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

determinations have traditionally relied on the 0 (modulo 2π) probability estimate for the three-phase structure invariants (Karle & Hauptman, 1956). Efforts to extend these techniques to larger structures have hinged upon obtaining more precise estimates for these phase invariants utilizing algebraic formulae (Karle & Hauptman, 1957; Vaughan, 1958; Hauptman, 1964; Hauptman, Fisher, Hancock & Norton, 1969; Duax, Weeks & Hauptman, 1972; Hauptman & Duax, 1972) or probabilistic formulae as applied to isomorphous replacement or anomalous-dispersion data (Hauptman, 1982*a*,*b*; Giacovazzo, 1983; Fortier, Moore & Fraser, 1985).

Standard tangent formula procedures are insufficient to produce stable solutions for macromolecular structures when phases are refined employing only the 0 (modulo 2π) invariant estimate. Several methods that have been shown to improve the stability and radius of convergence of the phase-refinement process include the active use of evaluated phase-invariant estimates in a modified tangent formula (Olthof & Schenk, 1982), or the use of computational methods that explicitly incorporate the structure-factor amplitude constraint inherent in the three-phase invariant relationship which is normally ignored (Sayre, 1974; Debaerdemaeker, Tate & Woolfson, 1985; Hauptman, 1988). An intriguing least-squares procedure that has been described and shown to produce stable phase solutions with smaller phase errors than the normal tangent formula makes use of linear, rather than trigonometric, phase-invariant equations (Woolfson, 1977). This linear least-squares phasing procedure cannot be employed unless one first has a method to evaluate unique estimates of the phase invariants, *i.e.* in the interval between 0 to 2π rad, and then is able to assign an integer modulo 2π fraction to this value, which would be consistent with a phase-based solution. With the advent of Hauptman's (1982*b*) anomalous-dispersion formula, unique estimates for the triples phase invariants within the interval 0 to 2π need not be contrived. A trial-and-error least-squares (TELS) procedure has been described (Han, DeTitta & Hauptman, 1991) by which a consistent set of 2π integers may be assigned to the linear phase-invariant equations based on the anomalous-dispersion estimates. This paper outlines an alternative method to resolve this 2π integer problem and determine the phases.

Linear phase relationships

A well known property of phase invariants is that their values are independent of the choice of structural origin within the same enantiomorph. Although the values of the individual phases, $\varphi_{\mathbf{h}}$, are affected by a shift of origin, $\Delta\mathbf{r}$,

$$F_{\mathbf{h}}(\text{new}) = F_{\mathbf{h}}(\text{old})\exp[2\pi i(\mathbf{h}\cdot\Delta\mathbf{r})], \quad (1)$$

$$\varphi_{\mathbf{h}}(\text{new}) = \varphi_{\mathbf{h}}(\text{old}) + 2\pi(\mathbf{h}\cdot\Delta\mathbf{r}), \quad (2)$$

the values of phase invariants, *e.g.*

$$\begin{aligned}\omega(\mathbf{h},\mathbf{k}) &= (\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}})_{\text{new}} \\ &= (\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}})_{\text{old}} + 2\pi(\mathbf{h}+\mathbf{k}+\mathbf{l})\cdot\Delta\mathbf{r}, \quad (3)\end{aligned}$$

remain unaffected since the vector sum $\mathbf{h}+\mathbf{k}+\mathbf{l}$ equals 0. The values of the individual phases, as defined by the real and imaginary components of $F_{\mathbf{h}}$, $A_{\mathbf{h}}$ and $B_{\mathbf{h}}$, respectively, have no absolute meaning in the linear-equation sense for any chosen origin, however, because one is free to specify the value of the phase correctly within any $N_{\mathbf{h}}$ times 2π interval

$$\varphi_{\mathbf{h}} = \tan^{-1}(A_{\mathbf{h}}, B_{\mathbf{h}}) + 2\pi N_{\mathbf{h}}. \quad (4)$$

