Acta Cryst. (1990). A**46**, 935-939

Combining Direct Methods with Isomorphous Replacement or Anomalous Scattering Data. VII. *Ab Initio* Phasing of One-Wavelength Anomalous Scattering Data from a Small Protein*

By Fan Hai-fu, Hao Quan, Gu Yuan-xin, Qian Jin-zi and Zheng Chao-de

Institute of Physics, Chinese Academy of Sciences, Beijing 100080, *People's Republic of China*

and Hengming Ke†

Department of Chemistry, Zhongshan University, Guangzhou, People's Republic of China

(*Received* 12 *March* 1990; *accepted* 3 *July* 1990)

Abstract

The previously proposed method [Fan, Han & Qian (1984). *Acta Cryst.* A**40**, 495-498; Fan & Gu (1985). *Acta Cryst.* A**41**, 280-284] has been improved and tested for *ab initio* phasing of the one-wavelength anomalous scattering data at 2 Å resolution from a known small protein, avian pancreatic polypeptide. Positions of the anomalous scatterers were found by the direct-method program *SAPI* and used to normalize the anomalous differences before calculating the phase doublets. 2108 independent reflections and about 790 000 $\sum_2$ relationships were involved in the phase derivation. The 'best phases' and the figures of merit obtained from the direct method were used to calculate a Fourier map, which revealed the essential feature of the structure.

Introduction

Information on anomalous scattering is important for the determination of protein structures. However, the one-wavelength anomalous-scattering (OAS) method yields two possible solutions to each reflection which is known as the problem of phase ambiguity. If a method can be found to resolve the ambiguity, the OAS method would be a useful technique in protein crystallography, since it is possible to solve a protein structure by either skipping the step of heavy-atom-derivative preparation if it contains suitable anomalous scatterers, or using only a heavy-atom derivative which may not be isomorphous to the native protein. Attempts have been made to resolve the phase ambiguity arising from the OAS technique by direct methods since the 1960's (Fan, 1965; Karle, 1966). More recently, Hauptman (1982), Giacovazzo (1983) and Karle (1984) have succeeded in deriving a large number of three-phase structure invariants from the error-free data of a model protein structure. However, up to the present there is no publication reporting a successful test on real data which leads to a Fourier map of reasonable quality. Apart from direct methods, Hendrickson & Teeter (1981) have proposed a method to resolve the phase ambiguity arising from the OAS technique by making use of the Sim (1959) distribution.

Using experimentally measured diffraction data, we have significantly improved the OAS phases of avian pancreatic polypeptide (aPP) by a method combining Sim (1959) and Cochran (1955) distributions with OAS. As for solving an unknown protein structure, the test was done in four steps: (1) determination of heavy-atom positions; (2) renormalization of anomalous differences; (3) determination of absolute configuration; and (4) derivation of phases. We describe here the details of the improvements to the direct method proposed in papers II and III of this series (Fan, Han & Qian, 1984; Fan & Gu, 1985), and compare our results with those from the method of OAS and from the combination of OAS with the Sim distribution.

Data

aPP is a globular protein containing 36 amino acids. It crystallizes with Zn^{2+} in space group $C2$ with unit-cell parameters $a = 34{\cdot}18$, $b = 32{\cdot}92$, $c = 28{\cdot}44$ Å, $\beta = 105{\cdot}30°$ and $Z = 4$. The structure was solved originally by the method combining single isomorphous replacement and OAS data (Blundell, Pitts, Tickle, Wood & Wu, 1981). In the following test, only the OAS data of the Hg derivative, which were collected with Cu $K\alpha$ radiation at 2·0 Å resolution, were used.

Determination of the heavy-atom sites

The use of direct methods to determine heavy-atom sites with OAS data has been proposed by Mukherjee, Helliwell & Main (1989). The principle of the method is as follows. The anomalous differences can be

---

* Supported in part by the National Natural Science Foundation of China.

† Present address: Department of Chemistry, Harvard University, Cambridge, MA 02138, USA.

