Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 7 May 2000
Accepted 17 September 2000

Estimation of individual phases from the four-phase structure invariants in the single isomorphous replacement case

Ning-Hai Hu,* Yong-Sheng Liu and Shu-Yun Wang

Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun 130022, People's Republic of China. Correspondence e-mail: hunh@ciac.jl.cn

The probability distribution of the four-phase invariants in the case of single isomorphous replacement has been developed to estimate some individual phases. An example of its application to obtain the phases having special values of 0, π or $\pm\pi/2$ is given for a known protein structure in space group $P2_12_12_1$. The phasing procedure includes the determination of starting phases and an iterative calculation. The initial values of starting phases, which are required by the formula, can be obtained from the estimate of one-phase seminvariants and by specifying the origin and enantiomorph. In addition, the calculations lead to two sets of possible phases for each type of reflection by assigning arbitrarily an initial phase value. The present method provides a possibility for the multisolution technique to increase greatly the number of known phases while keeping the number of the trials quite small.

1. Introduction

Attempts to solve *ab initio* macromolecular structures led to active approaches that integrate the techniques of direct methods, which have been remarkably successful in determining the structures of small molecules, with either isomorphous replacement or anomalous scattering. The probability distribution of the three-phase structure invariants (3PSIs) based on a combination of direct methods with single isomorphous replacement (SIR) data was pioneered by Hauptman (1982) and subsequently by Giacovazzo *et al.* (1988). Recently, the probabilistic approaches of the four-phase structure invariants (4PSIs) in the SIR case were initiated by Kyriakidis *et al.* (1996) and Giacovazzo & Siliqi (1996*a*,*b*). These results provide a theoretical basis for *ab initio* solution to macromolecular structures by using SIR data.

Although the 4PSI estimates have been used in the non-isomorphous replacement case to improve starting-set and figures-of-merit procedures (Schenk, 1973; De Titta *et al.*, 1975; Cascarano *et al.*, 1987) and, together with the 3PSIs, to solve some small macromolecular structures (Sheldrick, 1993; Burla *et al.*, 1999), their practical applications combining the SIR data remain problematic and need further research. It has been shown that some structure seminvariant phases having values of 0 or π can be directly obtained by a method that incorporates a Σ_1 relationship into the 3PSI distribution in the SIR case (Hu & Liu, 1995). This paper focuses on developing a method that uses the SIR data to estimate some individual phases from the 4PSI distribution based on four pairs of isomorphously related structure factors that we presented in a previous paper (Liu *et al.*, 1999). For space groups up to the orthorhombic system, the phases having special values of 0, π or $\pm\pi/2$ can be directly obtained by means of a sign probability. These phases could be seminvariants, the same as those defined by Σ_1, or non-seminvariants depending on the starting reflections. Therefore, the formulae are equivalent to the Σ_3 relationship of the small-molecular structure determination (Hauptman & Karle, 1953). The method has been tested with the experimental SIR data from a known protein cytochrome c_{550} and its $PtCl_4^{2-}$ derivative (Timkovich & Dickerson, 1976).

2. Theory and method

2.1. The probability distribution of the 4PSIs in the SIR case

Assuming that R, φ and S, ψ are the magnitudes and phases of the normalized structure factors from the native protein, $E = R\exp(i\varphi)$, and the heavy-atom derivative, $G = S\exp(i\psi)$, respectively, when a quartet of reciprocal-lattice vectors $\mathbf{H}$, $\mathbf{K}$, $\mathbf{L}$, $\mathbf{M}$ satisfies $\mathbf{H}+\mathbf{K}+\mathbf{L}+\mathbf{M}=0$, there are 16 4PSIs in the SIR case:

$$\begin{aligned}
\omega_1 &= \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{\mathbf{L}} + \varphi_{\mathbf{M}},\\
&\vdots\\
\omega_5 &= \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \psi_{\mathbf{L}} + \psi_{\mathbf{M}},\\
&\vdots\\
\omega_{16} &= \psi_{\mathbf{H}} + \psi_{\mathbf{K}} + \psi_{\mathbf{L}} + \psi_{\mathbf{M}}.
\end{aligned} \tag{1}$$

