Lead Article

Acta Cryst. (1991). B47, 145–154

Diffraction Investigation of the Atomic Structure of Matter*

The Ewald lecture delivered on 26 July 1990 at the XV Congress of the International Union of Crystallography, Bordeaux, France

BY B. K. VAINSHTEIN

Institute of Crystallography, Academy of Sciences of the USSR, Leninsky pr. 59, *Moscow* 117333, *USSR*

(*Received* 31 *August* 1990; *accepted* 19 *October* 1990)

Abstract

The outstanding role of P. P. Ewald in the development of X-ray diffraction theory is described. Principal schemes describing the relation between structure and diffraction pattern for various classes of objects – crystals, liquid crystals, macromolecules in solution, protein crystals – are presented. A brief survey of the method of electron diffraction structure analysis is given. The results of several X-ray structural investigations of liquid crystals are considered. The basic principles of structure derivation from diffraction data are described for macromolecules in solution. The possibilities of X-ray crystallography for proteins are illustrated by an investigation of catalases.

Introduction

The basic modern data describing the atomic structure of matter have been obtained by the use of diffraction methods – X-ray, neutron and electron diffraction.

At present more than 100 000 atomic structures of various natural and synthetic crystals – inorganic, metallic, organic and biological – have been determined. Diffraction methods have been applied to the study of the atomic structure of substances in a less ordered state – high-molecular-weight polymers, liquid crystals, amorphous substances and liquids, and isolated molecules in vapours or gases. This tremendous amount of material is used for both the solution of problems concerning the relation between the structure of a given substance in the crystalline state and its properties, crystal–physical and crystal–chemical generalizations, and the development of solid-state physics and physics of the condensed state in general. Investigation of biological substances has become of exceptional importance for molecular biology and medicine.

All these achievements would have been impossible without the rapid development of the theory of diffraction, structure analysis of crystals and less-ordered substances, the discovery of new experimental possibilities, and the development of the diffraction technique. Experimental investigations into the atomic structure of matter are based mainly on the kinematical theory of scattering, whereas the ideal and real structures of single crystals are analyzed on the basis of the dynamical theory.

Development of the theory of diffraction in the works of P. P. Ewald

Considering the development of diffraction theory, we must pay tribute to the great contribution made by an outstanding physicist and crystallographer Peter Paul Ewald. Working at Munich University under A. Sommerfeld and still a very young scientist, from 1911 onwards he investigated the transmission of electromagnetic radiation through crystals.

In 1912, through the experiments of von Laue, Fridrich and Knipping, the diffraction of X-rays in crystals was discovered. In 1913 the first in the series of Ewald's brilliant works appeared – 'Zur Theorie der Interferenzen der Röntgenstrahlen in Kristallen' (Ewald, 1913). He developed the theory, from the case in which the wavelength is much greater than the lattice period to the case of X-rays. He created a system of remarkable simplicity and clarity which we now call the Ewald sphere and Ewald construction (Fig. 1.). He wrote: "In einem Gitter mit den Teilungen π/a, π/b, π/c, ('reziprokes Gitter') schlage man um den Punkt (α, β, γ) die Kugel, welche durch den Nullpunkt des Gitters geht. Liegen auf der Kugelfläche noch andere Gitterpunkte (l_0, m_0, n_0), so treten im Kristall Wellen, mit maximaler Intensität auf, die die gleiche Richtung wie die Verbindungs-

* *Editorial note*: This invited paper is one of a series of comprehensive Lead Articles which the Editors invite from time to time on subjects considered to be timely for such treatment.

0108-7681/91/020145-10$03.00

© 1991 International Union of Crystallography

linien von (l_0, m_0, n_0) zum Mittelpunkt der Kugel haben."

Later Ewald considered the problems of X-rays passing through a crystal and laid the foundations of the dynamical theory of X-ray scattering, which, as we know, deals with the interference of all diffracted beams arising in the crystal. He developed these concepts in a number of later works (Ewald, 1916, 1917, 1920, 1921, 1924, 1933).

The problem of X-ray crystal optics was solved by analyzing the vibrations of resonators (dipoles) in the crystal lattice. The occurrence and propagation of elementary waves caused by these vibrations were considered as vibrations of the system proper. The system of vibrating dipoles and the waves passing through it are self-consistent. Ewald's theory explained the dispersion of the incident wave in the crystal, transfer to multi-wave scattering, and the concept of the dispersion surface, that is, practically all scattering phenomena in ideal crystals. Ewald also developed the theory of the reciprocal lattice and introduced the apparatus of Fourier transformation to the theory of diffraction (Ewald, 1933).

Below I shall touch upon several general principles of structure analysis and present the results of some theoretical and experimental investigations carried out by myself and my colleagues at the Institute of Crystallography of the USSR Academy of Sciences.

Diffraction from crystals

The most ordered atomic structure with a three-dimensional periodicity is represented by crystals, while the least ordered structure is characteristic of amorphous and liquid substances.

A schematic representation of the connection between the structure of the crystal $\rho(\mathbf{r})$ and scattering intensities I_{hkl} (at kinematic scattering) is given in Fig. 2. (Here and in similar schemes hereafter we assume that all functions describing a real experiment are taken into account and $I = |F|^2$ values have been obtained.) The discreteness of the I_{hkl} peaks is explained by the long-range order (three-dimensional periodicity) of the atomic structure of the crystal which exists despite local disorder, *i.e.* thermal displacements of atoms from their equilibrium positions (see Fig. 9). At registration of the intensities only $|F_{hkl}|^2$ can be measured, and the phases of the F_{hkl} values are lost. Mathematically intensity can be regarded as the Fourier transform of the Patterson function $Q(\mathbf{r})$.

Electron diffraction structure analysis (EDSA)

Analysis of the atomic structure of crystals by electron diffraction (ED) began its development as a method independent from X-ray analysis by the end of the 1940's–beginning of the 1950's (Pinsker, 1953; Vainshtein, 1956, 1964).

EDSA differs from X-ray structure analysis in that it deals with the strong interaction of electrons with the substance. Electrons are scattered at the electrostatic potential of a crystal created by positively charged nuclei and negatively charged electron shells, while X-rays just 'feel' the electron density of an object. The samples used are from 50 to 500 Å thick. This permits investigation of crystals and substances that cannot be obtained as single crystals in a high-dispersion state. Special electron diffraction cameras and electron microscopes are used for such investigations.

One of the basic features of high-energy ED (accelerating voltage ~60–100 kV) is the short wavelength of the electrons used – 0·05 Å or less. Therefore Ewald's sphere practically degenerates into a plane, and the electron diffraction pattern is the planar cross-section of the reciprocal lattice (Fig. 3).

