Acta Cryst. (1989). A**45**, 794–797

Application of the P_s-Function Method to Macromolecular Structure Determination

BY HAO QUAN AND M. M. WOOLFSON

Department of Physics, University of York, York YO1 5DD, England

(*Received* 23 *March* 1989; *accepted* 23 *June* 1989)

Abstract

The P_s function derived from anomalous-dispersion data [Okaya, Saito, & Pepinsky (1955). *Phys. Rev.* **98**, 1857–1858] has been tested with observed data for an Hg derivative of a small protein, avian pancreatic polypeptide [Glover, Moss, Tickle, Pitts, Haneef, Wood & Blundell (1985). *Adv. Biophys.* **20**, 1–12]. The P_s map was superimposed on the four Hg sites *via* a sum function and negative densities were eliminated from the resultant map. This map, with appropriate density inserted at Hg sites, closely resembles a map calculated with true phases; the two maps have a correlation coefficient of 0·67. For 2109 reflexions the unweighted mean phase error is 39·9° but with $|F_oF_c|$ weighting this reduces to 29·5°.

Introduction

Okaya, Saito & Pepinsky (1955) proposed the use of anomalous X-ray dispersion for the solution of crystal structures through interpretation of the function

$$P_s(\mathbf{u}) = \sum_{\mathbf{h}} [|F(\mathbf{h})|^2 - |F(\bar{\mathbf{h}})|^2] \sin 2\pi \mathbf{h} \,.\, \mathbf{u}. \quad (1)$$

The scattering factor for an atom in the case of anomalous scattering can be written as $f + if''$ where the real part, f, includes an anomalous component. The structure factor, $F(\mathbf{h})$, can be written as

$$F(\mathbf{h}) = \sum_{j=1}^{n} (f_j + if_j'') \exp 2\pi i \mathbf{h} \,.\, \mathbf{r}_j \quad (2)$$

and

$$|F(\mathbf{h})|^2 = F(\mathbf{h})^* F(\mathbf{h}) = \sum_{i=1}^{n} \sum_{j=1}^{n} (f_i + if_i'')(f_j - if_j'') \times \exp [2\pi i \mathbf{h} \,.\, (\mathbf{r}_i - \mathbf{r}_j)]. \quad (3)$$

Similarly

$$|F(\bar{\mathbf{h}})| = \sum_{i=1}^{n} \sum_{j=1}^{n} (f_i + if_i'')(f_j - if_j'') \times \exp [-2\pi i \mathbf{h} \,.\, (\mathbf{r}_i - \mathbf{r}_j)]. \quad (4)$$

From (3) and (4)

$$|F(\mathbf{h})|^2 - |F(\bar{\mathbf{h}})|^2 = 2i \sum_{i=1}^{n} \sum_{j=1}^{n} (f_i + if_i'')(f_j - if_j'') \times \sin 2\pi \mathbf{h} \,.\, (\mathbf{r}_i - \mathbf{r}_j). \quad (5)$$

Table 1. *$f_j'' f_i - f_i'' f_j$ values with respect to different i, j combinations*

		i	
		a.s.	n.s.
j	a.s.	0	$f_j'' f_i$
	n.s.	$-f_i'' f_j$	0

Combination of terms involving (i, j) and (j, i) gives

$$[(f_i + if_i'')(f_j - if_j'') - (f_i - if_i'')(f_j + if_j'')] \times \sin 2\pi \mathbf{h} \,.\, (\mathbf{r}_i - \mathbf{r}_j) = 2i(f_i'' f_j - f_j'' f_i) \times \sin 2\pi \mathbf{h} \,.\, (\mathbf{r}_i - \mathbf{r}_j). \quad (6)$$

Combining (5) and (6), we obtain

$$|F(\mathbf{h})|^2 - |F(\bar{\mathbf{h}})|^2 = 2 \sum_{i=1}^{n} \sum_{j=1}^{n} (f_j'' f_i - f_i'' f_j) \times \sin 2\pi \mathbf{h} \,.\, (\mathbf{r}_i - \mathbf{r}_j). \quad (7)$$