0108-7673/90/110935-05$03.00

Table 1. *Results of the determination of heavy-atom sites by a default run of the direct-methods program SAPI*

Peak	x	y	z	Height	Atom
1	0·2454	0·0000	0·2221	5302	Hg
	(0·2444	0·0000	0·2277)*		
2	0·2462	0·0000	0·3309	1788	
3	0·3126	0·1572	0·2843	1498	Zn
	(0·3134	0·1605	0·2967)*		
4	0·3127	−0·1575	0·2843	1479	Zn′
5	0·1914	0·1441	0·2752	960	
6	0·0306	0·0000	0·0985	950	

* Parameters in parentheses are the atomic positions originally obtained by Blundell *et al.* (1981).

expressed as (Blundell & Johnson, 1976)

$$\Delta_{\rm ano} = F^+ - F^- \sim 2F''_{\rm ano} \cos(\varphi_{\rm H} - \varphi''_{\rm ano})$$
$$= 2F''_{\rm ano} \cos \Delta\varphi_{\rm H}, \qquad (1)$$

where $\Delta_{\rm ano}$ can be obtained from the experimental anomalous differences, $\Delta_{\rm obs}$, by correcting the temperature and scaling factors B and K, *i.e.*

$$\Delta_{\rm ano} = \Delta_{\rm obs} K^{1/2} \exp[B(\sin\theta/\lambda)^2], \qquad (2)$$

$F''_{\rm ano}$ and $\varphi''_{\rm ano}$ are respectively the magnitude and phase contributed from the imaginary part of anomalous scattering

$$\mathbf{F}''_{\rm ano} = \sum_{j=1}^{N} i\Delta f''_j \exp(i2\pi\mathbf{H}\,.\,\mathbf{r}_j)$$
$$= F''_{\rm ano} \exp(i\varphi''_{\rm ano}), \qquad (3)$$

where $\varphi_{\rm H}$ is the phase of the corresponding structure factor contributed from the whole unit cell. From (1) it is evident that the largest $|\Delta_{\rm ano}|$'s will be mostly accompanied by large $F''_{\rm ano}$ and large $|\cos \Delta\varphi_{\rm H}|$. For the largest $|\Delta_{\rm ano}|$ we have approximately

$$|\Delta_{\rm ano}|/2 \sim F''_{\rm ano}. \qquad (4)$$

Hence a set of the largest $|\Delta_{\rm ano}|$ can be used to determine the positions of the anomalous scatterers. Direct methods are particularly suitable for this purpose, since they make use of only the largest E's. The whole set of experimental anomalous differences was first normalized by a Wilson plot with $\sum f_j^2$ replaced by $\sum \Delta f''^2_{\rm Hg} + \sum \Delta f''^2_{\rm Zn}$. A default run of the program *SAPI* (Yao, Zheng, Qian, Han, Gu & Fan, 1985; Fan, 1986) resulted in an E map which clearly showed peaks of the anomalous scatterers (Table 1). The first peak in the list corresponds to the Hg atom. The second peak is about 2 Å from the Hg atom and is situated at the first diffraction ripple of Hg. Hence this peak is negligible. The third and fourth peaks correspond to the Zn atom and its enantiomorph respectively. The latter is due to the centrosymmetric arrangement of Hg atoms. At this stage it is impossible to distinguish the true Zn atom from its enantiomorph. For a comparison, the final positions of the Hg and Zn atoms reported by Blundell *et al.* (1981) are cited in Table 1. It should be mentioned that similar results can be obtained also by an automatic Patterson analysis procedure in the *SAPI* program. However, if there are more heavy atoms in the asymmetric unit, direct methods will probably be superior to the Patterson method.