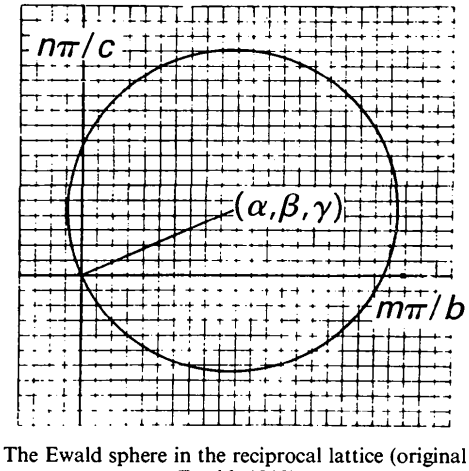

Fig. 1. The Ewald sphere in the reciprocal lattice (original picture, Ewald, 1913).

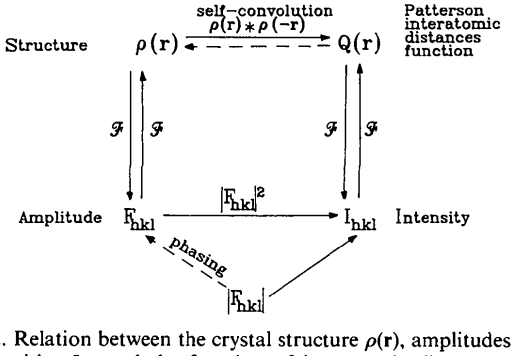

Fig. 2. Relation between the crystal structure $\rho(\mathbf{r})$, amplitudes F_{hkl}, intensities I_{hkl} and the function of interatomic distances $Q(\mathbf{r})$. The solid arrows indicate direct connection between the respective functions; dashed arrows denote the need to introduce special methods to obtain the structure $\rho(\mathbf{r})$. $\mathcal{F}$ is the Fourier transformation.

An ED pattern from a mosaic single crystal is shown in Fig. 4. ED patterns of the 'oblique' texture type (Fig. 5) are also very useful for structural analysis. Such patterns are formed when a sample has many crystallites with one and the same face parallel to the substrate, but which exhibit random azimuthal orientation around the axis perpendicular to the face. In this case the reciprocal lattice rotates around the axis indicated and presents a system of rings, the diffraction pattern – an oblique planar cross-section of the rings – see Fig. 6 (Pinsker, 1953; Vainshtein, 1956).

Procedures for ED pattern indexing and determination of the unit cell were established for structures with all kinds of symmetry, as well as for superstructures (Vainshtein, 1956, 1964; Zvyagin, 1967).

Intensity formulae for kinematic scattering were derived for ED patterns of all types (Vainshtein, 1956). The condition for the applicability of kinematical theory is:

$$\lambda|(\Phi_{hkl}/\Omega)|A \simeq 1. \quad (1)$$

where Φ_{hkl} is the structure amplitude for electrons, Ω is the cell volume, and A is the crystal thickness. The critical value of A for crystals with heavy atoms is about 50 Å, and about 200–400 Å for crystals with medium and light atoms.

The integral intensity I_{hkl} of spot hkl on a mosaic single-crystal ED pattern is:

$$I_{hkl}/J_0 S = \lambda^2|(\Phi_{hkl}/\Omega)|^2 A(d_{hkl}/\alpha) \quad (2)$$

where J_0 is the primary beam intensity, S is the irradiated area of the crystal, and α is the averaged spread of mosaic blocs. For thick crystals allowances

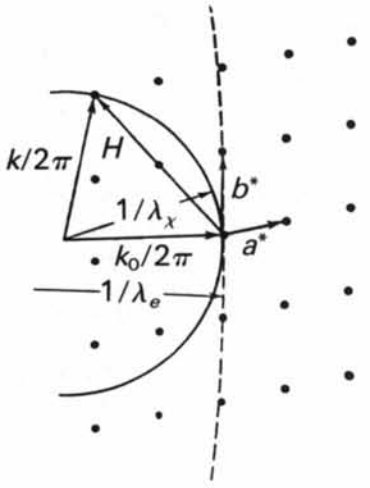

Fig. 3. Ewald construction for X-rays (solid sphere) and electrons (dotted sphere).

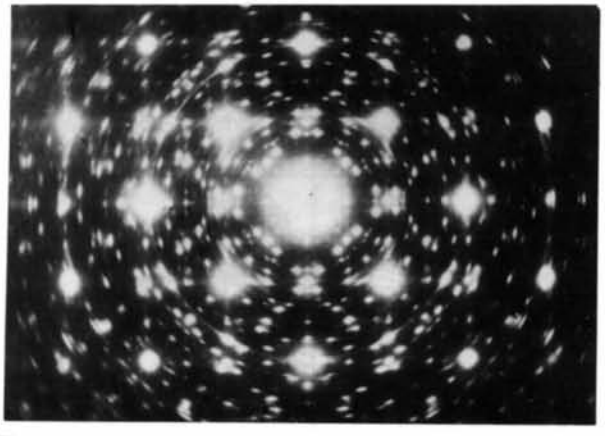

Fig. 4. Electron diffraction pattern (EDP) from a mosaic single crystal of Au, with an oriented AuOH film on it.

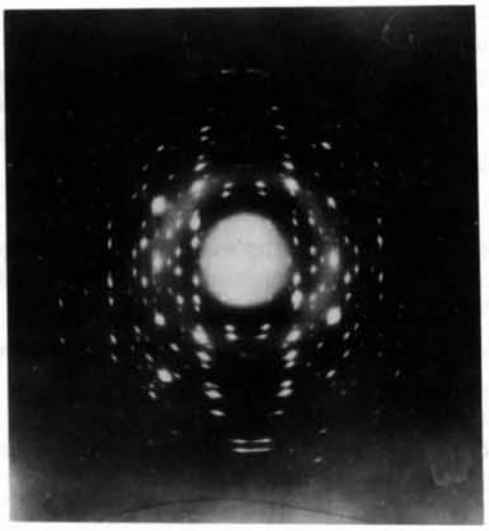

Fig. 5. Oblique texture EDP of Sb_2S_3.