Thus the P_s function has peaks of weight proportional to $(f_i'' f_j - f_j'' f_i)$ at positions $\mathbf{r}_j - \mathbf{r}_i$ [see, for example, Pepinsky, Robertson & Speakman (1961), pp. 273–277]. The information contained in this function is best appreciated by considering the case with m anomalous scatterers (a.s.) all of the same kind in the presence of normal scatterers (n.s.). For i = a.s. and j = a.s.

$$f_j'' f_i - f_i'' f_j = (f_j''/f_j - f_i''/f_i) f_i f_j. \quad (8)$$

which always equals zero because the ratio f_j''/f_j is independent of the site occupancy or the thermal motion. The $f_j'' f_i - f_i'' f_j$ values are listed in Table 1.

Thus there are positive peaks from each anomalous scatterer to each non-anomalous scatterer and negative peaks in the reverse direction. This antisymmetric pattern is inherent in the form of the P_s function. The P_s map is a highly deconvoluted vector function with only $m(n-m)$ positive peaks instead of the usual $n(n-1)$, where n is the number of atoms and m the number of anomalous scatterers in the unit cell.

The P_s function, considering the positive regions alone, contains information consisting of a degraded superposition of m images of the structure, each image having a different anomalous scatterer at the origin. The degradation is due to positive and negative peaks cancelling where there is any centrosymmetric arrangement in the overlapped images; clearly, if m

0108-7673/89/110794-04$03.00

is large and the structure is complex, a great deal of information is lost in this way.

Pepinsky (1956) and Pepinsky & Okaya (1957) proposed that a superposition method could be used to find a single image of the structure when the positions of the anomalous scatterers were known, and this was demonstrated to be so for small structures. However, since that time there have been great advances in experimental techniques and in computing facilities and we decided to explore the limitations of the P_s function by a test with a small known protein.

The test procedure

The structure chosen for the test was that of the hormone avian pancreatic polypeptide (APP), a small globular protein containing 36 amino-acid residues (Glover *et al.*, 1985). The native product gives X-ray data to 0·98 Å resolution but 2·04 Å anomalous-scattering data are avaliable for an isomorphous mercury derivative (Pitts, Tickle, Wood & Blundell, 1982). The native material forms crystals of space group $C2$ with $a = 34{\cdot}18$, $b = 32{\cdot}92$, $c = 28{\cdot}44$ Å and $\beta = 105{\cdot}30°$. In the Hg derivative there is one heavy atom per molecule, or four in the unit cell. A total of 2109 independent pairs of magnitudes, $|F(\mathbf{h})|$ and $|F(\bar{\mathbf{h}})|$, were available from anomalous-scattering measurements. By the use of anomalous differences

$$\Delta F(\mathbf{h}) = \| F(\mathbf{h})| - |F(\bar{\mathbf{h}})\| \qquad (9)$$

input into either the *MULTAN* (Main, Fiske, Hull, Lessinger, Germain, Declercq & Woolfson, 1980) or *SAPI* (Yao Jiaxing, Zheng Chaode, Qian Jinzi, Han Fuson, Gu Yuanxin & Fan Haifu, 1985) programs, the positions of the four Hg atoms can easily be found (Mukherjee, Helliwell & Main, 1989). These positions can then be used to compute a superposition map *via* a sum function, where the Fourier coefficients are

$$\chi(\mathbf{h}) = [|F(\mathbf{h})|^2 - |F(\bar{\mathbf{h}})|^2] \sum_{i=1}^{m} \exp 2\pi i \mathbf{h} \cdot \mathbf{R}_i, \qquad (10)$$

where $\mathbf{R}_i$ is the position of the ith anomalous scatterer. Under ideal conditions, with no information lost by positive-negative peak annihilation, the resultant map would show an image of the structure, with fourfold weight (assuming equal occupancy of anomalous scatterers) and correctly positioned with respect to a conventional crystallographic origin plus 12 displaced ghost images of the structure with unit weight. Equating all negative density to zero should give a map, Q_s, dominated by density representing the true structure.

We tried modifications of the procedure described above as follows:

(i) remove negativity of the P_s function before calculating a sum function,

(ii) density modification of the final sum function so that all density above a certain minimum level was changed to a fixed value. The rationale here was to reduce the effect of the ghost images while, at the same time, enhancing any slightly reduced regions of the main image.

