transition region are determined finally by dielectric anomalies (Vlokh et al 1977), the attainable specific rotation of the light polarization plane being as high as several degrees in the field of 1 kV.

Detection of forced electrogyration required a careful search for suitable crystals and refined experimental technique. Linear electrogyration has been observed, measured and studied in $\alpha$-HIO$_3$, LiIO$_3$, PbMoO$_4$ and several other crystals todate. The study of electrogyration in optically biaxial $\alpha$-HIO$_3$ crystals (class 222) was made difficult by the accompanying electrooptic and piezooptic effects. The experimental value of electrogyration coefficient $\gamma_{52}$ is $(4.33 \pm 0.70) \times 10^{-13}$ m/V for $\lambda = 632.8$ nm (Vlokh et al 1972).

Electrogyration in uniaxial crystals of classes 4, 6, 3, etc. is not so complicated by electrooptic and piezooptic effects if the electric field applied and the wave normal are directed along the optical axis (the crystal remains uniaxial). Therefore, experiments with LiIO$_3$ crystals belonging to class 6 (Vlokh et al 1975) were of key importance in proving the observability of the forced electrogyration. The effect was measured at $-140^\circ$C, since the crystal exhibited ionic conductivity at room temperature. The value of $(1.57 \pm 0.30) \times 10^{-13}$ m/V was obtained for electrogyration coefficient $\gamma_{33}$, the contribution of piezoelectric deformation being 17% of the value observed (figure 1).

The most noteworthy are the experiments on forced electrogyration in PbMoO$_4$ crystals, because, unlike the above crystals optically active even in the absence of the electric field, PbMoO$_4$ acquires gyrotropy only when the field is applied (the crystal is centrosymmetric, its symmetry group being 4/m). Electrogyration can be observed in pure form without the usually accompanying electrooptic and piezoeffects. Specific rotation in readily obtainable fields (~ 10 kV/cm) attains quite considerable values in

![Figure 1. Dependence of the increment of the specific rotation of polarization plane on the field strength $E_z$ for LiIO$_3$ crystals.](image-url)
this crystal, being as high as several degrees. The phenomenon is characterized by the dispersion (figure 2): at $E = 10 \text{kV/cm}$ $\rho = 1.5 \text{ cm}^{-1}$ for $\lambda = 700 \text{ nm}$ and $\rho = 5 \text{ cm}^{-1}$ for $\lambda = 400 \text{ nm}$. Coefficient $\gamma_{33}$ (the field and wave normal directed along the optical axis) is $(1.23 \pm 0.12) \cdot 10^{-12} \text{m/V}$ for $\lambda = 517 \text{ nm}$ (Vlokh et al 1975).

Quadratic electrogyration is described by a fourth-rank axial tensor and can occur only in acentric crystals (Vlokh et al 1970). It was experimentally observed in quartz crystals with electric field applied along crystallographic axes and plane-polarized wave propagating along the optical axis. The value of the quadratic electrogyration coefficient observed was $\beta_{31} = (4.5 \pm 0.34) \cdot 10^{-20} \text{m}^2/\text{V}^2$ (Vlokh 1971).

3. Magnetogyration

There is a number of well-known and well-investigated magnetic phenomena (Ramaseshan 1951). Magnetogyration is a new one consisting, by definition (Zheludev 1964), of the ability of anisotropic media to acquire (or to change) their optical activity $g_{ij}$ ($g_{ij}$ is a symmetric second-rank tensor) in the magnetic field, $H_k$:

$$g_{ij} = g_{ij}^0 + \delta_{ijk} H_k + Q_{ijk} H_k H_l + \ldots ,$$

Figure 2. Dispersion of induced optical activity of PbMoO$_4$ crystals in static field $E_z = 10 \text{kV/cm}$.

Figure 3. Block diagram of the experimental device used in studying Faraday effect and magnetogyration. 1,1'—He–Ne lasers; 2,2'—$\lambda/4$ plate in diagonal position; 3,3'—rotatable polarizer with limb; 4—sample under investigation; 5—permanent electromagnet; 6,6'—recording system; 7,7'—galvanometer.
where \( g_{ij}^0 \) is the gyration tensor in the absence of the magnetic field; and \( \delta_{ijk} \) and \( Q_{ijkl} \) are third- and fourth-rank polar and axial tensors describing linear and quadratic magnetogyration, respectively. Tensor \( \delta_{ijk} \) has nonvanishing components only in the case of non-centrosymmetric crystals and textures, its form being the same as that of the respective piezoeffect tensors. Magnetogyration as well as the natural optical activity can be described by a surface of the “twisted cylinder” symmetry, group \( \infty 2 \) (or by a set of such surfaces).

If magnetic field is applied to a non-centrosymmetric medium (particularly, to a nonmagnetic one), the rotation of the light polarization plane is determined by the two effects: Faraday effect and magnetogyration. The contributions of these effects, linear in field, can be separated due to the difference in their symmetry (Faraday effect is characterized by the symmetry of “rotating cylinder”, group \( \infty/m \)). One of the methods of their separation lies in measuring the rotation of the polarization plane for the light passed in opposite directions (figure 3). If both effects occur simultaneously, they have the same sign (i.e. add up) for one direction, and different signs (i.e. subtract from each other) for the opposite direction.

![Figure 4](image)

**Figure 4.** Dependence of the specific rotation of polarization plane on the magnetic field strength for CdS crystals. 1—wave vector is \(+ k\); 2—wave vector is \(- k\); 2’—curve 2 reflected with respect to \( H_z \)-axis.
Magnetogyration represents a magnetic analogue of electrogyration. In the simplest cases, the former consists of the appearance of optical activity induced by a magnetic field (axial action) applied along the exclusive polar direction (crystals of $3m$, $4mm$, $6mm$ symmetry), whereas the latter lies in the appearance of the optical activity induced by an electric field (polar action) applied along the exclusive axial directions (crystals of $6$, $4/m$, $6/m$ symmetry). In particular, by analogy with the occurrence of the spontaneous electrogyration in the phase transitions due to spontaneous polarization $P_s$ in the crystals of this symmetry, spontaneous magnetization $I_s$ in crystals of classes $3m$, $4mm$, $6mm$ results in the spontaneous magnetogyration. Magnetogyration is a consequence of the spatial dispersion, i.e. the dependence of the quantities characterizing the interaction of electromagnetic wave with medium on the wavelength (wave vector).

![Figure 5. Dependence of the specific rotation of polarization plane on the magnetic field strength for \( \text{Pb}_{x} \text{Ge}_{y} \text{O}_{z} \); \( \text{Eu}^{3+} \) crystals in different states of spontaneous electrogyration. 1, 2, 3, 4, 5 —at 20, 50, 120, 140 and 160°C, respectively; \( x \)—in polydomain crystals (plots are matched to the initial points of monodomain crystals, because in polydomain crystals $\rho = 0$ at $H = 0$); \( O, \bullet \)—in monodomain crystals at opposite directions of spontaneous polarization; 6—at 180°C (paraphase); 7—for quartz crystals at opposite directions of the wave vector (\( O, \bullet \)) and from the results of double propagation measurements (\( x \)).](image-url)
Figure 6. Temperature dependence of the increments of the specific rotation of polarization plane for Pb$_2$Ge$_2$O$_4$: Eu$^{3+}$ crystals at $H_z = 3.5$ kE. 1—in polydomain crystals (Faraday effect); 2,3—in monodomain crystals at the opposite signs of enantiomorphism (superposition of Faraday effect and magnetogyration). The Curie point of the crystal is 161°C.

(Agranovich and Ginzburg 1979). Magnetogyration is determined by the part of this dependence that appears due to or is changed by magnetic field.

Linear magnetogyration was first studied in the crystals of linear dielectrics LiIO$_3$ and SiO$_2$ (Vlokh 1981, 1982). For LiIO$_3$ crystals (symmetry class 6) the obtained values of coefficients $\alpha_{33}$ (for Faraday effect) and $\delta_{33}$ (for magnetogyration) were $(111.15 \pm 0.04) \cdot 10^{-11}$ and $(0.51 \pm 0.06) \cdot 10^{-11}$ rad/Oe, respectively. Quartz crystals were studied mainly to ascertain whether Faraday effect is not accompanied with magnetogyration if field $H$ is directed along z-axis (in third-rank polar tensor $\delta_{33} = 0$ for the crystals of class 32). In CdS crystals magnetogyration should appear in pure form if magnetic field applied is directed along the optical axis. In this crystal $\alpha_{33} = (501.8 \pm 0.3) \cdot 10^{-11}$ rad/Oe and $\delta_{33} = (4.4 \pm 0.5) \cdot 10^{-11}$ rad/Oe (figure 4).

In polydomain ferroelectrics with spontaneous polarization in centrosymmetric phase, induced linear magnetogyration does not occur (tgs, BaTiO$_3$, etc.). If original modification is optically active (e.g. in Rochelle salt), then polydomain crystal displays magnetogyration. In the cases where domain of opposite orientation is characterized by enantiomorphism of different signs (transitions with the change in symmetry) e.g. $2/m \rightarrow 2$ (tgs), $6 \rightarrow 3$ (lead germanate, Pb$_2$Ge$_2$O$_4$, etc.), crystals monodomainized by electric fields of opposite directions, being affected by magnetic field, show magnetogyration of opposite signs. Studies of magnetogyration in lead germanate crystals gave $\alpha_{33} = (151.5 \pm 0.2) \cdot 10^{-11}$ rad/Oe and $\delta_{33} = (2.5 \pm 0.3) \cdot 10^{-11}$ rad/Oe (figure 5). In the phase transition temperature region (near 162°C) magnetogyration in lead germanate demonstrates an anomalous behaviour and at higher temperatures vanishes (figure 6).

References

Agranovich V M and Ginzburg V L 1979 Crystalloptics with regard to spatial dispersion and exciton theory (Moscow: Nauka)


Ramaseshan S 1951 Proc. Indian Acad. Sci. A34 32, 97

Vlokh O G 1970 Ukr. Fiz. Z. (Kiev) 15 74

Vlokh O G 1971 Pis’ma v JETF, Moscow 13 118

Vlokh O G 1981 Ukr. Fiz. Z. (Kiev) 26 1623

Vlokh O G and Krupilnitskaya T D 1970 Kristallographia 15 587
Vlokh O G, Lazko L A and Nesterenko V Ya 1972 Kristallographia 11 1248
Vlokh O G, Zheludev I S and Klimov I I 1977 DAN SSSR 223 1391
Zheludev I S 1964 Kristallographia 9 501
Zheludev I S 1983 Symmetry and its applications energeatomizdat Moscow
Zheludev I S and Vlokh O G 1980 Diploma of the State committee of the USSR on investigations and
discoveries, No 211 Electrogyration phenomenon in crystals.
A theoretical consideration of the intermediate phase in $\alpha$-$\beta$ transition of quartz

N KATO  
Department of Crystalline Materials Science, Faculty of Engineering, Nagoya University, Chikusa-ku, Nagoya, Japan.

Abstract. Recently, a thermodynamically stable phase between $\alpha$ (low) and $\beta$ (high) phases of quartz has been found. This paper explains the three phases by a statistical theory of Bragg-William's type by assuming three states of Si$_4$O$_4$ tetrahedron along $a$-axis. Adjusting the interaction parameters one can predict the successive transitions of the first-order ($T_\alpha$) and the nearly second-order ($T_\beta$). Also with an assignment of $T_\alpha = 573^\circ$C, one obtains $T_\beta = T_\alpha + 1.2$ degree which fits to the experiment. The difference in energy between Dauphiné twin bonding and the normal bonding is estimated to be 0.525 eV.

Keywords. Phase transition theory; quartz.

1. Introduction

Recently, we found an intermediate phase between $\alpha$ (low) and $\beta$ (high) phases of quartz (Gouhara et al 1983). Figure 1 is the most direct illustration of the new phase. The figure is an x-ray oscillating topograph* (Schuttky 1965) near the transition temperature ($T_c = 573^\circ$C) and under a small thermal gradient. The differences in the intensity of different phases are mainly due to the difference in the degree of extinction. The details of the contrasts are given in Gouhara et al (1983).

The temperature region of the intermediate phase ($\iota\pi$) is about 2 degrees above $T_c$. Several evidences show that the transition between $\alpha$ and $\iota\pi$ is of first-order and the transition between $\iota\pi$ and $\beta$ is of second-order. Recently, distinct satellite spots were also observed near the strong Laue spots (Gouhara et al 1983). Analysis shows that $\iota\pi$ has a character of incommensurate phase**.

This paper intends to show that the successive phase transitions can be understood by statistical mechanics assuming three energy states of SiO$_4$ tetrahedron. Historically, Wannier (1959) proposed a theory of Bragg-William’s type assuming two energy states but the transition was predicted to be of the second-order. Recent experimental evidences favour the first-order transition. The present theory can predict that the transition between $\alpha$ and $\iota\pi$ is of first-order whereas the transition between $\iota\pi$ and $\beta$ can be of second-order or at least a weak first-order. Our experimental results agree with this prediction.

2. Free energy of the system

Let us consider the structure of quartz in figure 2, which illustrates only Si atoms in $\beta$-phase. Si locates in the middle of three equivalent $a$-axes of a unit cell. The crystal has

---

* This is essentially equivalent to the traverse pattern of Lang's technique.  
** Recently, a similar result obtained by means of neutron diffraction is reported (Dolino and Bachheimer 1983). Since no experiment curve is published we cannot compare their results with ours.
six-fold axes. In $\alpha$-phase, Si is displaced by $\pm u$ along a-axis. The crystal then transforms to the structure of three-fold symmetry. The phases corresponding to $\pm \beta$ are referred to $\alpha_1$ and $\alpha_2$, which are mutually equivalent. The macroscopic coexistence is traditionally called Daupine twin. Incidentally, the separation by faint white curves in the $\alpha$ phase of figure 1 can be interpreted to be the twinning of this sort, which was caused by thermal strain when the temperature of the specimen was decreased from $\beta$ phase.

In the present theory, it is assumed that Si can take three positions on a-axis; $u = 0$ and $\pm u$. The oxygen tetrahedron is associated to each Si. They also take three states corresponding to the positions of Si. Henceforth, the three states are denoted by the suffixes 0, 1 and 2.

Since Si–O bonding is homopolar, it is reasonable to take into account only the interaction between the nearest neighbours of tetrahedron units. The interaction
energy is listed in table 1. Besides, we assume the intrinsic energy of the unit, which are denoted by $e_0$ and $e_1 = e_2 = e$.

Assuming the mean-field approximation (Bragg–William’s approximation), one can write the total energy in the form

$$U/N = u_0 + \bar{e} n_0 + 2\tau n_0^2 + 4\bar{\mu} n_1 n_2.$$  \hspace{1cm} (1a)

In the special case of $n_1 = n_2$, it has the form

$$U/N = \bar{u}_0 + \bar{\mu} + (\bar{e} - 2\bar{\mu}) n_0 + (2\tau + \bar{\mu}) n_0^2.$$ \hspace{1cm} (1b)

Here, $N$ is the total number of the tetrahedron unit, and $n_i$ is the population ratio in the state $i$. Obviously, the condition

$$n_0 + n_1 + n_2 = 1$$ \hspace{1cm} (2)

must be satisfied. The (reduced) energy parameters $\bar{u}_0$, $\bar{e}$ etc. are given by

$$\bar{u}_0 = e + 2\tau,$$ \hspace{1cm} (3a)

$$\bar{e} = e_0 - e - 4\tau + 4\mu_0,$$ \hspace{1cm} (3b)

$$\bar{\tau} = \tau_0 + \tau - 2\mu_0,$$ \hspace{1cm} (3c)

$$\bar{\mu} = \mu - \tau.$$ \hspace{1cm} (3d)

Since $\alpha$-phase is stable at a sufficiently low temperature, $U(n_0 = 1, n_1 = n_2 = 0)$ must be larger than $U(n_0 = 0, n_1 = 1, n_2 = 0)$. This implies that

$$\bar{e} + 2\bar{\tau} > 0.$$ \hspace{1cm} (4)

| Table 1. Interaction energy of the nearest neighbours of SiO$_4$ tetrahedron. Symmetry is taken into account. |
|---|---|---|
| 0 | 1 | 2 |
| 0 | $\tau_0$ | $\mu_0$ | $\mu$ |
| 1 | $\mu_0$ | $\tau$ | $\mu$ |
| 2 | $\mu$ | $\mu$ | $\tau$ |
Also, one can anticipate the right pair \((i = j)\) has less interaction energy than the wrong pair \((i \neq j)\). Then, the following inequalities must be fulfilled.

\[
\bar{e} > 0, \; \bar{\tau} < 0, \; \bar{\mu} > 0
\]

The entropy \(S\) is the mixing entropy of \(Nn_0, Nn_1\) and \(Nn_2\) in \(N\) sites. Therefore, it has the expression

\[
S = k \log \left( \frac{(1/N)^N}{n_0! n_1! n_2!} \right),
\]

\[
= -kN[n_0 \log n_0 + n_1 \log n_1 + n_2 \log n_2],
\]

where \(k\) is the Boltzmann constant and Stirling’s approximation is used in the last expression.

In the thermodynamical equilibrium the Helmholtz-free energy \([F = U - TS]\) must be minimum with respect to two independent variables of \(n_0, n_1\) and \(n_2\). Various expressions are available as the equilibrium conditions. For example,

\[
n_0 \exp\left[\{ (\bar{e} - 4\bar{\mu}) - 4(\mid \bar{\tau} - \bar{\mu})n_0 \}/kT\right] = n_1 \exp\left[ -4\bar{\mu}n_1/kT \right],
\]

\[
= n_2 \exp\left[ -4\bar{\mu}n_2/kT \right].
\]

Wannier’s two-state theory (Wannier 1959) \((i = 1, 2)\) is a special case of \(\bar{e} > 4\bar{\mu}\). In this case, the equilibrium condition reduces to the last relation (7b). The phase transition occurs at

\[
T_w = 2\bar{\mu}/k.
\]

The present theory is more versatile. Depending on the energy parameters, the various types of the phase transitions are described. Since this paper intends to show the possibility of the successive transitions \((\alpha \to \gamma \to \beta)\), we shall not discuss the details and present only some numerical results in the next section. As to the formal discussion, it is worth referring to the paper of Nakano (1964) which dealt with a statistical system of unity spin.

3. Numerical analysis

Figure 3 is a series of the contour diagram of the free energy. Here, we shall define a normalized temperature

\[
R = T/T_w.
\]

At low temperature \((R = 0.65)\), a large minimum \(A\) and a small minimum \(B1\) are expected. Increasing temperature \((R = 0.71)\), the minimum \(A\) becomes shallow and \(B1\) becomes deep. Further, by increasing the temperature, \(A\) and \(B1\) disappear and another minimum \(B2\) appears.

To understand more precisely the behaviour of \(B1\) and \(B2\), the minimum and maximum positions of \(n_0\) are plotted against \(R\) under the condition of \(n_1 = n_2\). One example is shown in figure 4. The minimum below the line of \(n_0 = 1/3\) is not a real minimum but a saddle point. A phase transition is expected between \(T1\) and \(T2\) with hysteresis. In this sense, the transition is of first-order. If \(T1\) and \(T2\) are very close, it is nearly of second-order. This is the case of quartz.
In figure 3 and 4, for the sake of clear illustration, the energy parameters used are not real ones ($\bar{\varepsilon}/\bar{\mu} = 3.925; |\varepsilon|/\bar{\mu} = 1.96$). Nevertheless, one can identify the minima of $A$, $B1$ and $B2$ to the $\alpha$, $\beta$ and $\gamma$ phases, respectively.

After several trials, we finally obtained a set of parameters, which are realistic in the sense that the phase transitions occur successively. The parameters employed are 

$$\bar{\varepsilon}/\bar{\mu} = 3.817, |\varepsilon|/\bar{\mu} = 1.890.$$  

(10a, b)
Then the transition occurs at

\[ \alpha \rightarrow IP: \quad R_c = 0.6937, \quad (11a) \]

\[ IP \rightarrow \beta: \quad R_\beta = 0.69465. \quad (11b) \]

If \( R_c \) is assigned to the temperature \( T_c = 573^\circ C \), we shall see

\[ T_\beta - T_c = 1.2 \text{ degree}, \quad (12) \]

which is slightly less than the experimental value (Gouhara et al 1983).* By the definitions of \( T_w \) and \( R \) (equations (8) and (9)), we also see

\[ T_w = 946.5^\circ C, \quad (13) \]

and

\[ \bar{\mu} = 0.525 \text{ eV}. \quad (14) \]

4. Discussion

The present theory is no more than a parameter theory, although it is microscopic. The parameters employed, therefore, must be regarded merely as "effective ones". Nevertheless, it is worth discussing their physical implications, particularly of the numerical figures given by (10) through (14).

First, although the intrinsic energies (\( \varepsilon_0 \) and \( \varepsilon \)) were formally introduced, they must be practically null. Secondly, the main part of the interaction energies (\( \varepsilon \)) listed in table 1 is the bonding energy of Si–O–Si. Thirdly, the deformation of tetrahedron units is negligible. This is justified to some extent by the structural study of Young (1962).

The figures of \( \varepsilon \) were sought under the requirements given by (4) and (5). From these figures, with the use of a reasonable working hypothesis (\( \sigma = 0 \)), one can obtain

\[ \mu = \bar{\mu}, \quad \mu_0 = 0.954\bar{\mu}, \quad \tau_0 = 0.018\bar{\mu}, \quad (15a,b,c) \]

The modification with a finite \( \tau \) comparable with \( \tau_0 \) does not change the sequential order of the magnitudes in (15) which is quite reasonable. Incidentally, \( \mu \) is the bonding energy per bond of the interface between \( \alpha_1 \) and \( \alpha_2 \) states.

Very little information is available for criticizing the \( \bar{\mu} \) values of (14). It is reported (Klassen–Neklyudova 1964) that indentation with a steel ball bearing can produce twinning even at room temperature. A rough estimation of the required pressure \( P \) was 1000 kg/mm\(^2\) or more. The elastic energy absorbed per unit cell divided by six (the number of bonding) will trigger the twinning. Here again the deformation of the tetrahedron is neglected. Based on this model, one can estimate \( \mu \) by \( \alpha^2/12(1/C)P^2v \) (\( C \): elastic constant, \( 10^{11} \) dyne/cm\(^2\); \( v \): volume of unit cell), which amounts to \( \alpha^2 \times 6.24 \times 10^{-2} \) eV. Here, \( \alpha \) is introduced as a multiplication factor to the pressure mentioned above. Since \( \alpha = 3 \sim 5 \) is not unrealistic, the figure given by (14) is satisfactory at the present state of the art.

A question may arise for any theory of the present type; namely whether the state \( 0 \) is really stable in the potential curve. The present author believes that the physical reality is a continuous distribution of the tetrahedron unit along \( \alpha \)-axis under a self-consistent

* It is not difficult to improve the calculation.
potential field. Thus, the problem of whether the potential curve has double or triple minima is not significant. The double minimum theory of Wannier (1959) cannot predict properly the intermediate phase. The success of the present theory implies that the sojourn of the tetrahedron unit within a limited range of a-coordinate is significant. The physical interpretation of the present theory, therefore, must be different from that for the statistical system of unity spin.

Acknowledgements

The author expresses his thanks to Professors Nakano and Ishibashi for their critical discussion and Mr Gouhara and Mr Li for their continuous collaboration.

References

Dolino G and Bachheimer J P 1983 Solid State Commun. 45 259
Schuttky G H 1965 J. Appl. Phys. 36 2712
Young R A 1962 Defence documents contract (USA) report
Crystal and molecular structure of pyrrolidine dithio carbamate cobalt (III) chelate

GOPINATH KARTHA and KOTRA V KRISHNAMURTHY*
Roswell Park Memorial Institute, Buffalo, New York 14263, USA
*United States Testing Co., Richland, Washington 99325, USA

Abstract. Pyrrolidine dithio carbamate (pdtc) cobalt(iii) chelate has been used as a coprecipitation matrix for obtaining preconcentration factors of 100 or more for analytical estimation of trace metals in natural waters. This chelate, suitable for bivalent and trivalent ions, crystallizes as dark green crystals in monoclinic space group C2/c with unit cell parameters \(a = 28.89 \text{ Å}, b = 14.39 \text{ Å}, c = 16.17 \text{ Å} \) and \(\beta = 104.2^\circ\). X-ray crystallographic analysis of these crystals reveals that three pdtc are coordinated to a Co\(^{3+}\) atom through their sulphur atoms, the six sulphur atoms forming a distorted octahedra around the metal atom at an average distance of 2.26 Å. The thiocarbamate part of the ligands is nearly planar but the pyrrolidine rings are puckered.

Keywords. Pyrrolidine dithio carbamate; cobalt-chelate; crystal structure.

1. Introduction

The occurrence of heavy metals, especially lead and cadmium in the environment is now recognized as a major health problem. The abundance of these heavy metals in natural waters has been directly linked to the extent of their low level toxic interaction with tissues such as kidney. With the low ambient level of Cd (< 1 ppb) in public waters, available analytical methods are inadequate and the method of coprecipitation for preconcentrations of trace metals prior to their analytical determination by atomic absorption spectroscopy has special advantage in reducing large volumes of the sample to a few milligrams of the material. Coprecipitation of trace heavy metals on tris-pyrrolidine dithio carbamato cobalt (iii) has been developed earlier for the study of sea water by Boyle and Edmond (1975). Krishnamurthy and Reddy (1977) extended the procedure to include other natural waters, soil extracts and sediment digestates. Concentration factors ranging from 40 to 400 are obtained in this method and it is particularly useful for preconcentration of Pb, Cd, Ni and Cu. In this study the coprecipitation matrix has been the subject of investigation in view of its extraordinary thermal and photochemical stability under ambient conditions as also its versatility to retain several trace metals such as Cu, Cd, Pb, Ni, Fe and Cr but not Mn and Zn. The dithio carbamate ligand is also of great biological interest due to its action on the growth and viability of sarcoma cells and fibrocytes, use in the treatment of alcoholism and its function as a protectant against lethal effects of radiation (Powell 1954; Cohen et al 1957). The structural studies of these ligand metal complexes provides much needed information on the metal-sulphur distances for characterizing the heavy metal sulphur binding in the tissue. In this paper we report the results of our study on the crystal structure of the matrix cobalt (iii)-(pyrrolidine dithio carbamate).
2. Experimental

Several batch samples of the cobalt-chelate are prepared under the coprecipitation conditions used in the analysis of natural waters. The procedure is as follows: to one litre of di-ionized distilled water are added 5 ml of 6 M HNO$_3$ and 2.5 ml of Co(NO$_3$)$_2$ stock solution (~2000 μg Co (μ) per 10 ml) followed by 10 ml of NH$_4$OAc buffer solution (170 g NH$_4$OAc in 5000 ml of water purified by ADPC-CCl$_4$ extraction 3 times). The pH of the mixture should be ~ 4. A 2% solution of the precipitant, ammonium pyrrolidine dithio carbamate (APDC) in the meanwhile is purified by repeated CCl$_4$ extractions and 5 ml of this reagent are added to the cobalt mixture. The contents of the beaker are allowed to stand for 30 minutes with intermittent stirring when a yellowish green opalescence develops due to the formation of cobalt (II) pyrrolidine dithio carbamato chelate. The yellowish green colour gradually changes to green on standing due to ready oxidation of the chelate to a + 3 cobalt state (Gleu and Schwab 1950; Malissa and Schöffmann 1955). The cobalt (III) chelate thus obtained is filtered through a 0.45 millipore filter and air dried at room temperature to yield a green flaky material exhibiting creep properties. Several one liter batch precipitations are filtered through the same filter paper to increase yield and facilitate sample handling. Microanalytical results using Perkin-Elmer 240 elemental analyzer gave molecular formula CON$_3$S$_6$C$_{15}$H$_{24}$ (M.W. 497) corresponding to 1:3 chelation between cobalt and the ligand.

