(Sciences Chimiques). The data set of $Ni_{3-x}Te_2$ was kindly put at our disposal by Dr J. L. de Boer (University of Groningen, The Netherlands).

References

EVAIN, M., VAN DER LEE, A., MONCONDUIT, L. & PETŘÍČEK, V. (1994). *Chem. Mater.* **6**, 1776–1783.

JANNER, A. & JANSSEN, T. (1977). *Phys. Rev. B*, **15**, 643–658.

LEE, A. VAN DER, EVAIN, M., MONCONDUIT, L., BREC, R., ROUXEL, J. & PETŘÍČEK, V. (1994). *Acta Cryst.* B**50**, 119–128.

LEE, A. VAN DER, VAN SMAALEN, S., WIEGERS, G. A. & DE BOER, J. L. (1991). *Phys. Rev. B*, **43**, 9420–9430.

PÉREZ-MATO, J. M., MADARIAGA, G. & TELLO, M. (1986). *J. Phys. C*, **19**, 2613–2622.

PETŘÍČEK, V. (1994). *JANA94. Programs for Modulated and Composite Crystals.* Institute of Physics, Praha, Czech Republic.

PETŘÍČEK, V. & COPPENS, P. (1988). *Acta Cryst.* A**44**, 235–239.

PETŘÍČEK, V., GAO, Y., LEE, P. & COPPENS, P. (1990). *Phys. Rev. B*, **42**, 387–392.

SCHUTTE, W. J. & DE BOER, J. L. (1993). *Acta Cryst.* B**49**, 398–403.

WOLFF, P. M. DE (1974). *Acta Cryst.* A**30**, 777–785.

WOLFF, P. M. DE (1977). *Acta Cryst.* A**33**, 493–497.

WOLFF, P. M. DE, JANSSEN, T. & JANNER, A. (1981). *Acta Cryst.* A**37**, 625–636.

YAMAMOTO, A. (1981). *Acta Cryst.* A**37**, 838–842.

Acta Cryst. (1995). A**51**, 535–542

TDSIR Phasing: Direct Use of Phase-Invariant Distributions in Macromolecular Crystallography

BY DAVID A. LANGS, DONGYAO GUO AND HERBERT A. HAUPTMAN

Medical Foundation of Buffalo, 73 High Street, Buffalo, NY 14203, USA

(*Received* 15 *July* 1994; *accepted* 17 *January* 1995)

Abstract

A new strategy for employing three phase triples invariant estimates from Hauptman's single isomorphous replacement (SIR) and anomalous dispersion (SAS) joint probability distribution formulae is outlined which produces a single unique phase-invariant solution in the case where the positions of the heavy-atom scatterers is known. A similar but non-identical result is obtained for the phase invariants of a structure for which a molecular-replacement solution has been obtained. It is important to note that the values of the individual native/derivative phases can be determined directly from the probability distribution formulae without having to utilize the phase-invariant estimates in an active way. Elimination of the multisolution aspect of utilizing phase-invariant estimates should have important repercussions with regard to phasing macromolecular sets of derivatized data. Trial calculations based on experimentally measured 2.5 Å data for three derivatives of cytochrome c_{550} are encouraging. The average of the three SIR maps resolves a number of structural ambiguities seen in the published multiple isomorphous replacement (MIR) map obtained from eight derivatives.

Introduction

Probabilistic formulae to estimate the three phase triples invariants of macromolecular structures using SIR (Hauptman, 1982*a*) and SAS (Hauptman, 1982*b*) data were first derived by Hauptman more than ten years ago. Whereas a unique solution exists for the phase invariant within the interval 0 to 2π from the conditional probability distribution formula in the SAS case, only the sign of the cosine of the phase invariant, but not its modulus, was obtained from the analogous SIR distribution formula if the positions of the derivative heavy atoms were unknown. A more specific formula appropriate to the two-derivative MIR situation was developed (Fortier, Weeks & Hauptman, 1984) that improved the accuracy of the estimates. It was noted that, if the reciprocal-lattice vectors, rather than the atomic coordinates, were considered the primitive random variables in the derivation of the SIR distribution function, a slightly different result could be obtained (Giacovazzo, Cascarano & Zheng, 1988), but still only the mode of the phase invariant, 0 or π, could be determined without knowledge of the heavy-atom scatterers. A prior study had meanwhile indicated that both the sign and modulus of the cosine invariant within the full range of +1 to −1 was obtainable if the positions of the heavy-atom scatterers were known for the SIR case (Fortier, Moore & Fraser, 1985); improved phase-invariant estimates for the SAS case were also reported (Fortier, Fraser & Moore, 1986).

In this paper, we have reinvestigated Hauptman's original SIR work to discover that both the sine and cosine of the phase invariant can be obtained from the joint probability distribution function in a rather obvious way if the positions of the derivative heavy atoms are known. Moreover, unique values for the native and derivative phases can be directly obtained from the distribution function without resorting to a multisolution approach employing the phase-invariant estimates in an active way.

Background

The major result of Hauptman's (1982*a*) SIR paper was the derivation of the joint probability distribution of six magnitudes and phases corresponding to the normalized structure factors $E_{\mathbf{h}}$, $E_{\mathbf{k}}$ and $E_{\mathbf{l}}$ of the native and $F_{\mathbf{h}}$, $F_{\mathbf{k}}$ and $F_{\mathbf{l}}$ of the derivative structure, where $\mathbf{h}+\mathbf{k}+\mathbf{l}=0$.

