Acta Cryst. (1969). B25, 380

Refinement of Protein Phases with the Karle–Hauptman Tangent Formula

BY JON E. WEINZIERL, DAVID EISENBERG AND RICHARD E. DICKERSON

Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California, U.S.A.*

(*Received* 27 *February* 1968)

The Karle–Hauptman tangent formula has been tested as a means of refining phases previously determined from single and multiple isomorphous replacement. A test of the tangent formula, using exact phases of a model protein as a starting point, showed that the formula reproduces phase angles in good agreement with the true phases, even at only 4 Å resolution. When errors were introduced into the model phases, the tangent method refined back toward the true phases. This procedure was used to refine the phases of cytochrome *c* obtained from a double isomorphous analysis at 4 Å resolution, and some improvement in the electron density at the heme group was apparent. Trials with single isomorphous replacement phases for a hypothetical derivative of the model protein showed that the tangent formula can be used to resolve the single-derivative ambiguity. Applying this method to single-derivative phases for cytochrome *c*, an electron density map was obtained which appears to be only slightly inferior to the two-derivative map.

Introduction

The multiple isomorphous replacement (MIR) method of phase analysis, which made protein structure analysis possible, is still the standard method of finding protein phases. The trial and error method is manifestly impossible with molecules of this complexity, and Fourier refinement methods are useful only in the later high resolution stages. Patterson methods have proven useful only for finding heavy atom positions. The heavy atoms which occur naturally or can be added to protein molecules are not sufficiently heavy to dominate phasing. The anomalous scattering effect is smaller yet, and can be used only as an auxiliary to isomorphous replacement. The only significant variant to the MIR method has been the single isomorphous replacement (SIR) method with its inherent phase ambiguity broken by the use of anomalous scattering data (Kartha, 1961; Blow & Rossmann, 1961). Kartha showed that the SIR method alone is similar to vector superposition using the heavy atom positions, and that if the heavy atom constellation itself does not possess a center of symmetry, then the SIR map is built up from a number of images of the protein molecule in proper register, plus an equal number of images of the enantiomorph which add in a nonsystematic manner to give a high background noise level. Blow & Rossmann demonstrated that the map of hemoglobin produced by SIR and anomalous scattering, although considerably degraded from the seven compound MIR map, was still correct in gross features.

One class of phasing methods which has seen relatively little use with proteins is statistical phasing, of which the most powerful methods seem to be those developed by Karle, Karle & Hauptman (Karle, 1964). In the two decades since these methods appeared, they have shown their worth with increasingly complex structures, first centrosymmetric and then non-centrosymmetric. Proteins, however, have been relatively unexplored territory. C. L. Coulter (1965) has carried out model compound experiments along lines similar to some of those in this paper, and D. C. Phillips has carried out some unpublished experiments with lysozyme. The feeling has been prevalent that Karle–Hauptman methods could not produce protein phases *ab initio*, and that in any case they would not be applicable to data short of atomic resolution, or the vicinity of 2–1·5 Å. We have accepted the first restriction, but have found the second not to be true under proper conditions.

We have attempted to use the Karle–Hauptman tangent formula (equation 4·5 of Karle & Karle, 1966, hereafter referred to as K and K) in two ways, first as a means of refining phases produced by the MIR method, and then as a means of choosing between the two possible solutions of the SIR method. In both cases, the procedure has first been checked on a dummy protein ('modelglobin') having structure factors and phases calculated from the backbone chain of myoglobin out to and including β-carbon atoms, and then applied to real data for horse heart cytochrome *c*. The data used extended out to 4 Å and to 3·7 Å, respectively.