Conversely, if one has phase-invariant estimates, $\omega'(\mathbf{h},\mathbf{k})$, which are uniquely determined within any 0 to 2π interval, as may be obtained from Hauptman's (1982*b*) anomalous-dispersion estimate of the triple, it follows that there is an infinity of values of $\omega(\mathbf{h},\mathbf{k})$, consistent with specific integer choices for (4), which must be resolved before these equations can be solved linearly.

$$\begin{aligned}\omega(\mathbf{h},\mathbf{k}) &= \varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{\mathbf{l}} \\ &= \omega'(\mathbf{h},\mathbf{k}) + 2\pi(N_{\mathbf{h}}+N_{\mathbf{k}}+N_{\mathbf{l}}). \quad (5)\end{aligned}$$

From a solution standpoint, it is only necessary to determine one such set of integers, $N(\mathbf{h},\mathbf{k}) = (N_{\mathbf{h}}+N_{\mathbf{k}}+N_{\mathbf{l}})$, which are consistent with some chosen phase solution. For example, if one restricts the integer choices in (4) to some specified range, such that $\pi > \varphi_{\mathbf{h}} > -\pi$, it can be seen that $N(\mathbf{h},\mathbf{k})$ will be restricted to one of three permissible values, -1, 0 or $+1$, when $\omega'(\mathbf{h},\mathbf{k})$ is also uniquely specified within the same π to $-\pi$ interval. If the individual phase values are randomly distributed, $N(\mathbf{h},\mathbf{k})$ is expected to be equal to 0 for two thirds of the time and either $+1$ or -1 for the remaining third of the time.

Degrees of freedom

In an earlier application, Woolfson (1977) used linear phase-invariant equations to refine phase sets that were the tentative magic integer solutions of Karle–Hauptman determinants. In a side comment, he pointed out that even if one did not have trial values of the phases from which to estimate the nearest integers of (5), because one has N_t linear triples equations based on N_p unknown phases, where N_t normally exceeds N_p by a factor five or more, one is free to arbitrarily assign N_p of the N_t triples integers provided that each triple corresponds uniquely to one, and only one, of the N_p phases. Thus, in our situation, although we cannot *a priori* predict which of the triples integers will be 0 or ±1 to correspond with individual phases restricted to the range $\pi > \varphi_{\mathbf{h}} > -\pi$, we can still correctly define a sizable basis set of N_p of the N_t triples integers that are consistent with some other

permissible solution of the N_p phases. Each of these two points will be considered in the sequel that follows.

If one considers quadrupole relationships (Viterbo & Woolfson, 1973) among the triples invariant estimates

$$\begin{aligned}
\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} &= \omega'(\mathbf{h}, \mathbf{k}) + 2\pi N(\mathbf{h}, \mathbf{k}) \\
-\varphi_{\mathbf{h}} + \varphi_{\mathbf{p}} + \varphi_{\mathbf{q}} &= \omega'(-\mathbf{h}, \mathbf{p}) + 2\pi N(-\mathbf{h}, \mathbf{p}) \\
-\varphi_{\mathbf{k}} - \varphi_{\mathbf{p}} + \varphi_{\mathbf{r}} &= \omega'(-\mathbf{k}, -\mathbf{p}) + 2\pi N(-\mathbf{k}, -\mathbf{p}) \\
-\varphi_{\mathbf{l}} - \varphi_{\mathbf{q}} - \varphi_{\mathbf{r}} &= \omega'(-\mathbf{l}, -\mathbf{q}) + 2\pi N(-\mathbf{l}, -\mathbf{q}),
\end{aligned} \tag{6}$$