$$\begin{aligned}
P = {} & [1/\pi^6(1-\alpha^2)^3]R_1R_2R_3S_1S_2S_3 \\
& \times \exp\{-(1-\alpha^2)^{-1}(R_1^2+R_2^2+R_3^2+S_1^2+S_2^2+S_3^2) \\
& + 2\beta[R_1S_1\cos(\Phi_1-\Psi_1)+R_2S_2\cos(\Phi_2-\Psi_2) \\
& + R_3S_3\cos(\Phi_3-\Psi_3)] \\
& + 2\beta_1[R_1R_2S_3\cos(\Phi_1+\Phi_2+\Psi_3) \\
& + R_1S_2R_3\cos(\Phi_1+\Psi_2+\Phi_3) \\
& + S_1R_2R_3\cos(\Psi_1+\Phi_2+\Phi_3)] \\
& + 2\beta_2[R_1S_2S_3\cos(\Phi_1+\Psi_2+\Psi_3) \\
& + S_1R_2S_3\cos(\Psi_1+\Phi_2+\Psi_3) \\
& + S_1S_2R_3\cos(\Psi_1+\Psi_2+\Phi_3)] \\
& + 2\beta_0R_1R_2R_3\cos(\Phi_1+\Phi_2+\Phi_3) \\
& + 2\beta_3S_1S_2S_3\cos(\Psi_1+\Psi_2+\Psi_3)\}. \qquad (1)
\end{aligned}$$

Here, R_j and S_j are the magnitudes and Φ_j and Ψ_j the phases corresponding to the normalized structure factors E_j and F_j of the native and derivative structures. The values of the constants α and β_j are defined in terms of the mixed moments α_{mn} of $E_{\mathbf{h}}$ and $F_{\mathbf{h}}$ as indicated in the original paper.

Normally only the magnitudes of $E_{\mathbf{h}}$ and $F_{\mathbf{h}}$ are known. In the case where the positions of the heavy-atom scatterers are known, the magnitude $G_{\mathbf{h}}$ and phase $\xi_{\mathbf{h}}$ of the heavy-atom contribution to $F_{\mathbf{h}}$ can be computed. Assuming that the triad of magnitudes $E_{\mathbf{h}}$, $F_{\mathbf{h}}$ and $G_{\mathbf{h}}$ are sufficiently precise to allow the phasing triangle to be closed, the cosine values of the phase doublets $\varphi_{\mathbf{h}}-\psi_{\mathbf{h}}$, $\varphi_{\mathbf{h}}-\xi_{\mathbf{h}}$ and $\psi_{\mathbf{h}}-\xi_{\mathbf{h}}$ can be determined. Values of these cosines may also be estimated and refined through the difference Patterson without explicitly solving the derivative substructure (Kyriakidis, Peschar & Schenk, 1993). These cosines give us the magnitude, but not the signs, of the phase doublet angles $|\varphi_{\mathbf{h}}-\psi_{\mathbf{h}}|$, $|\varphi_{\mathbf{h}}-\xi_{\mathbf{h}}|$ and $|\psi_{\mathbf{h}}-\xi_{\mathbf{h}}|$. Now, if we chose to interpret (1) as relating the derivative ($R_j=F_{\mathbf{h}}$) and heavy-atom substructures ($S_j=G_{\mathbf{h}}$), so that values of ψ_j correspond to the known phases $\xi_{\mathbf{h}}$, the equation can be rewritten as

$$\begin{aligned}
P = {} & [1/\pi^6(1-\alpha^2)^3]R_1R_2R_3S_1S_2S_3 \\
& \times \exp\{-(1-\alpha^2)^{-1}(R_1^2+R_2^2+R_3^2+S_1^2+S_2^2+S_3^2) \\
& + 2\beta[R_1S_1\cos(\Delta_1)+R_2S_2\cos(\Delta_2)+R_3S_3\cos(\Delta_3)] \\
& + 2\beta_1[R_1R_2S_3\cos(\Phi\pm\Delta_3)+R_1S_2R_3\cos(\Phi\pm\Delta_2) \\
& + S_1R_2R_3\cos(\Phi\pm\Delta_1)] \\
& + 2\beta_2[R_1S_2S_3\cos(\Phi\pm\Delta_2\pm\Delta_3) \\
& + S_1R_2S_3\cos(\Phi\pm\Delta_1\pm\Delta_3) \\
& + S_1S_2R_3\cos(\Phi\pm\Delta_1\pm\Delta_2)] + 2\beta_0R_1R_2R_3\cos(\Phi) \\
& + 2\beta_3S_1S_2S_3\cos(\Psi_1+\Psi_2+\Psi_3)\}, \qquad (2)
\end{aligned}$$

where $\Delta_j=|\Phi_j-\Psi_j|$ are the known doublet magnitudes, Ψ_j are the known heavy-atom phases and $\Phi=\Phi_1+\Phi_2+\Phi_3$ is the invariant of interest to be determined. The conditional probability formula was obtained from (2) by permuting over the eight combinations of signs of the Δ_j values and was shown to yield four enantiomeric pairs of solutions of the form $\sum_4 A_j\cos(\Phi\pm\xi_j)$, which permitted a unique value of the cosine to be determined within the full interval from +1 to −1 (Fortier, Moore & Fraser, 1985), but not its related sine component. Similar formulae exist for SAS data except that all eight terms $A_j\cos(\Phi+\xi_j)$ are, in general, different (Fortier, Fraser & Moore, 1986).

This inability to determine the sine component in the SIR case is rather puzzling in view of the fact that an algebraic analysis based on correctly oriented, but misplaced, structural fragments does provide a unique and accurate estimate for the triples phase invariants of the partial structure within the interval of 0 to π rad, even though the absolute position of the fragment within the unit cell may be unknown (Langs, 1985). Similar algebraic formulae that uniquely determine the enantiomorph of the phase doublets $\Delta_{\mathbf{h}}$ (and hence the phase invariants) have been developed for the SIR, SAS and partial structural applications when the absolute positions of the heavy-atom scatterers or fragments are known (Fan, Han, Qian & Yao, 1984; Fan & Gu, 1985; Langs, 1986). In the latter reference, it was shown that extremely accurate sign determinations of the phase doublets could still be obtained, even if the cosine invariants were known only to their closest node, *i.e.* ±1. Trial calculations based on error-free SIR data for the rhombohedral *R*3 structure of pig insulin at 1.9 Å resolution showed that it was possible to determine the native phases to a precision of less than 6° with a single S replacement atom among the 1083 independent non-H atoms of the structure.