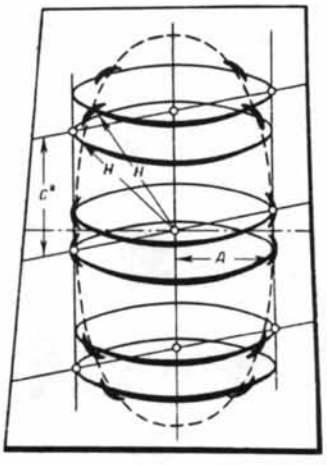

Fig. 6. A scheme of the oblique texture EDP formation – concentric rings in reciprocal space and their planar cross-section.

can be made for the dynamic character of scattering electrons. For the diffraction patterns from textures

$$\frac{I_{hkl}}{J_0 S} = \lambda^2 \left|\frac{\Phi}{\Omega}\right|^2 A \frac{L\lambda p}{2\pi R' \sin\varphi} \qquad (3)$$

where φ is the angle of inclination of the specimen plane to the beam, L is the specimen–detector distance, R' is the distance from the vertical axis of the EDP to reflection, and p is the multiplicity factor.

Making use of $|\Phi_{exp}|$ and the calculated phases, the Fourier synthesis of the potential can be obtained (Vainshtein & Pinsker, 1949; Vainshtein, 1954, 1956, 1964):

$$\varphi(\mathbf{r}) = \varphi(x,y,z) = \frac{1}{\Omega}\sum_{hkl} \Phi_{hkl} \exp[-2\pi i(\mathbf{r}\mathbf{H}_{hkl})]. \qquad (4)$$

Fig. 7 shows the Fourier synthesis of the potential (projection) for the $BaCl_2.H_2O$ structure, constructed for the first time from ED data.

Analysis of the atomic scattering amplitudes of electrons f_e reveals several interesting regularities.

It is known that the charge and potential are related by the Poisson equation

$$\Delta^2\varphi(\mathbf{r}) = -4\pi e[\rho_+(\mathbf{r}) - \rho_-(\mathbf{r})] \qquad (5)$$

whence we can derive the well-known Mott equation relating the atomic amplitudes for electrons, f_e, and for X-rays, f_x:

$$f_e(\sin\theta/\lambda) = (me^2/2h^2)[Z - f_x(\sin\theta/\lambda)]/(\sin\theta/\lambda)^2 \qquad (6)$$

where Z is the atomic number (electron cloud charge). It was established that f_e amplitudes [as well as peak heights of atomic potentials $\varphi_{at}(0)$ for electrons] are much less dependent on the atomic number Z than the atomic amplitudes for X-rays. For small $\sin\theta/\lambda$ values $f_e \simeq Z^{1/3}$, for large $\sin\theta/\lambda$ values $f_e \simeq Z$. Peak heights $\varphi_{at}(0) \simeq Z^{0.75}$ (Vainshtein, 1956, 1964). The weak dependence on Z allows one to determine the position of light atoms in the presence of heavy ones fairly easily. In particular, this is how H atoms were located in a number of compounds. In the diketopiperazine structure (Fig. 8) the peak-centre potential $\varphi_{at}(0)$ is: C = 151, N = 156, O = 160, H = 30 V (Vainshtein, 1954, 1955, 1960).

Atomic amplitudes f_e are highly sensitive to the ionic state of the atom. Referring back to expression (6), $f_x = Z'$ at low $\sin\theta/\lambda$, where Z' is the real charge of the electron cell. For ions $Z' = Z \pm n$, where n is the charge of the ion, f_e becomes $\sim \pm n$, and

$$f_e \underset{\sin\theta/\lambda \to 0}{\longrightarrow} \pm\infty. \qquad (7)$$

Thus, for Li_2O it was established that the chemical bond is not purely ionic, but partly covalent (Vainshtein, 1960) and for MgO $n = 0{\cdot}9$ (Avilov, Semiletov & Storozenko, 1989).

So far, electron diffraction has been applied to the structural investigation of many ionic crystals, of various inorganic (Cowley, 1953), organic (Dorset, 1976, 1983) and semiconductor compounds, carbides, nitrides (Pinsker & Imamov, 1981), various minerals and layer silicates (Zvyagin, 1967, 1989; Drits, 1987). Liquid crystals, amino acids (Chapyrina, Diakon, Donu & Bydnikov, 1986), polypeptides (Vainshtein & Tatarinova, 1967), and some other biostructures (Unwin & Henderson, 1975) have also been studied by EDSA.

I believe that the future of EDSA lies in direct precision measurement of the intensities, introduction, where necessary, of dynamical corrections, in

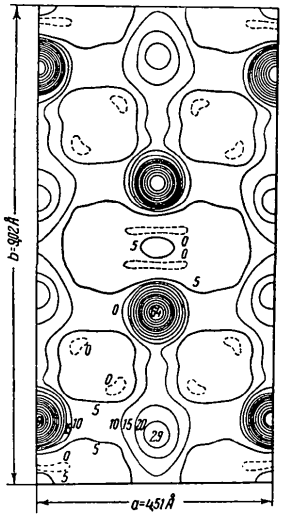

Fig. 7. Fourier synthesis of the potential of $BaCl_2.H_2O$ (projection along the c axis).

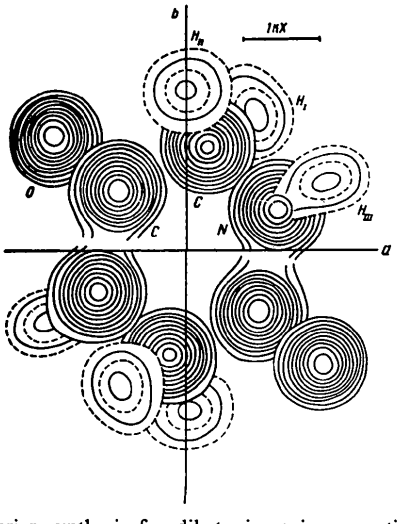

Fig. 8. Fourier synthesis for diketopiperazine – sections of the three-dimensional potential distribution.

the combination of electron diffraction and high-resolution electron microscopy, and in the use of novel techniques, particularly the converging-beam method.

The structure of liquid crystals

In crystals the positional function of atoms is a lattice: long-range order exists (Fig. 9*a*). Atoms are in fixed positions, though displaced from their equilibrium positions by thermal motion. Such distortions are called distortions of the first kind. The self-convolution of such a structure, $\rho(\mathbf{r})^*\rho(-\mathbf{r}) = Q(\mathbf{r})$, is given in Fig. 9(*b*). The last function – distribution of interpoint distances – is related to the lattice points, and the density of its peaks indicates the probability of displacement of atoms from their equilibrium positions.