The marginal conditional probability distributions of (1), given the eight magnitudes R_j and S_j ($j = 1, 2, 3, 4$), are given by

$$P_i(\omega_i|R_1, R_2, R_3, R_4, S_1, S_2, S_3, S_4) \cong K_i \exp(A_i \cos\omega_i), \quad i = 1, 2, \ldots, 16, \qquad (2)$$

where K_i is a normalizing constant. Considering $\sigma_{4P}\sigma_{2P}^{-2} \ll \sigma_{4H}\sigma_{2H}^{-2}$ in general, the term containing R_j is neglected in the reliability parameters A_i and we obtain the simplified expressions

$$\begin{aligned} A_1 &\cong 2\sigma_{4H}\sigma_{2H}^{-2}\Delta_{1R}\Delta_{2R}\Delta_{3R}\Delta_{4R}, \\ &\vdots \\ A_5 &\cong 2\sigma_{4H}\sigma_{2H}^{-2}\Delta_{1R}\Delta_{2R}\Delta_{3S}\Delta_{4S}, \\ &\vdots \\ A_{16} &\cong 2\sigma_{4H}\sigma_{2H}^{-2}\Delta_{1S}\Delta_{2S}\Delta_{3S}\Delta_{4S}, \end{aligned} \qquad (3)$$

where

$$\begin{aligned} \Delta_{jR} &= \sigma_{2H}^{-1/2}(C_j|F_{jD}| - |F_{jP}|), \\ \Delta_{jS} &= \sigma_{2H}^{-1/2}(|F_{jD}| - C_j|F_{jP}|), \quad j = 1, 2, 3, 4, \end{aligned} \qquad (4)$$

$\sigma_{2H} = \sum_H Z_j^2$, $\sigma_{4H} = \sum_H Z_j^4$, likewise for σ_{2P} and σ_{4P}, the subscripts P, D and H refer to protein, derivative and heavy-atom structures, respectively, and $C_j = I_1(x)/I_0(x)$ is a ratio of two modified Bessel functions (for details, see Giacovazzo & Siliqi, 1996*b*, and Liu *et al.*, 1999). Thus the reliability depends on $|\Delta|$ but not on $|E|$ or $|G|$. The distribution (2) is derived from four main pairs of structure factors, $E_{\mathbf{H}}$, $E_{\mathbf{K}}$, $E_{\mathbf{L}}$, $E_{\mathbf{M}}$ and $G_{\mathbf{H}}$, $G_{\mathbf{K}}$, $G_{\mathbf{L}}$, $G_{\mathbf{M}}$, neglecting the structure-factor pairs of three cross terms $\mathbf{H}+\mathbf{K}$, $\mathbf{H}+\mathbf{L}$ and $\mathbf{K}+\mathbf{L}$. It seems that (2) is a good approximation of the more complicated distribution derived from seven pairs of structure factors (Giacovazzo & Siliqi, 1996*b*) when it is used in the procedure described below.

2.2. The phase estimate formulae

If reciprocal-lattice vectors $\mathbf{H}_1$, $\mathbf{H}_2$, $\mathbf{H}_3$ satisfy

$$\mathbf{H}_1 - \mathbf{H}_2\mathbf{R}_i - \mathbf{H}_3\mathbf{R}_j + \mathbf{H}_3\mathbf{R}_k = 0, \qquad (5)$$

where $\mathbf{R}_i$, $\mathbf{R}_j$ and $\mathbf{R}_k$ are the 3×3 rotation-matrix components of the space-group symmetry operators, we construct the 4PSIs

$$\begin{aligned} \omega_1 &= \varphi_{\mathbf{H}_1} - \varphi_{\mathbf{H}_2\mathbf{R}_i} - \varphi_{\mathbf{H}_3\mathbf{R}_j} + \varphi_{\mathbf{H}_3\mathbf{R}_k}, \\ &\vdots \\ \omega_5 &= \varphi_{\mathbf{H}_1} - \varphi_{\mathbf{H}_2\mathbf{R}_i} - \psi_{\mathbf{H}_3\mathbf{R}_j} + \psi_{\mathbf{H}_3\mathbf{R}_k}, \\ &\vdots \\ \omega_{16} &= \psi_{\mathbf{H}_1} - \psi_{\mathbf{H}_2\mathbf{R}_i} - \psi_{\mathbf{H}_3\mathbf{R}_j} + \psi_{\mathbf{H}_3\mathbf{R}_k}. \end{aligned} \qquad (6)$$

It is noted that equations (6) represent a set of the two-phase structure seminvariants, *e.g.* $\Phi_2 = \varphi_{\mathbf{H}_1} - \varphi_{\mathbf{H}_2}$. The algebraic properties and the probabilistic estimation of Φ_2 as well as its applications have been extensively studied (Green & Hauptman, 1978*a*,*b*; Hauptman & Green, 1978; Giacovazzo, 1979; Giacovazzo *et al.*, 1979; Burla *et al.*, 1989). The following analysis aims at estimating individual phases *via* the two-phase seminvariants using SIR data.