Given the preceding observations, and supposing that one has a pair of derivative data sets for which the known positions of the heavy-atom scatterers or partial structure are sufficient to select the enantiomorph for the space group in question, how can this information be used to obtain a unique estimate for the phase invariant under the constraint that the magnitudes of the phase doublets are known?

The TDSIR formula

Here, we show for the SIR example that the expected values of both the sine and cosine of the phase invariant

$\Phi = \Phi_1 + \Phi_2 + \Phi_3 = \Psi_1 + \Psi_2 + \Psi_3 + \Delta_3 + \Delta_2 + \Delta_3$ can be determined from (2) as

$$\sin \Phi' = \left(\sum_{j=1}^{8} \sin \Phi_j P_j \right) \Big/ \sum_{j=1}^{8} P_j \qquad (3a)$$

$$\cos \Phi' = \left(\sum_{j=1}^{8} \cos \Phi_j P_j \right) \Big/ \sum_{j=1}^{8} P_j \qquad (3b)$$

by permuting over the eight combinations of signs allowed for the three Δ_j magnitudes. A unique value for the phase invariant may be obtained from

$$\Phi = \tan^{-1}(\sin \Phi' / \cos \Phi') \qquad (3c)$$

since it is not ensured that $\sin^2 \Phi' + \cos^2 \Phi'$ will always equal 1.0. A weight appropriate to the accuracy of this determination is given as

$$\begin{aligned} wt(\mathbf{h}, \mathbf{k}) &= \varepsilon[\cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} - \Phi)] \\ &= [\cos^2 \Phi' + \sin^2 \Phi']^{1/2}. \end{aligned} \qquad (4)$$

The weight (4) has values between 0.0 and 1.0 and cannot exceed 1.0. It will approach 1.0 when all eight estimates are in good agreement, as would be the case for zonal restricted phases, since all three Δ values must equal either 0 or 180°. A low weight would result if all eight estimates were in poor agreement with equal probabilistic reliability P_j.

Since it is known how the individual phases of the native/derivative are related to the individual heavy-atom phases $\Psi_{\mathbf{h}}$ within the summations (3*a*), (3*b*), *i.e.* $\Phi'_{\mathbf{h}} = \Psi_{\mathbf{h}} \pm \Delta_{\mathbf{h}}$, these phases may be obtained directly from the integration of the distribution as

$$\varphi_{\mathbf{h}} = \left[\sum_{\mathbf{h},\mathbf{k}} w(\mathbf{h}, \mathbf{k}) \sum_{j=1}^{8} (s_{\mathbf{h}} \Phi'_{j\mathbf{h}} + t_{\mathbf{h}}) P_j \right] \times \left[\sum_{\mathbf{h},\mathbf{k}} w(\mathbf{h}, \mathbf{k}) \sum_{j=1}^{8} P_j \right]^{-1}, \qquad (5)$$

where the outer summation is over the triples containing $\varphi_{\mathbf{h}}$ and the terms $s_{\mathbf{h}}$ and $t_{\mathbf{h}}$ in the inner summation are the sign and translational phase shifts applied to $\Psi_{\mathbf{h}}$ to relate it back to its standard form. Since $\varphi_{\mathbf{h}}$ is generally complex, (5) must be evaluated over its real and imaginary components to obtain the native phases. We shall refer to (5) as the TDSIR formula.

Similar formulae for (3*a*), (3*b*) exist for the molecular-replacement SIR (MRSIR) application except that, rather than permuting over the eight combinations of signs for the known phase doublet magnitudes, the integrations on Φ_1, Φ_2 and Φ_3 must be iterated over the full range of allowed values from 0 to π rad:

$$\sin \Phi = \int \sin \Phi_j \, wt_1 wt_2 wt_3 P_j \, \mathrm{d}\Phi_1 \, \mathrm{d}\Phi_2 \, \mathrm{d}\Phi_2 \qquad (6a)$$

$$\cos \Phi = \int \cos \Phi_j \, wt_1 wt_2 wt_3 P_j \, \mathrm{d}\Phi_1 \, \mathrm{d}\Phi_2 \, \mathrm{d}\Phi_2, \qquad (6b)$$

where the weights wt_j represent the expectation probability of the residual structure amplitude, $R_{\mathbf{h}} = E_{\mathbf{h}} - G_{\mathbf{h}}$ (fragment), which is not an observable magnitude from the molecular-replacement solution. First, it must be determined for each angle Φ_j if the phasing triangle can be closed, or whether $R_{\mathbf{h}}$ exceeds the scattering power of the remaining atoms of the structure. Second, if the triangle can be closed, $wt_{\mathbf{h}}$ represents the probability that $R_{\mathbf{h}}$ equals the computed value, *i.e.* $R_{\mathbf{h}} \exp(-R_{\mathbf{h}}^2)$ if the amplitude is non-centrosymmetric or $(2/\pi)^{1/2} \times \exp(-R_{\mathbf{h}}^2/2)$ if the amplitude is centrosymmetric, in which case Φ is integrated over its two permissible restricted values.

It follows that phase-invariant estimates analogous to formulae (3*a*), (3*b*) for any multiple of derivatized data sets may be expressed as

$$\cos \Phi = \left(\sum_{j1}^{8} \sum_{j2}^{8} \ldots \sum_{j'}^{8} \cos \Phi_j P_{j1} P_{j2} \ldots P_{j'} \right) \times \left(\sum_{j1}^{8} \sum_{j2}^{8} \ldots \sum_{j'}^{8} P_{j1} P_{j2} \ldots P_{j'} \right)^{-1} \qquad (7a)$$

$$\sin \Phi = \left(\sum_{j1}^{8} \sum_{j2}^{8} \ldots \sum_{j'}^{8} \sin \Phi_j P_{j1} P_{j2} \ldots P_{j'} \right) \times \left(\sum_{j1}^{8} \sum_{j2}^{8} \ldots \sum_{j'}^{8} P_{j1} P_{j2} \ldots P_{j'} \right)^{-1}, \qquad (7b)$$

provided that the pairs of heavy-atom substructures can be related to the same cell origin and are of the same relative hand. Often this may be established by Fourier methods which phase the difference amplitudes of one derivative with estimates of the native phases obtained using a second derivative (Stryer, Kendrew & Watson, 1964; Dickerson, Kopka, Varnum & Weinzierl, 1967). A more drastic method to resolve this problem could involve measuring data from crystals into which both heavy-atom labels have been soaked and inspecting the difference Patterson for the cross vectors between these sites. The computational accuracy of (7*a*), (7*b*) is expected to improve if the appropriate joint probability distributions for paired structures, *e.g.* two isomorphous derivatives (Fortier, Weeks & Hauptman, 1984) or a SAS/SIR combination (Peschar & Schenk, 1991), are employed.