The Co (PDTC)$_3$ is dissolved in methyl isobutyl ketone and recrystallized to yield dark green shining crystals. The crystals are monoclinic, C centered with $a = 28.89$ Å, $b = 14.39$ Å, $c = 16.17$ Å and $β = 104.2°$. The x-ray diffraction intensities were measured to a Bragg angle of 60° using nickel filtered Cu radiation ($λ = 1.542$ Å) on a General Electric 490 automatic diffractometer by stationary crystal stationary counter method (Furnas and Harker, 1955). The crystals were very small and poor scatterers and so 7116 diffraction data from two equivalent sets were measured and averaged to give 3506 unique reflections. Only about 2100 of these (60%) had intensities greater than 3σ(I) and were considered observed and used in structure refinement. The intensities were corrected for geometrical factors in the usual way. No absorption corrections were applied.

3. Structure determination

Systematic absence $hkl$ for $h + k$ odd and $hol$ for $l$ odd revealed C centering and a glide plane parallel to c axis. The intensity statistics favoured a centrosymmetric space group. Hence the structure analysis and refinements were carried out in space group C2/c. The measured density and volume of the unit cell indicated 12 molecules of Co (PDTC)$_3$ in the unit cell. In space group C2/c this implies one molecule in the general position and another in a special position on the two-fold axis. The location of the cobalt atoms were found from a sharpened Patterson map and all the remaining 37 atoms were found and refined by Fourier and least squares methods. Positional and anisotropic thermal parameters of all but the three atoms on the two fold axes were refined by block diagonal least squares and the usual R factors for 2105 reflections used in the structure determination and refinement is 12%. Fractional atomic coordinates and equivalent thermal parameters are given in table 1. Lists of observed and calculated structure factors are also included as supplementary material.
Table 1. Fractional coordinates and equivalent B values of atoms.

<table>
<thead>
<tr>
<th>Atom</th>
<th>X/a</th>
<th>Y/b</th>
<th>Z/c</th>
<th>B_eq</th>
</tr>
</thead>
<tbody>
<tr>
<td>COA</td>
<td>0.1673</td>
<td>0.4609</td>
<td>0.5303</td>
<td>5.6</td>
</tr>
<tr>
<td>COB</td>
<td>0.0000</td>
<td>0.0408</td>
<td>0.7500</td>
<td>5.4</td>
</tr>
<tr>
<td>S1A</td>
<td>0.1366</td>
<td>0.6072</td>
<td>0.5153</td>
<td>5.1</td>
</tr>
<tr>
<td>S2A</td>
<td>0.2108</td>
<td>0.5308</td>
<td>0.4498</td>
<td>4.4</td>
</tr>
<tr>
<td>S3A</td>
<td>0.2117</td>
<td>0.3299</td>
<td>0.5519</td>
<td>4.1</td>
</tr>
<tr>
<td>S4A</td>
<td>0.2220</td>
<td>0.4870</td>
<td>0.6574</td>
<td>4.7</td>
</tr>
<tr>
<td>S5A</td>
<td>0.1093</td>
<td>0.4033</td>
<td>0.5862</td>
<td>5.3</td>
</tr>
<tr>
<td>S6A</td>
<td>0.1144</td>
<td>0.4004</td>
<td>0.4158</td>
<td>3.9</td>
</tr>
<tr>
<td>S1B</td>
<td>0.0627</td>
<td>-0.0086</td>
<td>0.7785</td>
<td>4.8</td>
</tr>
<tr>
<td>S2B</td>
<td>0.0153</td>
<td>0.0258</td>
<td>0.8918</td>
<td>5.2</td>
</tr>
<tr>
<td>S3B</td>
<td>0.0487</td>
<td>0.1651</td>
<td>0.7510</td>
<td>5.3</td>
</tr>
<tr>
<td>NA</td>
<td>0.1813</td>
<td>0.7098</td>
<td>0.4158</td>
<td>5.5</td>
</tr>
<tr>
<td>NB</td>
<td>0.2790</td>
<td>0.3369</td>
<td>0.6987</td>
<td>5.6</td>
</tr>
<tr>
<td>NC</td>
<td>0.0374</td>
<td>0.3198</td>
<td>0.4642</td>
<td>4.9</td>
</tr>
<tr>
<td>ND</td>
<td>0.0953</td>
<td>-0.0684</td>
<td>0.9527</td>
<td>5.1</td>
</tr>
<tr>
<td>NE</td>
<td>0.0000</td>
<td>0.3216</td>
<td>0.7500</td>
<td>6.8</td>
</tr>
<tr>
<td>C1A</td>
<td>0.1772</td>
<td>0.6288</td>
<td>0.4549</td>
<td>4.6</td>
</tr>
<tr>
<td>C2A</td>
<td>0.1539</td>
<td>0.7942</td>
<td>0.4283</td>
<td>4.4</td>
</tr>
<tr>
<td>C3A</td>
<td>0.1710</td>
<td>0.8666</td>
<td>0.3728</td>
<td>5.7</td>
</tr>
<tr>
<td>C4A</td>
<td>0.1873</td>
<td>0.8082</td>
<td>0.3051</td>
<td>6.0</td>
</tr>
</tbody>
</table>

4. Discussion

The coordination geometry of the chelate complex and the atom numbering scheme are shown in figure 1. A stereoscopic drawing of the molecule in the general position is given in figure 2. There is another molecule in the asymmetric part of the unit cell which

![Figure 1](image-url)
Figure 2. The stereo view of the Co (PDTC)$_3$ molecule. This stereo pair may be conveniently viewed to see three dimensional arrangement with some practice or using simple hand held stereo viewers.

Figure 3. Average dimensions of a pyrrolidine dithio carbamate ligand. The carbamate part is planar but the pyrrolidine rings are puckered.

is very similar but sitting with the crystallographic two fold axis passing through cobalt, CIE and NE atoms.

Each molecule is formed by three PDTC ligands coordinated to the Co atoms through their sulphur atoms. The six sulphur atoms form a distorted octahedron around the Co atom with an average Co–O distance of 2.26 Å. The distortion from regular octahedron is imposed by the molecular geometry resulting in $S_1$–Co–$S_2$ angle of 77° while $S_1$–Co–$S_3$ and $S_1$–Co–$S_4$ are 169 and 96 degrees respectively. All atoms in the PDTC ligand lie very nearly in a plane with the cobalt atom. The dithio carbamate part is mostly planar, but the pyrrolidine part is puckered. The C–N and S–C bonds are appreciably shortened and show partial double bond character. The bond lengths and angles in the five independent ligands are very similar and figure 3 shows the average dimensions for these. These bond lengths and angles do not show significant deviations from other similar complexes (Merlino 1968; Brennen and Bernal 1969; Healy and White 1972) involving both divalent and trivalent metals. This seems to indicate that for dithio carbamate ligand, the contributions of various resonance structures are not much affected by the nature of the central atoms or the crystal geometry. These resonance stabilized hybrids presumably readily react with radiation induced radicals giving the dithio carbamate the ability to protect the sensitive biological targets from effect of radiations.

The pyrrolidine ring shows some interesting variations for the different PDTC ligand molecules in the asymmetric unit. The intra ring torsion angles for the five different pyrrolidine rings are given in table 2 as well as their pseudo rotation amplitude and phase angle (Altona and Sundaralingam 1972) parameters. It must be remembered that because of the space group symmetry of the crystal there are also an equal number of
Table 2. Torsion angles of pyrrolidine rings.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_0 ) (C(^\beta)-N-C(^\alpha)-C(^\beta))</td>
<td>-1</td>
<td>-8</td>
<td>16</td>
<td>-16</td>
<td>-16</td>
</tr>
<tr>
<td>( \gamma_1 ) (N-C(^\alpha)-C(^\beta)-C(^\beta))</td>
<td>25</td>
<td>22</td>
<td>-32</td>
<td>34</td>
<td>40</td>
</tr>
<tr>
<td>( \gamma_2 ) (C(^\alpha)-C(^\alpha)-C(^\alpha)-C(^\beta))</td>
<td>-39</td>
<td>-30</td>
<td>43</td>
<td>-41</td>
<td>-48</td>
</tr>
<tr>
<td>( \gamma_3 ) (C(^\beta)-C(^\beta)-C(^\beta)-N)</td>
<td>37</td>
<td>24</td>
<td>-29</td>
<td>28</td>
<td>40</td>
</tr>
<tr>
<td>( \gamma_4 ) (C(^\alpha)-C(^\alpha)-N-C(^\alpha))</td>
<td>-23</td>
<td>-9</td>
<td>7</td>
<td>-7</td>
<td>-16</td>
</tr>
<tr>
<td>Pseudo rotation amplitude</td>
<td>40</td>
<td>29</td>
<td>40</td>
<td>40</td>
<td>49</td>
</tr>
<tr>
<td>Pseudo rotation phase</td>
<td>268</td>
<td>254</td>
<td>66</td>
<td>244</td>
<td>252</td>
</tr>
</tbody>
</table>

rings for which the torsion angles of each ring are reversed in sign and hence have a pseudo rotation phase angle which is 180 degrees in advance. Except for the ring 1 which has an envelope conformation with the C\(^\beta\) atom out of the plane formed by the other four atoms (Kartha et al. 1974), all other pyrrolidine rings are significantly twisted and have conformations similar to those observed in other studies (Druyan et al. 1976; Kartha et al. 1982) also.

The sulphur atoms of the different ligands around the cobalt atom show contacts of 3.27 to 3.33 A and correspond to some of the shortest S–S contacts reported in the literature.

Acknowledgements

We wish to thank the assistance of Mr C T Lu in diffraction data collection and Dr K Bhandary with the preparation of the diagrams. Part of the work was done by one of us (kvk) in the division of laboratories and research, New York State Department of Health in Albany. We thank New York State Department of Health for support of this project.

References

Altona C and Sundaralingam M 1972 J. Am. Chem. Soc. 94 8205
Brennen and Bernal I 1969 J. Phys. Chem. 73 443
Furnas T C and Harker D 1955 Rev. Sci. Instrum. 26 449
Gleu K and Schwab R 1950 Angew. Chem. 62A 320
Powell A K 1954 Br. J. Cancer 8 529
Electron density distribution in potassium bis-(carbonato)cuprate(II)

E N MASLEN, N SPADACCINI and K J WATSON
Crystallography Centre, University of Western Australia, Nedlands, Western Australia 6009

Abstract. The electron density distribution in potassium bis-(carbonato)cuprate(II) has been analyzed using x-ray diffraction data from an earlier structure determination. While the copper-ligand geometry is close to square planar the deformation density near the metal is strongly asymmetric. There are local maxima near the copper atom along the line of the Cu-K vectors. These resemble features found in corresponding regions in normal length metal-metal bonds. The observation is consistent with the long range nature of the Coulomb potential associated with the potassium ion.

Keywords. Coulomb potential; long range force; electron density deformation.

1. Introduction

The Coulomb potential is slowly varying at long range. For a random assembly of charges with the same sign long range interactions make a dominant contribution to the total energy because of their number. Such terms are relatively unimportant in crystalline materials because macroscopically matter is electrically neutral.

Chemical bonding refers to the equilibrium resulting from the static interaction between neutral atoms in close proximity. The overlapping of charge clouds alters the balance between attractive and repulsive forces to give a net attractive term which dominates the energy. The overlap contributions decrease rapidly with increasing internuclear separation.

The bonding interaction perturbs the atomic electron density distribution and the consequent displaced charge interacts via the Coulomb potential with other more distant neighbouring atoms. It is sometimes believed that the latter effect on the electron density should be far smaller than the first; this assumption is not consistent with the different ranges over which the contributing forces are effective. This consideration is important for example in the choice of basis functions for crystal orbital calculations and for the interpretation of deformation densities. To demonstrate the relative significance of the long range of the Coulomb potential the deformation density in a crystal structure in which the long range forces can be easily visualized has been examined. The crystal structure of potassium bis-(carbonato)cuprate(II), K₂Cu(CO₃)₂, was reported by Farrand et al (1980). It contains alkali metal atoms interacting at medium range with oxygen atoms and at longer range with a highly polarizable transition metal atom. A preliminary examination of the electron density distribution was made by Figgis et al (1981).

2. Experimental

The data from the original structural study were used in the present investigation. The intensities were corrected for absorption using the program ABSCOR and reduced to
1693 independent observations. The refinement in the earlier charge density study was based on $I$ rather than $F$ to eliminate the bias caused by setting to zero those observations for which the experimentally determined intensities were negative. This is not appropriate when the optimum values of the structure factor moduli and their associated errors are required for the calculation of reliable difference densities (French and Wilson 1978). To treat the negative observations correctly the intensities and their estimated standard deviations were modified using a Bayesian statistics approach which combines prior knowledge of the non-negativity of the true intensities with the intensity distribution determined from the data set.

The positional and thermal parameters were redetermined in a weighted, full matrix least squares refinement using all 1693 independent reflections. The form factors for $K^+$, Cu, C and O by Cromer and Mann (1968) and their respective real and imaginary anomalous dispersion corrections (Cromer and Liberman 1970) were used and the data were weighted by $1/\sigma^2(F)$. The refinement converged giving the agreement factors shown in table 1. The final structure parameters are listed in table 2. Interatomic distances and angles are given in table 3.

3. Structure

The structure of $K_2Cu(CO_3)_2$ has been discussed by Farrand et al (1980). It consists of almost square planar copper-ligand oxygen moieties linked in a polymeric array by bridging carbonate ligands. The potassium cations are interspersed through the array. Although each copper atom is on a crystallographic two-fold rotation axis the geometrical structure exhibits local four-fold symmetry about this site. The angles between the Cu-O(1) and Cu-O(2) vectors are 88.56(6)° and 92.00(6)° and the bond lengths are 1.934(1) and 1.929(1) Å respectively. The copper-ligand atom chromophore deviates slightly from planarity in a tetrahedral sense with the copper atom lying 0.06 Å above and each symmetry related pair of oxygens lying 0.14 Å above and 0.16 Å below the least squares plane through these five atoms.

The geometry of each bridging carbonate anion differs significantly from trigonal

<table>
<thead>
<tr>
<th>Table 1. Experimental and refinement data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space Group</td>
</tr>
<tr>
<td>Cell Dimensions</td>
</tr>
<tr>
<td>a</td>
</tr>
<tr>
<td>b</td>
</tr>
<tr>
<td>c</td>
</tr>
<tr>
<td>U</td>
</tr>
<tr>
<td>Z</td>
</tr>
<tr>
<td>$\mu$(MoKα)</td>
</tr>
<tr>
<td>Diffractometer</td>
</tr>
<tr>
<td>Monochromator</td>
</tr>
<tr>
<td>Maximum 2θ</td>
</tr>
<tr>
<td>Independent reflections</td>
</tr>
<tr>
<td>Reflections used</td>
</tr>
<tr>
<td>Final R factor (on $F$)</td>
</tr>
<tr>
<td>Final weighted R factor (on $F$)</td>
</tr>
<tr>
<td>Weights used</td>
</tr>
</tbody>
</table>
Electron density distribution in $K_2Cu(CO_3)_2$

Table 2. Final atomic positional parameters expressed in fractional coordinates with estimated standard deviations in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>$x/a$</th>
<th>$y/b$</th>
<th>$z/c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu</td>
<td>0(0)</td>
<td>0(0)</td>
<td>0(0)</td>
</tr>
<tr>
<td>K</td>
<td>0.57809(4)</td>
<td>0.20005(2)</td>
<td>0.19615(9)</td>
</tr>
<tr>
<td>O(1)</td>
<td>0.08278(11)</td>
<td>0.09545(7)</td>
<td>0.01308(29)</td>
</tr>
<tr>
<td>O(2)</td>
<td>0.10135(11)</td>
<td>0.19980(7)</td>
<td>0.21335(35)</td>
</tr>
<tr>
<td>O(3)</td>
<td>0.05979(13)</td>
<td>0.08876(9)</td>
<td>0.37189(31)</td>
</tr>
<tr>
<td>C</td>
<td>0.08180(14)</td>
<td>0.12671(9)</td>
<td>0.20509(39)</td>
</tr>
</tbody>
</table>

Table 3. Interatomic distances (Å) and angles (°) with estimated standard deviations in parentheses.

<table>
<thead>
<tr>
<th>Bond distances</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu-O(1)</td>
<td>1.9343(13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu-O(2)</td>
<td>1.9290(13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-O(1)</td>
<td>1.3042(29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-O(2)</td>
<td>1.3108(20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-O(3)</td>
<td>1.2514(28)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interatomic distances</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu-K$^{ii}$</td>
<td>3.8649(11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu-K$^{iii}$</td>
<td>3.4892(9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cu-K$^{iv}$</td>
<td>4.0315(11)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bond angles</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>O(1)-Cu-O(2)</td>
<td>88.56(6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O(1)-Cu-O(2)$^i$</td>
<td>92.00(6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O(1)-C-O(2)</td>
<td>116.77(20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O(1)-C-O(3)</td>
<td>121.22(16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O(2)-C-O(3)</td>
<td>121.98(22)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

i. $(1/4-x, -1/4+y, -1/4+z)$  ii. $(-1/4+x, 1/4-y, -1/4+z)$  iii. $(3/4-x, -1/4+y, 1/4+z)$  iv. $(3/4-x, -1/4+y, -3/4+z)$

planar. The bridging bonds, C-O(1) and C-O(2), are elongated relative to the free C-O(3) bond and the O(1)-C-O(2) angle is compressed from the ideal 120° to 116.8(2)°.

One pair of K cations, related by the two-fold axis, lies only 0.27 Å from the Cu-ligand oxygen plane, with the Cu-K vectors of length 3.865 Å directed midway between the respective Cu-O(1) and Cu-O(2) bonds. Four other K cations surrounding the same copper atom (two above at 3.49 Å and two below at 4.032 Å) form with the copper atom a plane perpendicular to the Cu-ligand oxygen plane. All these distances between Cu and K are short enough that the possibility of significant electrostatic interactions between Cu and K must be considered.

4. Electron density

The difference density section in the Cu-ligand oxygen plane is shown in figure 1. While the immediate structural environment of Cu has pseudo four-fold symmetry the electron density distribution shows little trace of such symmetry. There are two pairs of
local maxima within 0.8 Å of the copper atom lying between the Cu-O(1) and Cu-O(2) bonds. Each member of a pair is related to the other by the two-fold axis but the pairs do not resemble each other topologically. These features are not consistent with the nearest neighbour geometry.

The difference map features can be accounted for in qualitative terms if the effects of next-to-nearest neighbours are considered. In figure 2 the difference density in the same plane extended to 5 Å from the copper atom is shown. The broader pair of peaks near the copper are directed towards the approximately in-plane K atoms. The electron-deficient regions near the copper atom are rotated approximately 15° from the respective copper-oxygen bonds.

Figure 3 shows the section through the four potassium cations above and below the Cu-ligand oxygen plane. Sharp maxima lie directly along the shortest Cu-K vectors, with more diffuse regions of excess density along the longest Cu-K vectors. Regions of electron density excess near the potassium atoms are directed approximately towards the copper atom.

Although the results must be viewed with caution because of the phase errors associated with the structure factors of this non-centrosymmetric structure, the features in the difference maps are consistent with a medium range electrostatic interaction between the copper and alkali metal atoms. The significance of this effect has been reinforced by the results of a charge density study of the centrosymmetric sodium analogue, Na$_2$Cu(CO$_3$)$_2$ (Maslen et al 1983). Features in that analysis, which are not hampered by potential difficulties of phasing, are similar to those described above. In Na$_2$Cu(CO$_3$)$_2$, the effects of next-to-nearest neighbours on the polarization of the metal electron density is observed on a number of difference density sections.
Electron density distribution in $K_2Cu(CO_3)_2$

Figure 2. Residual density in the section containing Cu, O(1) and O(2) extended to a distance of at least 5.0 Å from Cu. Contour interval is 0.2 eÅ$^{-3}$.

Figure 3. Residual density in section perpendicular to the plane in figure 1 containing the two-fold axis and the Cu and four K atoms. Contour interval is 0.2 eÅ$^{-3}$.
5. Conclusions

It has been shown that residual density near transition metal atoms for near-to-ideal structures can be described in terms of the effects of the nearest neighbour geometry (Marumo et al 1974). In the present case, however, such a procedure is inadequate and the interpretation must be extended to correlate the density features with next-to-nearest neighbour interactions. These features resemble those observed in normal length bonds between metal atoms (Wang and Coppens 1976; Mitschler et al 1978). This suggests that the Cu-K interaction might be regarded as a weak metal-metal bond. Ito and Higashi (1983) have recently reported similar features along the Al-Al vectors in LiAlB$_{14}$. The Al-Al distance in that structure is 2.92 Å.

The assertion that in these cases nearest neighbour and next-to-nearest neighbour interactions lead to effects of comparable magnitude reflects the slowly varying nature of the Coulomb potential. It is therefore important always to consider all interactions in the interpretation of residual density maps.

Acknowledgements

The financial support of the Australian Research Grants Scheme and the Research Committee of the University of Western Australia is gratefully acknowledged. One of us (ns) acknowledges receipt of a Commonwealth Postgraduate Research Award.

References

Crystallographic study of push-pull ethylenes

D ADHIKESAVALU, NIRUPA U KAMATH and K VENKATESAN
Department of Organic Chemistry, Indian Institute of Science, Bangalore 560 012, India

Abstract. From structural data of several crystal structure analyses of push-pull ethylenes it is found that C=C bond lengths in this class of molecules are significantly longer than the value for this bond in ethylene. With powerful donors and acceptors such as -NMe$_2$ and -COOMe groups respectively, the C=C bond length is as long as 1.460 Å. Correlations are observed between (i) C=C bond lengths and the twist angles about the C=C bond and (ii) C=C bond lengths and the torsional barrier measured by the DNNR technique.

Keywords. X-ray structure analysis; push-pull ethylenes; C=C bond lengths; rotational barriers.

1. Introduction

Ethylene substituted with electron donating and electron accepting groups in the vicinal positions is known as polarized or push-pull ethylenes.

The conjugative interaction between the donor and acceptor parts of the molecule through the ethylene π bond results in interesting changes in chemical and physical properties of this class of compounds. For example, in 1-dimethyl amino-2-acetyl ethylene (figure 1) the conjugation gives rise to resonating structures (I) and (II). The C=C bond in such systems are resistant to oxidation and reduction and to the addition of bromine under normal conditions (Jensen and Henricksen 1968). Whereas the barrier to rotation about the C=C bond in 1,2-dideuterioethylene is 62 kcal mol$^{-1}$, it is substantially reduced in polarized ethylenes (Sandström and Wennerbeck 1971; Shvo and Shanan-Atidi 1969).

There are two important factors which determine the conformation about the ethylene bond in this class of molecules: (i) intramolecular steric interaction between the substituents on the double bond, and (ii) loss of π-electron energy ($E_n$) as the rotation about the C=C bond increases.

The conjugation is maximum in the planar state and as such this will favour a planar geometry. The steric interaction between the donor and acceptor groups is minimum at the dihedral angle of 90° and so, a 90° twisted conformation will be favoured.

Push-pull ethylenes can be represented by the general formula I as shown below

Figure 1. Resonance structures of 1-dimethyl amino-2-acetyl ethylene.
where $D_1$ and $D_2$ represent electron-donating groups and $A_1$ and $A_2$ electron accepting groups.

The twist about the $C=C$ double bond is measured as the angle between the plane through the atoms $D(1), D(2)$ and $C(1)$ and the plane through the atoms $A(1), A(2)$ and $C(2)$.

Although results based on NMR and other techniques have been analysed in literature (Sandström 1983), no such detailed discussion of the x-ray results has been reported. In recent years, we have been investigating the molecular structures of a large number of push-pull ethylenes via x-ray crystallography and in this paper we discuss our findings.

2. Results and discussion

In table 1 are recorded the bond lengths of the ethylene bond and the twist angles for the push-pull ethylenes investigated in our laboratory (Adhikesavalu and Venkatesan 1981, 1982, 1983a, b; Nirupa Kamath and Venkatesan 1983) and those by Abrahamsson et al (1974). The present results of the push-pull ethylenes are based on the intensity data collected on a Nonius CAD-4 diffractometer and the final $R$-factors are in the range 4-5.8% with the exception of 14 in table 1 for which the $R$-value at the present stage of refinement is around 10%. In most of the cases investigated, the samples were kindly provided by Professor J. Sandström. Also recorded in this table are the values of the rotational barrier about the $C=C$ bond measured using the DNMR technique (Sandström 1983). It may be noted that in molecules 1 to 4 (table 1), whereas the donor groups differ, the acceptor groups remain the same.

Comparing 1 and 2, on the basis of electron donating capacity, the presence of two dimethyl amino groups in 1 would be expected to lead to a longer $C=C$ length than in 2 in which the donor groups are $-SMe$ and $-NMe$ groups. But this is not borne out by the x-ray results. The structural evidence shows that, due to steric interaction the N-methyl groups are twisted by $33^\circ$ in 1 whereas in 2 it is of lower magnitude ($23^\circ$) with respect to the plane defined by the two donor atoms and the olefinic carbon atoms. The twist about the $C=C$ bond is $29-3^\circ$ in 1 and $21-7^\circ$ in 2. Comparing the molecule 1 having an acyclic donor with 3 containing a cyclic donor, we observe that the $C=C$ bond length is, within experimental error, same although there is a difference of about $9^\circ$ in the twist angle. Cyclisation in 3 and 14 has reduced the steric interaction between the donor and acceptor parts of the molecule. Similarly, the larger twist in 4 ($31-5^\circ$) compared to that in 3 ($20-2^\circ$) and 14 ($22-9^\circ$) is due to the greater steric interaction in the planar state in 4. The average value of the exocyclic angle (i.e. $C_2-C_1-D_1$ and $C_2-C_1-D_2$) in 4 is significantly smaller ($120-6(5)^\circ$) than in 3 ($124-8(2)^\circ$). However, the significantly longer $C=C$ bond length in 4 (1.429(6) $\AA$) compared to that in 3 (1.407(3) $\AA$) cannot be due to the difference of about $11^\circ$ in twist angles. In this connection it is noteworthy that the difference in the twist angles between 5 and 7 is about $16^\circ$ (table 1) but there is no significant difference in the $C=C$ bond length. These observations seem to indicate that the hexahydropyrimidine ring possesses better electron donating ability than the five-membered imidazolidine ring.
### Table 1.

<table>
<thead>
<tr>
<th>S.No. &amp; Reference code</th>
<th>Structural formula</th>
<th>Rotational barrier about the C=C bond (kcal mol$^{-1}$)</th>
<th>Twist about the C=C bond ($^\circ$)</th>
<th>C=C bond length (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Me$_2$N /CN</td>
<td>29.3</td>
<td>1.407 (4)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Me$_5$ /CN</td>
<td>13.2</td>
<td>21.7</td>
<td>1.397 (8)</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>20.2</td>
<td>1.407 (3)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>31.5</td>
<td>1.429 (6)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Me$_2$N /CN</td>
<td>&lt;8</td>
<td>56.8</td>
<td>1.461 (2)</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>5.0</td>
<td>1.451 (3)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>&lt;8</td>
<td>72.9</td>
<td>1.468 (3)</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>&lt;8</td>
<td>62.6</td>
<td>1.484 (2)</td>
</tr>
<tr>
<td>9</td>
<td>Me$_5$ /CN</td>
<td>38.3</td>
<td>1.414 (4)</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>9.5</td>
<td>41.0</td>
<td>1.448 (4)</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>19.0</td>
<td>15.0</td>
<td>1.386 (7)</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>19.1</td>
<td>12.8</td>
<td>1.377 (3)</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>9.8</td>
<td>34.8</td>
<td>1.409 (4)</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>9.5</td>
<td>22.9</td>
<td>1.415 (8)</td>
</tr>
</tbody>
</table>

E.S.d's are given in parenthesis. References: X-ray work on 1-4, 6-9 (Adhikesavalu and Venkatesan); 5, 12-14 (Nirupa Kamath and Venkatesan). Rotational barriers on 5, 7, 8, 13 (Wennerbeck and Sandström 1972); 10, 11 (Abrahamsson et al 1974); 12 (Dreier et al 1978); 14 (Sandström et al 1977); 15 (Sandström and Wennerbeck 1978).
The barrier to rotation in 1 could not be measured due to symmetry reasons (Kessler 1968). Also, the barriers to rotation for 3 and 4 have not been reported. But, for similar systems the barrier is reported to be less than 10 kca\(\text{l}\)/mol (Sandströ\(\text{m}\) and Wennerbeck 1971). However, NMR studies have shown that the energy barrier about the C=C bond is less in molecules with electron donating groups present as part of a ring than in those with acyclic systems. The barrier to rotation in 2 seems to indicate that the transition state is stabilized better in this molecule compared to 1.