The conditional probability distribution of (6) can be obtained by substituting (6) into (2) given the six magnitudes R_j, S_j ($j = 1, 2, 3$). Let us now consider the distribution of ω_1 as an example, which is written as

$$P(\omega_1) = K_1 \exp[A_1 \cos(\varphi_{\mathbf{H}_1} - \varphi_{\mathbf{H}_2\mathbf{R}_i} - \varphi_{\mathbf{H}_3\mathbf{R}_j} + \varphi_{\mathbf{H}_3\mathbf{R}_k})]. \qquad (7)$$

For a given $\mathbf{H}_1$, when there exists a pair of reflections $\mathbf{H}_2$ and $\mathbf{H}_3$ satisfying (5), we have the probability distribution of $\varphi_{\mathbf{H}_1}$, $P_s(\varphi_{\mathbf{H}_1})$, from (7) for a single quartet. The total distribution of $\varphi_{\mathbf{H}_1}$ for all the reflection pairs ($\mathbf{H}_2$, $\mathbf{H}_3$) is given by the product of individual distributions,

$$\begin{aligned} P(\varphi_{\mathbf{H}_1}) &= \prod_{(\mathbf{H}_2,\mathbf{H}_3)} P_s(\varphi_{\mathbf{H}_1}) \\ &= K_1 \exp\left[\sum_{(\mathbf{H}_2,\mathbf{H}_3)} A_1 \cos(\varphi_{\mathbf{H}_1} - \varphi_{\mathbf{H}_2} + n\pi)\right], \end{aligned} \qquad (8)$$

where $n = 2\mathbf{H}_2\mathbf{t}_i + 2\mathbf{H}_3(\mathbf{t}_j - \mathbf{t}_k)$, and $\mathbf{t}_i$, $\mathbf{t}_j$ and $\mathbf{t}_k$ are the translation-vector components of the space-group symmetry operators. Let

$$\begin{aligned} \alpha' \cos\beta &= \sum_{(\mathbf{H}_2,\mathbf{H}_3)} A_1 \cos(\varphi_{\mathbf{H}_2} - n\pi), \\ \alpha' \sin\beta &= \sum_{(\mathbf{H}_2,\mathbf{H}_3)} A_1 \sin(\varphi_{\mathbf{H}_2} - n\pi). \end{aligned}$$

We have

$$P(\varphi_{\mathbf{H}_1}) = [2\pi I_0(\alpha')]^{-1} \exp[\alpha' \cos(\varphi_{\mathbf{H}_1} - \beta)], \qquad (9)$$

where

$$\alpha' = \left\{\left[\sum A_1 \cos(\varphi_{\mathbf{H}_2} - n\pi)\right]^2 + \left[\sum A_1 \sin(\varphi_{\mathbf{H}_2} - n\pi)\right]^2\right\}^{1/2}, \qquad (10)$$

$$\tan\beta = \sum A_1 \sin(\varphi_{\mathbf{H}_2} - n\pi) \Big/ \sum A_1 \cos(\varphi_{\mathbf{H}_2} - n\pi). \qquad (11)$$

Thus, $P(\varphi_{\mathbf{H}1})$ has a maximum at $\varphi_{\mathbf{H}_1} = \beta$ when α' is large. This implies that $\varphi_{\mathbf{H}_1}$ can be estimated by (11), which woks for all space groups, provided that appropriate phases $\varphi_{\mathbf{H}_2}$ are available.