Applications

The initial tests concerning the accuracy of the TDSIR formula were performed using ideal 2.5 Å data computed from the deposited coordinates of cytochrome c_{550} (Timkovich & Dickerson, 1976). The heme Fe atom was chosen to be the known heavy-atom position ($x = 0.2989$, $y = 0.4157$, $z = 0.2884$), E magnitudes were computed only over those reciprocal-lattice vectors

Table 1. *Representative list of five triples computed from the error-free* 2.5 Å *cytochrome* c_{550} *SIR data*

(*a*) List of the native (R) and heavy-atom (S_i) amplitudes and phases ($\varphi H/Ni$) and doublet phase values (Δ_i) in rad. Other data pertinent to the triple are its direct-methods A value, commonly defined as $2|E_{\mathbf{h}}E_{-\mathbf{k}}E_{\mathbf{k}-\mathbf{h}}|/(N)^{1/2}$, and its three phases, here identified by their serial numbers ($N1$, $N2$, $N3$) from a list sorted on decreasing magnitude of $|E|$. The phase invariant is defined as $\Phi = \varphi_{N1} + \varphi_{N2} + \varphi_{N3} + T$, where $\varphi_{-Ni} = -\varphi_{Ni}$ and T is a translational phase shift employed to relate the symmetry variant phases to their standard parent form.

A	$N1$	$N2$	$N3$	T	R/S_1	R/S_2	R/S_3	$\varphi H/N1$	$\varphi H/N2$	$\varphi H/N3$	Δ_1	Δ_2	Δ_3
1.53	10	−41	94	π	3.67	2.55	2.26	−1.70	1.43	2.28	0.49	0.17	3.09
					0.94	0.40	0.70	−2.19	−1.61	−0.92			
1.26	10	−65	196	0	3.67	2.37	1.76	−1.70	1.92	−0.82	0.49	1.88	0.40
					0.94	1.15	0.64	−2.19	2.48	−0.43			
1.16	10	−90	−239	π	3.67	1.96	1.93	−1.70	1.58	1.71	0.49	0.22	2.40
					0.94	1.51	1.36	−2.19	−1.36	0.69			
0.95	10	−207	324	0	3.67	1.99	1.68	−1.70	0.98	2.32	0.49	2.96	1.89
					0.94	0.94	0.95	−2.19	2.34	−2.07			
0.91	10	−221	365	π	3.67	2.02	1.52	−1.70	−3.03	0.39	0.49	3.11	0.05
					0.94	1.33	0.44	−2.19	−0.07	0.34			

(*b*) Values of the probability function and phases for each of the eight permitted sign permutations of the doublet phase angles. EXP is the calculated magnitude of the exponent of exp(EXP) in equation (2), P(rel) is rescaled so the permutation with the highest probability is unity, an asterisk (*) denotes the sign combination corresponding to the true phase values. The last line for each of the five triples entries gives the weighted average of the native phase invariant Φ and individual native phases Φ_i in rad. The actual computed TDSIR phases would represent a weighted average of phase indications over a large basis set of triples. In the worst-case senario, P(rel) would be the same for each of the eight sign permutations and the calculated phases would equate to the traditional SIR weighted average proposed by Kartha (1961).

$\pm\Delta_1$	$\pm\Delta_2$	$\pm\Delta_3$	P(rel)	EXP	Φ	Φ_1	Φ_2	Φ_3
−1	−1	−1	0.863	−26.16	1.40	−2.19	−1.26	−0.80
−1	−1	1	1.000	−26.02	1.29	−2.19	−1.26	−0.92
−1	1	−1	0.544	−26.62	1.75	−2.19	−1.61	−0.80
−1	1	1	0.632	−26.48	1.64	−2.19	−1.61	−0.92*
1	−1	−1	0.263	−27.35	2.39	−1.21	−1.26	−0.80
1	−1	1	0.293	−27.24	2.27	−1.21	−1.26	−0.92
1	1	−1	0.204	−27.60	2.74	−1.21	−1.61	−0.80
1	1	1	0.218	−27.54	2.62	−1.21	−1.61	−0.92
	Calc.				1.70	−1.97	−1.40	−0.86
−1	−1	−1	0.161	−24.59	2.91	−2.19	−0.04	−1.22
−1	−1	1	0.182	−24.46	−2.58	−2.19	−0.04	−0.43
−1	1	−1	1.000	−22.76	0.38	−2.19	2.49	−1.22
−1	1	1	0.593	−23.28	1.18	−2.19	2.49	−0.43*
1	−1	−1	0.203	−24.35	−2.39	−1.21	−0.04	−1.22
1	−1	1	0.401	−23.67	−1.59	−1.21	−0.04	−0.43
1	1	−1	0.499	−23.45	1.37	−1.21	2.49	−1.22
1	1	1	0.240	−24.18	2.16	−1.21	2.49	−0.43
	Calc.				0.99	−1.80	2.15	−0.88
−1	−1	−1	0.409	−25.98	1.61	−2.19	−1.36	0.69*
−1	−1	1	1.000	−25.08	0.14	−2.19	−1.36	2.17
−1	1	−1	0.281	−26.35	2.06	−2.19	−1.81	0.69
−1	1	1	0.872	−25.22	0.59	−2.19	−1.81	2.17
1	−1	−1	0.202	−26.68	2.60	−1.21	−1.36	0.69
1	−1	1	0.618	−25.57	1.12	−1.21	−1.36	2.17
1	1	−1	0.179	−26.81	3.05	−1.21	−1.81	0.69
1	1	1	0.424	−25.94	1.57	−1.21	−1.81	2.17
	Calc.				1.01	−1.86	−1.55	1.83
−1	−1	−1	0.247	−23.57	2.53	−2.19	1.99	0.43
−1	−1	1	1.000	−22.17	0.03	−2.19	1.99	−2.07
−1	1	−1	0.299	−23.38	2.18	−2.19	2.34	0.43
−1	1	1	0.959	−22.22	−0.33	−2.19	2.34	−2.07*
1	−1	−1	0.227	−23.66	−2.77	−1.21	1.99	0.43
1	−1	1	0.698	−22.53	1.01	−1.21	1.99	−2.07
1	1	−1	0.215	−23.71	−3.12	−1.21	2.34	0.43
1	1	1	0.853	−22.33	0.65	−1.21	2.34	−2.07
	Calc.				0.48	−1.76	2.17	−1.86
−1	−1	−1	0.951	−22.20	1.43	−2.19	−0.15	0.34
−1	−1	1	0.886	−22.27	1.53	−2.19	−0.15	0.44
−1	1	−1	1.000	−22.15	1.36	−2.19	−0.07	0.34*
−1	1	1	0.932	−22.22	1.46	−2.19	−0.07	0.44
1	−1	−1	0.508	−22.83	2.41	−1.21	−0.15	0.34
1	−1	1	0.486	−22.87	2.52	−1.21	−0.15	0.44
1	1	−1	0.527	−22.79	2.34	−1.21	−0.07	0.34
1	1	1	0.502	−22.84	2.44	−1.21	−0.07	0.44
	Calc.				1.77	−1.86	−0.11	0.39