Still, in nature there are many condensed atomic molecular systems having no three-dimensional periodicity: liquid crystals, liquids, polymers. In them short-range order is retained – the nearest neighbour distances have a certain mean value (Fig. 9*c*). The interpoint distances distribution function $Q(\mathbf{r})$ (see Fig. 9*d*) already reveals an increase in the peak width for the next-nearest neighbours, and at a certain distance (correlation radius) the probability of finding a particle becomes the same everywhere – the long-range order is absent. Such distortions are called distortions of the second kind (Hosemann & Bagchi, 1962; Vainshtein, 1966).

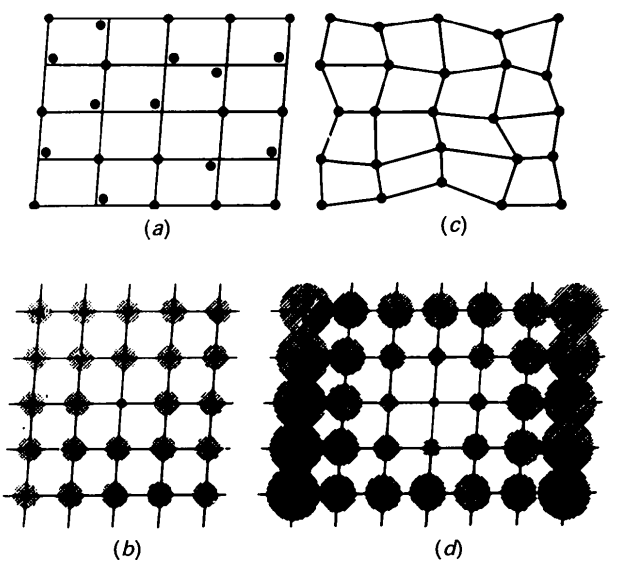

Fig. 9. Distortions of the first and second kind: (*a*) crystal lattice with long-range order and distortions of first kind; (*b*) self-convolution of distribution function (*a*); (*c*) structure with short-range order – distortions of the second kind; (*d*) self-convolution of distribution function (*c*).

The scattering intensity of such systems – for example, of liquid crystals (LC) – is a set of more or less diffuse peaks. The relation between the structure $\rho_{LC}(\mathbf{r})$ and intensity $I(\mathbf{S})$ is shown in Fig. 10. In this case the intensity depends on the constantly changing reciprocal-space vector $\mathbf{S}$, and not on the discrete set $\mathbf{S_H} = \mathbf{H}$, which is found in crystals (Fig. 2). Of course, here we have a more difficult problem than in the case of crystals. The positions of molecules are changing and, as the result, ρ_{LC} is averaged in time and space. The main structural feature of the liquid-crystal state is the approximately parallel array of molecules with high lability of contacts between them. Such packing determines their short-range order and can be characterized by the statisitical function $W(r)$ of side distances between the molecular centres and the function $\tau(z)$ of the distances between the molecular centres along the axis. The orientation function $f(\omega)$ of the molecular axis, which is statistical in character, should also be introduced. For liquids it is spherically averaged, and for high-molecular-weight polymers and liquid crystals it has a cylindrical symmetry as the result of averaging around the molecular axes of the LC. Accordingly, the distribution of intensity in reciprocal space $I = I(R,Z)$ is also cylindrically symmetrical. Fig. 11 shows the structural models of nematic and smectic LC and the optical diffraction from them which imitates well the observed X-ray patterns (Vainshtein & Chistyakov, 1975).

The nematic LC molecules (with their parallel arrangement preserved) occupy random positions along the z axis. In the smectic LC, the molecular axes are also nearly parallel to each other, but the molecules are packed in layers. The position of layers

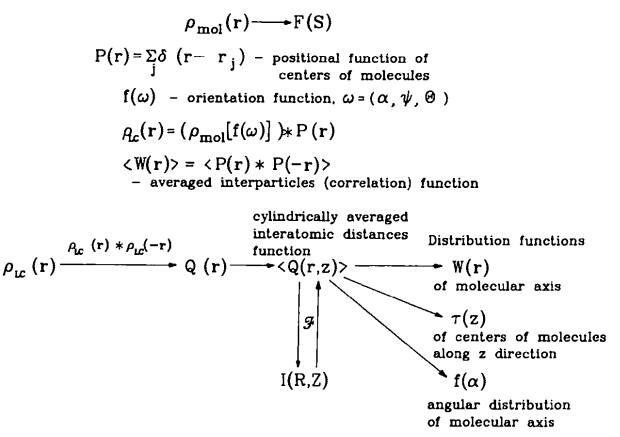

Fig. 10. A scheme illustrating scattering by liquid crystals ρ_{LC} and method of obtaining of the distribution functions. ρ_{mol}: the structure of one molecule; $W(r)$: distribution function of the projections of the molecular axis onto the base plane; $\tau(z)$: projection of the distribution of molecular centres onto the main axis.

is described by a one-dimensional pseudoperiodical function $\tau(z)$; the ordering of molecules in layers may be different. In smectic A LC the molecules are perpendicular to the plane of the layer. In smectic C LC the molecular axes are inclined to the plane of the layer.

From the intensity distribution along the equator $I(R)$ we can derive the distribution function of side distances between the atoms:

$$2\pi r Z_M(r) = 2\pi r Z_0 + 4\pi^2 r \int_0^\infty i(R) J_0(2\pi r R) R \mathrm{d}R. \quad (8)$$

As the molecular structure is usually known, from (1) and (2) we can obtain the function of intermolecular distances $W(r)$ – the projection of molecular axes on the plane perpendicular to the principal axes. The distribution of intensity $I(z)$ along the meridian gives us the function $\tau(z)$. The angular density of arc reflections determines the axial spread $f(\alpha)$.

An example of the radial function $W(r)$ of distances between the molecular axes is given in Fig. 12. This is a nematic LC – *p*-azoxyanisole (PAA), oriented by a constant electric field (4000 V cm^{-1}) (Vainshtein, Chistyakov, Kosterin & Chaikovsky, 1969).

The most complete information from the X-ray diffraction pattern of LC can be obtained by applying the cylindrically symmetric function of interatomic distances $Q(r,z)$ (MacGillavry & Bruins, 1948; Vainshtein & Chistyakov, 1975):

$$Q(rz) = 2\int\int |F(RZ)|^2 J_0(2\pi r R) \times \cos(2\pi z Z) 2\pi R \mathrm{d}R \mathrm{d}Z. \quad (9)$$

It contains both inter- and intramolecular distances between atoms. The $Q(r,z)$ function for ethyl *p*-(*p*-anisylamino)cinnamate (EAAC) is shown in Fig. 13. The packing of molecules into smectic layers is evident.

Some smectics have several phase modifications that can be explained by the unequal values of side interactions of long molecules in the smectics which possess different 'firmness' along the axis, and the corresponding difference of thermal vibrations. Thus, aliphatic 'tails' of such molecules are practically melted, which makes it possible for the layers to glide easily with respect to each other.