In fact we found no improvement with these modifications and the results we present are for a Q_s map derived by the first-described procedure.

Table 2. *Errors of Q_s map arranged in descending order of $|F_oF_c|$*

F_o: observed structure factor; F_c: calculated structure factor from Q_s map; NR: number of reflexions in the group; $|F_oF_c|$: minimum $|F_oF_c|$ in the group; WME: $|F_oF_c|$ weighted mean phase error; ME: mean phase error.

NR	$\|F_oF_c\|$	WME(°)	ME(°)
200	3432	23·87	23·09
400	2127	24·62	24·72
600	1483	25·93	27·64
800	1116	26·89	29·54
1000	869	27·44	30·54
1200	649	28·00	31·72
1400	489	28·41	32·76
1600	347	28·85	34·23
1800	201	29·16	35·69
2000	70	29·41	38·09
2109	0	29·47	39·93

Table 3. *Errors of Q_s map grouped in descending order of resolution*

NR: Number of reflexions in the group; RES: resolution range; WME: $|F_oF_c|$ weighted mean phase error; ME: mean phase error.

NR	RES (Å)	WME (°)	ME (°)
21	10·0-	40·34	70·81
8	9·0-10·0	15·63	24·25
8	8·0-9·0	15·03	44·50
20	7·0-8·0	17·07	39·55
27	6·0-7·0	29·48	43·74
63	5·0-6·0	23·00	38·51
128	4·0-5·0	33·70	41·13
361	3·0-4·0	28·42	36·43
1473	2·0-3·0	30·26	40·30

Results and concluding remarks

The first very clear conclusion was that the Q_s map, with density inserted at the Hg sites, bore a very strong resemblance to the true electron density; a conventional correlation coefficient was 0·67. We compared the phases of the Q_s map obtained by Fourier transformation with the true phases and we show this comparison in Table 2. The reflexions are taken in the order of the product $|F_o| \times |F_c|$ where F_o is the observed structure factor and F_c is the Fourier coefficient of the Q_s map. For all the 2109 reflexions available within the 2·04 Å sphere the mean phase error ($\langle|\Delta\varphi|\rangle$) was 39·93° but with $|F_oF_c|$ weighting this was reduced to 29·47°. From the results in the table it is evident that the mean phase error is strongly correlated with the weight being used.

We also analysed the phase errors from the resolution point of view. The results are listed in Table 3.

The very-low-resolution (within 10 Å) solvent-affected reflexions have quite large phase errors. However this situation can be improved by sharpening the data to E's but it turns out that the overall phase error remains unimproved. As is conventional in protein crystallography we excluded reflexions corresponding to resolution greater than 10 Å when computing the Q_s map.

Next we computed a map, Q_{oc}, with the observed structure amplitudes and phases from the Q_s map. This Q_{oc} three-dimensional map, plotted by *FRODO* (Jones, 1985), assembles the characteristics of the correct model. There is little doubt that the Q_{oc} map would quickly have led to a complete structure determination *via* a model-fitting exercise. We show in Fig. 1 a part of the *FRODO* Q_{oc} map, compared with the true model. Again, in Fig. 2 we present two sections of the Q_{oc} map, arbitrarily chosen at $y = 3/8$ and $y = 3/4$, compared with sections calculated with correct phases. For both kinds of presentation the agreement is by no means perfect but the general correspondence can be seen.

We conclude that with modern techniques of collecting optimized anomalous-scattering data, and of carrying out large-scale Fourier transformations quickly and easily even on modest computational facilities, the P_s function approach has much to offer. It is our intention to extend our testing to larger and more complex structures and to explore possible procedures for improving the quality of the acquired phase information.

We are indebted to Dr E. Dodson for providing programs for Fourier transformations and Mr Zhang Yongjian and Dr L. Refaat are thanked for their help in the computational work. We are also grateful to the Royal Society and Chinese Academy of Sciences for supporting the collaboration between the University of York and the Institute of Physics, Beijing and to the Science and Engineering Research Council for their general support of direct-methods work at York. We are also grateful for suggestions by referees which led to an improvement in the presentation of this work.