Table 2. *Summary of heavy-atom replacement sites reported for three of the eight deposited cytochrome* c_{550} *derivatives* (*Timkovich & Dickerson*, 1976), *which diffract to* 2.5 Å *resolution and which were used in the SIR triples phasing employing equation* (5)

The atomic scattering of the refined heavy-atom sites is defined as $Z(e^-)\exp[-(\beta_{11}h^2+\beta_{22}k^2+\beta_{33}l^2+\beta_{12}hk+\beta_{23}kl+\beta_{13}hl)]$.

Derivative	x	y	z	$Z(e^-)$	β_{11}	β_{22}	β_{33}	β_{12}	β_{23}	β_{13}
$PtCl_4^{2-}$	0.1110	0.0524	0.2097	84.2	3.02	1.05	2.55	0.04	1.34	2.32
$Pt(CN)_4^{2-}$	0.0979	0.4307	0.7934	86.8	5.87	0.97	5.09	−0.04	−1.35	4.65
	0.4690	0.4690	0.8628	53.9	5.35	1.01	1.82	0.10	−1.17	−1.67
	0.3074	0.2834	0.7649	39.3	3.71	0.94	1.46	0.66	0.99	1.34
	0.3999	0.0976	0.7727	17.4	4.20	1.13	7.67	0.0	0.0	0.0
UO_2^{2+}	0.0995	0.3012	0.7396	80.3	3.22	0.69	7.93	−0.80	0.28	5.50
	0.0169	0.3368	0.7014	7.1	3.42	0.92	6.24	0.0	0.0	0.0

that corresponded to measureable data within the 2.5 Å sphere as reported in the Brookhaven Database entry for this structure. Only 2994 of the 4160 data within 2.5 Å resolution were recorded, the majority of the missing data represented weak measurements that were discarded. A data file was prepared that contained the derivative, native and heavy-atom amplitudes and the phase of the heavy atom. 13 999 triples were generated for the largest 1500 native amplitudes ($E \geq 0.83$) that had A values $[2|E_{\mathbf{h}}E_{-\mathbf{k}}E_{\mathbf{k}-\mathbf{h}}|/N^{1/2}]$ greater than 0.25. Only 10 s of CPU time were required on a Silicon Graphics Iris workstation to read this triples file and compute the native phases by means of (5). The weighted (4) mean phase error between these computed values and the error-free native phases was 23.0° (r.m.s. 33.0°) as compared with 36.6° (r.m.s. 46.8°) if the native map was computed by using both SIR phase estimates (Kartha, 1961). Table 1 presents details of the calculation for each of the eight sign combinations for five typical triples relationships.

The second test involved using the actual experimentally measured native and derivative X-ray diffraction data that have also been deposited for the cytochrome c_{550} structure (Brookhaven Database file code rcyc1551). TDSIR phasing computations were performed for three of the eight derivatives that diffract to 2.5 Å resolution (Table 2), the remaining derivatives were reported to be isomorphous to only 4.0 Å resolution. The three derivatives, $PtCl_4^{2-}$, $Pt(CN)_4^{2-}$ and UO_2^{2+}, each have 2807, 2399 and 2886 reflections, respectively, in common with the native measured data.

The deposited F amplitudes were reported to be on a common absolute scale, such that the heavy-atom contribution could be calculated directly from the data in Table 2 and assumed to be on the same scale for the purpose of constructing the SIR phasing triangles and obtaining the cosines of the doublet phases. Normalized E values for the native and derivative data sets were scaled using an anisothermal scattering-factor model (Levy, Thiessen & Brown, 1970; Blessing & Langs, 1988). The largest 1500 E values were selected to generate approximately 15 000 triples for each of the three SIR data sets and the native phases were determined by (5). The mean phase error between the three computed native TDSIR and published MIR native phases were noticeably larger than had been noted for the error-free 2.5 Å data, *i.e.* 40.5, 46.9 and 42.6° (r.m.s. 59.6, 67.5 and 62.2°), respectively. The true phase errors are probably smaller than the values cited in view of the fact that this structure was not refined against the native data to provide phases that are more accurate than the initial MIR estimates.

The $PtCl_4^{2-}$, $Pt(CN)_4^{2-}$ and UO_2^{2+} TDSIR maps were individually inspected and compared with the published MIR map. Phases corresponding to the weighted mean of the three TDSIR refinements were then used to synthesize an averaged map to compare with the MIR map. The mean phase error between the averaged TDSIR phases and the MIR phases was 25.8°.