In liquid crystals composed of polar molecules bilayers may appear, the thickness of which is less than the double length of the molecule, $2L$, due to the interpenetration of aliphatic tails (Ostrovsky, 1989).

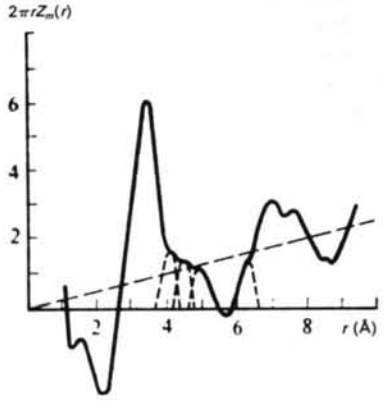

Fig. 12. Cylindrical distribution function of the molecular axis $W(r)$ for PAA oriented by electric field.

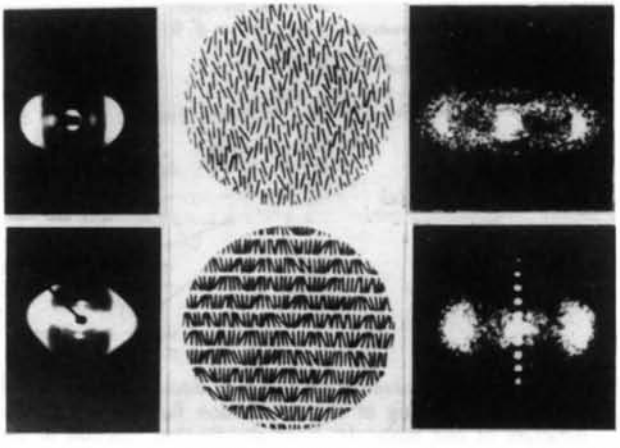

Fig. 11. Scattering by LC. Left: X-ray diagrams; centre: model of structure; right: optical diffraction pattern from model. Top: nematic structure; bottom: smectic structure.

Fig. 13. Two-dimensional cylindrical Patterson function $Q(r,z)$ of EAAC obtained by the optical diffraction method. Solid lines indicate the calculated $Q(r,z)$.

Structure of biomolecules in solution

Now I would like to discuss briefly small-angle scattering, taking the case of identical particles *i.e.* molecules in solution. In this case spherical averaging of the orientation function of molecules takes place. Molecules are far apart, so we can disregard the intermolecular interference. The relationship between the structure and intensity is shown in Fig. 14 (Vainshtein, Feigin & Svergun, 1982).

Intensities from all molecular orientations in reciprocal space are mixed and the $I(s)$ function [$s = 4\pi(\sin\theta/\lambda)$] is spherically symmetric and one-dimensional. From the intensity distribution of the zero peak, Guinier scattering invariants can be found – gyration radius R, molecular volume V, surface S, maximum size l_{max} (Guinier & Fournet, 1955). It seems that the problem of obtaining $\rho_{mol}(\mathbf{r})$ from $I(s)$ cannot be solved. Still, a solution, at least for some objects, has been found. In order to achieve this, precise measurement of the $I(s)$ curve for high values of S corresponding to scattering angles of 2θ is necessary (Fig. 15).

An important step towards solving the problem was made by Stuhrmann (1970), who proposed applying Fourier transformation in spherical coordinates and represented intensity with the aid of spherical harmonics. The electron density of the molecule $\rho(\mathbf{r})$ is expressed by the series

$$\rho(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \rho_{lm}(r) Y_{lm}(\theta,\varphi) = \sum_{l=0}^{\infty} \rho_l(\mathbf{r}), \quad (10)$$

where z, θ, φ are the spherical coordinates, Y_{lm} the spherical harmonics, ρ_{lm} the radial functions, and ρ_l the multipole densities.

The intensity of small-angle scattering is

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} |A_{lm}(s)|^2. \quad (11)$$

The radial functions $\rho_{lm}(r)$ and scattering amplitudes $A_{lm}(s)$ are connected by a Hankel transform of order l:

$$A_{lm}(s) = i^l (2/\pi)^{1/2} \int_0^{\infty} \rho_{lm}(s) j_l(sr) r^2 \mathrm{d}r, \quad (12)$$

$$\rho_{lm}(r) = (-i)^l (2/\pi)^{1/2} \int_0^{\infty} A_{lm}(s) j_l(sr) r^2 \mathrm{d}r. \quad (13)$$

Using certain restrictions, (11) can be rewritten as follows:

$$I(s) = 2\pi^2 \sum_{l=0}^{\infty} |A_l(s)|^2, \quad (14)$$

where $|A(s)|^2$ determines the radial function ρ_l by use of the Hankel transformation.

Svergun, Feigin & Schedrin (1982) (see also Feigin & Svergun, 1987) proposed a method for the reconstruction of $\rho(\mathbf{r})$ which consists of the following. First of all, it is necessary to know the symmetry of the object. For instance, in the case of cylindrical symmetry $\rho_{lm} = 0$ if $m = 0$, and $\rho_l(r) = \rho_{l_0}(r)$. Limiting conditions for the particle size $\rho(r) = 0$ if $r > R$ are also used. As the first approximation the function

$$\rho_l^{(0)}(r) = \Pi(r - R) = \begin{cases} 1, & r \leq R \\ 0, & r > R \end{cases} \quad (15)$$

is chosen. Then by gradual normalization of amplitudes A_l and experimental values for $I(s)$ we get $L + 1$ functions $\rho_l^{(k)}(r)$:

$$\tilde{A}_l^{(k)}(s) = A_l^{(k)}(s)[I(s)/I^{(k)}(s)]^{1/2} \quad (16)$$

$$\tilde{\rho}_l^{(k+1)}(r) = \tilde{\rho}_l^{(k)}(r)\Pi(r - R) = \begin{cases} \rho_l^{(k)}, & r \leq R \\ 0, & r > R. \end{cases} \quad (17)$$

Then, taking into account the limiting values of the density of the scattering object $\rho_{min} \leq \rho(r) \leq \rho_{max}$, we proceed to the next approximation of $A_l^{(k+1)}$ by (14), therefore approaching the true intensity $I(s)$. The process shows convergence, which is estimated in a way similar to that used for crystals, by an integral reliability factor R:

$$R_I = \int_{s_1}^{s_2} [I_{model}(s) - I(s)]^2 s^4 \mathrm{d}s \Big/ \int_{s_1}^{s_2} I^2(s) s^4 \mathrm{d}s. \quad (18)$$

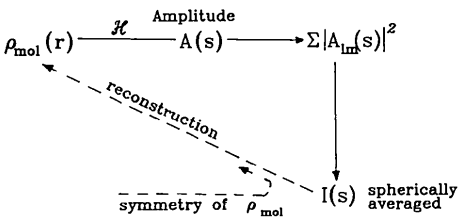

Fig. 14. Scattering scheme for macromolecules in solution (A: scattering amplitudes) and $\rho_{mol}(\mathbf{r})$ reconstruction.