for which

$$\begin{aligned}
&\omega'(\mathbf{h}, \mathbf{k}) + \omega'(-\mathbf{h}, \mathbf{p}) + \omega'(-\mathbf{k}, -\mathbf{p}) + \omega'(-\mathbf{l}, -\mathbf{q}) \\
&\quad + 2\pi[N(\mathbf{h}, \mathbf{k}) + N(-\mathbf{h}, \mathbf{p}) + N(-\mathbf{k}, -\mathbf{p}) \\
&\quad + N(-\mathbf{l}, -\mathbf{q})] = 0,
\end{aligned} \tag{7}$$

one has access to linear relationships (7) by which the individual $N(\mathbf{h}, \mathbf{k})$ can be determined provided that the number of quadrupole equations, N_q, exceeds N_t, the number of triples invariants integers. Given that the second derivative of (7) with respect to any of the integers is zero, there is a unique linear least-squares solution which is independent of the initial values of $N(\mathbf{h}, \mathbf{k})$ that are chosen. A unique *a priori* phase solution of the linear phase-invariant equations (5) also exists, which does not require the testing of numerous tentative trial sets of phases, as was inferred in the TELS paper (Han, DeTitta & Hauptman, 1991).

In closing this section, we point out that there are actually only $N_p - 3$ triples integers that can be arbitrarily assigned for a basis set of N_p phases when the quadrupoles:triples ratio is sufficiently high. In retrospect, this may have been anticipated from the structure of the quadrupole (6). If one had chosen the $\varphi_{\mathbf{h}}$, $\varphi_{\mathbf{p}}$ and $\varphi_{\mathbf{k}}$ phases to assign integers for $N(\mathbf{h}, \mathbf{k})$, $N(-\mathbf{h}, \mathbf{p})$ and $N(-\mathbf{k}, -\mathbf{p})$, respectively, one cannot assign an arbitrary integer for the fourth remaining triple of the quadrupole, $N(-\mathbf{l}, -\mathbf{q})$, even though none of its three constituent phases $(\varphi_{\mathbf{l}}, \varphi_{\mathbf{q}}, \varphi_{\mathbf{r}})$ has been previously used to assign an integer for another triple. This is because $N(-\mathbf{l}, -\mathbf{q})$ must equal the sum of the first three integers minus the sum of the four ω' values. Different triples, not linearly connected to previously assigned integer values, must be selected to satisfy these constraints. Thus, although a single quadrupole has six independent phases, only three phases may be used to arbitrarily assign integers for three of the four triples. Moreover, if one performs a convergence mapping (Germain, Main & Woolfson, 1970) on the quadrupoles of an overdetermined set of triples, one will always obtain $N_{\mathbf{p}} - 3$ 'origin defining' triples. This difference of three from the total number of N_p phases appears to be constant and independent of the fact that certain space groups may require fewer than three phases to define an origin for the triples basis set (the enantiomorph is always chosen by the handedness of the ω' estimates).

Convergence mapping (see Germain, Main & Woolfson, 1970) is a procedure that may be used to determine the reverse order of phase determination for a list of triples invariants, and consequently identify a small starting set of basis phases from which all of the other phases may be determined when the procedure is executed in the forward direction. In the above application, to determine the starting basis set of triples invariants for a set of quadrupoles equations, one sums a weight of reliability for each triple, based on the number of quadrupoles into which it enters. The triple having the lowest reliable weight is ordered to be solved last, and its quadrupole contributions to the remaining triples sums is subtracted. The procedure is repeated to identify the next triple that is to be eliminated from the quadrupoles list, which is the triple currently having the lowest weight of reliability among those triples that have not been eliminated, and the procedure repeated for the remaining triples until a block of $N_p - 3$ triples vectors is identified which have zero weight of reliability. We will refer to this meaning of 'convergence' as a CVG procedure to distinguish it from the use of convergence in the mathematical sense in other parts of this text, as meaning the asymptotic approach to a solution of a set of simultaneous equations by iterative methods.