Test of the tangent formula with model protein phases

The principles of Karle–Hauptman phase refinement are given in Karle (1964) and in K & K, and only those equations which were of direct use will be given here. The tangent formula is:

$$\tan \varphi_{\mathbf{h}} \simeq \frac{\sum_{\mathbf{k}} |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \sin (\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}})}{\sum_{\mathbf{k}} |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \cos (\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}})}, \qquad (1)$$

* Contribution No. 3656 from the Gates and Crellin Laboratories of Chemistry.

where **h** and **k** each represent Miller index triples. The E's in equation (1) are normalized structure factor magnitudes as given by:

$$|E_{\mathbf{h}}|^2_{\text{obs}} = \frac{|F_{\mathbf{h}}|^2}{\varepsilon \sum_j f_j^2 (\mathbf{h})} \quad . \tag{2}$$

$|F_{\mathbf{h}}|$ is the observed structure factor magnitude of reflection **h**, f_j is the atomic scattering factor for the jth atom in the unit cell and ε is a number which corrects certain reflections for space group extinctions and the presence of symmetry elements such as mirror planes. Equation (1) provides a way of finding a new, refined phase for each reflection in terms of the entire set of phases which were initially assumed on the right hand side of the equation.

Before setting out to refine sets of approximate phase angles, the accuracy of the tangent formula itself at low resolution was tested with the exact phases of modelglobin. Structure factors and phases for the 1239 reflections within the 4 Å limit were calculated using the polypeptide backbone atoms, β-carbons, iron and the heme group of myoglobin in its true unit cell. We are indebted to Dr John C. Kendrew for the myoglobin coordinates. These phases and normalized structure factors will be designated by φ_{obs} and E_{obs}.

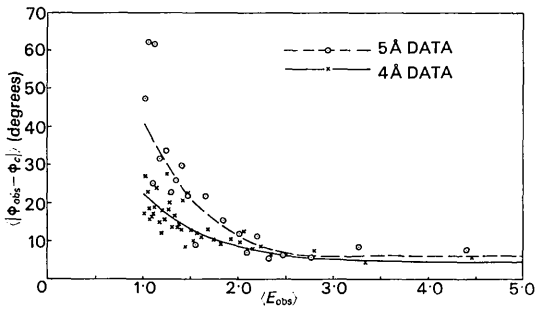

Fig. 1. The error in calculating phase angles of a model protein from the tangent formula, as a function of the normalized structure factor magnitude, using correct phases in the right hand side of the tangent formula. Each point represents the average over ten reflections.

The phase angles of the 410 reflections of modelglobin with E_{obs} greater than 1·0 were calculated from equation (1), using φ_{obs} and E_{obs} of all 1239 reflections on the right hand side. The results are shown in Fig. 1. The mean error in the calculated phase angle is only about 6° for the fifty reflections with the largest E_{obs}, but the error in phase increases as E_{obs} decreases. The tangent formula begins to fail at significantly higher E_{obs} values if the data are cut off at 5 Å. This suggests that too few contributors to formula (1) are available with low resolution data, and that the tangent formula will be more useful at 2 Å or higher resolution. Even at low resolution, however, the tangent formula can be used to improve the phases of the strongest reflections; these are often poorly determined by the MIR method because the ratio of $|f_H|$ to $|F_P|$ is small. At low resolution, the Karle–Hauptman and isomorphous replacement methods are complementary.

Several functions were tried as possible indices of the correctness of phases determined by the tangent formula, but none of these proved entirely satisfactory. One such function, called the 'unnormalized calculated E', was defined by analogy with the tangent formula:

$$|E_{\mathbf{h}}|_c = \left| \sum_{\mathbf{k}} |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \exp\{i(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}})\} \right| . \tag{3}$$

These E_c's asymptotically approached constant values as the number of terms in the tangent formula increased (Fig. 2). For modelglobin, the mean value of E_c/E_{obs}