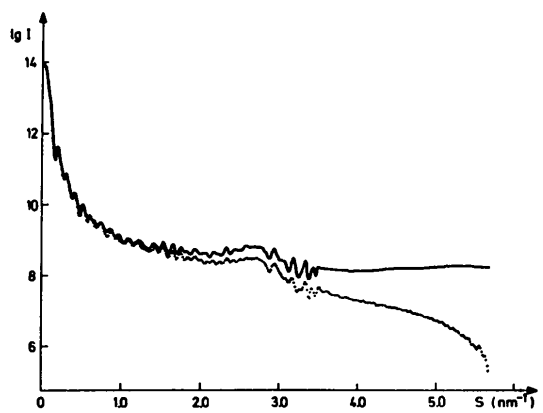

Fig. 15. The scattering curves of bacteriophage T7 in solution. The solid line indicates the experimental curve, the dotted line the calculated curve.

An example of the application of the procedure to the construction of an electron density map for bacteriophage T_7, with scattering curves I_{exp} and I_{cal}, is shown in Fig. 15. This bacterial virus possesses an approximately axial symmetric structure. The result of structure determination at 12 Å resolution is shown in Fig. 16. The phage head, tail, and details of the structure are clearly seen. Inside the head there is a protein core 24 nm in diameter. Regions with greater density probably correspond to DNA. The intensity curve calculated from this model (Fig. 15 – dashed line) is in good agreement with the experimental one.

Structure investigation of protein crystals: catalases

As protein crystals possess an extremely complex structure, the traditional methods of phase determination used for ordinary crystals cannot be used in the X-ray structure investigation of proteins (*cf.* Fig. 2).

The basic technique for such crystals is the method of multiple isomorphous replacement (MIR). It consists of measuring the intensity I_P of the protein crystal and I_{P+H_i} of the protein crystals with heavy-atom additions H_i. Heavy-atom positions are determined by comparing the Patterson functions Q_P and Q_{P+H_i}. Then, comparing $|F_P|$ and $|F_{P+H_i}|$, the phases of $|F_P|$ are determined (Green, Ingram & Perutz, 1954; see also Blundell & Johnson, 1976).

If the crystal possesses non-crystallographic symmetry, for example, when there are two molecules or equal parts of a molecule (subunits) in the asymmetric unit, it is of considerable help in the structure analysis of the protein. Another possibility is to take the solution flattening into account.

If it is known that molecules in the protein crystal P are similar in structure to those of protein P' already studied, the method of molecular replacement can be used. Comparison of Q_P and $Q_{P'}$, construction of rotation and translation functions, allows determination of the orientation and position of the molecules in ρ_P (Rossmann & Blow, 1962; Rossmann, 1990).

The process of protein structure investigation may be illustrated by the example of catalases. Electron microscopic data and three-dimensional reconstruction of beef liver catalase (BLC) showed that its molecule (M.w. 240 000) contains four subunits and possesses tetrahedral symmetry (Vainshtein, Barynin & Gurskaya, 1968). Later, we determined the structure of fungal catalase from *Pennicillium vitale* (PVC). MIR and non-crystallographic symmetry were used for the structure determination (Vainshtein, Melik-Adamyan, Barynin, Vagin & Grebenko, 1981). Crystals that yield resolution better than 2 Å were grown in a centrifuge at 25 000 g acceleration. PVC has molecular weight 290 000, space group $P3_121$, $a = 144{\cdot}6$, $c = 138{\cdot}8$ Å; the number of molecules in the unit cell $n = 3$, $\frac{1}{2}$ molecule per asymmetric unit. MIR was used to reveal and determine non-crystallographic symmetry. The molecule contains four subunits (Fig. 17). Only one of the 222 axes of the molecule coincides with the 2 axis of the space group, the other twofold axes determine the non-crystallographic symmetry.

A scheme of the elements of the tertiary structure of the PVC subunit is shown in Fig. 18. The polypeptide chain which makes up a PVC subunit contains 670 residues. The first 56 residues are situated away from the subunit globule and are involved in many contacts with amino-acid residues of neighbouring

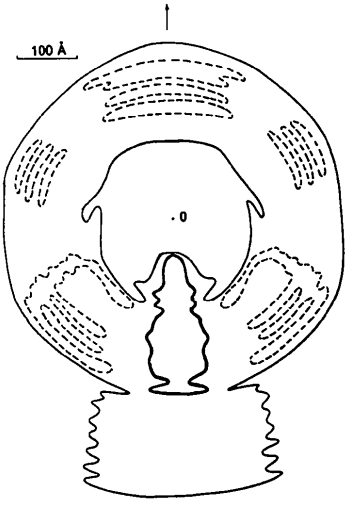

Fig. 16. The electron density map for bacteriophage T7; cross-section of cylindrical three-dimensional distribution. Solid line: level 0·38 Å^{-3} (hydrated protein); dashed line: 0·42 Å^{-3} (strongly hydrated DNA); thick line: 0·52 Å^{-3} (slightly hydrated DNA).

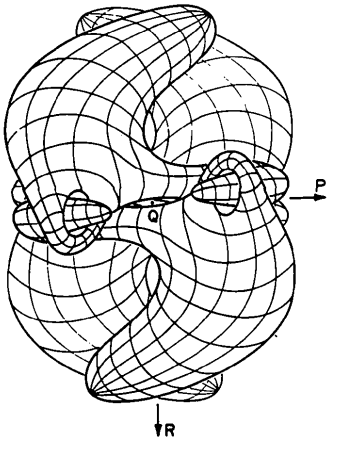

Fig. 17. Quarternary structure of the PVC molecule. P, Q and R are twofold axes.

subunits. The polypeptide chain forms three domains. The largest domain (I) consists of about 300 residues and contains a β-barrel of eight anti-parallel strands forming a surface of the closed near-cylindrical β-sheet, in which strands are interchanged with helical segments. Then the polypeptide chain has an irregular segment of about 70 residues connecting domain (I) with the smaller domain (II). This domain contains 70 residues forming four α-helices. The C-terminal domain (III) contains about 150 amino acids forming a sheet of five parallel β-strands and four α-helices which are above and below the β-sheet. Its topology is similar to that of flavodoxin.