The combined effect of the powerful electron donating \(-\text{NMe}_2\) group and the strong acceptor combination of acetyl and carbomethoxy groups in 5 to 8 produce a remarkable lengthening of the C=C bond (1.460(3) Å). The slightly smaller value of C=C bond length in 6 (1.451(3) Å) compared to the values obtained for 5, 7 and 8 may be attributed to the considerably smaller angle of twist in 6. It is pertinent to observe that the theoretical calculations by \textsc{mindo}/2 method (Dewar and Haselbach 1970) have shown that the C=C bond length extension in ethylene, on twisting about the C=C bond by 90° is only 0.03 Å. The C=C bond length calculated for 90° twisted state is 1.367 Å compared to the normal value of 1.336(2) Å (Kuchitsu 1966).

There would be severe steric interaction involving \(-\text{NMe}_2\) groups at the donor end and the acetyl groups at the acceptor end in the planar state in 7 and 8. To relieve this strain, the molecules are twisted 72.9° in 7 and 62.6° in 8 about the C=C bond. The carbomethoxy group introduces less steric strain compared to an acetyl group and the lesser twist in 8 may be due to this. In the absence of the \(-\text{NMe}_2\) group, we should expect the twist angle to be less, but the unexpectedly low value of 5° in molecule 6 is attributed to the two intramolecular N-H \ldots O bonds (figure 2). In fact the twist angle corresponding to the minimum energy conformation, calculated using semi-empirical methods (Ramachandran and Sasisekharan 1968), is 22°. However, when the contribution of the hydrogen bonding energy is introduced in the conformational energy calculations, the conformation with a twist angle of 8° is favoured.

$^1\text{H NMR}$ spectra of 7 under conditions of slow exchange shows two forms, one with diastereotopic and the other with homotopic \(-\text{COMe}\) protons. The first must be

![Figure 2. Intramolecular N-H \ldots O bonding in 6.](image-url)
degenerate $EZ-ZE$ pair, whereas the second may be $EE$ or $ZZ$ (figure 3). But the $EE$ form is less likely because of the strong repulsion to be expected between parallel dipoles as shown by the effect of solvent polarity and in particular by the  $asis$ (Sandström 1983).

In the solid state, 6 adopts the $ZZ$ conformation favoured by the intramolecular hydrogen bonding. In 5, 7 and 8 the acceptor groups assume $EZ$ conformation.

The alkylthio groups have been found to be poor donors and this is clearly seen in the shorter $C=C$ bond length ($1.369(7)$ Å) in 11. Comparison of the $C=C$ lengths obtained for 9 and 11, shows again that the thio group is a poor donor. As expected then, the ketene mercaptals being weak donors, the bond length in 12 is $1.377(3)$ Å. In both cases the rotational barrier has been measured and is of the order of 19 kcal mol$^{-1}$.

The combination of an acetyl and phenyl group in 13 as acceptor groups does not seem to produce the same degree of delocalization as the acetyl-carbomethoxy combination does as, for example, observed in 5 with the same donor substituents in both the molecules.

3. Conclusions

From the present x-ray crystallographic results the following significant conclusions may be drawn:

(i) The larger $C=C$ bond length observed in molecules 5 to 8 compared to those of 1 to 4 clearly shows that the $-\text{COMe}$, $-\text{COMe}$ or $-\text{COMe}$, $-\text{COOMe}$ substituents as acceptors are definitely better than the $-\text{CN}$ group as acceptors. Further, as noted

![Figure 3. Conformational possibilities of acetyl group.](image)
earlier, the hexahydropyrimidine ring appears to be a better donor than the imidazolidine ring.

(ii) In figure 4, the C=C bond lengths are plotted against the observed twist angle about the C=C bond. It is observed that the nature of the correlation for the molecules in which the donors are acyclic is different from that for those with cyclic donors. For the same angle of rotation, the C=C bond length tends to be longer in the molecules with cyclic donors than the ones with acyclic donors. However, for angles of rotation above 55°, the C=C bond lengths tend to be very nearly equal in both the cases. The deviation of molecule 6 from the curve (b) arises from the intramolecular N-H...O hydrogen bonding in this molecule. The observed difference depicted in figure 4 with acyclic (a) and cyclic (b) donors could be partly due to the difference in steric factors between these two classes of molecules although it is known from theoretical calculations (Dewar and Haselbach 1970) that the variation expected in C=C bond length from the twist about this bond is less than 0.03 Å. Thus the importance of the possible difference in the electronic factor between the cyclic and acyclic donors cannot be overruled.

(iii) It is reasonable to expect a correlation between \(d(C\text{sp}^2-N)\) and \(d(C=C)\). However, it is found from the scatter plot of C=C and C(sp^3)-N, that there is a correlation corresponding to cyclic donors molecules but not when the donors are acyclic (figure 5). In the molecules with acyclic donors such as -NMe_2 and -SMe, there is rotation about N-Me and S-Me bonds in order to overcome the steric interaction between the donor and acceptor groups. However, this is not possible to the same degree on the molecules with cyclic donors. This provides an explanation for the absence of correlation in molecules with acyclic donors.

(iv) It was pointed out earlier that the rotational barrier about the C=C bond in
Crystallographic study of push-pull ethylenes

Figure 5. Correlation plot of $d(Cp^2=N)$ vs $d(C=C)\ (\AA)$.

Figure 6. Plot of rotational barrier (kcal mol$^{-1}$) vs C=C bond length (Å).

polarized ethylenes is substantially reduced. From the x-ray results reported here it is seen that the C=C bond length can be as large as 1.460 Å. The rotational barrier value may be expected to decrease as the C=C bond length increases. The rotational barrier values for quite a few cases are not available from the NMR data either because the low torsional barriers of C=C double bond are sometimes below the limit for NMR technique ca 5 kcal mole$^{-1}$ for these type of compounds or due to experimental difficulties.

In figure 6 are plotted C=C bond lengths vs the rotational barriers. The donor and acceptor groups being different in these molecules, the intramolecular steric strains will be different. However, there is a remarkable correlation between C=C bond length and rotational barrier.
(v) If we make the reasonable assumption that the electronic effects are more important in lengthening the C=C bond, then from the crystallographic data recorded in table 1 for molecules 3, 6, 7, 8 and 10 we arrive at the conclusion that the electron withdrawing power of groups are in the order

\[-\text{COMe} \approx -\text{COOMe} > \text{Ph-C}=\text{O} > \text{CN}\.

(vi) It has been shown from a feasibility study (Trueblood and Dunitz 1983) that from the internal molecular motions in crystals an estimation of force constants, frequencies and barriers from diffraction data can be obtained. The rotational barriers from NMR measurements are available (Sandström 1983) for many of the polarized ethylenes and the x-ray data are based on diffractometric measurements. It was, therefore, tempting to use the thermal motion parameters for the calculation of the rotational barriers. From the point of view of Hirshfeld’s (1976) ‘rigid-bond’ test only three cases (2, 5, 13 in table 1) appeared reasonable for the estimation of rotational barriers. But it was very disconcerting to note that the rotational barrier values obtained even for these were widely different from those obtained from the NMR measurements. Although crystal environment may be responsible to some degree for the discrepancy it is clear that far more accurate intensity data than we have now are required to recover the pin from the haystack.

**Acknowledgements**

The authors thank the CSIR and the UGC for financial support. They are grateful to Prof. J Sandström, for many valuable comments.

**References**


Adhikesavalu D and Venkatesan K 1982 *Acta Crystallogr.* B38 855


Hirshfeld F L 1976 *Acta Crystallogr.* A32 239


Kessler H 1968 *Angew. Chem. Inter. Edit.* 7, No. 11 898

Kuchitsu K 1966 *J. Chem. Phys.* 44 906

Nirupa Kamath U and Venkatesan K 1983 (Unpublished results)


Sandström J 1983 in *Topics in stereochemistry*, (eds.) L Norman, Allinger, Ernest L Eliel and Samuel H Wilen Vol 14


Structural evidence for the radical scavenging mechanism of some aminothiol group of radioprotectants

KALYANI VIJAYAN, B M VEDAVATHI and A MANI
Materials Science Division, National Aeronautical Laboratory, Bangalore 560017, India

Abstract. Aminothiols constitute an important group of radioprotectants. The structures of a few well-known compounds belonging to the family of radioprotectants have been determined by single crystal x-ray diffraction methods. The sulphur and the amino nitrogen atoms are separated by two tetrahedral carbon atoms in these compounds. The gauche conformation of the sulphur and the nitrogen atoms and the consequent non-bonded intramolecular S...N interaction observed in some of the crystal structures appear to favour the hypothesis that the protective mechanism of these compounds is by free radical scavenging.

Keywords. Radioprotectants; aminothiols; crystal structure; radical scavenging; S...N interaction.

1. Introduction

It is well-known that on exposure to ionising radiation, biological systems incur severe damages affecting their properties and functions. Dale (1942) showed that the hazardous effects of radiation on biological systems may be reduced by administering some chemicals prior to exposure. Since then, a large number of such chemicals known as 'radioprotectants' have been identified. With the rapid advances in space technology, chemical radioprotection has gained importance and extensive research is being carried out to understand the mechanism of action of chemical radioprotectants. Some of the well-known theories of radioprotection involve (i) the method of radical scavenging (ii) formation of mixed disulphide (iii) production of tissue hypoxia (iv) donation of hydrogen atoms (v) binding or chelating to the metal ions in the biological system and (vi) introduction of energy traps. An essential pre-requisite for understanding the mechanism of radioprotection is the study of molecular geometry of the radioprotectants. In this context, we have analysed the structures of a few radioprotectants belonging to the aminothiol family by single crystal x-ray diffraction methods. The results obtained are critically examined with reference to a particular theory of radioprotection.

2. Aminothiol radioprotectants

Sulphur atoms occurring in the compounds in the aminothiol family could belong to a -SH group or to a disulphide bond or could be bonded to two carbon atoms. It is, however, believed that the non-sulphydryl compounds are converted into the appropriate sulphydryl forms on administration and that a free sulphydryl group is essential for the radioprotective action (Shapira et al 1957; Doherty et al 1957a; Schwartz and Shapiro 1960; Doherty and Shapiro 1963). Most aminothiol radioprotectants contain one or more amino groups, but they could be replaced by guanidyl groups...
without loss of activity. In fact, \( \beta \)-mercapto ethyl guanidine, a typical radioprotectant is known to be more active than \( \beta \)-mercapto ethylamine (Shapira et al 1957). The aminothiols have been known to possess a high degree of structural specificity with regard to their radioprotective property. In particular, the relative orientation of the amino (or guanidyl) and the thiol groups is believed to be important in their mechanism of action.

3. X-ray results

The bromide hydrobromide of S, 2-aminoethyl-isothiouronium (AET) (I), is an effective radioprotectant belonging to the aminothiol group. A perspective view of the AET molecule as derived from the x-ray analysis is given in figure 1 (Vijayan and Mani 1977). The terminal amino group and the sulphur atom in the AET molecule are trans about the connecting C–C bond. This conformation eliminates the possibility of any intramolecular S . . . N interaction.

When administered, AET is known to undergo the chemical changes shown in figure 2.

![Figure 1. A perspective view of the AET molecule.](image)

![Figure 2. Formation of 2-AT and GED from AET.](image)
The two end products, *viz*, 2-aminothiazoline (IV) (2-AT) and bis (2-guanidinoethyl) disulphide (V) (GED), are obtained *via* the cyclic intermediate II, which could either open to form mercaptoethyl guanidine (III), (MEG), at pH 7 or split off ammonia to form 2-AT at pH 2-5. At pH 7 and above, MEG is unstable and is readily oxidized by air to form the disulphide GED. GED, the crystal structure of which has been analysed (Vedavathi and Vijayan, to be published) is also known to protect against ionising radiation (Distefano *et al* 1959; Schwartz and Shapiro 1960).

A view of the GED molecule as seen along the S–S bond is given in figure 3. The two halves of the GED molecule, though chemically identical, have strikingly different structural characteristics. The sulphur and the nitrogen atoms S(2) and N(4) in one half of the GED molecule are gauche, leading to an intramolecular non-bonded S...N distance of 3.46 Å. However, in the other half of the molecule, the corresponding sulphur and nitrogen atoms S(1) and N(1) are trans about the connecting C–C bond.

Cystamine (VI) is obtained by reduction of the well-known aminothiol radioprotect-

\[
\begin{align*}
\text{S} &- \text{CH}_2\text{CH}_2\text{NH}_2 \\
\text{S} &- \text{CH}_2\text{CH}_2\text{NH}_2
\end{align*}
\]

ant cysteamine or β-mercapto ethylamine (MEA). Cystamine di HCl, the crystal structure of which has been studied (Vedavathi and Vijayan 1979), is known to provide the same

![Figure 3. A view of the GED molecule as seen along the disulphide bond.](image-url)
order of radioprotection as cysteamine itself (Doherty et al 1957a). The cystamine molecule is essentially two ethylamine groups connected by a central disulphide bond. The only difference between the chemical structures of cystamine and ged is the replacement of amino groups in the former by the guanidine groups in the latter. A perspective view of the cystamine molecule, as seen along the disulphide bond, is given in figure 4. As in the ged molecule, the chemically identical halves of the cystamine molecule have different structural characteristics. In one half, the sulphur and the nitrogen atoms, S(1) and N(1) are gauche leading to an intramolecular S ... N distance of 3.22 Å. In the other half of the molecule, the S(2) and the N(2) atoms are trans, eliminating any possible S ... N interaction.

As shown in figure 2, one of the end products obtained from aet is 2-at. Even though 2-at is known to be the toxic component, it has also been found to offer radioprotection (Distefano et al 1959; Shapira et al 1957). The crystal structure of the hydrochloride of 2-at has also been determined (Vedavathi and Vijayan 1981). However, the structural characteristics of 2-at are not relevant to the present discussion.

4. Discussion

4.1 S ... N interaction in the crystal structures

The striking, common features in the crystal structures of ged Br • HBr and cystamine di HCl are the gauche conformation of the sulphur and the nitrogen atoms and the consequent non-bonded S ... N interaction, observed in one half of the molecules. Similar gauche conformation and a non-bonded S ... N interaction have also been observed in the crystal structure of another aminothiol radioprotectant, cysteamine mono hydrochloride (mea • HCl), (Jandacek and Swartz 1970). The repeated occurrence of the non-bonded S ... N interaction in the three radioprotectants strongly suggests that these features may be related to their mechanism of protection.

4.2 Radical scavenging mechanism and S ... N interaction

The radical scavenging mechanism is based on the indirect action theory of radiation. It
is well-known that radiolysis of water produces the radicals $H^+$, $OH^-$ and $OOH^-$, which are free to interact with the biological system exposed to radiation and cause radiation damage. However, according to the radical scavenging theory, when a chemical radioprotectant such as an aminothiol is included in the system exposed to radiation, the radioprotectant preferentially competes for the radiation induced radicals from water, before the free radicals enter into any further action.

Doherty and Shapira (1957) have studied the radioprotective property of several aminothiols and have put forth the hypothesis that both cysteamine and cystamine are protective by virtue of their interaction with the radiation induced free radicals leading to the formation of resonance stabilised structures. According to them, the radical scavenging action of the aminothiol, say, cysteamine, \( \text{MEA} \), takes place as follows. The radioprotectant interacts with the free radicals $HOO^-$ and $OH^-$ to produce the radicals of $\text{MEA}$, as shown in (1).

\[
\begin{align*}
\text{MEA} + HOO^- & \rightarrow \text{MEA}^- + H_2O_2, \\
\text{MEA} + OH^- & \rightarrow \text{MEA}^- + H_2O.
\end{align*}
\]

(1)

The $\text{MEA}^-$ radicals may enter into further interaction with the free radicals from water and regenerate into $\text{MEA}$ itself (2).

\[
\begin{align*}
\text{MEA}^- + HOO^- & \rightarrow \text{MEA} + O_2, \\
\text{MEA}^- + OH^- & \rightarrow \text{MEA} + O.
\end{align*}
\]

(2)

The resonance structures of the radicals described in (1) are shown in figure 5.

It is obvious that in this mechanism of radical scavenging, the stability of the resonance hybrids is an important factor. This stability arises mainly from the interaction between the amino and the thiol groups, indicated by dotted lines in figure 5. The gauche conformation and the S . . . N interaction observed in the crystal structures of cystamine, cysteamine and GED appear to provide evidence in support of the formation of resonance stabilised structures and thus favour the hypothesis that these compounds act by the radical scavenging mechanism.

The stability of the resonance hybrids is also controlled by the length of the carbon chain between the amino group and the sulphur atom. In cystamine, cysteamine and GED, there are only two carbon atoms between the sulphur and the nitrogen atoms thus leading to the formation of a pseudo-five-membered ring which is known to be structurally stable.

Absence of the intramolecular S . . . N interaction in the crystal structure of \( \text{AET} \), a very effective aminothiol group of radioprotectant, is, however, consistent with the results suggesting that \( \text{AET} \), as such, is inactive and when administered, gets converted to its active form, mercaptoethyl guanidine, \( \text{MEG} \).

![Figure 5. Resonance structures of the \( \text{MEA} \) radical.](image-url)
Acknowledgement

The authors wish to thank Dr S Ramaseshan for providing the initial impetus and encouragement.

References

Dale W M 1942 Biochem. J. 36 80
Doherty D G, Burnett W T and Shapira R 1957a Rad. Res. 7 13
Doherty D G and Shapira R 1963 J. Org. Chem. 28 1339
Doherty D G, Shapira R and Burnett W T Jr. 1957b J. Am. Chem. Soc. 79 5667
Khym J X, Shapira R and Doherty D G 1957 J. Am. Chem. Soc. 79 5663
Schwartz E E and Shapiro B 1960 Rad. Res. 13 768
Shapira R, Doherty D G and Burnett W T Jr. 1957 Rad. Res. 7 22
Vedavathi B M and Vijayan K 1979 Curr. Sci. 48 1028
Vedavathi B M and Vijayan K (to be published)
Structure and interactions of aspirin-like drugs

M VIJAYAN
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

Abstract. A pre-requisite for the elucidation of the mechanism of action of aspirin-like drugs, which are believed to exert their pharmacological effects through the inhibition of prostaglandin biosynthesis, is an understanding of their molecular geometry, the non-covalent interactions they are likely to be involved in, and the geometrical and the electronic consequences of such interactions. This has been sought to be achieved through the x-ray analysis of these drug molecules and their crystalline complexes with other suitable molecules. The results obtained from such studies have been discussed in terms of specific typical examples. For instance, antipyrine can form metal and hydrogen-bonded complexes; phenylbutazone can form ionic complexes with basic molecules. Complex formation is accompanied by characteristic changes in the molecular geometry and the electronic structure in both the cases. The results obtained so far appear to indicate that the important common invariant structural features of the fenamates, deduced from crystal structures, are retained even when complexation takes place.

Keywords. Aspirin-like drugs; prostaglandin biosynthesis; non-covalent interactions; crystalline complexes; pyrazole derivatives; fenamates.

1. Introduction

Nonsteroid anti-inflammatory analgesics like salicylates, aniline derivatives, pyrazole derivatives and fenamates, which are effective in varying degrees against pain, inflammation and fever, have been among the most commonly used drugs for several decades. Hardly anything was known about the molecular mechanism of their pharmacological action until the early seventies when it was discovered that all these drugs, despite their widely different chemical structures, inhibit prostaglandin biosynthesis (Vane 1971). They are now believed to act through the inhibition of the system of enzymes responsible for the biosynthesis of prostaglandins (Flower 1974). In this context, a pre-requisite for the elucidation of the molecular details of the action of these compounds is a thorough understanding of (a) their molecular geometry, (b) the non-covalent interactions they are likely to be involved in and (c) the changes introduced in their molecular and electronic structure as a result of these interactions. One of the x-ray crystallographic projects in our laboratory seeks to achieve this through the x-ray analysis of these drug molecules as well as their crystalline complexes with other suitable molecules. The focus of attention in these studies has been on pyrazole derivatives and fenamates, and the structures analysed include drugs such as antipyrine (Singh and Vijayan 1973), amidopyrine (Singh and Vijayan 1976), metamizol (analgin or novalgin) (Krishna Murthy et al 1979), phenylbutazone (Singh and Vijayan 1977), oxyphenbutazone (Krishna Murthy and Vijayan 1981a), niflumic acid (Krishna Murthy and Vijayan 1979), meclofenamic acid (Krishna Murthy and Vijayan 1981b), flufenamic acid (Krishna Murthy et al 1982) and some of their complexes (Singh and Vijayan 1974, 1977; Dhanaraj and Vijayan 1983). These analyses have produced a wealth of information on the structure and interactions of the concerned drug
molecules, but we shall confine ourselves here to a few representative results to illustrate the type of information that could be obtained from studies of this nature.

2. Pyrazole derivatives

2.1 Pyrazolone derivatives

Pyrazole derivatives with aspirin-like action can be classified into two groups, namely, pyrazolone derivatives and pyrazolidinedione derivatives. Antipyrine (1-phenyl-2,3-dimethyl-5-pyrazolone) (I) was the first pyrazolone derivative to be introduced into medical practice, and the better known amidopyrine and metamizol (analgin or novalgin) are its derivatives. It is therefore perhaps appropriate to take antipyrine as an example of the drugs in this family.

The molecular structure of antipyrine, as derived from the x-ray analysis of its crystals, is shown in figure 1 (Singh and Vijayan 1973). The five-membered and the six-membered rings, connected through a N–C bond, are inclined with respect to each other at 52°. The two nitrogen atoms are pyramidal, and the phenyl and the methyl groups attached to them lie on the opposite sides of the plane of the five-membered
Structure and interactions of aspirin-like drugs

(A)

(B)

(C)

Figure 2. Canonical structures of antipyrine.

ring. To a good approximation, the molecule can be considered to be a resonance hybrid of the three canonical forms shown in figure 2. The observed bond lengths in the molecule can be satisfactorily explained, using standard bond length-bond order curves, if the contributions of A, B and C are taken to be 66, 22 and 12% respectively.

The crystal structures of a number of metal complexes of antipyrine are available and coordination is achieved in all of them through the carbonyl oxygen atom (Vijayan and Viswamitra 1966, 1967, 1968; Cingi et al. 1972; Brassy, et al. 1974a, b, c, d). The bond lengths and angles in the pyrazolone moiety in the metal complexes have similar values in general. But they differ systematically from those found in free antipyrine. It turns out that the observed bond lengths in the metal complexes can be accounted for if the contributions from the canonical forms A, B and C (figure 2) are taken to be 41, 37 and 22% respectively. Thus, while the neutral form predominates in free antipyrine, the polar forms do in the metal complexes. Another difference between the molecular geometries of antipyrine in the uncomplexed state and in the metal complexes pertains to the hybridisation state of the hetero nitrogen atoms. The two nitrogen atoms are more pyramidal in the free molecule than in the metal complexes. Consequently, the geometry of the molecule is substantially different in the two cases. This is illustrated in figure 3 which shows the uncomplexed molecule as well as that in a metal complex, viewed along comparable directions. This represents an interesting situation in which a perturbation at one end of the molecule, namely, association of a metal ion at the carbonyl oxygen atom, induces striking changes in its overall geometry.

In addition to metal complexes, antipyrine is known to form hydrogen-bonded complexes as well. The 1:1 complex between antipyrine and salicylic acid (generally

Figure 3. Two views of the molecule in free antipyrine and in its metal complexes along comparable directions (Singh and Vijayan 1973). Solid and broken lines correspond to free antipyrine and calcium hexa-antipyrine perchlorate respectively.
known as salipyrine) is typical of the latter. The structure of this complex, determined by x-ray crystal structure analysis, is shown in figure 4 (Singh and Vijayan 1974). Complexation is achieved through an intermolecular O–H⋯O hydrogen bond with a carboxyl oxygen of the salicylic acid molecule as the donor and the carbonyl oxygen of the antipyrine molecule as the acceptor. In the antipyrine molecule in the complex, the pyramidal nature of the hetero nitrogen atoms is intermediate between those observed in free antipyrine and in the metal complexes. The bond lengths in the pyrazolone moiety in the complex indicated the contributions of forms A, B and C to be 45, 34 and 21 %, respectively, suggesting thereby that the antipyrine molecule in salipyrine is more polar than in free antipyrine but less so compared to the molecule in the metal complexes.

It has been possible to define the molecular geometry of antipyrine in different states of association, from the studies outlined above. The molecule can take part in metal coordination or hydrogen bonding. Association is achieved by an interaction at the carbonyl group in both the cases and its effect is (i) to make the hetero nitrogen atoms more planar and consequently to introduce substantial changes in molecular geometry; and (ii) to increase the polar nature of the molecule. The interaction of a metal ion with the carbonyl group is expected to be stronger than that arising from a hydrogen bond and hence the planarity of the nitrogen atoms and the polar nature of the molecule are more pronounced in metal complexes than in salipyrine.

2.2 Pyrazolidinedione derivatives

Phenylbutazone (4-butyl-1,2-diphenyl-3,5-pyrazolidine-dione) (II) and its hydroxy
derivative oxyphenbutazone are the best known pyrazolidinediones with aspirin-like activity. The three-dimensional structures of the two crystallographically-independent molecules in the crystals of phenylbutazone are shown in figure 5 (Singh and Vijayan 1977). The two molecules have essentially the same structure except for the differences in the conformation of the butyl group. The two nitrogen atoms in the five-membered ring are pyramidal and the attached phenyl rings lie on opposite sides of the plane of the pyrazole ring. But for the presence of the butyl group, the molecule could have had two-fold symmetry about an axis passing through the tetrahedral carbon at the 4-position and the mid-point of the N–N bond. This symmetry is, however, prevented by the substitution of the butyl group at the tetrahedral carbon atom. The environments of the two phenyl groups are also now dissimilar and their orientations with respect to the five-membered ring are different.

The phenylbutazone molecule, being acidic, may be expected to interact with basic amino acid residues. Attempts were made to co-crystallize it with basic amino acids and their derivatives, but these attempts were not successful. It was, however, possible to prepare a 2:1 crystalline complex of phenylbutazone with the basic compound, piperazine. The ring carbon atom C(4) in phenylbutazone is deprotonated when complexation takes place. The structure of the resulting anionic molecule is shown in figure 6 (Singh and Vijayan 1977). The complex is stabilized by ionic and hydrogen bonded interactions with the cationic piperazine molecules. There are significant structural differences between the neutral phenylbutazone molecule and the deprotonated anionic molecule in the complex. The carbon atom C(4) is tetrahedral in the former whereas it is trigonal in the latter resulting in different orientations of the butyl group in the two cases. The excess electron in the anionic molecule is delocalised over the two C–O and the two C–C bonds whereas these bonds are double and single respectively in the neutral molecule, as can be seen from the bond lengths given in figure 7. As C(4) is sp² in the anion, the C(4)–C(7) bond lies in the plane of the five-membered ring and the butyl group has no preferred orientation with respect to this plane; therefore, unlike in the case of free phenylbutazone, the two phenyl rings are
Figure 6. Molecular geometry of the anionic phenylbutazone in its complex with piperazine (Singh and Vijayan 1977).