Various CVG mapping criteria can be explored to select the $N_p - 3$ basis integers for a particular structure. A CVG ordering based on $\beta_{\mathbf{h},\mathbf{k}}$, the minimum of the current total number of quadrupoles per triple at any point in the CVG procedure as described above,

$$\beta_{\mathbf{h},\mathbf{k}} = \sum_{N_{qi}} \mathrm{wt}_i, \tag{8}$$

which is weighted on the lack of closure of the quadrupoles,

$$\begin{aligned}
\mathrm{wt}_i = 0.5\{1 + \cos[&\omega'(\mathbf{h}, \mathbf{k}) + \omega'(-\mathbf{h}, \mathbf{p}) \\
&+ \omega'(-\mathbf{k}, -\mathbf{p}) + \omega'(-\mathbf{l}, -\mathbf{q})]\},
\end{aligned} \tag{9}$$

may be shown to be an effective method for this purpose. Here, the weights wt_i are close to 1.0 when the sum of the four ω' values equal zero (modulo 2π), which it must for error-free data, and 0.0 when the sum of the errors in the individual ω' estimates approach π (modulo 2π) when random Gaussian r.m.s. errors are added to the initial ω' estimates. Furthermore, it will be shown that the initial estimates of the remaining triples integers that one obtains by extension through this CVG map are often accurate enough to ensure a solution of (5) that will converge in a small number of cycles using the diagonal approximation to the full least-squares matrix. More so, especially, if the errors in the initial ω' estimates have been previously minimized utilizing the same quadrupole identities

$$\begin{aligned}
\exp[i\omega'(\mathbf{h}, \mathbf{k})] = \sum_{N_{qi}} \exp\{i[&-\omega'(-\mathbf{h}, \mathbf{p}) - \omega'(-\mathbf{k}, -\mathbf{p}) \\
&- \omega'(-\mathbf{l}, -\mathbf{q})]\}/N_{qi},
\end{aligned} \tag{10}$$

where equation (10) is complex valued and must be evaluated by a tangent formula expression based on its real and imaginary components summed over the number of quadrupole relationships N_{qi} into which $\omega'(\mathbf{h}, \mathbf{k})$ enters. This tangent formula refinement of the initial ω' values will be referred to as an ω' refinement, so as not to be confused with the least-squares-refinement procedures to determine the 2π integers of (7) and, subsequently, the individual phases by means of (5).

Given the above analysis, calculations were performed to determine whether in practice the triples integers were indeed solvable by these methods using reasonably precise experimental data.

Computational analysis

All of the calculations were performed on a Silicon Graphics Iris workstation. A 0.85 Å resolution data set from the triclinic structure of a synthetic non-peptidic enkephalin analog was initially used to develop the integerization procedure: ENKA, ethyl 2-{(2*S*)-3,4,4a,5,6,7a-hexahydro-6-oxo-2-(phthalimidomethyl)-2*H*-pyrano[2,3-*b*]pyrrol-7-yl}-4-methylvalerate, $C_{24}H_{30}N_2O_6$, *P*1, $Z = 3$, *i.e.* 96 non-light atoms (Smith & Krstenansky, 1994).

Experimentally derived ω' values for the native data of ENKA do not exist as the structure has no significant anomalous signal. This small low-symmetry structure was mainly chosen to simplify and scale down the size of the calculations during the initial testing and debugging of computer codes. The ω' values that were first used were error-free values computed from the known structure. Subsequent calculations employed ω' values to which random Gaussian-distributed r.m.s. errors of various magnitudes had been added. In these latter calculations, it is important to note that if the error, $\delta\omega'$, added to the true value of ω' caused the value to fall outside the bounds of π and $-\pi$, *e.g.* $\omega' + \delta\omega' > \pi$, it was reset within the π to $-\pi$ range by subtracting 2π rad such that its adjusted value actually differed from ω' by $2\pi - \delta\omega'$ rad rather than $\delta\omega'$ rad, which in the majority of cases would represent the smaller of the two values. This 'wrap around' penalty cannot be justifiably ignored in our calculations, and actually makes the task of obtaining a convergent solution much more difficult.