For space groups up to the orthorhombic system, n is an integer and, with the assumption that $\varphi_{\mathbf{H}_1}$ and $\varphi_{\mathbf{H}_2}$ have values of 0 or π, the probability distribution of $\varphi_{\mathbf{H}_1}$ obtained from (8) is given by

$$P(\varphi_{\mathbf{H}_1}) = [2\pi I_0(\alpha)]^{-1} \exp(\alpha \cos\varphi_{\mathbf{H}_1}), \qquad (12)$$

where

$$\alpha = \sum_{(\mathbf{H}_2,\mathbf{H}_3)} (-1)^n A_1 \cos\varphi_{\mathbf{H}_2}. \qquad (13)$$

The problem is reduced to the estimation of the sign of $\cos\varphi_{\mathbf{H}_1}$ by use of (12) and (13), that is, $\cos\varphi_{\mathbf{H}_1} = +1$ or -1 according as $\alpha > 0$ or $\alpha < 0$ when $|\alpha|$ is large. Similarly, if $\varphi_{\mathbf{H}_1}$ and $\varphi_{\mathbf{H}_2}$ have values of $\pm\pi/2$, then

$$P(\varphi_{\mathbf{H}_1}) = [2\pi I_0(\alpha)]^{-1} \exp(\alpha \sin\varphi_{\mathbf{H}_1}), \qquad (14)$$

where

Table 1
Comparison between the Σ_1 and Σ_3 estimates for 137 seminvariant reflections of the types $gg0$, $g0g$, $0gg$.

NC is the number of the reflections $\mathbf{H}_1$ having $P_\alpha \geq$ PM, α_n is defined by equation (19), PER (%) is the percentage of reflections whose phases are correctly estimated and MPE (°) is the mean phase error.

		Σ_1				Σ_3			
PM	$\langle\|\Delta\|\rangle$	NC	$\langle\alpha_n\rangle$	PER	MPE	NC	$\langle\alpha_n\rangle$	PER	MPE
0.50	0.793	137	1.00	70.8	52.6	137	1.00	80.3	35.5
0.60	1.053	75	1.76	84.0	28.8	92	1.33	93.5	11.7
0.70	1.137	58	2.19	89.7	18.6	74	1.58	97.3	4.9
0.80	1.226	45	2.66	93.3	12.0	57	1.90	96.5	6.3
0.90	1.306	32	3.38	96.9	5.6	41	2.31	100	0
1.00	1.287	9	7.33	100	0.0	3	4.44	100	0

$$\alpha = \sum_{(\mathbf{H}_2,\mathbf{H}_3)} (-1)^n A_1 \sin\varphi_{\mathbf{H}_2}. \tag{15}$$

The probability distribution of the other invariants ω_i can be obtained with a similar analysis. The difference is only in the expressions of the parameters A_i. For example, for ω_1 and ω_5, we have

$$\begin{aligned} A_1 &= 2\sigma_{4H}\sigma_{2H}^{-2}\Delta_{\mathbf{H}_1R}\Delta_{\mathbf{H}_2R}\Delta^2_{\mathbf{H}_3R} \\ A_5 &= 2\sigma_{4H}\sigma_{2H}^{-2}\Delta_{\mathbf{H}_1R}\Delta_{\mathbf{H}_2R}\Delta^2_{\mathbf{H}_3S}. \end{aligned} \tag{16}$$

With known phases $\varphi_{\mathbf{H}_2}$, we can calculate $\varphi_{\mathbf{H}_1}$ using the probability formula of ω_1 or ω_5 according to whether $R_{\mathbf{H}_3} > S_{\mathbf{H}_3}$ or $S_{\mathbf{H}_3} > R_{\mathbf{H}_3}$, which are the so-called large-modulus invariants (Hu & Liu, 1997).

2.3. Phasing procedure

Equations (13) and (15) require the initial value of $\varphi_{\mathbf{H}_2}$ before the phases of reflections $\mathbf{H}_1$ can be determined. In order to select starting reflections that should be involved in more quartet relationships (5) with higher reliability, a conventional convergence process is used (Germain *et al.*, 1970). First, the expected value of $\alpha_{\mathbf{H}_1}$ for each $\mathbf{H}_1$ reflection is calculated from all quartet relationships by means of

$$(\alpha_{\mathbf{H}_1})_{\rm exp} = \sum_{(\mathbf{H}_2,\mathbf{H}_3)} |A|[I_1(|A|)/I_0(|A|)]. \tag{17}$$

Then, the reflection with the smallest $(\alpha_{\mathbf{H}_1})_{\rm exp}$ is eliminated together with all the quartet relationships to which the reflection contributes. This process is repeated until it has converged and the eliminated reflections are recorded one by one. The reflections at the bottom of the convergence list are the candidates for starting reflections, from which those with large $|\Delta|$, R and S values are favourite. Taking the space group $P2_12_12_1$ as an example, the initial phase values of starting reflections are determined by the following considerations.