Figure 7. Bond lengths of the pyrazolidinedione moiety in the crystals of phenylbutazone and those of the phenylbutazone-piperazine complex.

structurally indistinguishable and are oriented with respect to the five-membered ring at nearly equal angles. Therefore, the two-fold symmetry of the molecule is almost restored on account of the deprotonation of C(4).

The results outlined above, which have been corroborated by those of the x-ray analysis of 1:1 complex between phenylbutazone and N-methyl piperazine (Toussaint et al 1974), indicate that C(4) is deprotonated, with accompanying changes in the molecule, when phenylbutazone interacts with a basic molecule or chemical group. Obviously then, deprotonation of the phenylbutazone molecule is likely to occur when it approaches carrier proteins or the prostaglandin synthetase system as a result of its interaction with basic amino acid residues. The consequent characteristic changes in the geometry and the electronic structure of the molecule is likely to be of importance in its protein binding properties and its inhibitory action against prostaglandin synthetase.
2.3 Fenamates

Fenamates (figure 8) form an important recently developed group of non-steroidal analgesics. The molecules of this family of compounds essentially consist of three planar groupings, namely, two six-membered rings and a carboxyl group. The important members of this family are mefenamic acid, flufenamic acid, meclofenamic acid and niflumic acid. The crystal structures of all these compounds have been determined (Krishna Murthy and Vijayan 1979, 1981a; Krishna Murthy et al 1982; McConnell 1973; McConnell and Company 1976). A striking common feature of all these structures is the coplanarity of the six-membered ring carrying the carboxyl group, the carboxyl group and the imino nitrogen atom. The internal N–H...O hydrogen bond with the imino nitrogen as the donor and one of the carboxyl oxygens as the acceptor, also exists in all the structures. This hydrogen bond, along with resonance interactions between the six-membered ring and the carboxyl group, presumably stabilise the coplanar arrangement. The variations in the geometry of the molecules arise almost exclusively on account of the differences in the orientation of the second six-membered ring with respect to the rigid coplanar moiety.

The crystal structure of a 1:1 complex between one of the fenamates, namely, niflumic acid, and ethanolamine has been determined recently (Dhanaraj and Vijayan 1983). As can be seen from the perspective view of the complex, shown in figure 9, the interactions between the two molecules involve primarily, though not exclusively, the deprotonated negatively charged carboxylate group of niflumic acid and the protonated positively charged amino group of ethanolamine. It is interesting to note that the invariant structural features of the fenamates observed in their crystal structures are retained in the complex also.

The work on fenamates, particularly that involving complexes, is still in progress. From the results obtained so far, it would appear that the only invariant features of the fenamates is the rigid coplanar geometry of the six-membered ring carrying the carboxyl group, the carboxyl group and the imino nitrogen atom, and also the internal hydrogen bonding interactions.
hydrogen bond connecting the imino and the carboxyl groups. The structure of the crystalline complex of niflumic acid with ethanolamine indicates that they are perhaps retained even when the fenamates interact with other molecules. It appears to be reasonable to assume that these invariant features are a requirement for their pharmacological action.

4. Conclusion

The approach involving crystalline complexes, which has been recognised to be of considerable utility in elucidating biologically and evolutionarily significant non-covalent interactions involving amino acids and peptides (Vijayan 1980; Sudhakar et al 1980; Salunke and Vijayan 1981; Vijayan 1983; Suresh and Vijayan 1983a, b) has been found to be useful in studying the structure and interactions of drug molecules as well. Using this approach, it has been possible to elucidate the molecular geometry of several pyrazole derivatives and fenamates with aspirin-like activity, in the free state as well as in states of association with other molecules. This approach has also provided valuable insights into the nature of the non-covalent interactions these drug molecules are likely to be involved in, and the geometric and the electronic consequences of such interactions. These insights are valuable in elucidating the molecular mechanism of action of aspirin-like drugs which are believed to exert their pharmacological effects through the inhibition of prostaglandin biosynthesis.

Acknowledgement

The work discussed here has been carried out in collaboration with the author’s past and present graduate students Dr T P Singh, Dr H M Krishna Murthy, Dr T N Bhat and Mr V Dhanaraj. He thanks them all. Financial support from the University Grants Commission and the Department of Science and Technology, India, is thankfully acknowledged. The author’s thanks are also due to Dr D M Salunke and Mrs N Nagaprabhavathy for help in the preparation of this paper.

References

Flower R J 1974 Pharmacol. Rev. 26 33
Salunke D M and Vijayan M 1981 Int. J. Peptide Protein Res. 18 348
Suresh C G and Vijayan M 1983a Int. J. Peptide Protein Res. 21 223; and the references therein
Suresh C G and Vijayan M 1983b Int. J. Peptide Protein Res. 22 129
Vane J R 1971 Nature New Biology 231 232
Vijayan M 1980 FEBS Lett. 112 135
Vijayan M and Viswamitra M A 1966 Acta Crystallogr. 21 522
Vijayan M and Viswamitra M A 1967 Acta Crystallogr. 23 1000
Structural changes in the monomeric despentapeptide (B30-B26) insulin crystal

RU CHANG BI*, ZBIGNIEW DAUTER+, ELEANOR DODSON, GUY DODSON, FEDERICO GORDIANO+, ROD HUBBARD, and COLIN REYNOLDS

Department of Chemistry, University of York, Heslington, York, YO1 5DD, England.
Universita di Napoli, Via Mezzocannone 4, 80134 Napoli, Italy.
* Permanent address: Institute of Biophysics, Academia Sinica, Peking, China.
+ Permanent address: Department of Biochemistry, Technical University of Gdansk, Gdansk, Poland.

Abstract. Insulin modified by the removal of its B chain C terminal residues is monomeric but remains substantially potent. The crystal structures of the beef and insulin (dpi) with two molecules in the asymmetric unit has been determined by x-ray analysis. The 3-dimensional structure of dpi proves to be generally similar to that of native molecule in 2 Zn insulin. More detailed comparison reveals that the slight differences in the two independent molecules of beef dpi are distributed uniformly throughout the structure in contrast to insulin in 2 Zn insulin, where the structural changes are concentrated in specific regions.

The loss of symmetry in the dpi crystal appears to be the inability of the A9 serine to pack effectively in the C2 cell. The efficient packing of the sheep dpi molecule whose crystal structure has also been determined and where A9 is glycine supports this conclusion.

Keywords. Insulin; despentapeptide insulin; crystal structure; monomer.

1. Introduction

Insulin molecule has been at the centre of biochemical research ever since its discovery in the early 1920s. Its medical importance in controlling sugar levels in the blood and alleviating diabetes has stimulated study of its chemical properties, its behaviour in solution and its biological effects. The molecule’s sequence, which was determined by Sanger and his colleagues, provided the first proper chemical description of proteins (Ryle et al 1955). Research into the hormone’s crystal structure was begun by Dorothy Hodgkin (then Crowfoot) in 1935 and consequently insulin crystals were the subject of the earliest protein crystallographic investigations (Hodgkin and Riley 1968).

A number of experimental techniques through which the phase could be derived were investigated over the years. One very important development in which Professor S. Ramaseshan played a leading part, was the exploitation of anomalous scattering differences produced by heavy atoms introduced into the crystal structure (Ramaseshan and Narayan 1981).

The 3-dimensional structure of the insulin molecule was first determined in 1969 (Adams et al 1969). In the 2 Zn insulin crystal, the molecule’s A chain of 21 residues was found to fold compactly with an N terminal helix extending from A1–A8 running adjacent and antiparallel to a C terminal helix between A12–A21. The B chain of 30 residues was seen to contain a central section of helix between B9–B19; from each end of this the N and C terminal arms extended around the A chain.
In 2 Zn insulin the hormone is organized as a hexamer. Since 1969 the crystal structures of another hexamer, 4 Zn insulin and a dimeric insulin (Cutfield et al. 1981) and, most recently, of a modified monomeric insulin have been established. Crystallographic studies on normal insulins are restricted to aggregated forms of the molecule owing to its property of self-assembly to dimer and hexamers at low concentrations. The obstacle to the study of the insulin monomer presented by its property of self-assembly has been overcome by the preparation and crystallisation of a modified molecule from which the 5 B chain C terminal residues (B26–B30) have been removed. These 5 residues complete the surface which is buried by dimer formation and provide a complementary pair of H bonds to the partner molecule. Their removal abolishes the molecule's capacity to dimerise and to hexamerise but leaves it with substantial potency (Shanghai Insulin Research Group 1976). The Chinese in their systematic analyses of those regions of insulin have studied the pig DPI molecules in solution and in the crystal co-crystallised with cadmium ions. Comparison between the pig and beef and sheep DPI crystal structures shows their conformations are generally very similar but there are significant differences which will need to be examined (D Stuart and D C Liang, private communication).

The structure of the monomer is of great interest for three principal reasons. First, the hormone is active as a monomer. Secondly, insulin's aggregation could bring about some rearrangements of the structure which in turn could affect our ideas of the relationship between the hormone's structure and activity. Thirdly the molecule possesses considerable flexibility when organised as a hexameric structure. Since there is evidence that conformational flexibility is important in the expression of the hormone's biological activity, crystallographic study of the monomer free of the constraints imposed by aggregation should give an insight into its structural and conformational behaviour relevant to its biological action. It has emerged that the mechanism by which the molecule adjusts is by relative movements of its helical structures; these behave as essentially rigid entities, linked by more flexible extended pieces of polypeptide chain (Chothia et al. 1983). This behaviour explains nicely the structural changes seen in the insulin molecules in the insulin hexamers where the differences are associated with the crystal-packing requirements forcing one molecule to alter. The despentapeptide insulin molecule provides a further and similar example of this phenomenon which will be the subject of this paper.

2. Experimental

2.1 Preparation and crystallisation of DPI

The DPI from beef and sheep insulin was prepared by controlled cleavage at the B25–B26 peptide bond using pepsin at 4°C according to the procedure described by Gattner (1975). The product was purified by column chromatography to homogeneity. The beef and sheep DPI samples was crystallized using citrate/tris buffer with (NH₄)₂SO₄ as a precipitant (Bi et al. 1983).

Several crystal forms of beef and sheep DPI have been obtained (Bi et al. 1983) but only two have been refined. Their main crystallographic details are given in table 1. The structures were initially determined by the molecular replacement method and subsequently refined by using the Agarwal (1978) fast Fourier least squares procedure. Because the resolution of the x-ray diffraction data in these studies is limited, the
calculated shifts are inaccurate and consequently the atomic positions have been constrained to conform to proper peptide geometry after each cycle (Dodson et al 1976).

The refinement of the beef DPI is now complete and the well-defined atoms are placed with an accuracy between 0.05 and 0.1 Å. Inevitably more poorly ordered atoms (including of course many of the solvent water molecules) are less accurately positioned; their errors range between 0.2–0.3 Å. The sheep DPI refinement is not yet finished but the well-defined protein atoms are nonetheless determined with sufficient accuracy for useful comparison with the beef DPI molecule.

2.2 Comparison of beef DPI and sheep DPI crystal organisation

In figures 1(a)–(d) the packing of the beef and sheep DPI molecules in their unit cells shows how one of the crystal axes (labelled A) in the sheep DPI crystal is not exactly obeyed in the beef DPI crystal. The reason for this is clear. At A9 beef DPI has a serine whose sidechains lie very near the now approximate axis of two fold symmetry (A) relating the molecules. In figure 1(c) two sheep DPI molecules related by the operation of the exact two fold axis (A) are shown in a view perpendicular to the axis. The two beef DPI molecules related by the now pseudo axis (A) in their unit cell are illustrated in figure 1(d) in a view equivalent to that in figure 1(c). The mismatching illustrated in the figures shows that in order to accommodate the bulk of the A9 serine on the axis, the equivalent structures in the two beef DPI molecules are displaced by about 4 Å along the direction of the b axis.

Inspection of the approximate diad A in figure 1(b) shows that the overall movement which is produced by the need to pack the A9 serines is however accompanied by practically no change in the conformation at A9 and only limited changes at A8 and A10. The ability of the seryl sidechain to resist the packing pressures is, in part at least, a consequence of the well-defined H bond between the A9 seryl OG and A10 NH, an interaction which is also present in 2 Zn insulin, together with the inherent stability of the helical organisation. In figure 2 the two beef DPI molecules related by the pseudo diad are overlapped, giving an impression of the internal changes in the molecules associated with the loss of symmetry.

2.3 Comparison between the structures of 2 Zn insulin and beef and sheep DPI

The comparison between the insulin molecule in 2 Zn insulin and the DPI molecules illustrated in figure 3 shows that they are all generally similar except at the B chain N terminal residues B1–B4. These residues extend away from the molecule in DPI in contrast to the situation in 2 Zn insulin, where they wrap around the A chain. There are also significant structural changes at B25 phenylalanine, a C terminal residue in DPI. In
Figure 1. Crystal packing of the DPI molecules shown as mainchain for all residues except A9 which in sheep DPI is glycine and beef DPI is serine. The two-fold axis in the sheep DPI are labelled A & B; B remains a crystal axis in beef DPI while its two-fold symmetry is broken.

(a) Sheep DPI. Note the water on the axis at A, H bonded to the 2-fold related A10 NH groups. The A & B chains of the two 2-fold related molecules are indicated; A9 is labelled 9.

DPI there are no contacts between B25 and the rest of the molecule; in 2 Zn insulin the B25 residue makes a well-defined H bond interaction to A19 O (Dodson et al 1979).

More detailed analyses (table 2) reveals that the DPI molecules from beef and sheep are much nearer to the molecule 1 structure (Chothia et al 1983) in 2 Zn insulin, although there is some relative movement in the A and B chains (see table 2). This molecule’s structure also occurs in 4 Zn insulin and in the dimeric pig and hagfish insulin crystals (Cutfield et al 1981; Reynolds and McCall 1983). For this reason and because of its larger number of internal H bonds, molecule 1 has been regarded as likely to be near the structure of the monomer, a proposal that the DPI analyses have quite confirmed.

Figure 4(a) shows the overlap of the two beef DPI molecules optimised on the B chain helix structure. It can be seen that while the B chain helices are very similar the A chain positioning relative to the B chain is evidently different in the two molecules, indicating there is some movement of the two chains (rms $\Delta = 0.63 \, \AA$). When the helical segments A1–A9 and A12–A20 are each optimally overlaid, their structural similarity becomes
Figure 1. (Contd.)
(b) Beef DPI. Note the close approach by the A9 seryl sidechains to the line of the local axis A. The A & B chains of the two molecules related by the local axis at A are indicated.

apparent from figures 4(b) and (c) and from the magnitudes of the rms discrepancies between the matched structures (table 2). The N terminal B chain residues B1–B5 correlate better with the A chain helical segments than with the B9–B19 helix. The disulphide bond at A7–B7 between the chains is presumably responsible for this linked behaviour. By contrast the structural changes in insulin in 2 Zn insulin, shown in figure 4(d), are focussed at A1–A8 and A19 & B25 sidechains.

This picture of roughly equal distribution of structural movement in the helices in the two beef DPI molecules by which they have adjusted to their crystal environment is confirmed by comparison with the sheep DPI. As mentioned earlier this molecule crystallises with only one molecule in the asymmetric unit permitted because its A9 residue is glycine which, having no side chain, allows symmetrical packing about the crystallographic two fold axis (see figure 1a).

Analysis of the two independent beef DPI
molecules listed in table 2 shows that they are almost equally different from the sheep DPI. There is however a significantly larger difference in the two beef and the sheep DPI molecules at the A1–A8 helix. In part this is because the A1 residue in sheep DPI is still not satisfactorily determined; there is furthermore rather low electron density at the residues A1–A4 suggesting they are less well-ordered than the subsequent structure. Other differences however occur at A5–A10 which apparently indicate this helix does alter in its internal structure unlike the others. There is on the other hand an important structural difference; the loss of the A9 serine in sheep insulin changes the main chain H bonding pattern (between A9 OG and A10 NH). Because the A10 NH no longer interacts with a serine OG (from A9 in beef DPI) it makes H bonds to a water molecule trapped on the local axis (A) between a pair of two fold related molecules (figure 1b). The adjustment of the peptide conformation needed to make this interaction then alters
the mainchain H bond contacts between A90 and B5 imidazole ND and the A8NH and A30 disturbing the structure of the helix.

3. Conclusion

The loss of two fold symmetry in the 2 Zn insulin hexamer is associated with the packing of the B5 histidines in adjacent hexamers stacked along the three-fold axis. These would intersect in the 2 Zn insulin crystal organisation if the molecules both had the conformation seen in molecule 1 (Cutfield et al 1981). It appears that in concert with structural changes at the A chain N terminals and at B28–B30 the two opposing B5 histidines can make satisfactory contacts in the 2 Zn insulin crystals but they loose two-
Figure 2. The overlap of the two beef DPI mainchain structures when related by the local diad A. The A and B chains are labelled. Note the movement in the A chain N terminal helix.

Figure 3. The pig 2 Zn insulin and beef DPI molecules all shown in an equivalent view which is perpendicular to the direction of the local axis in 2 Zn insulin. There are two crystallographically independent pig insulin molecules in 2 Zn insulin and two independent beef DPI molecules in the C2 beef DPI crystal, and in each case both structures are shown. Only selected sidechains are included to simplify the figure.

The general similarity in the chain folding of the pig 2 Zn insulin and the DPI molecules is clear; the changes in structure at A1–A9 in 2 Zn pig insulin (molecule 1) are evident.
**Table 2.** The rms discrepancies between the helical segments in Beef and sheep DPI and 2 Zn insulin.

<table>
<thead>
<tr>
<th>Insulin</th>
<th>Helical Segments</th>
<th>Beef DPI (1)</th>
<th>Beef DPI (2)</th>
<th>Sheep DPI</th>
<th>2Zn insulin molecule 1</th>
<th>2Zn insulin molecule 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>beef DPI (1)</td>
<td>B9–B19</td>
<td>0.29</td>
<td>0.29</td>
<td>0.24</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A1–A8</td>
<td>0.31</td>
<td>0.56*</td>
<td>0.27</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A12–A20</td>
<td>0.27</td>
<td>0.36</td>
<td>0.24</td>
<td>0.41</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>0.63</td>
<td>0.43*</td>
<td>0.88</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B9–B19</td>
<td></td>
<td>0.33</td>
<td>0.24</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>beef DPI (2)</td>
<td>A1–A8</td>
<td>0.60*</td>
<td></td>
<td>0.26</td>
<td>1.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A12–A20</td>
<td>0.36</td>
<td></td>
<td>0.24</td>
<td>0.28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>0.52*</td>
<td></td>
<td>0.61</td>
<td>1.27</td>
<td></td>
</tr>
<tr>
<td>sheep DPI</td>
<td>B9–B19</td>
<td></td>
<td>0.38</td>
<td>0.26</td>
<td>0.38</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A1–A8</td>
<td>0.59*</td>
<td></td>
<td>0.26</td>
<td>1.28*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A12–A20</td>
<td></td>
<td>0.42</td>
<td>0.24</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>0.77*</td>
<td></td>
<td>0.61</td>
<td>1.09*</td>
<td></td>
</tr>
<tr>
<td>2 Zn insulin</td>
<td>B9–B19</td>
<td></td>
<td></td>
<td></td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>molecule 1</td>
<td>A1–A8</td>
<td></td>
<td></td>
<td></td>
<td>1.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A12–A20</td>
<td></td>
<td></td>
<td></td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All</td>
<td></td>
<td></td>
<td></td>
<td>0.46*</td>
<td></td>
</tr>
</tbody>
</table>

* These overlapped segments include residues A2–A8 only owing to the poor definition of A1.
† These overlapped segments include residues A12–A20 and B9–B19 only owing to the large structural differences at A1–A5, B1–B8 and B27–B30.

**Figure 4.** The overlap of the helical segments in the two beef DPI molecules in the C2 crystal assymmetric unit illustrating their close structural similarity. Only some selected sidechains are shown. (a) B9–B19. (b) A1–A8. (c) A12–A20. (d) The overlap of the two insulin molecules in 2 Zn pig insulin carried out by matching the mainchain between A10–A20 & B1–B27. There is a close similarity in much of the two molecules' structure, except at A1–A9 and B5 and at tyrosyl & phenylalanyl sidechains A19 & B25. More selected sidechains are included than in a, b & c in order to illustrate that they also possess close similarity in conformation.
Local 2-fold axis between hexamers

Figure 5. The crystal packing around the B5 histidines in pig 2 Zn insulin. In accommodating the packing requirements, the histidine of molecule 2 has changed its conformation and makes different internal contacts to the B5 histidine in molecule 1 and no contacts with the adjacent molecule. The histidine of molecule 2 makes 2 well-defined internal H bonds (to A7 and A9 carbonyl O), and a good inter hexamer H bond to A90, which has allowed it to retain its original conformation while its companion alters in structure.

fold symmetry in doing so (figure 5). There are obvious parallels between the situation in 2 Zn insulin and the beef DPI crystals with its packing together of the A9 serines. But the structural response of the molecules in the DPI crystal is quite different. The 2 Zn insulin changes of structure are large but are limited essentially to A1–A5 and B28–B30 ( & B25 sidechains). The 2 Zn insulin hexamer probably contains exact two-fold symmetry in solution; the perturbations needed to allow satisfactory crystal packing have to overcome a pre-existing symmetrical structure. It thus appears that the integration of the hexamer prevents much deviation from symmetry except on the surface where the adjustments are relatively large. Those within the hexamer are much smaller, of the order 0.2–0.3 Å and the segments themselves change very little, by between 0.1 and 0.2 Å. The absence of structural change at A9 serine and the structural change in the beef DPI molecules in response to the requirements of crystal packing illustrates the expected greater flexibility of a monomeric insulin species. This property, however, derives mainly from the ability of the various segments of secondary structure, such as the helices, to move relative to one another through the flexibility of the extended pieces of main chain that link them (Chothia et al. 1983).

References

Crystal structure of despentapeptide insulin

Agarwal R C 1968 Acta Crystallogr. A34 791
Gattner H G 1975 Hoppe Seyler’s Z. Physio. Chem. 386 1397
Hodgkin D C and Riley D P 1968 In Structural chemistry and molecular biology (eds) A Rich and N Davidson (San Francisco: W H Freeman) p. 15
Reynolds C D and McCall M Private communication
Shanghai Insulin Group 1976 Sci. Sin. 19 351
X-ray structural studies on metal complexes with nucleotides and pyridoxal-amino acid Schiff bases. Models for platinum binding to DNA and pyridoxal catalyzed reactions

M DAMODARA POOJARY, S P SUDHAKARA RAO and H MANOHAR
Department of Inorganic and Physical Chemistry, Indian Institute of Science, Bangalore 560012, India

Abstract. X-ray structural studies on metal complexes with nucleotides and with pyridoxal-amino acid Schiff bases are briefly reviewed. The results with ternary metal nucleotide complexes show that the oxopurine nucleotides coordinate to the metal ion through the N(7) atoms of the bases in cis position. The relevance of this mode of binding is discussed in terms of the possible mechanism of action of the novel platinum drugs. On the basis of the studies on metal pyridoxal-amino acid Schiff base complexes, the variations in stereochemistry of the ligands in different metal complexes have been related to the catalytic activity of various metal ions in pyridoxal-catalyzed nonenzymatic reactions.

Keywords. X-ray crystallography; metal-nucleotide complexes; metal-Schiff base complexes; pyridoxal-amino acid Schiff bases; Pt(II)-DNA interaction.

1. Introduction

Metal ion complex formation is one of the prominent interactions in biological systems. Groups like nucleotides and pyridoxal-amino acid Schiff bases are important biological entities and their interactions with metal ions are a fertile field of research activity. We have initiated work on structural studies of ternary metal nucleotide complexes and metal complexes with pyridoxal-amino acid Schiff bases. In this paper, our recent results, as well as earlier work in this area, are briefly reviewed.

2. Models for platinum binding to DNA

2.1 Role of metal ions in the biochemistry of nucleic acids

It is well-known that metal ions play an important role in the biochemistry of nucleic acids (Eichhorn 1973). For example, the transfer of ATP phosphoryl, which is an enzymatic process essentially requires the presence of metal ions. Even the non-enzymatic transphosphorylation, which occurs between ATP and the orthophosphate ion, is catalysed by metal ions, notably Mn$^{2+}$, Cd$^{2+}$, Ca$^{2+}$. Similarly the cleavage of pyrophosphate from four deoxyribonucleotide triphosphates, dATP, dCTP, dGTP, dTTP, involved in the DNA-polymerase reaction, together with the formation of phosphodiester linkage requires the presence of divalent metal ions, mainly Mg$^{2+}$. It was also discovered that the transition metal ions like Fe$^{2+}$ and Cu$^{2+}$ are components of tobacco mosaic virus DNA, and they are strongly bound to the ligating sites in the nucleic acid. More recent investigations have shown that Zn$^{2+}$ is essential for the activity of
Escherichia coli polymerase 1. Moreover the in vivo role of metal ions in nucleic acid chemistry—DNA replication, transcription and translation; DNA denaturation and renaturation; RNA conformational properties; depolymerization of RNA; enzyme-metal-nucleic acid ternary species has been well recognized (Eichhorn 1973). Furthermore, the clinical success of certain transition metal complexes, e.g. cis-[Pt(NH$_3$)$_2$Cl$_2$] (Rosenberg et al 1969) and the speculation that these complexes may act by crosslinking DNA in vivo has promoted research activity in this area. Also the unique properties of metals are used as probes for nucleic acid chemistry, for structural investigations and for isolation of these biologically important molecules (Eichhorn 1973; Marzilli 1977; Hodgson 1977).

Although the importance of metal ions in nucleic acid processes is well established their mechanism of action is little understood. As these processes are selective in nature, the interaction of inorganic species with nucleic acids has also to take place through selective coordination. The polynucleotides offer three major ligating sites to metal ions: (i) the heterocyclic ring nitrogen atoms, especially N(1) and N(7) of purines and N(3) of pyrimidines, and the exocyclic functional groups of purine and pyrimidine bases; (ii) the phosphate oxygens of phosphodiester linkages; (iii) the ribose moiety. A knowledge of the metal ion binding sites for each of these component nucleotides (chart I), therefore, is of great value in assessing how the metal ions might interact with nucleic acids.

This led several workers to investigate metal complexes of nucleic acids and their

![Chart 1](chart1.png)
constituents: nucleotides, nucleosides and bases, while earlier studies in this direction
carried out in solution and using spectroscopic methods are significant, more
experimental data are necessary to determine with certainty the mode of binding of
metals and their effects on the structure and conformations of nucleic acids. Among the
various physical techniques suitable for studying substances in the crystalline state,
x-ray diffraction has proved to be a powerful tool to elucidate the structure of
complexes formed by transition metal ions with nucleic acid constituents. A great
wealth of information regarding binding sites, conformations, packing, hydrogen
bonding, etc. has been accumulated in nucleic acid constituents in their complexes
(Hodgson 1977; Marzilli 1977; Marzilli and Kistenmacher 1977; Gellert and Bau 1979;
Swaminathan and Sundaralingam 1979; Aoki 1981).