Renormalization of the anomalous differences

Equation (1) can be rewritten as

$$|\Delta\varphi_{\rm H}| = |\cos^{-1}(\Delta_{\rm ano}/2F''_{\rm ano})|. \qquad (5)$$

From (5) we can obtain a phase doublet for each reflection. Since the method is sensitive to $|\Delta\varphi_{\rm H}|$ and thus to $\Delta_{\rm ano}$, it is essential to obtain as accurately as possible the temperature and scaling factors for the experimental anomalous differences. This can be achieved by a renormalization procedure making use of the known positions of the anomalous scatterers. Substitution of (2) into (1) gives

$$\Delta^2_{\rm obs}/4F''^2_{\rm ano}$$
$$= \cos^2 \Delta\varphi_{\rm H} K^{-1} \exp[-2B(\sin\theta/\lambda)^2]. \qquad (6)$$

Averaging (6) over a small interval of $(\sin\theta/\lambda)^2$, it follows that

$$\langle \Delta^2_{\rm obs}/2F''^2_{\rm ano}\rangle = K^{-1} \exp[-2B(\sin\theta/\lambda)^2]. \qquad (7)$$

Here we assume that $\Delta\varphi_{\rm H}$ is uniformly distributed, hence $\langle\cos^2 \Delta\varphi_{\rm H}\rangle \sim \frac{1}{2}$. Once the positions of the anomalous scatterers are known, $F''^2_{\rm ano}$ can be calculated directly from (3), then according to (7) the temperature and scaling factors can be found by a Wilson plot with $\ln\langle\Delta^2_{\rm obs}/2F''^2_{\rm ano}\rangle$ as the ordinate and $(\sin\theta/\lambda)^2$ as the abscissa.

Determination of absolute configuration

In principle, the anomalous differences with their magnitudes and signs bear the information about the absolute configuration. However, in the determination of heavy-atom sites, only the magnitudes are used. This does not allow the absolute configuration of heavy atoms to be determined. Nevertheless, the probability formula (9) below will give different results when changing the heavy-atom configuration from one enantiomorph to the other. Furthermore, only the correct absolute configuration of the heavy atoms will lead to an interpretable result. Hence the problem of absolute configuration can simply be solved by trying both enantiomorphs. In the case of Hg-aPP this problem is further simplified by the centrosymmetric arrangement of the Hg atoms. With the known position of Hg, (9) will yield a set of phases corresponding to the correct absolute configuration.

Derivation of phases

The phase of a reflection with reciprocal vector **H** is expressed by

$$\varphi_{\mathbf{H}} = \varphi'_{\mathbf{H}} + \Delta\varphi_{\mathbf{H}}. \tag{8}$$

In the case of OAS, $\varphi'_{\mathbf{H}}$ is the same as φ''_{ano} of (3). The magnitude of $\Delta\varphi_{\mathbf{H}}$ can be calculated according to (5). The probability for $\Delta\varphi_{\mathbf{H}}$ to be positive is given by

$$\begin{aligned} P_+(\Delta\varphi_{\mathbf{H}}) = \tfrac{1}{2} + \tfrac{1}{2}\tanh \Bigg\{ &\sin|\Delta\varphi_{\mathbf{H}}| \\ &\times \Bigg[\sum_{\mathbf{H}'} m_{\mathbf{H}'} m_{\mathbf{H}-\mathbf{H}'} \kappa_{\mathbf{H},\mathbf{H}'} \\ &\times \sin(\Phi'_3 + \Delta\varphi_{\mathbf{H}'\mathrm{best}} + \Delta\varphi_{\mathbf{H}-\mathbf{H}'\mathrm{best}}) \\ &+ x \sin \delta_{\mathbf{H}} \Bigg] \Bigg\}, \end{aligned} \tag{9}$$

where $x \sin \delta_{\mathbf{H}}$ comes from the Sim (1959) distribution,

$$\begin{aligned} m_{\mathbf{H}} = \exp(-\sigma^2_{\mathbf{H}}/2)\{&[2(P_+ - \tfrac{1}{2})^2 + \tfrac{1}{2}] \\ &\times (1 - \cos 2\Delta\varphi_{\mathbf{H}}) + \cos 2\Delta\varphi_{\mathbf{H}}\}^{1/2} \end{aligned} \tag{10}$$

and

$$\tan(\Delta\varphi_{\mathbf{H}\mathrm{best}}) = 2(P_+ - \tfrac{1}{2}) \sin|\Delta\varphi_{\mathbf{H}}| / \cos \Delta\varphi_{\mathbf{H}}. \tag{11}$$