Discussion of results

The native map resultant from the TDSIR phase refinement of the 2.5 Å error-free SIR data was inspected on a Silicon Graphics Iris workstation using the graphics program *CHAIN* (Sack, 1988). Almost the entire structure, except for the occasional terminal sidechain atom, was seen to lie within density contoured at the 1.0σ level. Another notable feature was that no significant density occurred at the center of the heme ring where the Fe atom was arbitrarily introduced as the heavy-atom for the derivative structure (Fig. 1*a*). This indicates that the native TDSIR phases were not biased by the initial heavy-atom phases as was reported in earlier attempts to use Hauptman's SIR triples invariant estimates to phase macromolecular data (Furey, Chandrasekhar, Dyda & Sax, 1990). The published MIR map does not appear to be as clean as the error-free TDSIR map, even though it was computed with 2994 reflections as compared to only 1500 for the error-free model.

The electron densities observed for the three individual TDSIR data sets each have their good and limited features. The $PtCl_4^{2-}$ derivative, because of the unfortunate position of the Pt atom ($x \cong 0$, $y \cong 0$), will not have any strong heavy-atom scattering for the ($h+k=2n$, $k+l=2n+1$) and ($h+k=2n+1$, $k+l=2n$) parity groups and will have heavy-atom phases that are polarized towards real ($h+k=2n$, $k+l=2n$) and imaginary ($h+k=2n+1$, $k+l=2n+1$) phase values for the remaining two parity groups. Whereas the

α-helical segment from residues 108 to 118 and the region around Pro36 are very well defined and almost as clear and interpretable as that seen in the MIR map, the density defining the heme group is poor in comparison owing to severe diffraction ripples produced by the Fe atom, which dominates this region of the map (Fig. 1*b*).

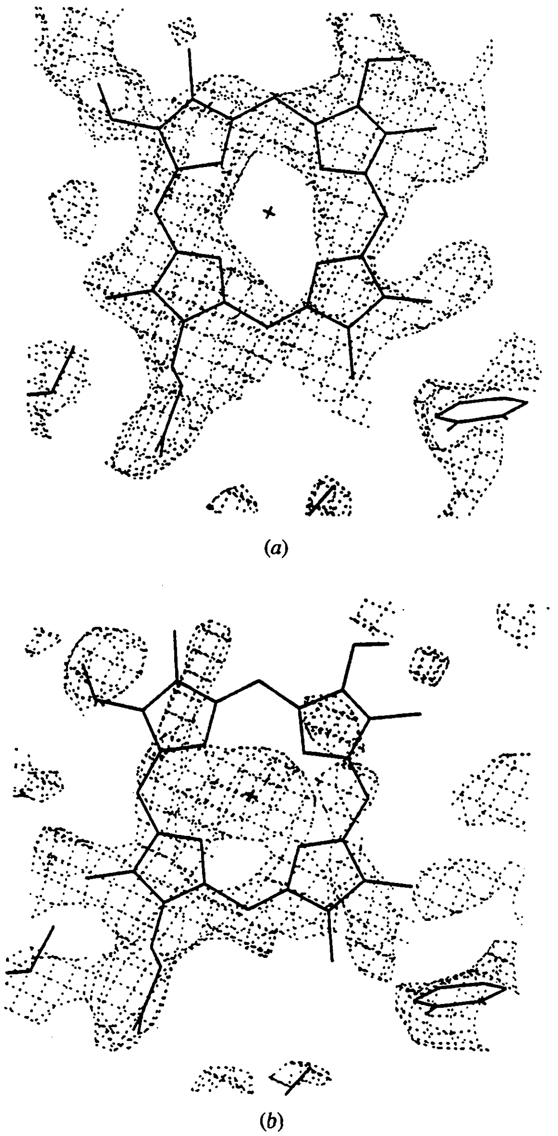


Fig. 1. (*a*) Detail of the heme region of the error-free 2.5 Å native map of cytochrome c_{550}. Note that the map is essentially featureless at the position of the replacement Fe atom. This tends to indicate that the native phases are not noticeably biased by the values of the phases of the Fe atom from which they have been computed. The map in this and subsequent figure illustrations are all contoured at the 1.0σ level. (*b*) Heme region computed from the native 2.5 Å data associated with the $PtCl_4^{2-}$ derivative of cytochrome c_{550} deposited by Timkovich & Dickerson (1976). The definition of the heme ring is noticeably affected by diffraction ripples from the Fe atom of the native structure owing to the special position of the Pt atom in the derivative structure.

Series-termination effects may contribute to this situation.

In contrast to this, the heme ring is well defined in both the $Pt(CN)_4^{2-}$ and UO_2^{2+} TDSIR maps. Regions centered around Trp70 and Pro36 for the UO_2^{2+} derivative are also quite good, but the $Pt(CN)_4^{2-}$ map is not quite as good as either of the other two derivatives. This may in part be due to the fewer number of data (2399) that were measured compared to the other two derivatives (2807, 2886).

In general, the quality of the three TDSIR maps does not approach that of the error-free TDSIR map for one primary reason: the phase doublet angles cannot be determined as accurately from the experimentally measured SIR data as it was for error-free SIR data, which assumed perfect isomorphism. In the TDSIR calculations, between 30 and 45% of the phase doublets had computed cosines in excess of ± 1.0 for the three derivative structures. This clearly indicates that the lengths of the native, heavy-atom and derivative amplitudes do not permit closure of the SIR phasing triangle. Consequently, those doublets for which $|\cos \Delta_{\rm h}|$ exceeded 5.0 were excluded from the refinement as $G_{\rm h}$, the heavy-atom amplitude, was so close to zero as to make its phase imprecise considering positional errors in the determination of the heavy-atom coordinates. Moreover, the zonal restricted phases must have doublet cosines which are either $+1.0$ or -1.0, but between 10 and 15% of the computed values were seen to fall in the mid-range between ± 0.30, such that it was safer to discard these data rather than risk selecting the wrong sign. Better protocols for estimating and weighting these empirically determined cosine doublets would probably improve the TDSIR results presented here.