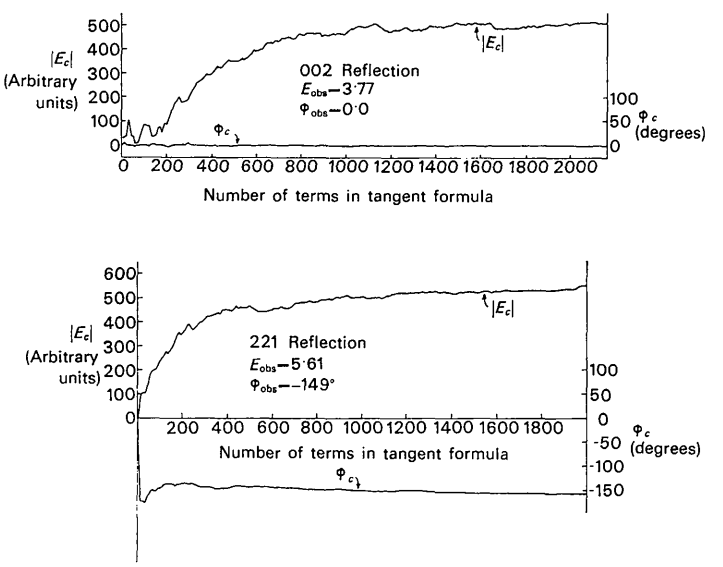

Fig. 2. The dependence of E_c and the calculated phase angle φ_c upon the number of terms used in the tangent formula. Terms were added in order of decreasing $E_{\mathbf{k}}$. (*a*) Centric 002 reflection. (*b*) Acentric 221 reflection.

for the 410 reflections with the largest E_{obs} was found to be 120, but values of this ratio for individual reflections ranged from 10 to 290. No correlation could be found between this ratio and the error in the corresponding phase from the tangent formula, except that when the ratio was less than 40, the phase angle was usually greatly in error. For this reason the Karle & Karle *R index* (equation 5·2 of K & K) would not seem to be useful here as a criterion of phase quality at low resolution.

The effect of the number of terms used in the tangent formula upon the calculated phase angle and E_c is shown in Fig. 2 for a centric and an acentric reflection. The terms were added in order of decreasing E_k. These two reflections are substantially determined by the top 300 terms. In general, however, at least 700 terms were needed to calculate phases of reflections with E_{obs} values of 1·0 or greater.

Another possible measure of phase correctness is based on the phase probability function defined by Karle & Karle (1966):

$$P_{(\varphi'_h)}=[2\pi I_0(\alpha)]^{-1}\exp\{\alpha\cos(\varphi'_h-\varphi_h\}\,. \qquad (4)$$

This expression gives the probability of correctness of a chosen phase angle φ'_h as a function of the coefficient α and the distance of φ'_h from the phase calculated from the tangent formula, φ_h. The sharpness of the probability distribution increases as α rises. The coefficient α is proportional to the product of observed and calculated E_h values:

$$\alpha=\frac{2\sigma_3}{\sigma_2^{3/2}}|E_h|_c|E_h|_{obs}\,, \qquad (5)$$

where the sigmas are defined in terms of the atomic numbers, Z_j, as:

$$\sigma_x=\sum_j Z_j^x\,. \qquad (6)$$

A plot of α against the error in determining a phase angle is shown in Fig. 3. Except for very large values of α, the correlation between α and the correctness of a phase is quite weak, and it thus appears that α is not much better than E_c/E_{obs} as a measure of phase quality. The variance of φ_h, suggested as an index of phase correctness by Karle & Karle (1966), could be calculated easily from α and equation 3·33 of K & K, but because of the lack of correlation shown in Fig. 3, this was not done.

In the refinements of cytochrome *c* described below, the appearance of the heme group provided the best quality criterion. The progress of refinement was monitored by calculating an electron density map after every 100 or so phases had been refined with the tangent formula. The refinement of modelglobin was judged by comparing refined phases with exact phases.