As mentioned earlier, the mixed ligand complexes have been recognised as the low-
molecular weight representatives for the active catalytic centres occurring in vivo. A
ternary complex involving a metal ion, a nucleotide and a secondary ligand such as 2-2'-
bipyridine can mimic enzyme-metal ion-substrate interactions in biological systems
(Naumann and Sigel 1974). Moreover it has been suggested that in solution, in the
presence of a π-aromatic amine, metal-nucleotide interaction takes place predomin¬
nantly through phosphate oxygens, while the base of the nucleotide does not take part
in coordination (Naumann and Sigel 1974; Sigel 1975; Mitchell and Sigel 1978). This
observation is attributed to the discriminating qualities of π-aromatic amines to the
metal ions, which in turn is one of the tools employed by nature to achieve selectivity.
Hence one is tempted to predict that much of the specificity and many of the properties
so far only associated with metal ion complexation in macromolecular biological
systems are inherent already in relatively simple ternary systems.

On the other hand, an entirely different binding mode has been observed in some
ternary platinum oxopurine nucleotide complexes (Aoki 1981). The nucleotides
coordinate to the metal ion in cis position through the N(7) atoms of the purine bases.
Such a type of binding has been observed (Poojary and Manohar 1982, 1983) in ternary
Co^{3+}, Ni^{2+} and Cu^{2+} oxopurine complexes. The importance of these compounds as
models for platinum binding to DNA is discussed below.

2.2 Anti-cancer activity of cis-DDP
Since the discovery of the anti-tumour activity of cis-[Pt(NH_{3})_{2}Cl_{2}] (cis-DDP)
(Rosenberg et al 1969), considerable research has been focussed on elucidating the
mechanism of activity of these drugs. The following common features which link
structure to reactivity have emerged from these studies (Rosenberg 1980): (i) The
complexes have both inert and labile ligands and only the latter ligands undergo
labilization quickly in reactions with biological molecules. (ii) The exchangeable
leaving groups should be in the cis configuration. (iii) The complexes should be neutral.
(iv) The geometry of the complexes is either square planar or octahedral. (v) Two cis
monodentate or one bidentate leaving groups are required. (vi) The leaving ability of
these labile groups should be moderate. (vii) The leaving groups should be separated
approximately by 3-4 A.

Studies on cis-DDP and its analogues clearly suggest that activity of these drugs
depends on the cis oriented labile ligands in the coordination plane of platinum. In
aqueous solution these labile ligands may be easily displaced, and it has been speculated
that the resulting cis (NH_{3})_{2}Pt^{2+} moiety is the active species which interacts with
biological target molecule (Barton and Lippard 1980). Studies have also shown DNA as cellular target for cis-DDP. In binding to DNA, cis-DDP could block replication, which would suppress rapid growth of cancerous tissues.

2.3. Possible mechanisms of action of cis-DDP

A variety of structural models for the ligating site in DNA for cis-DDP have been proposed. Studies have shown that N(7) nitrogen atom of purine and N(3) nitrogen atom of pyrimidines are the primary sites of coordination for the platinum atom. However, this unidentate coordination is not sufficient to explain the observed differences in activity of cis and trans isomers. Potentiometric studies (Macquet and Theophanides 1975) show that on binding of cis-DDP to DNA both the chloride ions are released, while the release of one chloride ion is observed on binding of trans isomer. Thus N(7)-O(6) chelation was proposed on the basis of bifunctional mode of coordination. The trans isomer could not bind to these sites simultaneously. However, geometrical considerations required for such chelation have led many to suggest that such a structure is unlikely.

Other possible bifunctional coordination modes include the formation of either intra or interstrand crosslinks. The crosslinking of base residues on opposite strands of DNA would severely hinder replication. Even though interstrand crosslinking of DNA does occur in the presence of cis-DDP, the frequency of this event is too low to account for its antitumour activity (Roberts and Thomson 1979). At present it appears more probable that the site of interaction involves an intrastrand crosslinking. The observation that the nonbonded distance between the chlorine ions in cis-DDP is 3.4 Å, the base pair stacking distance suggests that intrastrand crosslinking is chemically more reasonable. Such a crosslink would necessarily result in an unstacking of bases. The base tilting required to accomplish this crosslink would lead to local denaturation. This intrastrand crosslink could be formed by platinum binding to N(7) nitrogen atoms of the adjacent guanine bases.

2.4 Intrastrand crosslinking models

Even though information about the binding of cis-DDP to the polymer is scarce, some recent studies on monomers allow interesting speculations to be made. One increasingly mentioned mode of binding to the cis isomer is intrastrand crosslinking. Goodgame et al (1975) described a cis-Pt(NH$_3$)$_2^+$ complex of 5'-IMP, cis-[Pt(5'-IMP)$_2$(NH$_3$)$_2$]$_2$$^2$⁻, wherein Pt(II) binds through equatorial sites to two 5'-IMP ligands which are related by a crystallographic two-fold axis. Similar type of complexes were reported with different inert ligands like ethylenediamine and trimethylenediamine (Kistenmacher et al 1978; Bau and Gellert 1978; Kistenmacher et al 1979; Marzilli et al 1980). Most of these complexes are nonstoichiometric. All these structures bear a strong resemblance to the structure of monosodium salt of 5'-IMP (Rao and Sundaralingam 1969), where a water molecule lies on a crystallographic 2-fold axis and links, via hydrogen bonds, two symmetry-related nucleotides. In the isomorphous platinum complexes this water site is occupied by the Pt(II) moiety and the hydrogen bonding scheme present in the sodium salt is replaced by two cis coordination bonds. In terms of the possible effect on a polynucleotide structure due to intrastrand crosslinking, the structural similarities displayed by these 5'-IMP salts are particularly important. Chiang et al (1978) showed that there is a strong competition between the crystal packing forces.
operative in the solid and the distortion of the basic structure due to the simultaneous binding of the two 5'-IMP bases by Pt(II) moiety. In fact if one assumes that the crystal packing forces operative in the 5'-IMP complexes approximate those found in the base-stacked polynucleotide, the above result suggests that intrastrand crosslinking by reagents such as cis DDP will place a significant strain on the conformational properties of the polynucleotide.

All the platinum-5'-IMP complexes discussed above are 1:2 complexes. The binding of metal to the oxopurine nucleotides in cis position has also been displayed by 1:1 ternary metal-nucleotide complexes containing ethylenediamine and metal ions like Co(III) and Ni(II) (Poojary and Manohar 1982). Again these complexes are isomorphous among themselves and isostructural with platinum complexes and sodium salts of oxopurine nucleotides. For comparison, structures of the sodium salt, the platinum and cobalt 5'-IMP complexes are given in figure 1. The geometry of the nucleotides around Pt(II) and Co(III) is the same. An interesting feature of the Co(III) and Ni(II) complexes is that the ethylenediamine molecule has been displaced from the metal centre, to which the nucleotides coordinate. This results in one metal ion interacting only with ethylenediamines and the other with nucleotides. The remaining octahedral coordination sites are occupied by waters. Both the metal ions lie on different two-fold

![Figure 1](image-url). A comparison of structures of (a) monosodium salt of 5'-IMP, (b) [Pt(5'-IMP)$_2$(NH$_3$)$_2$]$^+$ and (c) [Co(5'-IMP)(en)(H$_2$O)$_3$]$^+$. Horizontal and vertical lines represent two-fold rotation axes. Broken lines indicate hydrogen bonds. (Figures 1(a) and (b) are reprinted with permission from Chiang et al 1978; Figure 1(c) with permission from Poojary and Manohar 1982).
axes. Chlorine ions which neutralize the charge in the Co(III) complex are disordered in the vacant site between the molecules. The ethylenediamine molecules are also disordered.

The 5'-GMP complexes of Co(III) and Ni(II), like the 5'-IMP complexes, are based on the sodium salt of 5'-GMP (Katti et al 1981). In the sodium salt of 5'-GMP, a sodium ion links the two nucleotide moieties. This sodium ion is replaced by the metal ions Co(III) and Ni(II). Recently a similar binding has been observed (Poojary and Manohar 1983) in a ternary Cu-5'-IMP complex, containing imidazole groups. This complex is isomorphous with cis [Pt(5'-IMP)(NH$_3$)$_2$]$^{2-}$. The imidazole moieties are in cis position with a partial occupancy of 0.4, while for the rest of the time these sites are replaced by waters. The cis binding of nucleotides to Cu(II) ion in the presence of a π-aromatic amine (imidazole) is at variance with the results of crystallographic studies on ternary complexes containing other π-aromatic amines (Aoki 1981) and contrary to the conclusions derived from solution studies mentioned earlier.

3. Model compounds for pyridoxal catalyzed reactions

3.1 Role of metal ions

Enzymes dependent on pyridoxal phosphate (PLP) (the cofactor form of vitamin $B_6$) catalyze a variety of metabolic reactions of amino acids such as decarboxylation, transamination, racemization and carbon-carbon bond cleavage. Many nonenzymatic model reactions also proceed by similar mechanisms in the presence of pyridoxal (PL) and a suitable metal ion (Holm 1973). A general theory on the mechanism of nonenzymatic reactions essentially involves the following steps: (i) Formation of a Schiff base with subsequent labilization of bonds attached to the α-carbon atom due to electron displacement from the α-carbon atom to the electronegative ring nitrogen through a conjugated system of double bonds, (ii) Release of H$^+$, COOH$^+$ or R$^+$ from the α-carbon atom depending upon their orientation with respect to the extended π-system (aldimine plane) to produce a transitional Schiff base, (iii) Localization of the lone pair of electrons of the heterocyclic nitrogen and subsequent protonation either at the α-carbon atom for racemization and decarboxylation or at the formyl carbon for transamination, (iv) Hydrolysis of the carbon-nitrogen bond to give the products.

The mechanism for the first stage of transamination reaction is given in chart 2. The imines 2 and 4 have been detected spectrophotometrically, but direct evidence supporting the existence of intermediate 3 is lacking. It has been suggested that the role of the metal ion is to stabilize the imines 2 and 4 through chelation in a tridentate fashion, which provides a planar conjugated system via 3 and increases the inductive effect away from the α-carbon atom (Metzler and Snell 1952). Thus, it is now believed that the metal ion serves as a trap for the preformed Schiff base.

It is of interest to note that different metal ions have varied catalytic influence. The efficiency towards PL transamination was found to be in the order Co$^{2+}$ < Ni$^{2+}$ < Zn$^{2+}$ < Fe$^{3+}$ < Fe$^{2+}$ < Al$^{3+}$ < Cu$^{2+}$, from the investigation of the influence of metal ions on the rate of the transamination of pyridoxamine with α-ketoglutarate (Longenecker and Snell 1957). As it is conjectured that the metal ion has no kinetic role in the Schiff base formation, apparently the cause for varied catalytic influence of different metal ions lies in the stereochemistry of metal-Schiff base complex. From pH-metric titrations (Christensen 1957), the expected downward shift in the pK value
of the pyridine nitrogen in Ni(II), Mn(II) and Zn(II) chelates was not observed and this result was attributed mainly to an unstable metal-phenolic oxygen bond in solution. But in the solid state, the crystal structures of [Mn(PL-DL-valine)$_2$] (Willstätter et al. 1963), [Ni(PL-DL-valine)$_2$] (Capasso et al. 1974) and [Zn(PL-L-valine)$_2$] (Capasso et al. 1974) have shown that metal-phenolic oxygen bond is equally as strong as other metal-donor bonds. Capasso et al. (1974) speculated that the puckering of the chelated Schiff bases could play a decisive role in catalytic activity. In [Ni(PL-glycine)$_2$] and [Cu(PLP-glycine) (H$_2$O)] complexes also contrasting features have been observed with respect to the stereochemistry (vide infra), but not in relation to the strength of the metal-donor bonds (Rao and Manohar 1983).

3.2 Stereochemistry of metal pyridoxal-amino acid complexes

Because of the tridentate nature of the Schiff base ligand, complexes with 2:1 and 1:1 ligand metal stoichiometry can be formed. The heterocyclic nitrogen may be protonated according to the charge neutralization requirement. In 1:1 complexes, confined to copper(II) only, approximate square-pyramidal coordination geometry exists, in which three positions in the basal plane are occupied by the aldimine ligand. In [Cu(PL-DL-valine)] (Dawes et al. 1982) the fourth basal position and the fifth apical position are occupied by a pyridine nitrogen and a hydroxymethyl oxygen respectively of different neighbouring molecules, whereas in [Cu(PLP-DL-phenylalanine) (H$_2$O)] (Bentley et al. 1968) they are filled by a water molecule and phosphate oxygen of another molecule. In [Cu(PL-o-phospho-DL-threonine)] (Aoki and Yamazaki 1980) the phosphate oxygen from a neighbouring molecule acts as the fourth donor which results in a dimeric structure, the fifth being a water molecule. The copper complex with PL-L-histidine has recently been reported (Dawes and Waters 1982) to be trimeric with the histidine nitrogen of another ligand of the same trimer and a water molecule being in the fourth and fifth coordination positions respectively.
In \([\text{Cu(PLP-glycine)} (\text{H}_2\text{O})]\) (Rao and Manohar 1983) a polymeric structure is observed, the axial site being occupied by a phosphate oxygen of a neighbouring molecule translated along \(a\) axis; a water molecule binds at the remaining equatorial site (figure 2). There are three more water molecules in the asymmetric unit which link the two polymeric chains through hydrogen bonds.

In 2:1 complexes, chelation takes place on an octahedral edge so that the two imine nitrogens are \textit{trans} and the two phenolate and carboxylate oxygens are mutually \textit{cis} to each other. From a structural point of view, such complexes are interesting in that the \textit{bis} (tridentate) complexes have the \textit{abf} structure for which two enantiomorphous forms \(A\) and \(B\) are possible (chart 3). The presence of an asymmetric centre in the condensed amino acid generates the diastereoisomeric species, \(A(\text{LL}) \equiv B(\text{DD}), A(\text{DD}) \equiv B(\text{LL})\) and \(A(\text{DL}) \equiv B(\text{LD})\) where \(A\) and \(B\) stand for the two possible enantiomorphous forms of the complex, \(D\) and \(L\) refer to the configuration of the two Schiff base ligands.

However, it has been observed that either the enantiomorphous pair \(A(\text{LL}), B(\text{DD})\) (or \(A(\text{DD}), B(\text{LL})\)) are present in the solid state as found in \([\text{Mn(PL-DL-valine)}_2]\) (Willstater \textit{et al} 1963) and \([\text{Ni(PL-DL-valine)}_2]\) (Capasso \textit{et al} 1974) or only the \(A(\text{LL})\) diastereoisomer as observed in \([\text{Zn(PL-L-valine)}_2]\) (Capasso \textit{et al} 1974). The fact that complexes such as \(A(\text{DL})\) or \(B(\text{DL})\) are not formed, as expected, with racemic
Schiff bases with the same yield as \( A(LL) \) and \( B(DD) \) indicate that the formation of crystalline complexes is stereospecific. In \([\text{Ni(PL-glycine)}_2]\) complex (Rao and Manohar 1983), only enantiomers \( A \) and \( B \) are present but not those arising from the chirality of the amino acid, as glycine is nonchiral.

\([\text{Ni(PL-o-phospho-DL-threonine)} (\text{H}_2\text{O})_2]\) (Aoki and Yamazaki 1980) is probably the only noncopper 1:1 transition metal pyridoxal-amino acid complex reported in the literature. For this complex, three octahedral coordination sites are occupied by the tridentate ligand and the remaining three by water molecules. The phosphate unit of the amino acid forms an intramolecular hydrogen bond with the axial water molecule.

Long et al (1980) reported a binuclear iron(III) complex, \([\text{Fe(N-(PL)}_2\text{alanine)} \text{ClO}_4]_2\), with an unusual \( \text{H} \) derivative. It is proposed that the ligand has resulted from the formation of Schiff base between PL and pyridoxamine and the subsequent nucleophilic attack of the Schiff-base imino group upon the \( \alpha \)-keto carbon atom of pyruvic acid. In the dimer, two Fe(III) ions are bridged by two \( \alpha \)-keto oxygens of pyruvic acid groups. Both Fe(III) ions have a highly distorted octahedral geometry with the other coordination sites of each ion being occupied by two phenolic oxygens, one carboxylate oxygen and one imine nitrogen.

### 3.3 Reactivity

It may be expected that the stereochemistry and ring strain of the Schiff base ligand could explain the rate and specificity of reactions, in particular, the lower and higher activities of the Ni(II) and Cu(II) chelates respectively. Mechanisms of pyridoxal-amino acid reactions postulated involve electron transfer via the extended \( \pi \)-orbital system. This requires the Schiff base ligand to be planar. However, it has been observed that in metal chelates the ligand itself is not planar and the distortion from planarity is mostly due to a rotation of carboxy group around the C(9)–C(10) bond (figure 2). This nonplanarity is more pronounced in Ni(II), Zn(II) and Mn(II) complexes than in Cu(II) chelates. For example, the dihedral angles between rings I and II and II and III (figure 3) are 4° and 0.7° respectively in \([\text{Cu(PLP-glycine)} (\text{H}_2\text{O})]\), whereas they are as high as 18.6° and 35.9° in \([\text{Ni(PL-glycine)}_2]\) (Rao and Manohar 1983). These differences in ligand distortions may be attributed to the different metal-donor distances for different metals. Copper ion, having a mean metal-donor distance less than 2 \( \text{Å} \)
occupies the center of a nearly planar tridentate ligand. The increased coordination bond lengths in the Mn(II), Ni(II) and Zn(II) complexes (2.00–2.15 Å) result in increased ring strain which determines the puckering of the chelated Schiff bases. Thus, the Schiff base ligated to the Cu(II) ion has the least distortion which is essential for facile electron transfer and this explains the higher activity of copper chelates.

The various reactions involving α-amino acids in model systems involve the formation of 1:1 and 2:1 aldime complexes followed by the breaking of a, b or c bond in the condensed amino acid portion (figure 3). Dunathan’s hypothesis (1966) states that the bond perpendicular to the extended π-system (which optimizes orbital interactions) is labilized most. Accordingly for [Ni(PL-α-phospho-DL-threonine) (H₂O)₃] and [Cu(PL-α-phospho-DL-threonine)], Aoki and Yamazaki (1980) considered the orientation of α-C-hydrogen atom and compared this to the relative activity. The α-C-hydrogen atom is positioned almost in the PL-aldimine plane, the torsion angles C(7)-N(2)-C(8)-H being 24° for the Ni complex and 33.5° for the Cu complex. This observation led them to rationalize that the Cu(II) ion under acidic conditions and the Ni(II) ion lack catalytic activity (for the β-elimination reaction).

Acknowledgement

This article is dedicated to Prof S Ramaseshan on the occasion of his sixtieth birthday. We wish him many more years of fruitful scientific activity. Two of the authors (MDP and SP) thank the UGC and Department of Atomic Energy, India, for financial support.

References

Bau R and Gellert R W 1978 Biochimie 60 1040
Christensen H N 1957 J. Am. Chem. Soc. 79 4073
Eichhorn G L 1973 Inorganic biochemistry (ed.) G L Eichhorn (Amsterdam: Elsevier) Chapters 3 and 4
Gellert R W and Bau R 1979 Met. Ions Biol. Syst. 8 57
Holm R H 1973 Inorganic biochemistry (ed.) G L Eichhorn (Amsterdam: Elsevier) Chapter 31
Longenecker J B and Snell E E 1957 J. Am. Chem. Soc. 79 142
Macquet J P and Theophanides T 1975 Bioinorg. Chem. 5 59
Metal nucleotide and metal-Schiff base complexes

Metzler D E and Snell E E 1952 J. Am. Chem. Soc. 74 979
Mitchell P R and Sigel H 1978 J. Am. Chem. Soc. 100 1564
Naumann C F and Sigel H 1974 J. Am. Chem. Soc. 96 2750
Poojary M D and Manohar H 1983 (submitted for publication)
Rao S P S and Manohar H 1983 (unpublished work)
Roberts J J and Thomson A J 1979 Prog. Nucleic Acid Res. Mol. Biol. 22 71
Sigel H 1975 Angew. Chem. Int. Edn 14 394
Swaminathan V and Sundaralingam 1979 CRC Crit. Rev. Biochem. 245
Macromolecular crystallography research at Trombay

K K KANNAN and R CHIDAMBARAM
Neutron Physics Division, Bhabha Atomic Research Centre, Bombay 400085, India

Abstract. Neutron diffraction studies of hydrogen positions in small molecules of biological interest at Trombay have provided valuable information that has been used in protein and enzyme structure model-building and in developing hydrogen bond potential functions. The new R-5 reactor is expected to provide higher neutron fluxes and also make possible small-angle neutron scattering studies of large biomolecules and bio-aggregates. In the last few years infrastructure facilities have also been established for macromolecular x-ray crystallography research. Meanwhile, the refinement of carbonic hydrases' and lysozyme structures have been carried out and interesting results obtained on protein dynamics and structure-function relationships. Some interesting presynaptic toxin phospholipases have also been taken up for study.

Keywords. Protein crystallography; enzyme mechanism; neutron diffraction; x-ray diffraction; carbonic anhydrase; lysozyme; presynaptic toxins.

1. Introduction

The function of biological macromolecules is often dependent on the ability of these molecules to transfer protons efficiently, which invariably involves active site hydrogen bonding. The advent of neutron diffraction has enabled the accurate determination of hydrogen atom positions in small molecules of biological interest like amino acids and peptides and the stereochemistry of the hydrogen bonds in these structures (Ramanadham and Chidambaram 1978). These studies have provided valuable information for use in protein and enzyme structure model-building and have also enabled the development of potential functions for hydrogen bonds between peptide groups (Chidambaram et al 1970) to be used in protein conformation studies. Neutron diffraction of biological molecules have been studied at Trombay using the neutron beams from the CIRUS reactor and a computer-controlled neutron diffractometer (Momin et al 1978). The latter and an x-ray diffractometer based on the same design are incidentally the first (and so far the only) computer-controlled diffractometers built in the country. The commissioning in 1984 of the 100-MW research reactor R-5 which will have several times the neutron flux of the CIRUS reactor will facilitate study of biological molecules using neutron beams. This reactor with a built-in cold source will also have a facility for small-angle neutron scattering (K R Rao, Private Communication) that can be used for studying large biomolecules and bio-aggregates using the powerful contrast-matching method with deuterium substitution.

X-ray diffraction techniques are most suited for studying globular protein and enzyme structures though some neutron diffraction studies have also been attempted at Brookhaven and Grenoble (Chidambaram 1981). The study of conformation of proteins was initiated by Prof. G N Ramachandran in India and he and his colleagues have made outstanding contributions which are too well-known to be recounted here.
The need for experimental studies in India of macromolecular structures especially protein structures, by x-ray diffraction methods has been felt for sometime. In 1978, when there was little effort being made in India in this direction, we decided to initiate a programme of study of enzyme structure and function by x-ray diffraction methods. From the instrumentation and computational knowhow generated from our earlier crystallography work and the wide interest in related fields in other divisions in BARC working on the biochemistry aspects of proteins and other macromolecules, our group seemed particularly well-suited to undertake this research.

1.1 Why protein structure?

Proteins are one of the most important components of any living system and carry out all the vital catalytic functions of the organisms. Protein enzymes exhibit a remarkable specificity towards their substrate. This specificity of the enzyme is mostly due to the spatial geometry of the active site of the enzyme, which is determined by the three-dimensional structure of the enzyme itself. Thus an understanding of the molecular aspect of enzyme catalysis necessarily requires the understanding of the three-dimensional structure of the enzyme. The elucidation of the enzyme mechanism is a difficult process and requires the study of the three-dimensional structure of the enzyme at different pH values and also of the enzyme-inhibitor and the enzyme-substrate complex (Blake et al 1967; Lipscomb et al 1968; Blow et al 1969; Kannan et al 1977b). The determination of the active site geometry requires the availability of accurate coordinates of the atoms that make up the protein structure. The structure of the solvent molecules that are important components of the live enzyme has also to be determined with a good degree of precision. A knowledge of the mode of inhibitor binding is also essential for an understanding of drug interaction with proteins (Kannan et al 1977c).

2. Protein structure determination

The all-important tertiary structure of the protein can be determined only by single crystal x-ray diffraction analysis. The method of isomorphous replacement, i.e. use of different heavy atom derivatives of the protein, has proved to be the single most important method to elucidate the structure of proteins (Green et al 1954; Blundell and Johnson 1976). Where complete or partial structure of a related protein is available, the application of the molecular replacement methods, i.e. the use of rotation and translation function (Rossmann 1972), may be fruitful in arriving at the structure.

The application of either the method of isomorphous replacement or the molecular replacement method results in an electron density map calculated from

$$\rho(x, y, z) = \frac{1}{P} \sum_{h} \sum_{k} \sum_{l} |F_{o}(hkl)| \exp i[2\pi(hx + ky + lz) - \alpha(hkl)],$$

where $|F(hkl)|$ is the structure amplitude of the reflexion $(hkl)$ and $\alpha(hkl)$ is the phase angle calculated by the methods mentioned above. The electron density maps thus obtained are then interpreted either in an optical comparator (Richards 1968) or in a graphics display (Jones 1978) and the structure of the protein built up. This is of course a very critical operation and the success of the interpretation depends on the quality of the electron density maps and the availability of at least partial amino acid sequence of the protein (Kannan 1977). The quality of the electron density map is dependent on the
quality of the diffraction data collected, i.e. the accuracy of the measurement of the structure amplitude for the native protein and the heavy atom derivatives. The quality of the map is dependent also on the quality of the phases $\alpha(hkl)$. These are entirely dependent on the quality of and the number of heavy atom derivatives used in the calculation in the case of the isomorphous method and on the degree of structural homology between the known and unknown structure in the case of the molecular replacement method.

It is important to stress that the quality of the crystals used and their stability to x-ray radiation are primary requisites to a successful solution of the protein structure. Thus crystallization of the native protein and the heavy atom derivatives assumes added significance as these are by no means well-established techniques like the rest of the protein structure determination.

The structure once determined as outlined above forms the basis for a more thorough analysis. The coordinates obtained are subjected to crystallographic refinement by a combination of least squares methods, and most important of all, inspection and interpretation of difference electron density maps where the amplitude $|F(hkl)|$ is now either $(|F_{\text{obs}}(hkl)| - |F_{\text{cal}}(hkl)|)$ or $(|2F_{\text{obs}}(hkl)| - |F_{\text{cal}}(hkl)|)$. During refinement the solvent molecules are also interpreted and their coordinates incorporated to get a complete tertiary structure. It must be stressed here that without locating the solvent molecules which are part of the protein structure, the determination of the tertiary structure of a protein is incomplete. Another important purpose of the refinement is to identify and include those amino acid residues which were not interpretable or wrongly identified during the course of the interpretation of the isomorphous or molecular replacement electron density maps.

A study of a protein structure is in itself very rewarding. However, the use of the structure results towards understanding functions of the protein is still more important and should be a natural goal for the protein crystallographer. This requires the availability of the accurate native protein structure at different pH values and also the accurate structure of the protein-substrate and protein-inhibitor complexes. Thus it is always necessary to obtain the structure of the protein under these varying conditions, a not so trivial task in terms of data collection and computation and electron density map inspection and not least the assimilation and interpretation of the biochemical-biophysical and enzyme kinetic data with reference to the structure of the protein in the different environments. It is thus apparent that a successful macromolecular work requires a well organised infrastructure facility in terms of powerful x-ray diffraction equipment, a well-equipped biochemical laboratory facility, automatic diffractometry or film recording equipment, automatic film measuring facility and not least of all good high speed computing facility and associated computer programs for doing the varied calculations needed for protein structure-function-inhibition analysis.

In the subsequent sections we will discuss the infrastructural facility at Trombay we have set up from 1979 onwards and also the results of the macromolecular analysis we have undertaken here.