The first question that had to be resolved for the ENKA calculations dealt with the number of E values that had to be chosen in order to obtain a sufficient quadrupoles:triples overdeterminancy for (7) to yield reliable results. A calculation based on the 1000 largest E values greater than 1.34 produced 28 474 triples and 174 884 quadrupole relationships for an overdeterminancy ratio of 6.14. It took 9 s to generate these triples and another 1.5 min to generate the quadrupoles and index their constituent triples with their position numbers in the original triples list. It was noted that, if the ratio of quadruples to triples was much less than six, moderate r.m.s. phase errors in the determination of the ω' estimates were often sufficient to produce unreliable results. That is, although (7) is an exact algebraic identity, it may be numerically imprecise due to errors in the ω' estimates, especially if the number of contributors is small.

It was obvious that brute-force full-matrix inversion procedures for evaluating the triples integers *via* (7) were out of the question. Given that it took 15 min to invert a 400×400 matrix on an Iris Indigo workstation, and that computation times increase proportional to the cube of the size of the matrix, it was estimated that it would require more than ten years to evaluate the triples integers in this way. However, since the matrix of the phase-invariant equations is quite sparse, and a large initial starting set of integers may be arbitrarily assigned, it was hoped that more efficient procedures to solve (7) could be found.

If we began with the correct integer choices (+1, 0, −1) for a unique set of N_p integers corresponding to a non-singular matrix, it was possible to assign initial zero values for the remaining integers and verify that the larger shifts indicated by a diagonal-matrix least-squares refinement tended to correctly indicate the +1 and −1 integers. Using error-free ω' values, it was possible to determine the values of the 28 474 integers in five iterative least-squares cycles that required less than 2 min of CPU time and, moreover, all 28 474 integers were determined correctly. It was also possible to use a diagonal-matrix least-squares method to solve the 28 474 equations (5) for the values of the 1000 basis phases in seven iterative cycles that required a total time of only 9 s. Starting with an initial zero value for the 1000 phases, the r.m.s. phase error rapidly reduced from 12° at the end of the first cycle and converged to a mere 0.05° at the end of the seventh cycle. This number represents the round-off error of the computer since the final phase error $\langle|\delta\varphi|\rangle$ is expected to be zero since it is related to $\langle|\delta\omega'|\rangle$, the average phase error of ω', by

$$\langle|\delta\varphi_{\mathbf{h}}|\rangle = \langle|\delta\omega(\mathbf{h}, \mathbf{k})'|\rangle / N_t(\mathbf{h})^{1/2}. \qquad (11)$$

Thus, even if the average error in ω' is large, say 60°, one might anticipate a phase-refinement convergence of $\sim 6.5°$, given that each phase is determined by an average of 85 triples contributors.

The use of a pre-conditioned Krylov algorithm adapted for sparse matrices gave significantly better results than the diagonal-matrix least-squares method of solution, but neither method was found to be effective when the chosen N_p basis integers were all assigned a zero value or randomly assigned +1, −1 and 0 values corresponding to the expected 1/6, 1/6, 2/3 distribution for these integers. If the N_p basis integers are not assigned their correct (+1, 0, −1) values, many of the

remaining integers must have values outside the ±1 range and the diagonal matrix and pre-conditioned Krylov procedures are not sufficient to determine if a large calculated shift should correspond to a magnitude of 1, 2 or larger. A full-matrix procedure will clearly resolve this dilemma, but it remains to be shown whether a more efficient matrix-solution method can achieve the same goal in a more reasonable time frame.

At this point, it was realized that the integerization problem could be solved through the quadrupoles CVG map in a rather straightforward manner without having to resort to standard matrix-solution methods.