(i) The phase values of the seminvariants, *e.g.* $0gg$, $g0g$, $gg0$ (g = even, u = odd) are taken from the estimate results of one-phase seminvariants (denoted as Σ_1, see Appendix *A*).

(ii) The reflections that specify the origin and enantiomorph are chosen, *e.g.* $gu0$, $u0g$, $0gu$, $ug0$, and their phase values can be assigned exactly.

Table 2
Starting reflections used for the Σ_3 estimates for non-seminvariants.

	Type	*hkl*	φ	Δ	*R*	*S*
Origin- and enantiomorph-fixing reflections	$g\,u\,0$	4,13,0	0	1.29	1.12	1.57
	$u\,0\,g$	506	π	1.49	1.54	2.07
	$0\,g\,u$	0,16,1	0	1.86	2.56	3.16
	$u\,g\,0$	340	$-\pi/2$	−2.32	2.01	1.02
Other reflections	$g\,0\,u$	609	$\pi/2$	1.29	1.42	1.93
	$0\,u\,g$	014	$-\pi/2$	2.55	1.48	2.38
	$u\,u\,0$	7,17,0	$-\pi/2$	−1.49	2.87	1.98
	$u\,0\,u$	901	$\pi/2$	1.52	1.28	1.83
	$0\,u\,u$	057	$-\pi/2$	1.85	4.72	5.17

(iii) The phases belonging to the remaining parity types of reflections in three principal zones with special values, such as $g0u$, $0ug$, $uu0$, $u0u$, $0uu$, cannot be uniquely estimated in the case that the initial values of starting phases are unavailable. But it is possible to obtain two sets of probable phase values for each type of reflection by assigning a random value to a starting phase. For example, for $g0u$ type, a starting reflection may be chosen to be arbitrarily given a value $\pi/2$ or $-\pi/2$. Starting from this reflection, the other $g0u$ reflections can be phased so that we have one set of phases for all reflections belonging to this type. The other set of probable phases of the $g0u$ reflections are just those from the estimated values plus π.

Once the initial values of the starting phases are determined, the α value for each $\mathbf{H}_1$ can be calculated by (13) or (15) and thus the sign of $\cos\varphi_{\mathbf{H}_1}$ or $\sin\varphi_{\mathbf{H}1}$ obtained. The associated probability of the sign is calculated by means of

$$P_\alpha \cong \tfrac{1}{2} + \tfrac{1}{2}\tanh(\alpha_n), \tag{18}$$

where

$$\alpha_n = |\alpha|/\langle|\alpha|\rangle_N, \tag{19}$$

and N is the number of $\mathbf{H}_1$ reflections. Then, a new set of starting phases is obtained by selecting the reflections with higher P_α values and used in further cycles of the calculations. An iterative procedure is carried out in such a way until each sign does not change any longer.

3. Test calculations

The test calculations were made using the experimental diffraction data from the protein cytochrome c_{550}, molecular weight ~14500 Da, and its $PtCl_4^{2-}$ derivative at 2.5 Å resolution (Timkovich & Dickerson, 1976). 2807 pairs of independent diffraction data corresponding to the derivative were used. The crystal belongs to space group $P2_12_12_1$. This is a typical space group in which all reflections in the three principal zones have special phase values of 0, π or $\pm\pi/2$. These reflections are fit for phasing by the method described above (denoted as Σ_3).

There are 137 seminvariant reflections of the types $0gg$, $g0g$, $gg0$ in the data set. 10 of the $0gg$ reflections, 5 of the $g0g$ and 12 of the $gg0$ were chosen as starting phases in terms of the convergence results. The initial values of these starting phases

Table 3
Estimate results of the reflections belonging to the types specifying origin and enantiomorph.