3. Infrastructure at Trombay

3.1 Instrumentation and laboratory facilities

The protein crystallography group has a modest biochemical facility for the rudimentary but necessary protein purification work and has also access to the
instrumentation at the Biochemistry Division of the Bhabha Atomic Research Centre. The equipment set up by the group consists of a locally manufactured cold cabinet for maintaining the temperature of samples during preparation at 4°C, an automatic fraction collector designed and fabricated by the Central Workshops and the Reactor Control Division, BARC (Koppikar et al Personal communication), apart from other essential instruments like pH meters, gel electrophoresis apparatus and a binocular wide field microscope for crystal viewing during crystallization and also for crystal mounting for x-ray diffraction work. Many of the crystallization apparatuses (Blundell and Johnson 1976) like zeppezaur capillaries, vapour diffusion chambers for normal mode and hanging drop mode have also been made by the group. The group has successfully crystallized proteins and reported the first ever protein crystallization and crystal data in India using these and other facilities like dialysis and solution interfacing in melting point capillaries. Snake venom toxins Notexin and Notechis II-5 have been crystallized at Trombay and are currently under investigation by x-ray crystallographic methods (Kannan et al 1981).

The x-ray diffraction equipment consists of a stabilized x-ray generator designed and fabricated by the scientists at BARC, Bombay and VEC, Calcutta and a high brilliance rotating anode (the latter is also used for energy-dispersive high-pressure diffraction research). The VEC-BARC generator uses the Phillips fine focus x-ray tubes with copper anode. For photographic data collection, there is a precession camera and an Arndt-Wonacott oscillation camera. It is also being proposed to modify an indigenous microprocessor-controlled x-ray diffractometer design to suit the crowded reciprocal lattices of macromolecular crystals.

The x-ray diffraction films are processed on a computer-controlled two-dimensional microdensitometer, which is also suitable for processing radiographs and other images on films. The map interpretation is done on a three-dimensional graphics display system available with the Computer Section at BARC. The group uses mostly the Prime 450 with the Computer Section at BARC and the VAX 11/780 at TIFR for their current computational needs.

### 3.2 Computer programs

Macromolecular structure determination and refinement requires a large number of computer codes. Often the programs used by the group have been available on computers other than those accessible to them. This has necessitated the conversion of a large number of computer codes for the protein group's day-to-day needs. A major programming effort has been that for the online control of the microdensitometer by the PDP 11/34 computer under the RSX11M operating system. The group has written the driver routine (Pal et al 1983) for the control and also converted Allan Wonacott's film processing programs written originally for a Nova computer system to the PDP 11/34 system under the RSX11M operating system using the driver developed at BARC. These include programs to index still photographs, processing of precession, oscillation photographs and absorption streak photographs (Pal et al 1983) and programs for film scaling and Lorentz and polarization corrections.

The data obtained from the data processing programs above are being handled by the Protein program package kindly made available to us by Dr. Steigeman, Max Planck Institute for Biochemistry, Munich, Germany for the VAX 11/780 computer. This also includes, among other things, the structure factor and electron density
calculation programs. This program also forms the essential interface to the map manipulation programs so essential to use on the graphics system. The maps thus prepared are transmitted to the PDP 11/34 through a computer link from the Prime 450 computer, the software for which has been written at BARC (R S Mundada and R K Talwar, Private Communication) for interpretation on the graphics with the Frodo programme package (Jones 1978).

The refinement of the heavy atom derivatives and the calculation of the isomorphous phases are possible through a number of programs available to the group (Jarup et al 1970) and also on the VAX 11/780 computer with the Munich program system.

The protein crystallography group has made effective use of the restrained least squares (Hendrickson and Konnert 1980) and the constrained-restrained least squares (CORELS) (Sussman et al 1977) methods to refine the protein structures described below (Kannan and Ramanadham 1981; Ramanadham et al 1980, 1981; Ramanadham 1982). The program systems used for this purpose have been implemented on several computers. All the above programs have been used extensively to refine the human carbonic anhydrase B and triclinic hen egg white lysozyme structures. Interpretation of the electron density maps also requires a complete set of contour maps drawn on transparent plastic sheets, as a number of sections of the unit cell are stacked together in a suitable way. This is done by the program PLUTO which also gives the molecular skeleton plotted and superposed on the contour maps. This program is operational on Prime 450. A fast fourier transform (FFT) program to calculate the electron density sections which is input to PLUTO is also operational on Prime 450.

For the use of the molecular replacement method the fast rotation function program of Crowther (Rossmann 1972), the slower rotation function program of Rossmann (1972) as also the translation function program are available. Stereo-drawings of the molecular model or selected parts of the molecule are obtainable by the use of the program ORTEP2.

4. Structure refinement and function of proteins

We give below a summary of the protein crystallographic studies carried out at Trombay at the same time as the necessary infrastructure facilities for experimental macromolecular crystallography research were being built up.

4.1 Presynaptic toxins

A programme of structure-function studies on presynaptic toxin phospholipases has been initiated at Trombay. The presynaptic toxins are widely distributed in nature and are present in good quantities in the venom of some snakes. They are broadly classified into monomeric toxins and multimeric toxins. They also possess phospholipase A2 activity and the amino acid sequence shows very large homology with snake venom, bovine and porcine phospholipases. The Australian tiger snake (Notechis Scutalus Scutalus) is a rich source of the presynaptic toxins of the monomeric type. There are a number of fractions (Halpert and Baker 1976) which show toxicity and phospholipase activity and also some which have only sequence homology but no toxicity or phospholipase activity. Three of these proteins are under investigation at Trombay. They are Notechis II-4 or Notexin, Notechis II-1 and Notechis II-5. Of these Notechis II-1 shows no toxicity and Notexin is more toxic than Notechis II-5. These toxins are
basic proteins with 119 amino acid residues cross-linked by several disulphide bridges. Notechis II-5 was the first protein to be crystallised and characterised in India and was reported by Kannan et al (1981). The homology with porcine and bovine phospholipases will be used and the structure solution attempted by molecular replacement methods. It may be possible to arrive at the toxicity centre of these toxins by comparing their three-dimensional structures which will help in finding the mechanism of presynaptic toxin action. The unit cell dimensions of these two toxins (Kannan et al 1977a) are

Notexin : $a = b = 75.03$ Å, $c = 49.04$ Å, $\alpha = \beta = 90^\circ$ and $\gamma = 120^\circ$. 6 molecules in a trigonal unit cell with space group $P3_121$ or $P3_221$

Notechin II-5 : $a = 146$ Å, $b = 43.5$ Å and $c = 39.0$ Å, $\alpha = \beta = \gamma = 90^\circ$. 8 molecules in an orthorhombic unit cell with space group $P2_12_12_1$.

4.2 Lysozyme

Lysozyme was the first enzyme and the second protein whose structure was resolved at high resolution by x-ray diffraction technique (Blake et al 1965). The enzyme lyses polysaccharide-like polymers of N acetyl glucosamine and N-acetyl muramic acid.

Lysozyme is a monomer of 129 amino acids of molecular weight 14 600. There are four disulphide bridges. Hen egg white lysozyme has been crystallised in a number of crystallographic forms. Blake et al (1965) investigated the structure of the tetragonal form of the enzyme crystals. Jensen and colleagues (Ramanadham et al 1980) have studied the triclinic modification for which x-ray diffraction data extends to about 1 Å resolution. The initial structure in the triclinic unit cell was derived by the application of molecular replacement methods and subsequently the structure was refined by difference fourier methods (Ramanadham et al 1980). The structure was then refined by Ramanadham et al (1981) by restrained least squares method.

The restrained least squares refinement of triclinic hen egg-white lysozyme at 2 Å resolution was initiated and carried out to an $R$-factor of 0.179 at the University of Washington, Seattle, by Jensen and colleagues (Ramanadham et al 1980). This work has been continued at Trombay in collaboration with the Seattle group. The present $R$-factor is 0.169.

The restrained least squares method utilizes the known geometry of the different protein groups and introduces various geometrical restraints in the least squares procedure as a set of additional observations, thus improving the overdeterminacy ratio.

The quantity minimised is

$$\Phi = \Phi_1 + \Phi_2 + \Phi_3 + \Phi_4 + \ldots,$$

where $\Phi_1$ is the conventional structure factor term in the least squares procedure.

$$\Phi_1 = \sum (|F_o| - |F_c|)^2,$$

$\Phi_2, \Phi_3, \text{etc.}$ are terms for the different restraints for planar groups, atomic distances, chiral volumes, non-bonded contacts, isotropic thermal parameters, conformational parameters, symmetry features and so on (Hendrickson and Konnert 1980).

In the refinement of Lysozyme (Ramanadham et al 1981) 4961 parameters were adjusted by the restrained least squares procedure. There were 1001 protein and 239 solvent atoms and one scale factor as against 7131 independent $F_{obs}$ structure
amplitude values within 2 Å resolution and about 3000 restraints. The root mean square deviation in the distances between different atoms from ideal values were less than 0.05 Å for the final model.

Protein crystals suffer from static and dynamic disorder and it would be difficult to assign a meaning to the temperature factor in a conventional sense. However it has been noticed that the temperature factor of the protein atoms in triclinic lysozyme was very low compared to many other protein structures. The average B value for the molecule is about 8.0 Å², much smaller than in tetragonal lysozyme and this seems to be related to a larger fraction of the water molecules being ordered in this structure (Ramanadham et al. 1981).

Mn²⁺, Co²⁺ and Gd³⁺ have been found to bind to the active site of Lysozyme (Kurachi et al. 1975 and Teichberg et al. 1974). It has also been observed that divalent copper ions inhibit lysozyme. The binding of Cu²⁺ to tetragonal lysozyme has been reported by Teichberg et al. (1974), Ramanadham et al. (1981) have studied the binding of Cu²⁺ to triclinic lysozyme at Trombay using the 2 Å data collected at University of Washington. The structure of the Cu²⁺ lysozyme complex was refined by restrained least squares method using the refined coordinates of the triclinic lysozyme structure to a crystallographic R-factor of 16.9 % after two cycles of refinement. Ramanadham et al. (personal communication) have not observed any significant deviations of the protein atoms on copper binding. They had observed that one Cu²⁺ ion was found in the vicinity of Glu 35 in the active site and a second copper ion was bound to the enzyme in the vicinity of His 15 residue.

4.3 Human carbonic anhydrase

Mammalian erythrocytes are a rich source of carbonic anhydrase enzyme which reversibly catalyses the interconversion of bicarbonate to carbon dioxide and water. The human erythrocytes contain, apart from the major component haemoglobin, two carbonic anhydrase isoenzymes, form B (hcab) and form C (hcac). The two isoenzymes are homologous to each other with about 260 amino acid residues and an essential Zn²⁺ ion and have a molecular weight of 30000. There are no disulphide bridges in these enzymes. The isoenzymes differ in their catalytic efficiency. Aromatic sulphonamides which are specific inhibitors of these enzymes also exhibit differing inhibitory properties. The enzymes, however, are highly efficient in the catalysis of CO₂ bicarbonate interconversion with a turnover rate of 10⁶ moles of CO₂ per mole of enzyme per sec. The mechanism of action of these fascinating enzymes are still hotly debated by the different workers in the field (Lindskog et al. 1971).

The three-dimensional structure of hcab (Kannan et al. 1975) and hcac have been known for some time (figure 1) (Kannan et al. 1971 and Liljas et al. 1972). Kannan et al. (1981) have been refining the structure of hcab at 2 Å resolution to get the complete tertiary structure of the enzyme including the solvent structure in the crystals. The solvent structure may play an important role in the mechanism of action of the enzyme. The enzymes also show a different folding pattern compared to many of the other protein structures known. The molecule is twice as large as lysozyme discussed above. The diffraction data had been collected at Uppsala, Sweden on photographic films using precession cameras (Kannan et al. 1975) during the course of the high resolution structure determination. The coordinates of hcab had been measured from a Kendrew Watson skeletal model and the model adjusted to conform to stereochemical standards.
using the model building program (Diamond 1971). The crystallographic $R$-factor was about 50%. The model had been fitted to the isomorphous map in a 3D graphics display by one of us at Max Planck Institute for Biochemistry, Munich, Germany using the programme \textit{frodo} (Jones 1978) and the resultant model had been refined with the real space refinement program (Diamond 1974) on a Siemens 4004 computer at Munich. The $R$-factor was 39.5% for the 14578 observed reflections in the 5 to 2 A sphere. The weak unobserved reflections had been dropped at this stage.

The structure was then refined by Kannan and Ramanadham (1981) at Trombay using the restrained least squares programs, discussed earlier in the section on lysozyme, to 36.5% $R$-value for the 3 A resolution data and later to 32.3% for the 14524 reflection in the 5 to 2 A sphere where individual isotropic temperature factors were used in the refinement. However the improvements in $R$-factors after 8 cycles of least squares refinement tapered off. The model was now compared with the electron density maps and wrong orientation of the residues corrected. The three-dimensional graphics facility at the Wallenberg Laboratory, Uppsala, Sweden was used during a visit by one of us for this purpose in collaboration with Dr T A Jones (Kannan \textit{et al} 1982). The model was refitted to the isomorphous electron density map in the graphics system. Kannan \textit{et al} (1982) have reported that the model required refitting in a number of regions and also a number of side chains and even main chain atoms dropped from the refinement due to difficulties in fitting to the isomorphous electron density map. Of the total of 2200 protein atoms they had included only about 1850 atoms in the constrained-restrained least squares procedure \textit{corels} with one overall temperature factor. The $R$-factor had dropped after 2 cycles from 45.5% for the refitted model to 36% for the 6 to 2 A data.
They then fitted the model to a \((|2F_o| - |F_c|)\) electron density map in the graphics system. In this process a number of residues which were left out from the previous calculation were located and included in further refinement. In all they had done alternatively 4-model fitting sessions to the electron density maps on the interactive graphics system and 8 cycles of corels. At this stage only 4 residues at the amino terminus, some side chains especially lysines and all the solvent atoms were not included. The R-factor had dropped to 31\% with one overall temperature factor and to 26-7\% while refining two group temperature factors for every residue in the last two cycle (Kannan et al 1982).

The refinement has since been continued at Trombay. About 60 solvent molecules located within hydrogen bonding distance to relevant proteins groups from \((|F_o| - |F_c|)\) electron density map were included and individual restrained isotropic temperature factors refined for all atoms using the restrained least squares procedure outlined in the section on lysozyme. After 2 cycles the R-factor had dropped to 22.7\%. The refinement using corels also gave an almost identical result giving a R-factor of 22.5\%. An inspection of the \((|F_o| - |F_c|)\) electron density map resulted in 40 more solvent atoms being located, within hydrogen bonding distance to the relevant protein group or solvent atoms located in the earlier \((|F_o| - |F_c|)\) map. The interactive three-dimensional graphics facility became available at this time at Trombay and the frodo program (Jones 1978) system was made operational on the PDP 11/34 computer system connected to the graphics system by one of us. A \((|2F_o| - |F_c|)\) map was then used to fit the refined coordinates and the solvent atoms (Kannan et al, personal communication). They have reported correcting the wrongly oriented side chains, especially the forked residues, for a number of side chains. They also located another 60 solvent atoms from an inspection of the \((|2F_o| - |F_c|)\) and the \((|F_o| - |F_c|)\) maps. A number of side chain conformational angles were also corrected to near-staggered orientation in conformity with the electron density on the graphics system. It was observed by them that none of the least squares refinement programs was capable of correcting such errors and often forcing the residues to false minima. Thus only manual intervention could rectify such faults. After three more cycles of restrained least squares refinement the R-factor had dropped to 19.5\% (figure 2) for the reflections in the 5 to 2 Å sphere. All the 15524 reflection in the 10–2 Å sphere were then used by them, and a \((|2F_o| - |F_c|)\) map and a \((|F_o| - |F_c|)\) map computed and used on the graphics system for further improvement in the model-electron density fittings and solvent localization. Kannan et al (personal communication) were able to identify two more of the amino terminal residues, all but two lysine side chains and an arginine side chain as also about 150 solvent atoms either hydrogen-bonded to the protein or to the solvents, located earlier. Thus about 300 solvent atoms have been added to the protein structure so far.

They also have found some interesting solvent structure in the HCA B molecule. They found that the catalytically important Zn^{2+} has a solvent coordinated to it and is within hydrogen bonding distance to Thr 199. There is a solvent molecule bridging Glu 106 and Tyr 7 by hydrogen bonding to them. Glu 106 was also found hydrogen-bonded to Thr 199. A solvent hydrogen-bonded by Tyr 6 was also found hydrogen-bonded to His 64 located in the active site of the enzyme. Another interesting feature of the solvent structure seems to be their participation in breaking pleated sheet hydrogen bonding. A few such solvents have been located by Kannan et al (private communication).

Another important feature reported by Kannan et al is regarding the temperature factors. Internal residues and residues in the pleated sheet structure have very low
temperature factor, in some instance 2.0 Å², and residues on the surface of the molecule, especially lysines, and other polar residues have very high temperature factors, sometimes as high as 30 Å². The solvents also have varying temperature factor from 6 Å² to 60 Å² depending on their location and hydrogen bonding. The average B-value for all the protein atoms is 12 Å².

The refined coordinates of \textit{hcab} has been used to calculate \(|2F_o| - |F_c|\) maps for a number of inhibitor complexes and one metal replacement for 2 Å²⁺ at Trombay. The results of these investigation agree well with those reported earlier by Kannan \textit{et al.} (1977b and 1977c) and Kannan (1980a and 1980b). The mechanistic proposal made by Kannan \textit{et al.} (1977b) would still be valid with minor modification in the light of the solvent structure found during the course of the refinement.

5. Conclusion

Neutron beam research at Trombay on small biomolecules will be enlarged in scope with the advent of the higher-flux R-5 reactor to include small angle neutron scattering studies of large biomolecules and bio-aggregates. We are now also fully equipped at Trombay to undertake macromolecular crystallography research with x-rays. The experimental facilities have been installed and a complete library of computer programs established. Meanwhile a number of protein structures have been refined and interesting results obtained on protein dynamics and structure-function relationships. We have also taken up the study of several new proteins.
Macromolecular crystallography research at Trombay

References

Chidambaram 1981 Int. J. Quant. Chem. 20 449
Diamond R 1971 Acta Crystallogr. 21 253
Diamond R 1974 J. Mol. Biol. 82 371
Halpert J and Eaker D 1976 J. Biol. Chem. 251 7343
Jones T A 1978 J. Appl. Crystallogr. 11 268
Kannan K K 1977 Molecular biology, biochemistry and biophysics: Advanced methods in protein sequences determination (ed.) S B Needleman (Berlin: Springer-Verlag) vol. 25 p 75
Kannan K K 1980b in Biophysics and physiology of carbon dioxide (eds) H Bastels, G Bauer and G Cross (Berlin: Springer-Verlag) p 164
Kannan K K, Lovgren S, Cid-Dresdner H, Petef M and Eaker D 1977a Toxicon 15 434
Kannan K K and Ramanadham M 1981 Int. J. Quant. Chem. 20 199
Kannan K K, Petef M, Fridberg K, Cid-Dresdner H and Lovgren S 1977b FEBS Lett. 73 115
Kurachi K, Sieker L C and Jensen L H 1975 J. Biol. Chem. 250 7663
Pal P K, Chakravarty S and Kannan K K 1983 Abstract submitted to international summer school on crystalographic computing, Kyoto, Japan
Ramanadham M 1982 Special symposium on Molecular biophysics and biocrystallography in honour of Prof. G N Ramachandran, Madras India Abstract A9
Ramanadham M, Seiker L C, Jensen L H and Birknes, B J 1981 X11th congress and general assembly of the international union of crystallography, Ottawa, Canada, Abstract 2.1
Richards F M 1968 J. Mol. Biol. 37 225
Rossmann M C 1982 The molecular replacement method, A collection of papers on use of non-crystallographic symmetry (New York: Gordon and Breach) p 4
Teichberg V I, Sharaon N, Moult J, Smilansky A and Yonath A 1974 J. Mol. Biol. 87 357
On the polymorphism of the smectic A phases of highly polar compounds

N V MADHUSUDANA
Raman Research Institute, Bangalore 560080

Abstract. The smectic A phase is characterized by a layered arrangement of rod-like molecules, with the average orientation of the molecules normal to the layers and a liquid-like organization within the layers. When the molecules have a strongly polar end group (like \(-\text{C}═\text{N}, -\text{NO}_2\), etc.), neighboring molecules prefer an antiparallel orientation. This results in a bilayer structure, with a layer spacing much larger than the molecular length. In many compounds, the structure of the bilayer is sensitive to temperature (or pressure). This results in some interesting phenomena: (a) reentrant phases, as for example, the appearance of a nematic phase below the temperature range of existence of the A phase; (b) a number of polymorphic forms of smectic A; and (c) an 'induced smectic A' phase when the compounds with the highly polar terminal groups are mixed with weakly polar compounds even when neither compound by itself exhibits the A phase. This article gives a brief summary of recent results on all these aspects.

Keywords. smectic A; polymorphism; reentrant phases; bilayer smectics; induced smectic phases.

1. Introduction

The smectic phases of liquid crystals are characterized by a layered arrangement of rod-like molecules (for a general introduction to liquid crystals, see Chandrasekhar, 1977; de Gennes 1975). Smectic A is the simplest kind of smectic phase with a liquid-like organisation within the layers, and the average orientation of the long axes of the molecules (indicated by a unit vector \(\mathbf{n}\) called the director) normal to the layers. A drawing of the structure is shown in figure 1. However this figure is purely schematic. In reality, the periodicity along the \(z\)-axis is very well described as a sinusoidal density wave (McMillan, 1971; de Gennes 1972) given by

\[
\rho = \rho_0[1 + \text{Re}\{\psi e^{i(q_0z + \phi)}\}]
\]  

Figure 1. An idealised schematic drawing of the arrangement of molecules in the smectic A phase.
where \( q_0 = 2\pi/d, \) \( d \) being the layer spacing and \( \phi \) a phase factor which fixes the position of the layers and \( |\psi| \) the amplitude of the density wave (figure 2). Indeed even the second order maximum is hardly seen in the x-ray diffraction pattern from most of the compounds in the A phase (Schaetzing and Litster 1979). The first order reflection is also not a true Bragg spot but has a power law decay of the intensity as one moves away from \( q_0 \). This is a consequence of the well-known Landau-Peierls instability of an one-dimensional periodic order (Landau and Lifshitz 1959; Peierls 1934).

The chemical structure of a typical mesogenic compound with rod-like molecules has a central aromatic core which is quite rigid and a relatively flexible alkyl chain at one or both ends (figure 3). Experimentally it is well known that the lower homologues (i.e., compounds with shorter alkyl chains) exhibit only the nematic (N) phase which is characterised by a purely orientational long range order. The smectic phase makes its appearance when the alkyl chain is relatively long. The first member exhibiting the A phase shows a relatively weak A–N transition, but as the chain length is increased, the strength of the transition increases and beyond a certain chain length, the A phase goes over directly to the isotropic phase. These observations can be understood by assuming that the strongest intermolecular attractive interactions are confined to the aromatic cores of the molecules. This would naturally lead to the formation of a layered arrangement of molecules with relatively long alkyl chains at both ends. This is the physical idea behind the molecular theory of the A phase proposed by McMillan (1971). Assuming that the translational and orientational orders are coupled, he showed that the A–N transition can be of second order character if the orientational order at the transition point is sufficiently high.

As we saw earlier, the order parameter characterising the A phase has two components viz the amplitude and the phase of the density wave (equation (1)). This is rather analogous to the order parameter of a superconductor. Further, because of the layered arrangement, the A phase does not allow a curl \( \mathbf{n} \) distortion which requires a change of layer spacing. This is analogous to the expulsion of a magnetic field by a

---

**Figure 2.** A realistic distribution of molecules in the A phase (from Schaetzing and Litster 1979).

**Figure 3.** Chemical structure of a typical mesogenic compound.
superconductor. The close analogy between the (second-order) A–N transition and superconductor-normal metal transition was recognised by de Gennes (1972) who wrote the relevant Landau-Ginzburg free energy functional. Using this analogy, he predicted the divergence of twist and bend elastic constants as $T_{AN}$ (the A–N transition point) is approached from the nematic side. [For a review of the experimental situation up to 1978 see Chandrasekhar and Madhusudana 1978. For recent results see Als-Nielsen 1981]. He also predicted a reduction in $T_{AN}$ due to a curl n type of distortion (which was only recently confirmed experimentally; Madhusudana and Srikanta 1982). However the analogy is not exact. For example, it has been found that $\xi_\parallel$ and $\xi_\perp$, the correlation lengths parallel and perpendicular to the director diverge with different critical indices as $T_{AN}$ is approached. The theory of A-N transition still continues to be a very active area of investigation (see for example, Als-Nielsen 1981).

In this article we shall confine our attention to some interesting features exhibited by the smectic A phase when the compound has a highly polar end group. Many compounds have been synthesised in recent years which have either a cyano (–C≡N) or nitro (–NO$_2$) end group, both of which have a longitudinal dipole moment of ~ 4 Debyes. Till about a decade ago the longitudinal dipoles were thought to play only a minor role in stabilising the nematic and smectic A phases, especially after it was recognized that Born's theory of the nematic phase (Born 1916) which is based entirely on interactions between permanent dipoles is inadequate (Maier and Saupe 1959). We proposed that for minimising the mutual interaction energy, neighbouring molecules with strong longitudinal dipoles should favour an antiparallel arrangement (Madhusudana and Chandrasekhar 1973; Madhusudana et al 1977). This was amply confirmed by x-ray and neutron scattering studies (Leadbetter et al 1975, 1979; Lyden and Coakley 1975) which showed that in alkyl (or alkoxy) cyanobiphenyls the smectic A phase is characterized by a layer spacing which is ~ 1.4 times the molecular length (figure 4). As we pointed out earlier, the attractive interaction between neighbouring molecules would favour aromatic regions to overlap, and the antiparallel correlation

Figure 4. Schematic diagram of antiparallel local structure in pentyl-cyanobiphenyl (Leadbetter et al 1975).
brought about by the strong dipole moments would then lead to a structure as shown in figure 4. Another way of describing such a structure would be to say that two such molecules would form an antiparallel ‘pair’, which should be a relatively stable unit because of frustration effects which any additional molecule would be subjected to in the neighbourhood of such a pair. An interesting consequence of the formation of such pairs is that they are much more symmetric than the individual molecules which are highly asymmetric because of the presence of the alkyl chain only at one end of the aromatic core. The pairs have chains at both ends and the McMillan argument regarding the conditions of formation of the smectic A phase is now applicable for the pairs and thus many such compounds with cyano end groups exhibit the A phase (de Jeu 1982). The A phase thus formed is referred to as a bilayer A phase for obvious reasons. Even if the chain length is too short to form the A phase, such compounds exhibit a strong smectic-like short range order in the nematic phase, the short range order having bilayer characteristics. The structure of the bilayer should depend on the chemical structure of the aromatic core. Further, in any given case the bilayer structure is an extra degree of freedom, which can be influenced by variations in temperature, pressure or, in the case of mixtures, the composition. This in turn leads to several interesting phenomena which will be discussed in the following sections.

2. Reentrant phases

Cladis (1975) discovered that some binary mixtures of compounds with cyano end groups exhibit the following sequence of transitions as the mixture is cooled from the isotropic phase:

\[
\text{isotropic} \rightarrow \text{nematic} \rightarrow \text{smectic A} \rightarrow \text{nematic}.
\]

The translational order of the A phase is lost both when it is heated and cooled. The nematic phase occurring at temperatures lower than the range of stability of the smectic A phase is called the ‘reentrant’ nematic (\(N_R\)) phase, in analogy with similar reentrant phases in certain superconductors doped with magnetic impurities (Riblet and Winzer 1971) and the solid-liquid \(^3\text{He} \) transition (Anderson et al. 1963).