A triples CVG map for the 174 484 quadrupoles was computed from which 997 triples integers were determined as the starting basis set, instead of the 1000 anticipated values. It was easy to verify that if one had error-free ω' values, one could assign the basis integers any values and indeed correctly determine a self-consistent set for the remaining 27 477 integers without a single error. Any number of simple recursive schemes can be outlined by which the integers may be reset to the (+1, 0, −1) range such that the diagonal-matrix refinement method would be sufficient to determine the phases from initial zero values as described earlier.

The calculations were next repeated using ω' values, which incorporated random Gaussian-distributed r.m.s. phase errors of varying magnitudes. The results of refining these ω' estimates to minimize the initial error are reported in Table 1. CVG maps were redetermined using $\beta_{\mathbf{h},\mathbf{k}}$ weights reflecting the lack of closure of the quadrupoles (9). The integer extension procedure continued to obtain correct integer values for the vast majority of the 28 474 triples as long as the r.m.s. phase error in the quadrupole-refined values of ω' did not exceed 45°.

We next wanted to test these methods using a more realistic example of empirically determined ω' estimates. In this sequel, we re-examined the TELS analysis of the 1.5 Å cytochrome c_{550} data that was reported earlier (Han, DeTitta & Hauptman, 1991). This second data set was obtained from the deposited native data of the orthorhombic $P2_12_12_1$ form of this 134 amino acid structure (Timkovich & Dickerson, 1976). An anomalous-dispersion data set was simulated by adding the anomalous signal of the $PtCl_4^{-2}$ derivative to the native measured data. A total of 31 360 triples were generated from 532 Friedel pairs of data that had E magnitudes greater than 1.28, the ω' estimates were obtained directly from Hauptman's anomalous-dispersion formula. The total number of quadrupoles relating these triples was 368 323 for an overdeterminacy of 11.75. The average r.m.s. error of the triples invariant estimates was 84.96°, and all attempts to minimize this error by ω' refinement failed. This negative result may have been anticipated from our experience with the ENKA data presented in Table 1. It was necessary to proportionately reduce this error to less than 70° in order to achieve a refinement that reduced the initial r.m.s. error to 20°, as is reported in Table 2. Here we note that the average r.m.s. error of the refined ω' values is inversely correlated with an empirically derived weight, wt[$\alpha(\mathbf{h},\mathbf{k})$], corresponding to the phase-invariant consistancy or relative 'alpha' value of the quadrupole sum

Table 1. *Progress of the refinement of ω' values for ENKA, which include random Gaussian r.m.s. phase errors of various magnitudes*

Nine trials were examined which had initial modeled r.m.s. phase errors between 10 and 80°. The phase errors listed are an average of five separate trials. The refinement fails to converge if the initial error exceeds 75°, as is indicated in trial 9 where the final r.m.s. phase error diverges to the random expected value of $\pi/3^{1/2}$ rad. When the procedure converges, the final r.m.s. phase error is usually reduced to about a sixth of the initial starting value.

Trial	Initial r.m.s. $\delta\varphi$	First cycle	Final cycle	Number of cycles
1	10	3.1	1.7	4
2	20	6.0	3.1	4
3	40	18.7	6.4	5
4	50	26.6	8.2	5
5	60	50.2	9.9	6
6	65	58.1	11.3	8
7	70	65.7	13.6	12
8	75	72.5	19.3	16
9	80	78.6	103.9	>20

$$\text{wt}[\alpha(\mathbf{h},\mathbf{k})] = \left\{\left[\left(\sum_{N_{qi}}\cos\Delta\right)^2 + \left(\sum_{N_{qi}}\sin\Delta\right)^2\right]^{1/2} \Big/ N_{qi}\right\}^n, \quad (12)$$