See Table 1 for the notation.

	g u 0					*u* 0 *g*				
PM	$\langle\|\Delta\|\rangle$	NC	$\langle\alpha_n\rangle$	PER	MPE	$\langle\|\Delta\|\rangle$	NC	$\langle\alpha_n\rangle$	PER	MPE
0.50	0.48	62	1.00	75.8	43.5	0.64	18	1.00	72.2	50.0
0.60	0.57	46	1.31	80.4	35.2	0.76	13	1.36	69.2	55.4
0.70	0.64	37	1.56	81.1	34.0	0.88	10	1.68	80.0	36.0
0.85	0.67	26	1.97	84.6	27.7	1.03	6	2.31	83.3	30.0
0.95	0.74	19	2.28	89.5	18.9	1.05	4	2.93	100	0

	0 *g u*					*u g* 0				
PM	$\langle\|\Delta\|\rangle$	NC	$\langle\alpha_n\rangle$	PER	MPE	$\langle\|\Delta\|\rangle$	NC	$\langle\alpha_n\rangle$	PER	MPE
0.50	0.73	52	1.00	75.0	45.0	0.68	49	1.00	85.7	25.7
0.60	0.88	34	1.50	82.4	31.8	0.91	31	1.54	96.8	5.8
0.70	1.03	24	2.00	95.8	7.5	1.07	23	1.97	100	0
0.85	1.28	16	2.70	100	0	1.26	17	2.42	100	0
0.95	1.25	13	3.09	100	0	1.46	11	3.11	100	0

were assigned as for those from the Σ_1 results. In the iterative calculations, the reflections having $P_\alpha > 0.90$ were used in the next cycle. The final results are listed in Table 1, together with the results from the Σ_1 formula. It is evident that the Σ_3 estimates lead to better results than those from Σ_1, as shown by the mean phase error (MPE) of 35.5 *versus* 52.6°.

Table 2 illustrates the other types of starting phases, apart from the seminvariants, which were chosen from the bottom of the convergence list with larger $|\Delta|$, *R* and *S* values. The upper four in Table 2 are the reflections that fix the origin and enantiomorph. Their phases were given as the same as those of the known structure for the sake of comparison. The estimated results for 181 reflections belonging to the types of fixing the origin and enantiomorph are shown in Table 3. Although the estimate accuracy varies from type to type with minimum MPE 25.7° for *ug*0 and maximum 50.0° for *u*0*g*, it seems to be quite satisfactory with total MPE 39.8° and correctness 77.9%.

The starting phases of the types *g*0*u*, 0*ug*, *uu*0, *u*0*u* and 0*uu* are also listed in Table 2. Their initial phases were arbitrarily chosen from two possible values $\pm\pi/2$. In the present test, we took the values of the known structure. The iterative calculations gave a set of estimated results for 207 reflections belonging to the five types with total MPE 37.4° and correctness 79.2%. The results are shown in Table 4. At the same time, we obtain the other set of possible phases for each type of reflection, that is, the estimated value plus π.

The statistical results of a total of 525 reflections in the three principal zones are listed in Table 5. It is observed that, while an MPE of 37.7° and correctness of 79% are obtained for all 525 reflections, 91% of the phases are correctly estimated and the MPE is reduced to 16° for 315 reflections having $P_\alpha \geq 0.65$. It is noted that the percentage of the correctly estimated phases is just comparable with the $\langle P_\alpha\rangle$ value. Clearly, the reliability relies mainly on the increase of $|\Delta|$ values, indicating that the contribution from heavy atoms in the derivative plays an important role.

Table 4
Estimate results of the other five types of reflection.

See Table 1 for the notation.

	g 0 *u*					0 *u g*				
PM	$\langle\|\Delta\|\rangle$	NC	$\langle\alpha_n\rangle$	PER	MPE	$\langle\|\Delta\|\rangle$	NC	$\langle\alpha_n\rangle$	PER	MPE
0.50	0.61	24	1.00	83.3	30.0	0.74	42	1.00	78.6	38.6
0.60	0.75	17	1.36	88.2	21.2	0.88	32	1.29	90.6	16.9
0.70	0.83	14	1.59	92.9	12.9	1.00	26	1.50	92.3	13.8
0.80	0.83	13	1.66	92.3	13.8	1.06	20	1.77	90.0	18.0
0.95	0.94	6	2.27	83.3	30.0	1.36	10	2.44	90.0	18.0

	u u 0					*u* 0 *u*				
PM	$\langle\|\Delta\|\rangle$	NC	$\langle\alpha_n\rangle$	PER	MPE	$\langle\|\Delta\|\rangle$	NC	$\langle\alpha_n\rangle$	PER	MPE
0.50	0.47	71	1.00	74.6	45.6	0.90	23	1.00	82.6	31.3
0.60	0.56	47	1.47	80.9	34.5	1.03	17	1.33	88.2	21.2
0.70	0.66	34	1.91	94.1	10.6	1.20	12	1.74	83.3	30.0
0.80	0.71	30	2.09	96.7	6.0	1.24	10	1.97	90.0	18.0
0.95	0.83	17	2.84	100	0	1.45	6	2.52	100	0