Inspection of the averaged map from the three TDSIR phase refinements revealed structural details that were not well resolved in the MIR map. This leads us to

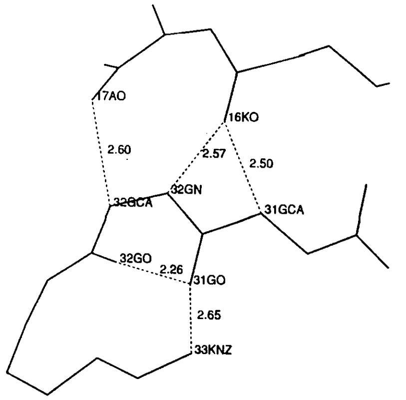
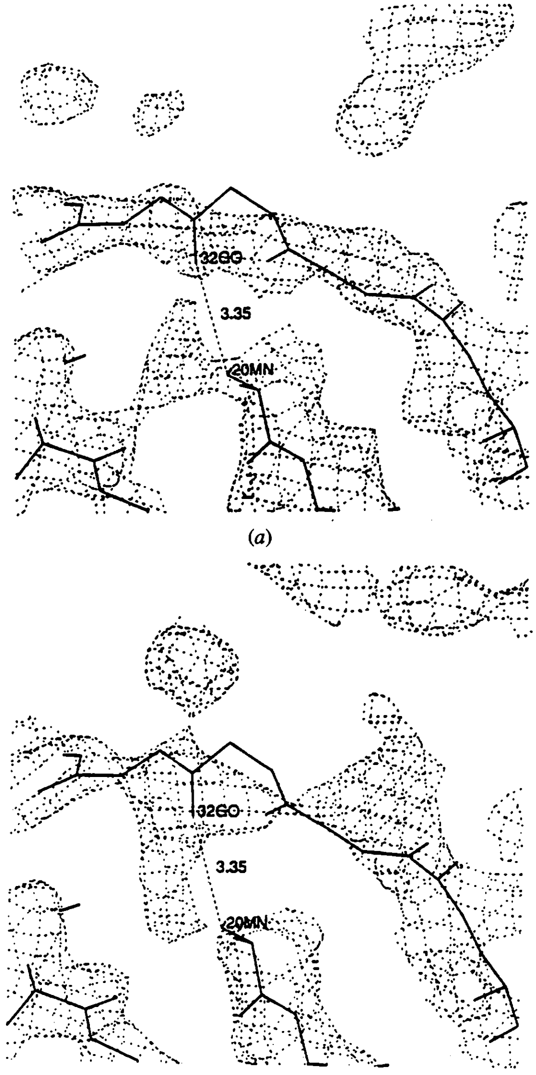


Fig. 2. Wire-frame model depicting the close contacts in the vicinity Gly31–Gly32 taken from the PDB coordinate model that was fitted to the MIR map based on eight derivatives. There are four interstrand contacts that range between 2.50 and 2.65 Å in addition to the 2.26 Å contact between the O31 and O32 carbonyl O atoms.

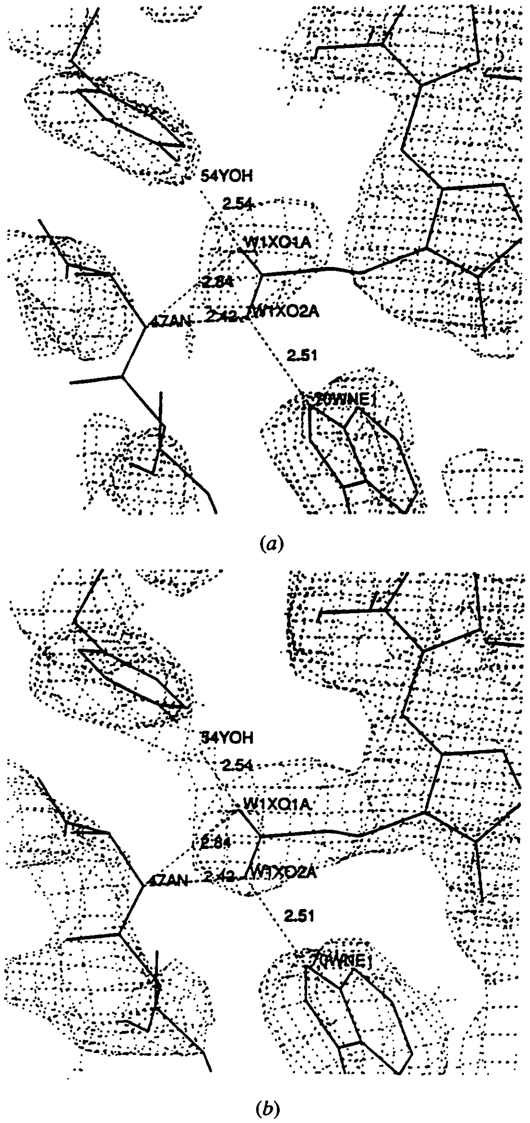
believe that the results obtained by using the TDSIR formula offer a significant improvement over those obtained by the standard MIR methods. The deposited coordinates for the cytochrome c_{550} structure have not been well refined as only 57% of the residues lie in the most favored regions of the Ramachandran φ, ψ plot. There are numerous untenable close contacts in the structure as well. Glycine residues 31 and 32 have an O· · ·O contact of 2.26 Å as well as four other interstrand contacts that range between 2.50 and 2.66 Å (Fig. 2). The MIR density (Fig. 3*a*) offers no strong clues as how to remodel this region of the protein. The averaged TDSIR map (Fig. 3*b*), however, strongly suggests how these residues should be refit to alleviate these bad contacts, as well as form an additional hydrogen bond between the keto-O atom of Gly32 and the peptide amide of Met20.