where $\Delta = -\omega'(-\mathbf{h},\mathbf{p}) - \omega'(-\mathbf{k},-\mathbf{p}) - \omega'(-\mathbf{l},-\mathbf{q})$, $n = [N_{qi}(\text{min.})/N_{qi}]^{1/2}$ and $N_{qi}(\text{min.})$ is the value of N_{qi} associated with that triple which has the smallest value for N_{qi} among the 31 360 triples. The values of wt(α) may be somewhat better fit in the range of 1.0 to 0.0 by perturbing the value of the exponent n. Table 2 lists the average r.m.s. error sorted into 20 equal groups on the decreasing magnitude of wt(α), both after the first and eleventh cycles of ω' refinement, at which time the r.m.s. phase error in the ω' values had been reduced to 19.9°. Two additional cycles of ω' refinement using lack-of-closure weights (9) permitted a further reduction to 10.6°, after which time the equations were successfully integerized from the quadrupoles CVG map as was described in the ENKA calculations. Note that, because of the $P2_12_12_1$ translational symmetry of the cytochrome c_{550} structure, the triples phase relationships often contain a fractional shift of π when the phases $\varphi_{\mathbf{h}i}$ are re-expressed in terms of their standard parent form

$$\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{\mathbf{l}} + \pi = \omega'(\mathbf{h},\mathbf{k}) + 2\pi N(\mathbf{h},\mathbf{k}) \quad (13)$$

such that $\omega'(\mathbf{h},\mathbf{k}) - \pi$ must replace the value of $\omega'(\mathbf{h},\mathbf{k})$ in the $P1$ ENKA example (5). It is necessary to make this transformation prior to solving the linear equations for the individual phase values.

Table 2. *Progress of the ω' refinement for cytochrome c_{550}*

The 31 360 triples are analyzed in 20 groups of 1568 each. Entries on the left-hand side of the table are sorted on decreasing magnitude of the A value of the anomalous-dispersion estimate, while those to the right of columns five and seven are sorted on decreasing magnitude of $\beta_{\mathbf{h},\mathbf{k}}$ after one and eleven refinement cycles. Column one gives the shell number, column two lists the average A value of the shell, column three gives the average cosine expected from the $I_1(A)/I_0(A)$ statistic, column four reports the average value of the cosine of ω_{true}. Columns five and seven record the average value of wt[$\alpha(\mathbf{h},\mathbf{k})$] after one and eleven cycles. The root mean square value of the residual phase error between ω_{true} and ω_{calc} after the first and eleventh refinement cycles are listed in columns six and eight, respectively. The r.m.s. phase error at the end of the 11th cycle was 19.9°. Two additional final cycles employing lack-of-closure weights (9) further reduced this error to 10.6°.

Shell	⟨A value⟩	I_1/I_0	cos ω_{true}	wt(α_1)	R.m.s.	wt(α_{11})	R.m.s.
1	1.838	0.673	0.625	0.99	16.0	0.98	15.1
2	1.197	0.513	0.515	0.96	17.8	0.96	16.5
3	1.004	0.449	0.421	0.95	18.5	0.95	17.4
4	0.888	0.406	0.409	0.94	20.5	0.94	18.3
5	0.802	0.372	0.351	0.93	21.7	0.93	18.1
6	0.738	0.347	0.358	0.91	24.1	0.92	19.4
7	0.685	0.325	0.335	0.90	25.5	0.92	18.5
8	0.639	0.305	0.321	0.89	27.3	0.91	18.4
9	0.599	0.288	0.274	0.88	29.0	0.90	20.3
10	0.564	0.272	0.261	0.87	31.7	0.89	19.6
11	0.531	0.257	0.245	0.84	34.3	0.88	20.7
12	0.500	0.243	0.236	0.82	36.0	0.87	20.9
13	0.472	0.230	0.224	0.80	41.7	0.86	20.8
14	0.446	0.218	0.223	0.78	44.7	0.85	21.6
15	0.420	0.206	0.214	0.74	49.5	0.84	22.7
16	0.394	0.194	0.200	0.70	55.5	0.82	22.0
17	0.368	0.181	0.191	0.65	61.7	0.81	22.9
18	0.338	0.167	0.158	0.57	68.2	0.79	22.8
19	0.305	0.151	0.130	0.47	74.6	0.76	23.7
20	0.254	0.128	0.152	0.30	78.2	0.68	26.9