The values of $\sigma^2_{\mathbf{H}}$ were obtained approximately using the relation*

$$\sigma^2_{\mathbf{H}} = D^2 / 2F''^2_{\mathrm{ano}}, \tag{12}$$

where D^2 is the mean square error of the 'lack of closure', as defined by Blow & Crick (1959). For details of (9), the reader is referred to paper III of this series (Fan & Gu, 1985).

The 2108 reflections within the limit of 2·0 Å resolution were used in the phase derivation. For the calculation of $|\Delta\varphi_{\mathbf{H}}|$ three different normalization procedures were tested. Procedure I used $\sum \Delta f''^2_{\mathrm{Hg}} + \sum \Delta f''^2_{\mathrm{Zn}}$ to replace F''^2_{ano} in (7). Procedure II used only the Hg position to calculate F''^2_{ano} and procedure III made use of both Hg and Zn. The temperature and scaling factors so obtained are listed in Table 2. Note that the 'temperature factor' here does not have the usual physical meaning. A negative value indicates an increase in the anomalous-scattering effect in the high-angle region. Three sets of $|\Delta\varphi_{\mathbf{H}}|$ were calculated according to the values of B and K in Table 2. Each set was independently treated in the phase derivation by the following procedure. All the $P_+(\Delta\varphi_{\mathbf{H}})$ were set to 1/2 at first, and then a new set of $P_+(\Delta\varphi_{\mathbf{H}})$ was calculated using (11), (10) and (9). The results were substituted back into (11) and (10) to obtain an improved set of $\Delta\varphi_{\mathbf{H}\mathrm{best}}$ and $m_{\mathbf{H}}$. The 'best phases' were then calculated as $\varphi_{\mathbf{H}\mathrm{best}} = \varphi'_{\mathbf{H}} + \Delta\varphi_{\mathbf{H}\mathrm{best}}$. The minimum value of $\kappa_{\mathbf{H},\mathbf{H}'}$ accepted for the direct-method calculation was 0·1. This yields ~790 000 $\sum_2$ relationships. For each set of data the above calculations took ~20 min of CPU time in a microVAX-II

Table 2. *Temperature and scaling factors obtained by different normalization procedures*

Procedure	B (Å^2)	K
I	12·95	0·772
II	−3·77	0·932
III	−6·40	0·688

Table 3. *Results of phase derivation in descending order of* $|P_+ - \frac{1}{2}|$

per: percentage of reflections with their signs correctly determined.
err: average error of phase in degrees, calculated according to the differences between the direct-method phases and the true phases from the final structure model obtained by Blundell *et al.* (not including the water molecules).
Procedures I, II and III are based on different temperature and scaling factors as listed in Table 2.

	Procedure					
Number of reflections	I		II		III	
	per	err	per	err	per	err
200	94·0	25·3	97·0	19·4	98·0	18·0
400	91·5	30·5	93·8	25·9	96·0	22·6
600	91·0	31·3	92·5	28·4	94·8	25·2
800	91·3	31·4	92·0	29·5	93·9	26·2
1000	90·8	31·8	91·3	31·1	93·0	27·5
1200	90·1	33·3	90·8	31·7	92·3	28·7
1400	89·6	33·4	89·4	33·4	91·2	30·2
1600	88·9	34·9	88·1	35·7	90·3	31·8
1800	86·9	37·7	85·2	39·6	88·5	34·6
2000	85·3	40·9	85·8	41·1	86·6	38·0