The heme group in catalase is positioned deep inside the molecule in the large domain. Fig. 19 shows the structure of the active site. On the proximal side a close contact with the iron atom is made by a tyrosine residue whose phenolic group occupies the fifth coordination position. On the distal side a histidine residue is nearest to the iron atom (Vainshtein, Melik-Adamyan, Barynin, Vagin, Grebenko, Borizov, Bartels, Fita & Rossmann, 1986).

The structure of beef liver catalase has also been determined (Fita, Silva, Murthy & Rossmann, 1986). Comparison of the structures of PVC and BLC turned out to be very interesting (Melik-Adamyan *et al.*, 1986). The conformation of the polypeptide chain in the first two domains and especially around the active site was found to be fairly similar. Surprisingly, it turned out that the 'flavodoxin' domain (III) in the C-terminal part of PVC subunit is completely absent in BLC. This can probably be explained by processes which occurred during the evolution of these proteins.

The structure of another heme-containing bacterial catalase – from *Micrococcus lysodeikticus* (MLC) – was also investigated. MLC has a two-domain structure like BLC, and not a three-domain structure like PVC. The tertiary structure of domains (I) and (II) was fairly similar to that of BLC and PVC (Yusifov *et al.*, 1989).

It seemed interesting to check whether the catalase structures established were universal for all types of catalases. Therefore we investigated an extremely thermophilic bacterial catalase from *Thermus Thermophilis* (TTC) (Vainshtein, Melik-Adamyan, Barynin, Vagin & Grebenko, 1985), M.w. = 206 000. The space group is cubic, $P2_13$, $a = 133{\cdot}4$ Å. Electron

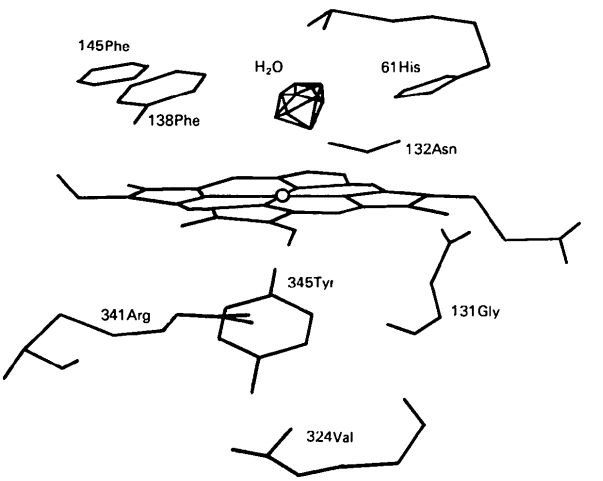

Fig. 19. Active centre of PVC. The heme group is in the centre.

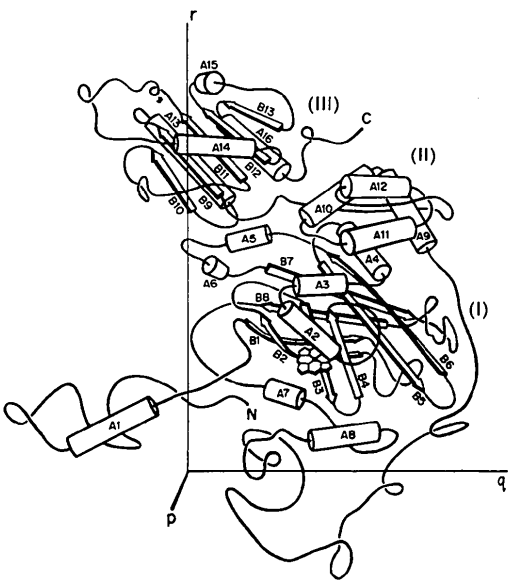

Fig. 18. Tertiary structure of one PVC subunit: domains (I) (with heme), (II) and (III). α-Helices are shown by cylinders, β-strands by arrows.

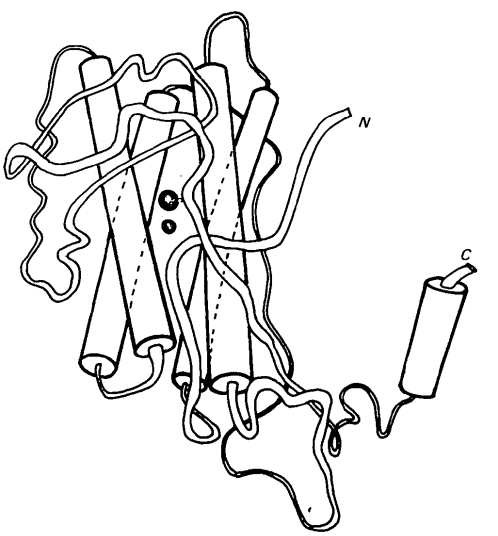

Fig. 20. Tertiary structure of the T-catalase subunit and two Mn atoms inside.

density maps at 2·85 Å resolution showed that the molecule is a hexamer, and allowed the course of the polypeptide chain in the subunit of TTC to be established and also revealed the active site. The framework of the subunit consists of four almost parallel long α-helices. In the high-density region, located in the middle of the helices, two Mn atoms are at the active site of TTC (Fig. 20). The packing mode of the central helices allows catalase T to be classified as a four-helical protein. This class of proteins contains hemoerythrine, some cytochromes, apoferritin, and a protein of TMW.

Thus it was established that catalases which are functionally similar fall into at least two different – chemically and structurally – classes.

This investigation of catalases allows the structure of a protein molecule to be interpreted and its enzymatic activity explained.

Summing up, I would like to emphasize once again that diffraction studies are undergoing rapid development and give important new results. The development of the theory and experimental techniques will render most interesting results on the atomic structure of crystals and the condensed state of matter in general.

Many scientists took part in the investigations I have mentioned, and I am grateful for the aid and assistance I have had from them for many years. I would especially like to pay tribute to the late Professor Z. G. Pinsker in whose laboratory I started my investigations by electron diffraction and Professor I. G. Chistyakov who worked with me on liquid crystals. I would also like to thank L. I. Tatarinova, B. B. Zvyagin, E. A. Kosterin, L. A. Feigin, D. I. Svergun, D. M. Kheiker, V. R. Melik-Adamyan, B. B. Barynin and A. A. Vagin for their cooperation.

References

AVILOV, A. S., SEMILETOV, S. A. & STOROZENKO, V. V. (1989). *Sov. Phys. Crystallogr.* **34**, 110–114.