	0 *u u*				
PM	$\langle\|\Delta\|\rangle$	NC	$\langle\alpha_n\rangle$	PER	MPE
0.50	0.78	47	1.00	83.0	30.6
0.60	0.94	33	1.39	97.0	5.5
0.70	1.03	25	1.74	100	0
0.80	1.13	19	2.13	100	0
0.95	1.25	15	2.42	100	0

4. Conclusions

We have shown that the 4PSI distributions are used to estimate some individual phases based on the SIR data, in a fashion similar to the Σ_3 formula of Hauptman & Karle (1953). The method, which works in principle for all space groups, has been applied to a known protein structure with experimental diffraction data in space group $P2_12_12_1$ to estimate some phases having special values of 0, π or $\pm\pi/2$. An important step in applying the method is to acquire appropriate starting phases since, at least, one known phase is necessary for determining the other phases that are of the same parity type as the starting phase. The formula proved to be effective even though one starting phase was used for each type of reflection.

The reflections for starting phases may come from three sources: (i) Σ_1-determined phases; (ii) origin- and enantiomorph-fixing reflections; and (iii) other special reflections that may have one or other of a pair of values such as 0, π or $\pm\pi/2$ according to their types. The example described in this paper gives the following conclusions. The Σ_1-determined seminvariant phases can be markedly improved by the Σ_3 formula. The calculations starting from the origin- and enantiomorph-fixing reflections have the result that a large number of other reflections of the same type are phased. This would enhance the effects of the origin- and enantiomorph-fixing reflections in further phase extension and refinement by basic triple relationships. In addition, the two sets of possible phases can be obtained for each type of reflection by assigning arbitrarily an initial phase value to one of this type of reflection. Thus, all

Table 5
Statistical results for all the reflections with special phase values (0, π or $\pm\pi/2$).

See Table 1 for the notation.

PM	$\langle\|\Delta\|\rangle$	NC	$\langle\alpha_n\rangle$	$\langle P_\alpha\rangle$	PER	MPE
0.50	0.681	525	1.00	0.75	79.0	37.7
0.60	0.837	362	1.38	0.85	88.1	21.4
0.65	0.895	315	1.55	0.88	91.1	16.0
0.75	0.985	250	1.85	0.93	94.0	10.8
0.85	1.059	199	2.14	0.96	94.5	9.9
0.95	1.169	134	2.63	0.98	97.0	5.4
1.00	1.660	19	4.75	1.00	100	0

types of reflection yield 2^n possible sets of phases, where n is the number of parity types to which the reflections belong. In the present test, for example, we obtain 2^5 sets of phases for the 207 reflections in the five types ($g0u$, $0ug$, $uu0$, $u0u$, $0uu$) by permutation, whereas these reflections will give rise to an extremely large amount of possible phase sets if each of these reflections is assigned two values in the multisolution process. Therefore, in the phasing procedures, the method described here can be used to drastically reduce the number of possibilities that need to be examined in the time the number of the initial phases increases greatly.

APPENDIX *A*
Estimate formula for one-phase structure seminvariants (Σ_1) in the SIR case

Substituting the invariant $\omega_1 = \varphi_{\mathbf{H}_1} - \varphi_{\mathbf{H}_2\mathbf{R}_i} + \varphi_{\mathbf{H}_2\mathbf{R}_j}$ into the 3PSI distribution of Hauptman (1982) in the SIR case, we obtain the probability distribution for a fixed $\varphi_{\mathbf{H}_1}$,

$$P(\varphi_{\mathbf{H}_1}) = [2\pi I_0(\alpha)]^{-1} \exp(\alpha \cos \varphi_{\mathbf{H}_1}), \tag{20}$$

where reciprocal-lattice vectors $\mathbf{H}_1$ and $\mathbf{H}_2$ satisfy $\mathbf{H}_1 - \mathbf{H}_2\mathbf{R}_i + \mathbf{H}_2\mathbf{R}_j = 0$,