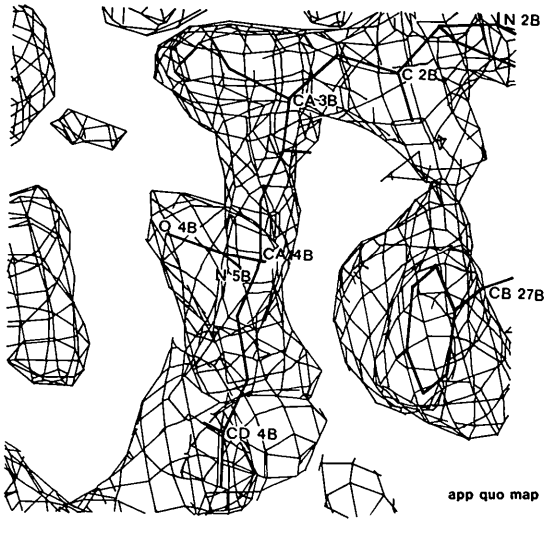

Fig. 1. A part of the three-dimensional Q_{oc} map for Hg-avian pancreatic polypeptide (APP) compared with the true model.

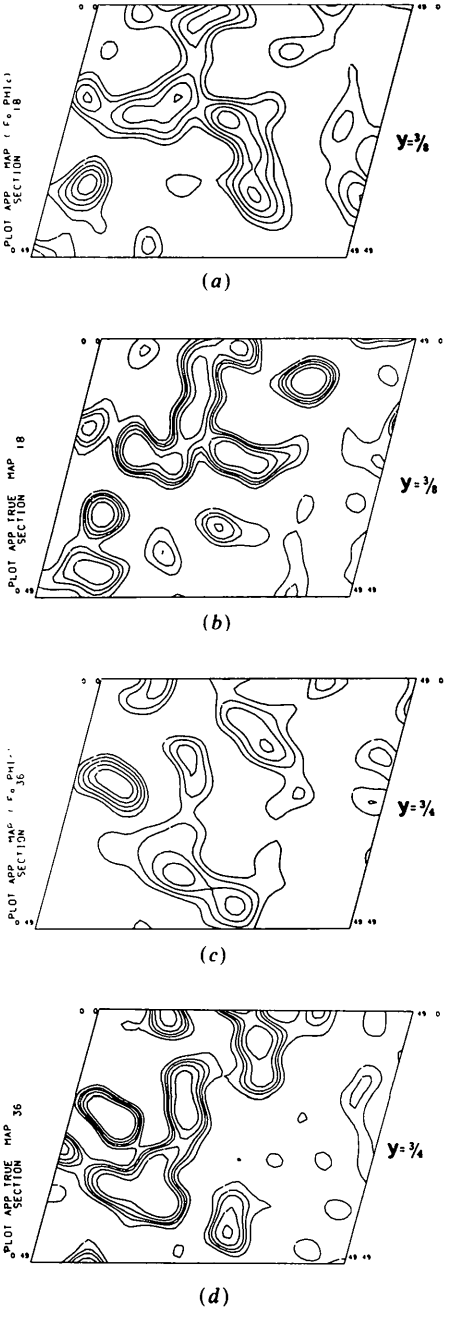

Fig. 2. Representations of electron density for Hg-APP for: (*a*) $y = 3/8$ with phases from the Q_s function; (*b*) $y = 3/8$ with correct phases; (*c*) $y = 3/4$ with phases from the Q_s function; and (*d*) $y = 3/4$ with correct phases.

References

GLOVER, I. D., MOSS, D. S., TICKLE, I. J., PITTS, J. E., HANEEF, I., WOOD, S. P. & BLUNDELL, T. L. (1985). *Adv. Biophys.* **20**, 1-12.

JONES, T. A. (1985). *Methods Enzymol.* **115**, 157-171.

MAIN, P., FISKE, S. J., HULL, S. E., LESSINGER, L., GERMAIN, G., DECLERCQ, J.-P. & WOOLFSON, M. M. (1980). *MULTAN80. A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Univs. of York, England, and Louvain, Belgium.