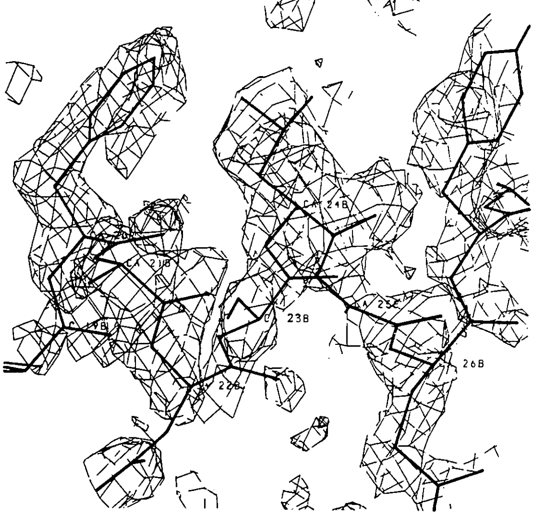


Fig. 1. Portion of the electron-density map calculated from the method combining the direct method and OAS (procedure III). The molecular model is from the refined structure of aPP (Blundell, Pitts, Tickle, Wood & Wu, 1981).

---

* In paper II of this series, we gave $\sigma^2_{\mathbf{H}} = D^2/4F''^2_{\mathrm{ano}} \sin^2 \Delta\varphi_{\mathbf{H}}$ which is based on the approximation $D^2/\sigma^2_{\mathbf{H}} \simeq (\mathrm{d}F_{\mathbf{H}}/\mathrm{d}\alpha_{\mathbf{H}})^2$ while, in this paper, (12) is based on the approximation $D^2/\sigma^2_{\mathbf{H}} \simeq \langle(\mathrm{d}F_{\mathbf{H}}/\mathrm{d}\alpha_{\mathbf{H}})^2\rangle$. We consider that the latter is more reasoable.

Table 4. *Comparison of results from different phasing methods*

aPP 2·0 Å OAS data.

F_{min} minimum F_{obs} in the group; d_{min} minimum d spacing in the group; N number of reflections in the group; A for all kinds of reflections; C for the centric reflections.

I: phases were obtained from the average of the two possible phases for each reflection.

II: phases selected from the two possibilities using the Sim distribution as criterion.

III: phases were selected from the two possibilities using the combination of Sim distribution and Cochran distribution as criterion.

IV: phases were obtained by choosing (according to the true structure model) the correct one from the two possibilities for each reflection.

(*a*) Results in descending order of F_{obs}

			Average phase error in degrees							
	N		I		II		III		IV	
E_{min}	A	C	A	C	A	C	A	C	A	C
500	43	8	70·6	90·0	56·5	90·0	45·6	67·5	24·4	0·0
400	109	13	64·1	90·0	44·7	69·2	37·5	55·4	24·0	0·0
300	271	34	66·5	90·0	40·2	42·4	34·7	37·1	23·5	0·0
200	662	49	66·5	90·0	40·5	51·4	36·6	40·4	26·0	0·0
100	1514	116	69·0	90·0	40·6	48·1	39·1	41·9	27·5	0·0
0	2108	185	72·1	90·0	39·1	37·9	38·3	35·0	28·0	0·0

(*b*) Results in descending order of *d* spacing

			Average phase error in degrees							
	N		I		II		III		IV	
d_{min} (Å)	A	C	A	C	A	C	A	C	A	C
10	21	6	69·9	90·0	73·3	77·1	62·1	51·4	31·0	0·0
9	29	9	72·1	90·0	73·1	100·0	48·5	40·0	26·0	0·0
8	36	12	72·5	90·0	66·5	90·0	46·7	45·0	23·6	0·0
7	57	16	73·0	90·0	61·2	101·0	46·1	56·3	24·4	0·0
6	84	19	70·0	90·0	54·4	94·7	43·4	56·8	25·6	0·0
5	147	29	70·4	90·0	46·1	68·3	39·9	43·4	24·5	0·0
4	276	47	67·0	90·0	42·9	57·4	39·1	49·8	23·1	0·0
3	640	82	66·7	90·0	39·0	46·1	36·7	43·9	23·3	0·0
2	2108	185	72·1	90·0	39·1	37·9	38·3	35·0	28·0	0·0

computer. The reflections in each set were arranged in descending order of $|P_+ - \frac{1}{2}|$ and then accumulated into ten groups. The results are listed in Table 3. It is obvious that procedure III gives the best results.