(*a*)

(*b*)

Fig. 3. (*a*) Detail of the Gly31–Gly32 region of the MIR map, which reveals a narrow sausage of density into which the O31 and O32 carbonyl atoms have been fitted. (*b*) Map obtained from averaging the TDSIR results of the three derivatives. The five untenable close contacts shown in Fig. 2 can be eliminated by readjusting the Gly31–Gly32 residues to position O31 and O32 into two new lobes of density which are *trans* related off the main chain. This new orientation permits O32 to form a hydrogen bond to the amide N atom of Met20.



(*a*)

(*b*)

Fig. 4. (*a*) MIR map illustrating the heme carbonyl environment in the vicinity of Asn47, Tyr54 and Trp70. Close contacts of 2.42, 2.51, 2.54 and 2.82 Å to the carbonyl O atoms are evident. The break in the main-chain density at Asn47 seen in this 1.0σ map also persists when the map is recontoured at the 0.5σ level. (*b*) The combined TDSIR map in the region of the heme carbonyl. The carbonyl side-chain density is well defined and indicates a small corrective shift in the position of the carbonyl terminus, which lengthens the hydrogen bonds to Tyr54 and Trp70. A more surprising detail of this map is that it clearly indicates that the Asn47 main-chain position should be translationally shifted about 1 Å to the lower left in order to fit it in density that corresponds to good residue geometry.

A similar situation exists in that Asn47 lies in a break in the main-chain density contoured at the 1σ level of the MIR map (Fig. 4*a*). This break in density persists even when the MIR map is recontoured at the 0.5σ level. The amide N atom of Asn47 forms two close hydrogen-bond contacts of 2.42 and 2.82 Å to the γ-carbonyl group of the heme; the carbonyl O atoms also have hydrogen-bond contacts of 2.51 and 2.54 Å to $N^{\varepsilon 1}$ of the indole of Trp70 and the side-chain hydroxyl of Tyr54. The density connecting the carbonyl group to the heme is broken. Inspection of the averaged TDSIR map indicates that Asn47 should be translationally shifted ~1.0 Å to reposition the residue in density (Fig. 4*b*). The heme side-chain density to the carbonyl is connected and suggests that its position should be swung about 10° to lengthen the hydrogen-bond contacts to Tyr54 and Trp70.

In summary, unique estimates for the phase invariants and phases may be obtained from the SIR joint probability distribution in the case when the positions of the heavy-atom scatterer are known. The new TDSIR phasing procedure appears to provide more accurate SIR phases than those obtained by the weighted centroid of the two phase estimates. The quality of the individual TDSIR maps, however, may still not be sufficiently improved to allow an unambiguous chain trace at 2.5 Å resolution if the degree of isomorphism is such as to make the determination of the phase doublets difficult. The TDSIR method has the potential to improve the phases of an MIR determination to allow a better fitting of the modeled structure.

Research support from NIH grant GM-46733 is gratefully acknowledged.

References

BLESSING, R. H. & LANGS, D. A. (1988). *Acta Cryst.* **A44**, 729–735.

DICKERSON, R. E., KOPKA, M. L., VARNUM, J. C. & WEINZIERL, J. E. (1967). *Acta Cryst.* **23**, 511–522.

FAN, H. & GU, Y. (1985). *Acta Cryst.* **A41**, 280–284.

FAN, H., HAN, F., QIAN, J. & YAO, J. (1984). *Acta Cryst.* **A40**, 489–495.

FORTIER, S., FRASER, M. E. & MOORE, N. J. (1986). *Acta Cryst.* **A42**, 149–156.

FORTIER, S., MOORE, N. J. & FRASER, M. E. (1985). *Acta Cryst.* **A41**, 571–577.

FORTIER, S., WEEKS, C. M. & HAUPTMAN, H. (1984). *Acta Cryst.* **A40**, 646–651.

FUREY, W. JR, CHANDRASEKHAR, K., DYDA, F. & SAX, M. (1990). *Acta Cryst.* **A46**, 560–567.

GIACOVAZZO, C., CASCARANO, G. & ZHENG, C. (1988). *Acta Cryst.* **A44**, 45–51.

HAUPTMAN, H. (1982*a*). *Acta Cryst.* **A38**, 289–294.

HAUPTMAN, H. (1982*b*). *Acta Cryst.* **A38**, 632–641.

KARTHA, G. (1961). *Acta Cryst.* **14**, 680–686.

KYRIAKIDIS, C. E., PESCHAR, R. & SCHENK, H. (1993). *Acta Cryst.* **A49**, 359–369.

LANGS, D. A. (1985). *Acta Cryst.* **A41**, 578–582.

LANGS, D. A. (1986). *Acta Cryst.* **A42**, 362–368.

LEVY, H. A., THIESSEN, W. E. & BROWN, G. M. (1970). Proc. Am. Crystallogr. Assoc. Meet., New Orleans, LA, USA. Abstract B6.

PESCHAR, R. & SCHENK, H. (1991). *Acta Cryst.* **A47**, 428–440.

SACK, J. S. (1988). *J. Mol. Graphics*, **6**, 224–225.

STRYER, L., KENDREW, J. C. & WATSON, H. C. (1964). *J. Mol. Biol.* **8**, 96–104.

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

Acta Cryst. (1995). A**51**, 542–547

Efficient Methods for the Linearization and Solution of Phase-Invariant Equations

BY DAVID A. LANGS

Medical Foundation of Buffalo, 73 *High Street, Buffalo, NY* 14203, *USA*

AND FUSEN HAN

Upjohn Laboratories, The Upjohn Company, Kalamazoo, MI 49001, *USA*

(*Received* 26 *July* 1994; *accepted* 23 *January* 1995)

Abstract

This paper describes a linear least-squares procedure, whereby, through quadrupole relationships, the 2π integers that linearize sets of unique phase-invariant estimates can be determined. It is subsequently shown that the phase solutions for these linear equations can be obtained, even for basis sets of thousands of phases, without having to either build or invert the full least-squares matrix. The final r.m.s. phase errors achieved by this method can typically be less than 5 or 10°.

Introduction

Crystallographic phase invariants have played a central role in the determination of structures by direct phasing methods. Tangent formula methods for small-molecule

Acta Crystallographica Section A
ISSN 0108-7673 © 1995