Discussion

The use of a quadrupole-based triples CVG map to evaluate the integers for the linear least-squares triples invariant equations of the ENKA trial structure is very encouraging. ω' refinement typically reduces the initial r.m.s. phase-invariant error by a sixfold factor as seen by comparing columns 2 and 4 of Table 1. All 28 474 triples integers could be correctly determined from the CVG map as long as the final refined error in ω' was less than 20°. Note that this procedure rapidly deteriorates and fails to converge for trial 9 of Table 1 when the initial r.m.s. phase error added to ω' exceeds 75°.

Similar computations utilizing empirically determined ω' values for the cyctochrome c_{550} test structure failed to converge because the initial r.m.s. errors in these values averaged 85°. A convergent ω' refinement could only be obtained after artificially reducing this error to less than 70° in a proportional manner, as is indicated in Table 2. Once the overall average r.m.s. phase error was reduced to 10.6°, all 31 360 triples integers could be correctly determined from the CVG map without a single error. Given that there are an average of 177 triples for each phase, the final individual phase error from solving the linear ω equations (5) should approach $10.6/177^{1/2}$ or about 0.80° by (11). The actual average r.m.s. phase error of the 532 individual phases determined from the linearized equations was 6.2°, compared to 39.1° using the TELS method. The fact that the phase refinement did not converge to its expected limit of 0.80° indicates that the r.m.s. errors from the ω' refinement were not normally distributed but were highly correlated with the values of the individual phases. These results may represent a significant improvement as to what was reported earlier, but it still remains to be demonstrated whether the ω' refinement/CVG mapping technique will be useful in many practical applications or whether the number of structures that can provide initial ω' estimates with a r.m.s. error less than ~70° is too restrictive.

This work was supported in part by NIH grant 46733.

References

DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* A**41**, 286–290.

DUAX, W. L., WEEKS, C. M. & HAUPTMAN, H. (1972). *Acta Cryst.* B**28**, 1857–1863.

FORTIER, S., MOORE, N. J. & FRASER, M. E. (1985). *Acta Cryst.* A**41**, 571–577.

GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* B**26**, 274–285.

GIACOVAZZO, C. (1983). *Acta Cryst.* A**39**, 585–592.

HAN, F., DETITTA, G. & HAUPTMAN, H. (1991). *Acta Cryst.* A**47**, 484–490.

HAUPTMAN, H. (1964). *Acta Cryst.* **17**, 1421–1433.

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

HAUPTMAN, H. (1982*b*). *Acta Cryst.* A**38**, 632–641.

HAUPTMAN, H. (1988). Proc. Am. Crystallogr. Assoc. Meet., Philadelphia, Pennsylvania, USA. Abstract R4.

HAUPTMAN, H. & DUAX, W. L. (1972). *Acta Cryst.* A**28**, 393–395.

HAUPTMAN, H., FISHER, J., HANCOCK, H. & NORTON, D. (1969). *Acta Cryst.* B**25**, 811–814.

KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635–651.

KARLE, J. & HAUPTMAN, H. (1957). *Acta Cryst.* **10**, 515–524.

OLTHOF, G. J. & SCHENK, H. (1982). *Acta Cryst.* A**38**, 117–122.

SAYRE, D. (1974). *Acta Cryst.* A**30**, 180–184.

SMITH, G. D. & KRSTENANSKY, J. (1994). Unpublished results.

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

VAUGHAN, P. A. (1958). *Acta Cryst.* **11**, 111–115.

VITERBO, D. & WOOLFSON, M. M. (1973). *Acta Cryst.* A**29**, 205–208.

WOOLFSON, M. M. (1977). *Acta Cryst.* A**33**, 219–225.