$$\alpha = \sum_{\mathbf{H}_2} (-1)^n A_1, \tag{21}$$

$$A_1 = 2\sigma_{3P}\sigma_{2P}^{-1.5} R_{\mathbf{H}_1} R_{\mathbf{H}_2}^2 + 2\sigma_{3H}\sigma_{2H}^{-1.5} \Delta_{\mathbf{H}_1 R} \Delta_{\mathbf{H}_2 R}^2, \tag{22}$$

$$n = 2\mathbf{H}_2(\mathbf{t}_j - \mathbf{t}_i). \tag{23}$$

For $\omega_4 = \varphi_{\mathbf{H}_1} - \psi_{\mathbf{H}_2\mathbf{R}_i} + \psi_{\mathbf{H}_2\mathbf{R}_j}$, we have a similar result with

$$\alpha = \sum_{\mathbf{H}_2} (-1)^n A_4, \tag{24}$$

$$A_4 = 2\sigma_{3P}\sigma_{2P}^{-1.5} R_{\mathbf{H}_1} R_{\mathbf{H}_2}^2 C_{\mathbf{H}_2}^2 + 2\sigma_{3H}\sigma_{2H}^{-1.5} \Delta_{\mathbf{H}_1 R} \Delta_{\mathbf{H}_2 S}^2. \tag{25}$$

$\varphi_{\mathbf{H}_1}$ can be estimated by equations (21) or (24) according to whether $R_{\mathbf{H}_2} > S_{\mathbf{H}_2}$ or $S_{\mathbf{H}_2} > R_{\mathbf{H}_2}$ when $|\alpha|$ is large (refer to §2 for notation).

This work was supported by the National Natural Science Foundation of China (No. 29773045).

References

Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Spagna, R. (1999). *Acta Cryst.* A**55**, 991–999.

Burla, M. C., Giacovazzo, C. & Polidori, G. (1989). *Acta Cryst.* A**45**, 99–104.

Cascarano, G., Giacovazzo, C. & Viterbo, D. (1987). *Acta Cryst.* A**43**, 22–29.

De Titta, G. T., Edmonds, J. W., Langs, D. A. & Hauptman, H. (1975). *Acta Cryst.* A**31**, 472–479.

Germain, G., Main, P. & Woolfson, M. M. (1970). *Acta Cryst.* B**26**, 274–285.

Giacovazzo, C. (1979). *Acta Cryst.* A**35**, 296–305.

Giacovazzo, C., Cascarano, G. & Zheng, C.D. (1988). *Acta Cryst.* A**44**, 45–51.

Giacovazzo, C. & Siliqi, D. (1996*a*). *Acta Cryst.* A**52**, 133–142.

Giacovazzo, C. & Siliqi, D. (1996*b*). *Acta Cryst.* A**52**, 143–151.

Giacovazzo, C., Spagna, R., Vickovic, I. & Viterbo, D. (1979). *Acta Cryst.* A**35**, 401–412.

Green, E. A. & Hauptman, H. (1978*a*). *Acta Cryst.* A**34**, 216–223.

Green, E. A. & Hauptman, H. (1978*b*). *Acta Cryst.* A**34**, 230–241.

Hauptman, H. (1982). *Acta Cryst.* A**38**, 289–294.

Hauptman, H. & Green, E. A. (1978). *Acta Cryst.* A**34**, 224–229.

Hauptman, H. & Karle, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal*. *ACA Monograph* No. 3. Pittsburgh: Polycrystal Book Service.

Hu, N.-H. & Liu, Y.-S. (1995). *Acta Cryst.* A**51**, 520–524.

Hu, N.-H. & Liu, Y.-S. (1997). *Acta Cryst.* A**53**, 161–167.

Kyriakidis, C., Peschar, R. & Schenk, H. (1996). *Acta Cryst.* A**52**, 77–87.

Liu, Y.-S., Hu, N.-H. & Wang, S.-Y. (1999). *Acta Cryst.* A**55**, 396–398.

Schenk, H. (1973). *Acta Cryst.* A**29**, 77–82.

Sheldrick, G. M. (1993). *Crystallographic Computing 6. A Window on Modern Crystallography*, edited by H. D. Flack, L. Parkanyi & K. Simon, pp. 100–110. IUCr/Oxford University Press.

Timkovich, R. & Dickerson, R. E. (1976). *J. Biol. Chem.* **251**, 4033–4046.