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*. New York: Academic Press.

CHAPYRINA, L. F., DIAKON, I. A., DONU, S. V. & BYDNIKOV, S. S. (1986). *Problemy Sovremennoy Bioorganicheskoy Khimii*, pp. 228–234. Novosibirsk: Sibirskoye Otd Nauka.

COWLEY, J. M. (1953). *Acta Cryst.* **3**, 516–529.

DORSET, D. L. (1976). *Acta Cryst.* A**32**, 207–215.

DORSET, D. L. (1983). *Ultramicroscopy*, **12**, 19–27.

DRITS, V. A. (1987). *Electron Diffraction and High-Resolution Electron Micoscopy of Mineral Structures*. Heidelberg: Springer-Verlag.

EWALD, P. P. (1913). *Phys. Z.* **14**, 465–472.

EWALD, P. P. (1916). *Ann. Phys. (Leipzig)*, **49**, 1–38, 117–143.

EWALD, P. P. (1917). *Ann. Phys. (Leipzig)*, **54**, 519–556, 557–597.

EWALD, P. P. (1920). *Z. Phys.* **2**, 332–342.

EWALD, P. P. (1921). *Z. Kristallogr.* **56**, 129–156.

EWALD, P. P. (1924). *Z. Phys.* **30**, 1–19.

EWALD, P. P. (1933). *Handbuch der Physik*, Vol. 23, Part 2, pp. 207–476. Berlin: Springer.

FEIGIN, L. A. & SVERGUN, D. I. (1987). *Structure Analysis by Small Angle X-ray and Neutron Scattering*. New York, London: Plenum Press.

FITA, I., SILVA, A. M., MURTHY, M. R. N. & ROSSMANN, M. G. (1986). *Acta Cryst.* B**42**, 497–515.

GREEN, D. W., INGRAM, V. M. & PERUTZ, M. F. (1954). *Proc. R. Soc. London Ser. A*, **25**, 287–300.

GUINIER, A. & FOURNET, G. (1955). *Small-Angle Scattering of X-rays*. New York: Wiley.

HOSEMANN, R. & BAGCHI, S. N. (1962). *Direct Analysis of Diffraction by Matter*. Amsterdam: North-Holland.

MACGILLAVRY, C. H. & BRUINS, E. M. (1948). *Acta Cryst.* **1**, 156–158.

MELIK-ADAMYAN, W. R., BARYNIN, V. V., VAGIN, A. A., BORISOV, V. V., VAINSHTEIN, B. K., FITA, I., MURTHY, M. R. N. & ROSSMANN, M. G. (1986). *J. Mol. Biol.* **188**, 63–72.

OSTROVSKY, B. I. (1989). *Sov. Sci. Rev. A Phys.* **12**, 85–146.

PINSKER, Z. G. (1953). *Electron Diffraction*. London: Butterworth.

PINSKER, Z. G. & IMAMOV, R. M. (1981). *Indian J. Pure Appl. Phys.* **19**, 926–932.

ROSSMANN, M. G. (1990). *Acta Cryst.* A**46**, 73–82.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.

STUHRMANN, H. B. (1970). *Acta Cryst.* A**26**, 297–306.

SVERGUN, D. I., FEIGIN, L. A. & SCHEDRIN, B. M. (1982). *Acta Cryst.* A**38**, 827–835.

UNWIN, P. N. T. & HENDERSON, R. (1975). *J. Mol. Biol.* pp. 425–440.

VAINSHTEIN, B. K. (1954). *Tr. Inst. Kristallogr. Akad. Nauk SSSR*, **9**, 259–276.

VAINSHTEIN, B. K. (1955). *J. Phys. Chem. USSR*, **29**, 327–344.

VAINSHTEIN, B. K. (1956). *Structurnaya Electronografiya*. Moscow: Izd. Akad. Nauk USSR.

VAINSHTEIN, B. K. (1960). *Q. Rev. Chem. Soc.* **14**, 105–132.

VAINSHTEIN, B. K. (1964). *Structure Analysis by Electron Diffraction*. Oxford: Pergamon Press.

VAINSHTEIN, B. K. (1966). *Diffraction of X-rays by Chain Molecules*. Amsterdam: Elsevier.

VAINSHTEIN, B. K., BARYNIN, V. V. & GURSKAYA (1988). *Dokl. Akad. Sci. USSR*, **182**, 569–572.

VAINSHTEIN, B. K. & CHISTYAKOV, I. G. (1975). *Problemy Sovremennoy Krystallografii*, pp. 12–26. Moscow: Izd. Nauka.

VAINSHTEIN, B. K., CHISTYAKOV, I. G., KOSTERIN, E. A. & CHAIKOVSKY, V. M. (1969). *Mol. Cryst. Liq. Cryst.* **8**, 457–470.

VAINSHTEIN, B. K., FEIGIN, L. A. & SVERGUN, D. I. (1982). *Acta Phys. Acad. Sci. Hung.* **53**(1/2), 105–119.

VAINSHTEIN, B. K., MELIK-ADAMYAN, W. R., BARYNIN, V. V., VAGIN, A. A. & GREBENKO, A. I. (1981). *Nature (London)*, **293**, 411–412.

VAINSHTEIN, B. K., MELIK-ADAMYAN, W. R., BARYNIN, V. V., VAGIN, A. A. & GREBENKO, A. I. (1985). *J. Biosci.* **8**(Suppl.), 471–479.

VAINSHTEIN, B. K., MELIK-ADAMYAN, W. R., BARYNIN, V. V., VAGIN, A. A., GREBENKO, A. I., BORISOV, V. V., BARTELS, K. S., FITA, I. & ROSSMANN, M. G. (1986). *J. Mol. Biol.* **188**, 49–61.

VAINSHTEIN, B. K. & PINSKER, Z. G. (1949). *Dokl. Akad. Nauk SSSR*, **64**, 49–52.

VAINSHTEIN, B. K. & TATARINOVA, L. I. (1967). *Conformation of Biopolymers*, Vol. 2, pp. 569–582. London, New York: Univ. of Madras.

YUSIFOV, E. F., GREBENKO, A. I., BARYNIN, V. V., MURSHUDOV, G. N., VAGIN, A. A., MELIK-ADAMYAN, W. R. & VAINSHTEIN, B. K. (1989). *Sov. Phys. Crystallogr.* **34**, 1451–1456.

ZVYAGIN, B. B. (1967). *Electron Diffraction Analysis of Clay Mineral Structure*. New York: Plenum Press.

ZVYAGIN, B. B. (1989). *Metody Structurnogo Analisa*, pp. 205–216. Moscow: Nauka.