Subsequently Cladis et al. (1978) found that some pure compounds exhibited the reentrant nematic phase at elevated pressures and a little later pure compounds were found which exhibited the reentrant phase even at atmospheric pressure (Madhusudana et al. 1979; Hardouin et al. 1979; Subramanya Raj Urs and Sadashiva 1982). During the past 3 years a large number of compounds have been synthesized which exhibit reentrant phases. (For a recent review on the chemistry of these compounds, see Tinh 1983). Table 1 gives a representative list of compounds which exhibit reentrant phases. It is seen that some compounds not only exhibit a reentrant nematic phase, but also a reentrant smectic A (\(A_R\)) phase which is obtained on cooling the \(N_R\) phase.

Though one might expect that changes in the bilayer structure should be responsible for the occurrence of the \(N_R\) phase, x-ray investigations do not reveal any universal trend in the thermal variation of the bilayer spacing close to the \(AN_R\) transition point. In some cases which exhibit only the \(N_R\) phase below the A phase, the bilayer expands as the sample is cooled to \(T_{AN_R}\) (Chandrasekhar et al. 1979). This expansion is also reflected in the low frequency dielectric dispersion which arises due to the flipping of the molecules about their short axes, the corresponding activation energy changing...
Table 1.

<table>
<thead>
<tr>
<th>STRUCTURAL FORMULA AND NAME</th>
<th>SEQUENCE OF TRANSITIONS AS THE COMPOUND IS COOLED</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \text{\parbox{0.2\textwidth}{\begin{align*} \text{H} &amp; \text{-C} = \text{C} - \text{C}<em>{11} \text{H}</em>{21} \text{O} - \text{O} - \text{CN} \ \text{OH} \end{align*}}} ]</td>
<td>173.5 N 56.8 A_d 32.5 N_r</td>
</tr>
<tr>
<td>[ \text{\parbox{0.2\textwidth}{\begin{align*} \text{H} &amp; \text{-C} = \text{C} - \text{C}<em>{12} \text{H}</em>{23} \text{O} - \text{O} - \text{CN} \ \text{CH}_{3} \end{align*}}} ]</td>
<td>1148 N 138.5 A_d 59.8 N_r</td>
</tr>
<tr>
<td>[ \text{\parbox{0.2\textwidth}{\begin{align*} \text{H} &amp; \text{-C} = \text{C} - \text{C}<em>{13} \text{H}</em>{27} \text{O} - \text{O} - \text{CN} \ \text{CH}_{3} \end{align*}}} ]</td>
<td>1283 N 248 A_d 138 N_r 94.5 A_i</td>
</tr>
<tr>
<td>[ \text{\parbox{0.2\textwidth}{\begin{align*} \text{H} &amp; \text{-C} = \text{C} - \text{C}<em>{14} \text{H}</em>{29} \text{O} - \text{O} - \text{CN} \ \text{CH}=\text{CH} \end{align*}}} ]</td>
<td>1257 N 257 A_d 87 C. 70 N_i</td>
</tr>
</tbody>
</table>

distinctly (Ratna et al. 1979). The activation energy of the N_r phase itself is much larger than in the normal N phase. But in compounds exhibiting the A_r phase also, the trend is quite different. The bilayer spacing decreases rapidly from \( \sim 1.25 \) \( \text{nm} \) to \( \sim 1.1 \) \( \text{nm} \) as the temperature of the A phase is lowered to \( T_{A_{AN}} \). In the reentrant smectic A phase there is a temperature independent layer spacing which is \( \approx l \), i.e., the A_r phase is a monolayer smectic (Hardouin and Levelut 1980). Moreover, the A_r phase gives rise to a diffuse scattering of x-rays indicative of a two-dimensional superlattice type of modulation of the layers with periodicities of \( \sim 70 \) \( \text{Å} \) in the plane of the layer and a periodicity of \( \sim 4 \) layers along the z-direction (figure 5). The activation energy of the low frequency

Figure 5. Schematic diagram of the structure of the reentrant smectic phase. The arrows represent the cyano-end groups of the molecules.
dielectric dispersion in the $A^\text{r}$ phase is nearly twice that of the bilayer $A$ phase (Benguigui and Hardouin 1981). Further, the $N_{A^\text{r}}$ phase has fluctuations of both $A_4$ and $A^\text{r}$ phases, as revealed by x-ray and dielectric studies.

Most of the compounds exhibiting the reentrant phases have three benzene rings with some bridging groups in between in the aromatic core excepting two compounds which have only two benzene rings (see table 1). The length of the core appears to play an important role and Tinh (1983) has concluded that the core has to have a length lying between 16 Å and 22 Å for the compounds to exhibit the reentrant phase.

There have been some attempts to develop a theory for the occurrence of the $N_{A^\text{r}}$ phase on the basis of the antiparallel orientations of neighbouring molecules. Assuming that associated pairs are formed in the medium, Cladis (1979) argued that as the temperature decreases (or equivalently the pressure increases) the stability as well as the number of such pairs should increase. The associated pairs are obviously much less polar than the individual molecules. If it is now assumed that the lateral interactions leading to the formation of layers are greater if the molecules are effectively more polar; it is clear that the smectic phase stability is decreased as the pairing is more efficient. Further, for the aromatic part of the molecular pairs, it will not be sterically advantageous to remain in the layers beyond a certain density. Equivalently, one can assume that there is an optimum density for the stability of the smectic $A$ phase of the bilayer kind. Pershan and Prost (1979) have used this idea in developing a Landau theory of the reentrant nematic-bilayer smectic $A$ transition. By adding a term of the form $
abla\rho|\psi|^2$ to the usual Landau theory of the $A-N$ transition where $|\psi|$ is the smectic order parameter, and expanding $
abla\rho$ to the leading power in $(\rho - \rho_0)$ where $\rho_0$ is the optimum density, they can account for both $A-N$ and $A-N_{A^\text{r}}$ transitions. Further, using this model, they can account for the experimental observation that the bilayer $A$ phase is bounded by the $N$ phase in the PT plane (Cladis et al 1977; for a more detailed account of the high pressure studies, see the article by Shashidhar 1983). Indeed the $A$ phase is also bounded in phase diagrams of binary mixtures of two cyano compounds, if only one of the components exhibits the $N_{A^\text{r}}$ phase (figure 6). This can also be accounted for by using the Landau model if $\nabla\rho$ is taken to be a function of both density and composition (Pershan and Prost 1979). Thermodynamic models to explain such nonlinear phase diagrams have also been developed (Clark 1979; Klug and Whalley 1979; Heppke and Schneider 1979; Billard 1979).

Longa and de Jeu (1982) have given a detailed molecular theory of the $A-N_{A^\text{r}}$ transition. By treating the system as a mixture of monomers and dimers interacting through both attractive and hard-core repulsive interactions, they essentially work out the requirements for the occurrence of the optimum density that was referred to in the previous paragraph. They show, using the framework of the McMillan model of the $A$ phase, that unfavourable packing of the dimers in planes leads to $N_{A^\text{r}}$ phase for $d/l \simeq 1.3-1.4$. We might also mention here a paper by Hida (1981) who has developed a lattice theory of the $A-N_{A^\text{r}}$ transition by assuming that dimer pairs occupy an fcc lattice. Allowing for vacancies in such a lattice, he shows that reentrance is possible for certain model parameters.

Another model for the occurrence of the reentrant phases was proposed by Berker and Walker (1981). It is based on the competition between short-range positional order in the layers and long-range antiferroelectric order which arises due to interactions between the dipole moments. This is called the ‘frustrated spin-gas’ model, since, as we mentioned earlier, frustration is inherent in a triangular (or equivalently hexagonal)
Polymorphism of the smectic A phases of highly polar compounds

Figure 6. Phase diagram of the binary mixtures of 4-cyano-benzylidene-4'-alkoxyanilines. The lines are calculated on the basis of a thermodynamic theory (Heppke and Schneider 1979).

Figure 7. The frustrated spin model for the occurrence of $N_R$ phase. A triangular lattice (b) in which two bonds are stronger than the third can form a network. As the density is increased, the molecules would like to be in a close packed triangular lattice (a), thus bringing in 'frustration' and hence a transition to the $N_R$ phase.

Close packing of the molecules in the layers: if two molecules have already formed an antiparallel pair, a third molecule which would try to sit on a triangular lattice with respect to the first two would be subjected to 'frustration' in respect of its orientation. The authors propose that this can be relieved by packing the molecules in a triangular lattice such that there are two strong bonds and a weak bond, thus leading to the formation of a network (figure 7), and this obviously favours the formation of a (bilayer) smectic A phase. As the density is increased, the molecules have to come closer, and pack in a regular triangular lattice. Frustration effects will destroy the network and
the system goes over to the re-entrant nematic phase. By assuming attractive interactions between the alkyl chains, a reentrant monolayer smectic A phase is predicted to occur with a ferroelectric order in the layers. However, it may be pointed out that according to this model, the bilayer would have a long range antiferroelectric order in the networks, and this should have profoundly influenced both the static and dynamic parts of dielectric anisotropy. The measurements mentioned earlier do not appear to support this model.

Prost (1980) points out that the reentrant nematic phase occurring between the bilayer and monolayer A phases has fluctuations of both types; and it is thus a result of the mutual exclusion of the two types of smectic phases which are competing for condensation. The reentrant phenomenon in this model comes out as one of the possibilities in a system in which two lengths are important, the other possibilities being different modifications of the smectic A phase which will be discussed in the next section.

On the experimental side, some of the more recent findings include a sequence NACN$_N$ (Tinh et al 1982b; The Halle group has found many mixtures with such a sequence, see Pelzl et al 1981), a (tilted) smectic C phase appearing between the A and N$_R$ phases (see table 1) and a more exotic triple reentrant behaviour which will be referred to in the next section.

3. Polymorphism of smectic A

X-ray investigations of compounds with cyano end groups have shown various other remarkable phenomena in recent years. For example, in pentyl cyanoterphenyl, a compound which exhibits only a nematic phase, short-range correlations corresponding to two incommensurate wavelengths (at $\pi /l$ and $\approx 1.5 \pi /l$) have been found (Brownsey and Leadbetter 1980), just as in the N$_R$ phase separating A and A$_R$ phases discussed in the previous section.

$4$-$n$-Pentylphenyl-$4'$-($4''$-cyanobenzoxyloxy)benzoate (hereafter called DBS, see table 2) has two ester linkage groups in the aromatic core which are arranged such that their dipole moments are essentially antiparallel to the dipole moment of the cyano end group. This compound has an NA transition which occurs at $\approx 110^\circ$ below $T_{NI}$, and according to the McMillan criterion (McMillan 1971) should have had only a second order NA transition. Actually, however, it has a moderately large first order NA transition. Meyer and Lubensky (1976) accounted for this by developing a mean field theory of the A phase with a coupling between the first and second translational order parameters. Indeed as we shall see presently, DBS and similar compounds have been found to possess a bilayer spacing which is close to twice the molecular length, which in turn give rise to a rich variety of smectic A polymorphism.

The first direct evidence for the coupling between the first and second fourier components of the density wave was obtained in a mixture of DBS with a weakly polar compound, viz terephthal-bis-butyylaniline (TBB$A$) which exhibits a monolayer A phase (Sigaud et al 1979). Calorimetric and x-ray investigations clearly indicate a first order transition from a monolayer phase (A$_1$) to a bilayer phase whose layer spacing is $\approx 2l$ (A$_2$ phase). Figure 8 shows the proposed molecular arrangement in the two phases. Pure DB7, a higher homologue DBS (see table 2) exhibits at higher temperatures of its A phase a layer spacing of $\approx 1.6l$ with a coexisting diffuse spot corresponding to $\approx l$. 
Table 2.

<table>
<thead>
<tr>
<th>COMPOUND or MIXTURE</th>
<th>SEQUENCE OF TRANSITIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>nC₅H₁₁—O—OC—O—OC—CN</td>
<td>83.4 °C, N₁₃₄ A₁₁₇ A₂</td>
</tr>
<tr>
<td>4-n-pentylphenyl-4′-(4-cyanobenzyloxy)benzoate (DB₅)</td>
<td>16.6 °C, N₁₂₃ N₁₇₂ A₄₆₈ A₂</td>
</tr>
<tr>
<td>C₄H₉—N=CH—CH=N—C₅H₉</td>
<td>83.4 °C, N₁₃₄ A₁₁₇ A₂</td>
</tr>
<tr>
<td>terephthal-bis(-p-butylole) (TBBA)</td>
<td>16.6 °C, N₁₂₃ N₁₇₂ A₄₆₈ A₂</td>
</tr>
<tr>
<td>nC₅H₁₅—O—OC—O—OC—CN</td>
<td>46.7 °C, N₁₄₅ A₁₁₂₅ A₁₁₆ A₂</td>
</tr>
<tr>
<td>4′-(4-n-pentylstyryl) phenyl-4′-cyanobenzoate (C₅ stilbene)</td>
<td>46.7 °C, N₁₄₅ A₁₁₂₅ A₁₁₆ A₂</td>
</tr>
<tr>
<td>nC₅H₁₅—O-OC—O-OC—NO₂</td>
<td>54.4 °C, N₁₄₅ A₁₁₂₅ A₁₁₆ A₂</td>
</tr>
<tr>
<td>4-n-heptylphenyl-4′-nitrobenzyloxybenzoate (DB₇N₂O₂)</td>
<td>N₉₅ A₁₉₀ A₂</td>
</tr>
<tr>
<td>nC₅H₁₃—O—OC—O—OC—CN</td>
<td>N₁₃₉.₅ A₁₂₄.₅ A₂</td>
</tr>
<tr>
<td>4-n-decylphenyl-3-methyl-4′-(4′-cyanobenzyloxy) benzoate (10 PNCBB)</td>
<td>N₁₃₉.₅ A₁₂₄.₅ A₂</td>
</tr>
<tr>
<td>nC₅H₁₃O—O—OC—O—OC—NO₂</td>
<td>122.4 N₁₉₅ A₉₁₅₈ A₁₀₂</td>
</tr>
<tr>
<td>(DB₉N₂O₂)</td>
<td>127 N₁₂₄ A₁₁₈ S₉₆ C₂</td>
</tr>
</tbody>
</table>

Figure 8. The molecular arrangement in (a) the A₁ phase, and (b) the A₂ phase.

This phase with \( l \leq d \leq 2l \) is called the A₉ phase. (All the bilayer A phases of compounds exhibiting reentrant nematic phases belong to this class.) As the temperature is lowered, a phase transition occurs such that in the lower temperature phase, there are two Bragg spots corresponding to \( d \approx l \) and \( d \approx 2l \), i.e. there is a commensurate lock-in of the two wavelengths in the lower temperature A₂ phase.

Yet another type of smectic A (known as smectic antiphase, A) is found in some pure compounds and mixtures. For example, some mixtures of DB₅ and C₅ stilbene (table 2) show a high temperature A₁ phase (with \( d = 2\pi/q₂ \approx l \)) with a diffuse scattering at wave vector \( q₁ \) which is split such that \( q₂/q₁₂ \approx 2 \). On cooling this compound, a phase transition occurs such that \( q₂ \) is not altered but diffuse maxima corresponding to \( q₁ \) get condensed. As a result, a superlattice long range order with
\[ 2\pi /q_{1x} \approx 130 \text{ Å} \] develops in the *lateral* direction (figure 9). Cooling the sample further, yet another transition takes place such that a collinear commensurate locking of the two wavelengths occurs along the \( z \)-direction. In other words, the lateral superlattice collapses in favour of an \( A_1 \) phase. Thus the sequence of transitions on cooling this mixture from the nematic phase is \( N - A_1 - \tilde{A} - A_2 \). Some pure compounds are also now known to exhibit the \( \tilde{A} \) phase (table 2).

We have recently studied many compounds with the basic chemical structure similar to that of the DBn series, except for the presence of a lateral methyl or methoxy substituent. Indeed the resulting 4-cyanophenyl-3'-methyl-4'-(4''-\( n \)-alkylbenzoyloxy)benzoates (called \( n \)PMCBB) are obtained by interchanging the two end groups of \( n \)PMBB compounds, some of which exhibit the \( N_\perp \) phase (tables 1 and 2). 10 PMCBB exhibits a large increase of its bilayer spacing as the temperature is lowered, with a jump of \( \sim 0.4 \text{ Å} \) at \( T_{AN} \approx T \approx 15° \) (figure 10) (Madhusudana *et al* 1982b). The second order reflection is hardly seen in the high temperature phase but shows up strongly in the lower temperature \( \tilde{A} \) phase and the transition corresponds to an \( A_1 - A_2 \) transition.

The physical interpretation (Madhusudana *et al* 1982c) of this phenomenon can be understood from figure 11. In 12 cpmbb, which exhibits the \( N_\perp \) phase, the two ester group dipole moments are parallel to the dipole moment of the cyano group, and neighbouring molecules are arranged as in figure 11a, to minimise the mutual interaction energy arising from dispersion as well as dipolar forces. On the other hand in 12 PMCBB, such an arrangement (figure 11b) is not likely to lead to a minimum in the interaction energy. Though mutual dispersion interactions should favour this configuration, it is clear that the dipolar energy between the cyano group and one of the ester groups of the neighbouring molecule would be repulsive. The molecules would then prefer the arrangement shown in figure 11c so as to make the dipolar interaction energy highly negative. Indeed this energy (\( \sim \mu^2 / r^3 \)) is a few times the thermal energy \( k_B T \).

However, this interaction is confined to one *end* of the molecule and the structure is
Figure 10. Temperature variation of the layer spacing of 10 PMCBB. The $A_2-A_3$ transition occurs with a jump in the layer spacing at the temperature corresponding to the dashed line. The region around this transition point is shown on a magnified scale in the inset.

Figure 11. Schematic diagrams showing the disposition of various dipolar groups of a pair of (a) $n$-CPMBB, (b) $n$-PMCBB molecules with an overlap of the aromatic cores and (c) a pair of $n$-PMCBB molecules with an overlap of the polar end groups.
hence rather fragile, especially since the molecules have a relatively long alkyl chain and the lateral methyl group. As the temperature is raised, the structure easily breaks up and more molecules form associations as in figure 11b which should still have a negative interaction energy. As a consequence the layer spacing rapidly decreases with increase of temperature. If the energy difference between the arrangements of figures 11b and 11c is $w$, the number of molecules having configuration $c$ is $\propto \exp \left( -\frac{w}{kT} \right)$ and one can calculate the temperature variation of the layer spacing (Guillon and Skoulios 1983).

If we replace the cyano end group of 10 PMCBB with a nitro group, the trend of the thermal variation of the layer spacing becomes somewhat different (figure 12; Madhusudana et al 1982a). The sign of the curvature of the variation is opposite to that of 10 PMCBB (figure 10) and further, there is no phase transition in the smectic range. The longitudinal component of the dipole moment of the NO$_2$ group is practically the same as that of CN group. However the lateral components of NO bonds should make the mutual interactions in configuration $c$ of figure 11 somewhat repulsive and hence lead to the difference in the trend of layer spacing variation.

The dielectric properties of such compounds also are in conformity with the model given above. For example, as shown in figure 13, the dielectric anisotropy changes sign, becoming negative at lower temperatures of the A phase of 10 PMNBB, as more and more molecules take the configuration of figure 11c, and hence the contribution of the strong end group dipole moment to the dielectric constant parallel to the director is reduced.

Figure 14 gives a schematic diagram of the expected x-ray diffraction pattern for different modifications of the A phase. A comprehensive theory to describe the various polymorphic forms of the A phase and transitions between them has been developed by Prost (1980, 1981). He has written down a general Landau free energy expression with two coupled order parameters which describe the tendency of the molecules to condense at two (generally incommensurate) wavelengths. The variety of A phases observed is a consequence of the competition between the two wavelengths. According to this model, each of these A phases is in effect a ‘frustrated’ smectic, in which the competition between the two wavelengths is sorted out in a specific manner. Thus, for instance, the A phase (figure 9) is essentially an escape from incommensurability when

![Figure 12. Temperature variation of the layer spacing of 10 PMNBB.](image-url)
Polymorphism of the smectic $A$ phases of highly polar compounds

Figure 13. Temperature variations of the dielectric constants and anisotropy of 10 PMNBB.

Figure 14. Schematic diagram of the expected x-ray diffraction spots from different modifications of smectic $A$ phase (Levelut et al. 1981).

it is too large, i.e. $l'$ is much different from $2l$. Then the dipolar heads switch from one layer to another periodically, thus developing a transverse two-dimensional order. As we saw at the end of the previous section, Prost has also described the reentrant phenomenon ($Na_dNa_aNa_1$) in the same framework, but including the effect of fluctuations. We end this section by pointing out a remarkable result recently found by Tinh et al. (1982a) in a compound similar to the DBn series, except that the end cyano group is replaced by a nitro group. 4-nonyloxyphenyl-4'-nitro-benzoyloxy benzoate (table 2) exhibits the following sequence on cooling:

$$1Na_dNa_rNa_drNa_rNa_1S(?)C_2.$$
Though many other compounds with similar structure exhibit only A₂ and A̅ phases, this particular compound with a nonyloxy end chain rather unexpectedly shows two reentrant nematic phases.

4. Induced smectic phases

When two nematicogenic compounds are mixed, one would expect that the mixtures also exhibit only the nematic mesophase. However, when one of the components has the strongly polar cyano or nitro group and the other component has no such end group, in certain composition ranges the mixture actually exhibits an 'induced' smectic phase which is usually of the A type (Dave et al. 1966; Pelzl et al. 1968; Park et al. 1975). Such a phase diagram is displayed in figure 15 (Moodithaya and Madhusudana 1979). Thus permanent dipole-induced dipole interactions appear to be important for the occurrence of the induced phase. Further, there is experimental evidence that a charge transfer complex formation takes place between the molecules of the two kinds (Sharma et al. 1980), the highly polar component acting as an acceptor while the other component acts like a donor. Such an interaction appears to lead to the formation of a layered arrangement characteristic of smectic phases. (In some cases even the smectic E phase which has orthorhombic symmetry is induced, see for example, Schneider and Sharma 1981).

If the highly polar compound itself exhibits a bilayer A̅₄ phase, then the mixtures of such a compound with some weakly polar compounds exhibit a maximum as well as a minimum in the A–N transition boundary (Srikanta and Madhusudana 1982). Figure 16 illustrates such a phase diagram. The maximum which occurs at ~ 50 mole % of the two components, corresponds to the induced A phase. It is obvious from the occurrence

![Figure 15. Phase diagram of mixtures of (2-hydroxy)-p-ethoxy-benzylidene-p'-butylaniline (OHEBBA) with 7CB.](image-url)
of the minimum in the AN transition boundary that the induced A phase is essentially incompatible with the bilayer A\textsubscript{d} phase of the highly polar component. Indeed x-ray studies show that for compositions rich in the weakly polar component, the layer spacing corresponds to an appropriate average molecular length of the two components (Engelen \textit{et al} 1979; Srikanta and Madhusudana 1982). In other words, it is an A\textsubscript{1} phase.

The layer spacing in the A\textsubscript{d} phase of \textsc{npoob} is \( \approx 1.2 \) \( \lambda \). For compositions between the minimum and maximum, the medium has a mixture of monolayer and bilayer species. This coexistence appears to smear out any possible A\textsubscript{d}-A\textsubscript{1} transition which could in principle occur near the minimum of the AN boundary. However, as the minimum is approached from either side, the thermal expansion coefficient of the layer spacing increases significantly (figure 17). We can understand this by assuming that the interaction energy between the strongly polar and weakly polar components (including contributions from both dipole-induced dipole interactions and charge transfer complex formation) is only slightly higher than the interaction energy which gives rise to antiparallel correlations between two neighbouring \textsc{npoob} molecules. Consequently, as the temperature is increased, the strongly polar-weakly polar pairs break up allowing \textsc{npoob} molecules to form more antiparallel pairs (We note that at the composition corresponding to the minimum in the AN boundary, there are 3 \textsc{npoob} molecules for every \textsc{oh-ebba} molecule). This naturally leads to a considerable thermal expansion of the layers. The dielectric studies on the same system confirm this mechanism (Srikanta and Madhusudana 1982). Thus this system also manifests the effect of a competition between two characteristic lengths at the molecular level.

In conclusion, strong longitudinal dipole moments in rod-like molecules give rise to an antiparallel correlation between neighbouring molecules. Depending on the detailed
chemical structure of the molecules, this in turn gives rise to a 'bilayer' of a thickness which lies between 1 and 2 molecular lengths. The coupling between this length and the molecular length which would usually characterize a density wave in the medium gives rise to a variety of interesting phenomena in the smectic phases formed by such systems.

Acknowledgement

The author is grateful to Prof. S. Chandrasekhar for useful suggestions.

References

Chandrasekhar S and Madhusudana N V 1978 in Progress in Liquid Physics (ed.) C A Croxton (Chichester: Wiley) Ch. 14
Cladis P E 1975 Phys. Rev. Lett. 35 48
Clark N A 1979 J. Phys. (Paris) 40 C3-345
Polymorphism of the smectic A phases of highly polar compounds


de Gennes P G 1972 *Solid State Commun.* 10 753

de Jue W H 1982 *Solid State Commun.* 41 529


Hardouin F, Sigaud G, Achard M F and Gasparoux H 1979 *Phys. Lett.* A71 347


Klug D D and Whalley E 1979 *J. Chem. Phys.* 71 1874

Landau L D and Lifshitz E M 1959 *Statistical physics* (London: Pergamon)


Longa L and de Jue N H 1982 *Phys. Rev.* A26 1632


Madhusudana N V, Savithramma K L and Chandrasekhar S 1977 *Pramana* B 22


Maier W and Saupe A 1959 Z. *Naturforsch.* 14a 882


Pelzl G, Demus D and Sackmann H 1968 Z. *Phys. Chem.* (Leipzig) 238 22


Schneider F and Sharma N K 1981 Z. *Naturforsch.* 36a 1086


XANES and EXAFS of copper compounds: Studies of copper carboxylates with metal-metal bonds and of the complex formed by *Pseudomonas aeruginosa*†

P R SARODE, G SANKAR and C N R RAO*

Solid State and Structural Chemistry Unit, Indian Institute of Science, Bangalore 560012, India.

Abstract. X-ray absorption near edge structure (XANES) of copper compounds with copper in 1⁺, 2⁺ and 3⁺ states has been studied. Extended x-ray absorption fine structure (EXAFS) has been employed to determine bond distances and coordination numbers in several model copper compounds. Employing both XANES and EXAFS, the structure of the copper complex formed by the micro-organism *Pseudomonas aeruginosa* has been shown to be square-planar with the Cu–O distance close to that in cupric glucuronates and cupric acetylacetonate. EXAFS has been shown to be useful for studying metal-metal bonds in copper carboxylates.

Keywords: XANES; EXAFS; copper compounds; x-ray spectra.

1. Introduction

X-ray absorption near edge structure (XANES) and extended x-ray absorption fine structure (EXAFS) have become powerful tools for studies in elucidating the structure of various compounds (Belli *et al* 1980; Parthasarathy *et al* 1982; Sankar *et al* 1983). XANES provides information about the nature of coordination in metal compounds, while EXAFS gives bond distances besides the coordination numbers. These tools are especially useful for investigating biological compounds, amorphous materials and other systems which may not be available in the form of single crystals for crystallographic studies. In this paper we describe the results of our investigation of copper compounds employing both XANES and EXAFS. After having verified the results from these techniques with respect to model compounds, we have used the techniques to elucidate the structure of copper complex formed from the slime of the *Pseudomonas aeruginosa*. This micro-organism accumulates copper to a considerable extent and little is known about the actual structure of the copper complex so formed (Payne *et al* 1981). We have also examined the use of EXAFS to study copper carboxylates containing metal-metal bonds. These studies have been able to establish square-planar coordination in the copper complex formed from *Pseudomonas aeruginosa* and also provides the metal-oxygen bond distances in the complex as well as in model copper glucuronates.