MUKHERJEE, A. K., HELLIWELL, J. & MAIN, P. (1989). *Acta Cryst.* A**45**, 715-718.

OKAYA, Y., SAITO, Y. & PEPINSKY, R. (1955). *Phys. Rev.* **98**, 1857-1858.

PEPINSKY, R. (1956). *Rec. Chem. Prog.* **17**, 145-189.

PEPINSKY, R. & OKAYA, Y. (1957). *Phys. Rev.* **108**, 1231-1232.

PEPINSKY, R., ROBERTSON, J. & SPEAKMAN, J. (1961). Editors. *Computing Methods and the Phase Problem in X-ray Crystal Analysis.* Oxford: Pergamon.

PITTS, J. E., TICKLE, I. J., WOOD, S. P. & BLUNDELL, T. L. (1982). *Sov. Phys. Crystallogr.* **27**, 56-62.

YAO JIAXING, ZHENG CHAODE, QIAN JINZI, HAN FUSON, GU YUANXIN & FAN HAIFU (1985). *SAPI85. A Computer Program for Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Institute of Physics, Academia Sinica, Beijing, People's Republic of China.

Acta Cryst. (1989). A**45**, 797-801

The Monte Carlo Simulation of Random Stacking Faults in Close-Packed Structures

BY B. I. NIKOLIN AND A. YU. BABKEVICH

Institute of Metal Physics, Academy of Sciences of the Ukrainian SSR, 36 Vernadsky str., 252142 Kiev 142, USSR

(*Received* 22 *February* 1989; *accepted* 23 *June* 1989)

Abstract

A new approach to the estimation of the concentration of random stacking faults in close-packed structures (and also multilayers) is presented. It is based on the Monte Carlo computer simulation of the arrangement of stacking faults in a crystal, given by an appropriate h-k sequence. Thus the corresponding intensity (structure-factor) distribution along the streaked reciprocal-lattice rows may be calculated from nearly the same expression as for a perfect multilayer structure. In particular, good agreement is observed with the computations on the basis of the intensity equations derived for several particular cases. Some peculiarities in the diffracted intensity distribution of crystals with multilayer structures containing random stacking faults of different types, or having different dimensions of the hexagonal unit cell, are pointed out.

1. Introduction

Stacking faults (SF) are frequently observed in close-packed structures. In some cases they are expected to be randomly distributed, *i.e.* the spacing between them is random. Random stacking faults (RSF) may result in one or more of the following diffraction effects: shift, broadening, asymmetry of the diffraction maxima and redistribution of the integrated intensity. Pertinent information about the types of SF as well as their concentrations in crystals can be obtained from a comparison of theoretically predicted diffraction effects with those visible on X-ray diffraction patterns. The theory of the intensity of X-ray diffuse scattering by the simplest close-packed structures such as h.c.p. ($2H$), f.c.c. ($3C$) and $4H$ containing random faults has been well developed by Wilson (1942), Paterson (1952), Christian (1954), Johnson (1963), Lele, Anantharaman & Johnson (1967) and Lele, Prasad & Anantharaman (1969). Because RSF are also observed in multilayer (long-period) polytype-like structures [see Verma & Krishna (1966), Nikolin (1984), Sebastian & Krishna (1987) and literature quoted therein], attempts have been made to construct a more general diffraction theory by Kakinoki & Komura (1965), Kakinoki (1967), Rushits & Mirzaev (1979), Kagan, Unikel' & Fadeeva (1982) and Berliner & Werner (1986). Nevertheless, the Kakinoki & Komura (1965) and Kakinoki (1967) approach, where the correlation between s neighbouring layers must be taken into account, becomes exceedingly complex even for structures with comparatively low periodicity, since 2^{s-1}-order matrices are necessary. Later, the theory was developed by Rushits & Mirzaev (1979), who studied the stacking disorder due to deformation faults only. The Kagan *et al.* (1982) method treated the problem for crystals of any symmetry group and complexity but with low defect concentrations.

In the following, we will give a technique to derive the intensity distribution of any given close-packed structure with arbitrary content of RSF of all existing types. Its distinguishing feature consists of a random arrangement of SF's by the Monte Carlo simulation computer program.

0108-7673/89/110797-05$03.00