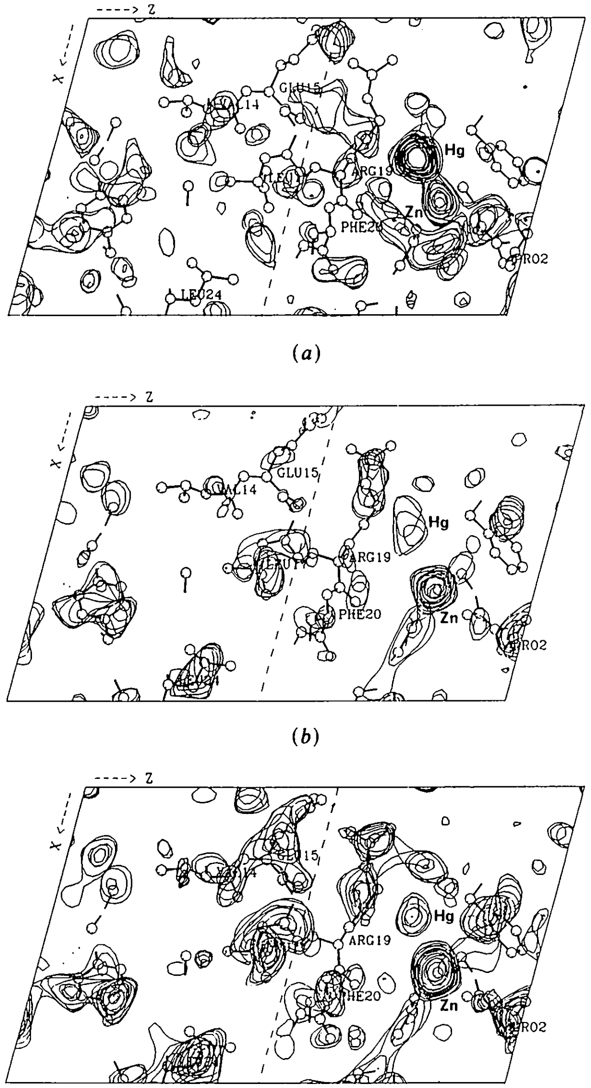
Results and discussion

It is seen in Table 3 that the values of $|P_+ - \frac{1}{2}|$ are in good correlation with the percentages of correctly determined signs and the average phase errors. Hence $|P_+ - \frac{1}{2}|$ can be used as a measure of reliability for the calculated phases when the structure is still unknown. A Fourier map was calculated with the coefficients $m_H F_H \exp(i\varphi_{H\text{best}})$ obtained from procedure III. The reflections up to 2·0 Å resolution were used. Part of the map is plotted with the true structure model by FRODO (Jones, 1985) in Fig. 1. Electron density in the map is essentially continuous, especially for the α-helix (residues 14 to 31). Most positions of the C_α atoms as well as some carbonyl groups can be located easily. However, there are not many recognizable side chains, and the electron density is poor for residues 10 to 15 and for the N and C termini. The correlation coefficient between the map phased by the direct method and that calculated from the true structure is 0·70. This indicates a reasonably good similarity. On the whole, the Fourier map is marginally good for polyalanine trace.

Comparison of different methods

In principle, several methods can break the phase ambiguity arising from the OAS technique. The first and the simplest method is to average the two possible phases for each reflection. The second one is to use the Sim distribution as a criterion to make a choice between the two possibilities. The third uses, instead of the Sim distribution, a combination of Sim distribution and Cochran distribution. The method presented in this paper belongs to the third category. All the above methods have been compared using the OAS data of aPP at 2 Å resolution. Results are listed in Table 4. The results accumulated according to F_{obs} are listed in (*a*), while those accumulated according to *d* spacing are listed in (*b*). Obviously, the third method gives the best results.