2. Experimental

Analar samples of CuSO₄·5H₂O and CuCl₂·2H₂O were obtained commercially. All other copper (I) and (II) compounds were prepared by standard methods (Brauer 1963).

† Communication No. 228 from the Solid State and Structural Chemistry Unit
* To whom all correspondence should be addressed.
Ba$_2$Cu$_2$O$_5$ was prepared by the method of Arjomand and Machin (1975). Cupric glucuronates and copper complex formed from *Pseudomonas aeruginosa* were prepared by the methods already reported (Payne *et al* 1981).

X-ray absorption spectra were recorded using a bent-crystal spectrograph. The absorbers of the compounds were prepared by spreading the finely-powdered specimen on cellophane adhesive tape. The energy scale was fixed by using the emission lines of tungsten which appeared due to the impurity in the target. A Carl-Zeiss MD 100 densitometer was used to obtain the traces of these photographically recorded spectra. In order to obtain finer details of the spectra, 200 $\times$ magnification scans were taken for each compound. The linear reverse dispersion on the traces in each case was 0.35 eV/mm. The uncertainty in the edge and its near structure at about 50 eV was $\pm$ 0.5 eV and beyond 50 eV it was about 1 eV. We have been able to observe the fine structure beyond the edge up to about $\sim$ 550 eV.

2.1 Data reduction and analysis

To extract the extended x-ray absorption fine structure modulation, $\chi(E) = (\mu - \mu_0)/\mu_0$, where $\mu$ is the observed absorption coefficient and $\mu_0$ is the absorption coefficient modulation in the absence of neighbouring atoms) from the raw absorption spectra, background subtraction was done using a method due to Lytle *et al* (1975). This involves the fitting of a polynomial (Victoreen function) to the pre-edge portion of the raw data and extrapolating it over the whole range of the absorption data and subtracting it. After this pre-edge subtraction, a second degree polynomial is fitted through the centre of EXAFS oscillations. This curve is assumed to be the $\mu_0$ vs energy curve and subtracted from the data and the difference is then divided by it in order to get $(\mu - \mu_0)/\mu_0$. This procedure gives the normalized absorption coefficient as a function of photon energy. For structural analysis, the normalized fine structure was obtained as a function of photoelectron wavevector, $k$, (in $\text{A}^{-1}$), defined by the relation $k = [0.262467 (E - E_0)]^{1/2}$ eV, where $E$ is the photon energy and $E_0$ the threshold energy at the absorption edge. The data thus obtained was multiplied by the Hanning window function (Vla *et al* 1979) in order to minimize the termination ripples. Subsequently, it was multiplied by a weighting function, $k^3$, to compensate for amplitude reduction as a function of $k$. The integrations (integration limits $k = 3.5$ to $k = 11.6 \text{A}^{-1}$) for the Fourier transforms, both for the real and imaginary components, were done numerically using Simpson’s rule (Cramer *et al* 1978). The modulus of the Fourier transform was obtained as a square root of the sum of the real and imaginary parts.

The extended x-ray absorption fine structure associated with the $K$-absorption edge was shown to have (in terms of structural parameters) the following general form (Stern 1974; Ashley and Doniach 1975):

$$
\chi(k) = \frac{\mu - \mu_0}{\mu_0} = \frac{1}{k} \sum_{j} \frac{N_j}{R_j^2} |f_j(\pi, k)| \exp (-\sigma_j^2 k^2) \exp (-2R_j/\lambda) \sin [2kR_j + \phi_j(k)].
$$

(1)

In this expression, $N_j$ is the number of scatterers ($j$) at distance $R_j$ from the absorber; $f_j(\pi, k)$ is the scatterer’s electron backscattering amplitude; $\phi_j$ is the total phase shift; $\sigma_j^2$ is the mean square deviation of $R_j$, $\lambda$ is the mean free path and $k$ is the photoelectron wave vector. A separate modulation or wave in the absorption coefficient is produced by each coordination shell of atoms of approximately the same distance $R_j$ from the
xanes and exafs of copper compounds

absorbing atom. The observed fine structure results from the sum of waves from the various coordination shells. The object of the data analysis procedure is to decompose the observed spectrum into its component waves and determine the phase and amplitude of each component. This in turn can yield information about the distance in the coordination shells and the number of atoms in them.

A least-squares curve fitting analysis was employed to elucidate finer details of the structure of compounds. The background subtracted absorption data were first Fourier-transformed and then the region of interest in FT plot was Fourier-filtered to isolate the first and second coordination shells. Six parameter fits were performed on the data using a parametrized form of equation (1)

\[ \chi_{\text{param}} = C_0 \cdot \exp \left( C_1 k^2 / k^2 \right) \sin \left( a_0 + a_1 k + a_2 / k \right). \] (2)

Fits were done by least-squares adjustment of parameters \( C_0, C_1, C_2, a_0, a_1 \) and \( a_2 \) using a general optimization technique (Himmelblau 1972). The function minimized in the fitting procedure is given by

\[ F = \left[ \frac{\sum k^2 \left( \chi_{\text{expt}} - \chi_{\text{param}} \right)^2}{N} \right]^{1/2}, \] (3)

where \( \chi_{\text{param}} \) is the parametrized exafs function, \( \chi_{\text{expt}} \) is Fourier filtered exafs function, \( N \) is the number of data points in the fit and the sum is over the number of data points. \( F \) is a measure of the goodness of the fit and is lower for a better fit.

The curve fitting procedure for determining the distance \( R \) to a shell of atoms in an unknown compound is as follows. exafs data from a model compound whose crystal structure is known is fitted with equation (2) by adjusting all the six parameters. The standard parameters \( C_1, C_2, a_0 \) and \( a_2 \) are then used as fixed constants in fitting (2) to data from the unknown compound by adjusting only \( C_0 \) and \( a_1 \). If the phase shift \( \phi(k) \) is assumed to have the form \( \phi(k) = a_0 + a_1 k + a_2 / k \), then comparison of the theoretical equation (1) with the parametrized equation (2) shows that \( a_1 = 2R + \phi \). The linear phase parameter \( \phi \) is determined from the standard fit since \( R \) is known, and then \( \phi \) is used to calculate \( R \) from the fit of the unknown. After determining \( R \) for the unknown, the number of atoms \( N \) can be estimated by \( N = (C_0 / C_{\text{ex}}) (R_s / R) \) where \( N, R \) and \( C_0 \) refer to the unknown and \( N_s, R_s \) and \( C_{\text{ex}} \) refer to the model compound.

3. Results and discussion

3.1 xanes of typical copper compounds

Figures 1 and 2 show plots of K-absorption edges of representative of copper compounds with copper in \( 1^+, \ 2^+ \) and \( 3^+ \) states; the figure also includes absorption edges of a few copper complexes. In the case of Cu\(^{2+}\) compounds (figure 1) we are able to make spectral assignments on the basis of the assignments available for other transition metal compounds (Sankar et al 1983; Shulman et al 1976). The small pre-edge peak at about 8977.5 eV is due to the vibronically activated \( 1s \rightarrow 3d \) transition. The next feature in the spectra appearing as a shoulder is attributed to the forbidden \( 1s \rightarrow 4s \) transition, which appears to become allowed due to the mixing of the 4p and 4s orbitals. The highest energy peaks in the spectra arise from the allowed \( 1s \rightarrow 4p \) transition which ultimately merges into the continuum. The consistency in the position of the peaks due to the \( 1s \rightarrow 3d \) transition is noteworthy; the energies of \( 1s \rightarrow 4s \) and \( 1s \rightarrow 4p \) transitions
Figure 1. Copper K-edge absorption spectra of (a) cupric oxide; (b) copper propionate; (c) copper citrate; (d) sodium copper glucuronate; (e) barium copper glucuronate; (f) magnesium copper glucuronate; (g) copper complex from Pseudomonas aeruginosa; (h) cesium copper trichloride; (i) cupric chloride dihydrate; (j) Ba$_2$Cu$_2$O$_5$.

Figure 2. Copper K-edge absorption spectra of CuFeS$_2$; CuCl, CuBr and CuI.

show a similar behaviour. Energies of these three electronic transitions for several Cu$^{2+}$ compounds are listed in table 1. The energy range for the 1s $\rightarrow$ 3d transition is 8977.3–8978.3 eV with an average of 8977.8 eV. The standard deviation of this average for the Cu$^{2+}$ compounds studied is only 0.6 eV. The energy range for the 1s $\rightarrow$ 4s transition is 8985–8987 eV with an average 8986 eV and a standard deviation of 0.8 eV. The average energy for 1s $\rightarrow$ 4p is 8996.5 eV.

Absorption edges of a few Cu$^{1+}$ compounds are shown in figure 2. The first peak which appears as a shoulder, in the spectrum of the Cu$^{1+}$ compounds is due to the 1s $\rightarrow$ 4s transition, the 1s $\rightarrow$ 3d transition being absent in this d$^{10}$ system. The 1s $\rightarrow$ 4s transition of Cu$^{1+}$ comes at a lower energy than Cu$^{2+}$. The energy of the 1s $\rightarrow$ 4s transition of Cu$^{1+}$ is 8982.6 eV and this does not overlap with either the 1s $\rightarrow$ 3d or 1s $\rightarrow$ 4s range found in Cu$^{2+}$ compounds, thus providing a clear distinction between the edge spectra of the compounds with copper in the two different oxidation states. As mentioned earlier, it is noteworthy that there is no 1s $\rightarrow$ 3d transition in Cu$^{1+}$ compounds and this again enables the distinction between Cu$^{1+}$ and Cu$^{2+}$ spectra.
XANES and EXAFS of copper compounds

Table 1. X-ray K-absorption edge data on copper compounds.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Transition energies (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1s → 3d</td>
</tr>
<tr>
<td>Copper ammonium sulphate monohydrate</td>
<td>8978-0</td>
</tr>
<tr>
<td>Cupric oxide</td>
<td>8978-2</td>
</tr>
<tr>
<td>Anhydrous copper propionate</td>
<td>8977-8</td>
</tr>
<tr>
<td>Copper sulphate pentahydrate</td>
<td>8977-4</td>
</tr>
<tr>
<td>Copper citrate dihydrate</td>
<td>8977-9</td>
</tr>
<tr>
<td>Cesium copper trichloride</td>
<td>8977-3</td>
</tr>
<tr>
<td>Cupric chloride dihydrate</td>
<td>8977-6</td>
</tr>
<tr>
<td>Sodium copper glucuronate</td>
<td>8977-5</td>
</tr>
<tr>
<td>Barium copper glucuronate</td>
<td>8977-4</td>
</tr>
<tr>
<td>Magnesium copper glucuronate</td>
<td>8977-4</td>
</tr>
<tr>
<td>Copper complex formed from <em>Pseudomonas aeruginosa</em></td>
<td>8978-3</td>
</tr>
</tbody>
</table>

Differences in the spectra of compounds with copper in the $2^+$ and $3^+$ states are small compared to the differences between the spectra of Cu$^{1+}$ and Cu$^{2+}$ compounds. Ba$_2$Cu$_2$O$_5$ containing copper in the Cu$^{3+}$ state shows a pre-edge transition (figure 1) with a peak at 8985 eV, about 8 eV above the highest energy found in Cu$^{2+}$ compounds. The peak position of the 1s → 4s transition is found at 8995 eV, 10 eV higher than that of Cu$^{2+}$ compounds.

3.2 EXAFS of typical copper compounds

In figure 3 we show background subtracted spectra of cupric oxide, anhydrous copper propionate, copper ammonium sulphate monohydrate and copper thiosalicylate. In figure 4 we have shown the Fourier-transformed spectra in R-space of these compounds. These are the four-model compounds used to identify near neighbours by determining the phase shift for Cu–O, Cu–Cu, Cu–N and Cu–S atom pairs. Except for the copper propionate, each of these has a major peak in the FT plot (single-shell model compounds)*. This intense peak is isolated using a suitable window function (Teo 1981) by retransforming it to k-space and then fitted with (2) by adjusting all six parameters $C_0, C_1, C_2, a_0, a_1$ and $a_2$. The first two peaks in the copper propionate transform were retransformed to k-space and fitted with the sum of two waves in two-shell fit. The first shell contains the amplitude and phase parameters from the cupric oxide fit, adjusting only $C_0$ and $a_1$, while all the six parameters of the second shell corresponding to Cu–Cu atom pair were adjusted. Table 2 lists the results of the fits of these model compounds including the fitting function value from (3), and the parameters $C_0, C_1, C_2, a_0, a_1,$ and $a_2$ from (2). In figure 5, we have shown the fitted and experimentally observed (Fourier-filtered) EXAFS spectra of these model compounds. The curve-fitted spectra obtained from the parameterized form of EXAFS equation matches very well with the observed Fourier-filtered spectra. According to Citrin et al (1976) and Cramer et al (1978) all these phase ($a_0, a_1, a_2$) and amplitude ($C_0, C_1, C_2$) parameters can be transferred to the

* The term single shell refers to structures where the absorbing atom is surrounded by a set of identical scatterers at nearly equal distances.
system in which the absorbing atoms have the same environments to predict the structural parameters. As a first test of transferability of these Cu-X parameters, these phase shifts and amplitudes were used to determine other crystallographically known compounds. The compounds chosen for this test were copper sulphate pentahydrate, copper acetylacetonate and copper phthalocyanine all of which possess square-planar co-ordination (Bacon and Curry 1962; Basu et al 1962; Harrison and Assour 1964) and bond distances comparable to those in CuO. The Fourier-transformed spectra for these compounds are shown in figure 6. To obtain structural information from the FT spectra of these compounds, the observed peak position is simply modified by previously determined phase shift to obtain a predicted distance. The co-ordination number is then obtained from the observed peak height. The predicted and crystallographically observed structural results are summarized in table 3.

For curve-fitting analysis of the spectra, $C_0$, and $a_1$ were floated while $C_1, C_2, a_0$ and
Figure 4. Fourier transforms of the four model copper compounds. Curves shows the magnitude only. Transform range: 3-5-11.6 Å⁻¹, k³ scaling.

3.3 XANES and EXAFS studies of copper glucuronates and the copper complex formed by the slime of Pseudomonas aeruginosa

Edge spectra for sodium-copper glucuronate, barium copper glucuronate, magnesium copper glucuronate and a copper complex isolated from the slime of the Pseudomonas aeruginosa bacteria are shown in figure 1. These spectra show a weak peak in the Cu²⁺ 1s → 3d region. The average energy of the 1s → 3d peak in these complexes is 8977.6 eV. The next feature due to 1s → 4s transition appearing as a shoulder is at 8985.1 eV. Both these transition energies are well within the range of Cu²⁺ compounds (see table 1). We can therefore conclude that copper in all these four copper complexes are in Cu²⁺ state. Accordingly these complexes show characteristic ESR signals due to Cu²⁺ (Payne et al...
Table 2. Curve-fitting phase and amplitude parameters for a $k$ range of 3.5–11.6 Å$^{-1}$. (Optimized function, $x_{\text{param}} = [C_0, \exp (-C_1k^2/|k|)]; \sin (a_0 + a_1k + a_2/k); k^6$ weighting).

<table>
<thead>
<tr>
<th>Standard atom pair(a)</th>
<th>Value of the fitting function(b)</th>
<th>$C_0$</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>$a_0$</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>Coordination number, $N$</th>
<th>Bond distance, $R$ (Å)</th>
<th>Phase shift parameter, $\phi^{(e)}$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cu–N</td>
<td>0.89</td>
<td>0.824</td>
<td>-0.0146</td>
<td>1.930</td>
<td>1.430</td>
<td>3.659</td>
<td>21.54</td>
<td>4</td>
<td>2.05$^{(e)}$</td>
<td>-0.441</td>
</tr>
<tr>
<td>Cu–O</td>
<td>0.92</td>
<td>0.410</td>
<td>-0.0190</td>
<td>1.380</td>
<td>1.560</td>
<td>3.476</td>
<td>21.84</td>
<td>4</td>
<td>1.955$^{(d)}$</td>
<td>-0.434</td>
</tr>
<tr>
<td>Cu–S</td>
<td>0.48</td>
<td>2.369</td>
<td>-0.0141</td>
<td>2.365</td>
<td>0.621</td>
<td>3.949</td>
<td>30.51</td>
<td>4</td>
<td>2.20$^{(e)}$</td>
<td>-0.451</td>
</tr>
<tr>
<td>Cu–Cu</td>
<td>0.68</td>
<td>2.089</td>
<td>-0.0146</td>
<td>2.900</td>
<td>3.760</td>
<td>5.052</td>
<td>38.57</td>
<td>1</td>
<td>2.578$^{(f)}$</td>
<td>-0.104</td>
</tr>
</tbody>
</table>

(a) Cu–N from copper ammonium sulphate, Cu–O from cupric oxide, Cu–S from copper thiosalicylate and Cu–Cu from copper propionate; 
(b) From equation (3); (c) Mazzi (1955); (d) Asbrink and Norrby (1970); (e) Srivastava and Nigam (1972); (f) Simonov and Malinovskii (1970); 
(g) Calculated from $\phi = a_1 - 2R$
Furthermore, the Cu(2p\textsubscript{3/2}) peaks in x-ray photoelectron spectra of these complexes exhibit characteristic satellites just as in CuO indicating the presence of Cu\textsuperscript{2+} ions (Payne et al 1981). It is well-known that near-edge structure gives information on the site symmetry of absorbing atoms in materials (Belli et al 1980; Grunes 1983). Since all the spectral features in the above complexes are essentially similar to those of Cu\textsuperscript{2+} model compounds where the coordination geometry of copper is square-planar, we may conclude that these cupric complexes are likely to have square-planar coordination. Thus, the ESR spectra of these copper complexes show \( g_{\parallel} \) and \( g_{\perp} \) values similar to cupric acetylacetonate (Payne et al 1981) confirming that the complexes are square-planar.
We shall now discuss the results of EXAFS analysis of these complexes. In (1), the total phase shift, $\phi$, can be treated (Cramer and Hodgson 1979) as the sum of absorber, $\phi_a$, and scatterer phase, $\phi_s$ shifts, $\phi(k) = \phi_a(k) + \phi_s(k)$. The absorber phase shifts $\phi_a(k)$ is the result of the photoelectron propagating out of and back through the central atom potential. Reflection of the photoelectron by the scattering atom potential causes the scatterer phase shift, $\phi_s$. For a series of compounds with the same absorber, differences in total phase shift are caused by the different $\phi_s$ of the scattering atoms involved. The total phase shifts obtained from the experimentally measured EXAFS data as a function of wavevector are used to predict the type of atoms surrounding the absorbing atom.

In the present investigation, the total phase shifts were obtained by subtracting $2Rk$ from $a_0 + a_1k + a_2/k$ (for copper oxide, copper propionate, copper ammonium sulphate and copper thiosalicylate), $k$, ranging from 3.5 and 11.6 (Å$^{-1}$). In figure 7 we have plotted these empirical, pair-wise phase shifts for Cu–O, Cu–S, Cu–N and Cu–Cu; the curve for Cu–C pair could not be given since we do not have a model compound with carbon neighbour within 3Å range. In figure 8, we have plotted the phase shifts for the pairs taken from theoretically calculated and tabulated data of Teo and Lee (1979) and Lee et al (1977). Back scattering phase shifts for nitrogen are not given in the tables of these authors; the phase shifts for the Cu–N atom pair (figure 8) was obtained by interpolation as suggested by Teo and Lee (1979). We see from figures 7 and 8 that Cu–O, Cu–N, Cu–Cu and Cu–S phase shifts are all similar. The differences in phase shifts for O, Cu, N, S, permit straightforward differentiation between these back-scattering atoms. In figure 7, we have shown the dotted curve for the copper complex.
Table 3. Calculation of single-shell distances and coordination numbers. Model compound: CuO.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Value of fitting function</th>
<th>Bond distance</th>
<th>Coordination number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>FT distance</td>
<td>Curve-fit distance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Å)</td>
<td>(Å)</td>
</tr>
<tr>
<td>Copper sulphate pentahydrate</td>
<td>0.62</td>
<td>2.00</td>
<td>1.97</td>
</tr>
<tr>
<td>Copper acetylacetonate</td>
<td>0.79</td>
<td>1.96</td>
<td>1.94</td>
</tr>
<tr>
<td>Copper phthalocyanine</td>
<td>0.92</td>
<td>1.85</td>
<td>1.84</td>
</tr>
<tr>
<td>Sodium copper glucuronate</td>
<td>0.91</td>
<td>1.99</td>
<td>1.97</td>
</tr>
<tr>
<td>Barium copper glucuronate</td>
<td>0.83</td>
<td>1.96</td>
<td>1.96</td>
</tr>
<tr>
<td>Magnesium copper glucuronate</td>
<td>0.87</td>
<td>1.95</td>
<td>1.96</td>
</tr>
<tr>
<td>Copper complex formed from Pseudomonas aeruginosa</td>
<td>0.94</td>
<td>1.99</td>
<td>1.96</td>
</tr>
</tbody>
</table>

* Values in parenthesis are from x-ray diffraction data.
+ Coordination number determined from Fourier transform analysis, $N = R^2MN_0/M_0R_s^2$, where $M_0$, $N_0$, and $R_s$ are the observed peak height of the FT curve, coordination number and bond distance for a model compound respectively. Primed symbols refer to the unknown compound.
Figure 7. Experimentally determined copper-backscatterer phase shifts vs. photoelectron wave vector, \( k \). \( 2\pi \) has been subtracted from the Cu–Cu phase shifts.

Figure 8. Theoretically calculated copper-backscatterer phase shifts vs. photoelectron wave vector, \( k \). \( 2\pi \) has been subtracted from the Cu–Cu phase shifts.
formed from Pseudomonas aeruginosa. We see that the curve overlaps with that for the Cu–O atom pair once again establishing that copper in this complex has near neighbour oxygen atoms (in the first co-ordination shell); synthetic copper glucuronates also exhibit similar curves. After identifying the type of surrounding of copper in these complexes, we will now determine the bond distances and co-ordination numbers from EXAFS data.

In figure 9 we have shown the Fourier transform spectra of sodium copper glucuronate, barium copper glucuronate, magnesium copper glucuronate and complex obtained from Pseudomonas aeruginosa. The FT spectra retransformed to k-space and fitted to the parametrized equation (2) varying the amplitude parameter, $C_0$ and phase shift parameter $\alpha_1$, while the other four parameters were fixed constant. The results of these analysis are summarized in table 3. As expected the co-ordination number in all these complexes is four establishing thereby that they are all square-planar. We readily see that the distances in the copper glucuronates and in the complex from Pseudomonas aeruginosa are quite similar to those of copper acetylacetonate.

Since the analysis of the slime produced by the bacteria Pseudomonas aeruginosa without copper present, using HPLC (Magee 1981) clearly shows that glucuronic acid is a major constituent, we suggest that the nature of the complex is itself copper glucuronate. The possible structures of the copper complex using the Haworth structure for the uronic acid is shown in chart I. In structure I, copper is co-ordinated
through the ether oxygen of carbon 1 and the aldehydic oxygen. In structure II, copper is coordinated through the hydroxyl oxygens of carbon 3 and 4.

3.4 *Carboxylates with Cu–Cu bonding*

In order to determine whether a peak in the Fourier-transformed spectra corresponding to backscattering from a metal is sufficiently strong compared to those due to oxygen neighbours to be easily distinguished from them, EXAFS data were collected on few compounds with various metal-metal distances. Anhydrous copper propionate, copper acetate monohydrate, copper benzoate trihydrate, copper citrate dihydrate and copper salicylate tetrahydrate have Cu–Cu distances of 2.578 (Simonov and Malinovskii 1970), 2.614 (Brown and Chidambaram 1973), 3.15 (Inoue et al 1965), 3.242 (Mastropaolo et al 1976) and 3.728 A (Inoue et al 1964) respectively. The Fourier-transformed spectra of these are shown in figure 10. In this figure the arrows show peaks corresponding to the Cu–Cu atom pairs. The propionate transform shows the Cu–Cu peak at 2.151 A which means that the linear phase shift in R space for Cu–Cu pair is 0.427 A. Applying this phase shift to the other transforms of the remaining three compounds, we have determined the Cu–Cu distances in these compounds. These distances are given in table 4. In this table we have also given the distances and coordination numbers obtained by curve fitting the data to equation (2). There is a good agreement between the distances and coordination numbers determined from Fourier transform and those obtained from curve-fitting. There is also a fair agreement between
XANES and EXAFS of copper compounds

Figure 10. Fourier transforms of anhydrous copper propionate, copper acetate monohydrate, copper benzoate trihydrate, copper citrate dihydrate and copper salicylate tetrahydrate.

Table 4. Calculation of two-shell scatterer distances and numbers in copper carboxylates. Model compound: Cu(CH₃CH₂COO)₂.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Value of the fitting function</th>
<th>Radial distance</th>
<th>Coordination number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cq</td>
<td>a₁</td>
<td>FT distance (Å)</td>
<td>Curve-fit distance (Å)</td>
</tr>
<tr>
<td>Copper acetate monohydrate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) 4Cu–O at 1.969 Å</td>
<td>0.75</td>
<td>1.984</td>
<td>4.5</td>
</tr>
<tr>
<td>(ii) 1Cu–Cu at 2.614 Å</td>
<td>2.532</td>
<td>2.591</td>
<td>1.3</td>
</tr>
<tr>
<td>Copper benzoate trihydrate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) 4Cu–O at 1.94 Å</td>
<td>0.98</td>
<td>1.950</td>
<td>4.3</td>
</tr>
<tr>
<td>(ii) 1Cu–Cu at 3.15 Å</td>
<td>0.980</td>
<td>3.180</td>
<td>0.8</td>
</tr>
<tr>
<td>Copper citrate dihydrate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) 4Cu–O at 1.946 Å</td>
<td>0.81</td>
<td>1.947</td>
<td>3.9</td>
</tr>
<tr>
<td>(ii) 1Cu–Cu at 3.242 Å</td>
<td>0.967</td>
<td>3.240</td>
<td>0.7</td>
</tr>
<tr>
<td>Copper salicylate tetrahydrate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(i) 4Cu–O at 1.92 Å</td>
<td>0.99</td>
<td>1.93</td>
<td>3.9</td>
</tr>
<tr>
<td>(ii) 1Cu–Cu at 3.728 Å</td>
<td>0.971</td>
<td>3.74</td>
<td>0.8</td>
</tr>
</tbody>
</table>
these data and the crystallographic data reported in literature.

Except for the acetate and the propionate where the Cu–Cu distance is less than 2.80 Å, there is no strong metal-metal peak in the Fourier transform of the benzoate or the citrate or the salicylate. Since the exafs amplitude is strongly dependent on the thermal motion of the backscatterers with respect to the absorbing atom, it is possible that at lower temperature metal-metal components can be more clearly distinguished in the FT spectra.

Acknowledgement

The authors thank the Department of Science and Technology, Government of India for support of this research and Prof. R J Magee for interesting them in the Cu complex formed by Pseudomonas aeruginosa.

References

Ashley C A and Doniach S 1975 Phys. Rev. B11 1279
Cramer S P and Hodgson K O 1979 Prog. Inorg. Chem. 25 1
Inoue M, Kishita M and Kubo M 1964 Inorg. Chem. 3 239
Magee R J 1961 Personal communication
Mazzi F 1955 Acta Crystallogr. 8 137
Srivastava U C and Nigam H L 1972 Curr. Sci. 41 81
Teo B K 1981 in EXAFS spectroscopy: Techniques and applications (eds) B K Teo and D C Joy, Chapter 3 (New York: Plenum)
Teo B K and Lee P A 1979 J. Am. Chem. Soc. 101 2815