Three Fourier maps were calculated using the phases from the above three methods. They are shown in Figs. 2(*a*), (*b*) and (*c*). In the average OAS map (Fig. 2*a*), Hg and Zn are well located in the density although they deviate more than 1 Å from the true positions, in contrast to no density for the protein. In addition to the accurate locations of Hg and Zn atoms, the map from the combination of the Sim distribution and OAS shows most electron density around the protein (Fig. 2*b*), but the map is not interpretable because of the poor connection of density. Compared to Fig. 2(*b*), density in the map from the combination of the direct method and OAS (Fig. 2*c*) is much stronger and its continuity is better. In this map, the shape of density for the α-helix is obvious, and also many C_α and carbonyl group positions can be clearly located.



(*a*)

(*b*)

(*c*)

Fig. 2. Comparison of electron-density maps from three methods: (*a*) averaged OAS, (*b*) OAS plus the Sim distribution, and (*c*) a combination of the direct method and OAS. The maps are projected along the y axis for a region of fractional $y=0{\cdot}11$ to $0{\cdot}21$, and a range of $x=0$ to $0{\cdot}5$ and $z=-0{\cdot}5$ to $0{\cdot}5$ is plotted. The first contour level of all three maps is about one times the background average. Molecular skeletons are plotted as balls and sticks. Two strong peaks are labelled as Hg and Zn. Note that Hg has a true position of $y=0$ which is outside the region of the plot $y=0{\cdot}11$ to $0{\cdot}21$, while Zn ($y=0{\cdot}16$) is at the centre of the y region of this plot.

Concluding remarks

It has been proved that the method presented in this paper is marginally good for *ab initio* phasing a small protein like aPP, using the experimental OAS data at 2 Å resolution. Application of this method to large proteins needs further testing.

In most cases, the present method will be capable of deriving a large number of reliable initial phases. These phases may serve as a good starting set for further phase extension/refinement techniques.

An important part of this work was done in the Department of Physics, University of York, England. We thank Professor M. M. Woolfson for his encouragement and helpful discussions and Dr Wan Zu-li for his help in calculating the *FRODO* map. We are indebted to Professor T. L. Blundell for making the diffraction data of aPP available and to Professor N. Yasuoka for his suggestion of making use of the atomic parameters of the native aPP deposited in the Protein Data Bank. Our gratitude is also due to the British Royal Society and the Chinese Academy of Sciences for supporting collaboration between our laboratories in Beijing, China and York, England.

References

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794-802.
Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*, p. 338. New York: Academic Press.
Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P. & Wu, C. W. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 4175-4179.
Cochran, W. (1955). *Acta Cryst.* **8**, 473-478.
Fan, H.-F. (1965). *Acta Phys. Sin.* **21**, 1114-1118.
Fan, H.-F. (1986). *Rigaku J.* **3**, No. 1, 25-30.
Fan, H.-F. & Gu, Y.-X. (1985). *Acta Cryst.* **A41**, 280-284.
Fan, H.-F., Han, F.-S. & Qian, J.-Z. (1984). *Acta Cryst.* **A40**, 495-498.
Giacovazzo, C. (1983). *Acta Cryst.* **A39**, 585-592.
Hauptman, H. (1982). *Acta Cryst.* **A38**, 632-641.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107-113.
Jones, T. A. (1985). *Methods Enzymol.* **115**, 157-171.
Karle, J. (1966). *Acta Cryst.* **21**, 273-276.
Karle, J. (1984). *Acta Cryst.* **A40**, 4-11.
Mukherjee, A. K., Helliwell, J. R. & Main, P. (1989). *Acta Cryst.* **A45**, 715-718.
Sim, G. A. (1959). *Acta Cryst.* **12**, 813-815.
Yao, J.-X., Zheng, C.-D., Qian, J.-Z., Han, F.-S., Gu, Y.-X. & Fan, H.-F. (1985). *SAPI85. A Computer Program for Automatic Solution of Crystal Structures from X-ray Diffraction Data.* Institute of Physics, Chinese Academy of Sciences, Beijing, China.