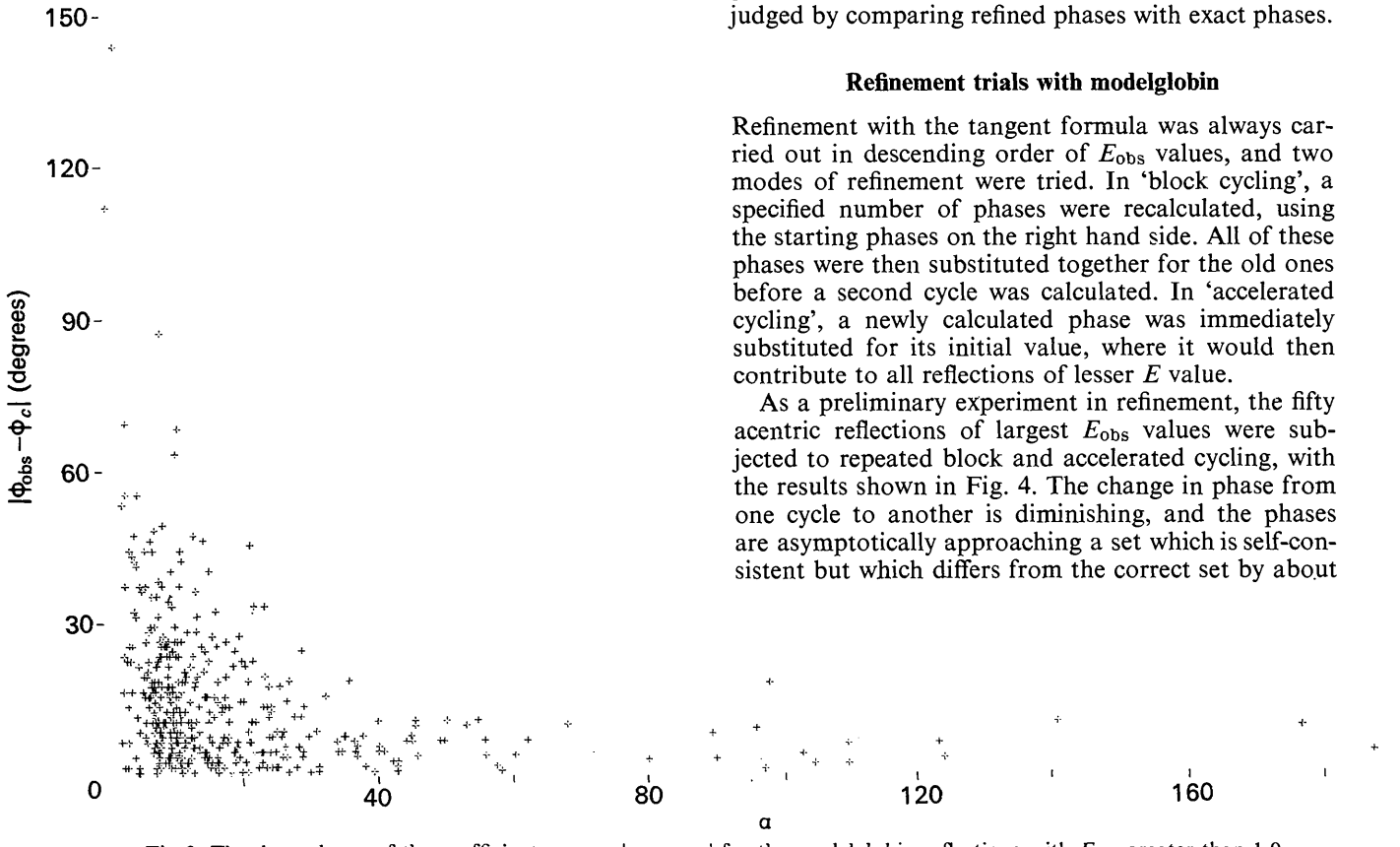

Fig. 3. The dependence of the coefficient α upon $|\varphi_c-\varphi_{obs}|$ for the modelglobin reflections with E_{ops} greater than 1·0.

Refinement trials with modelglobin

Refinement with the tangent formula was always carried out in descending order of E_{obs} values, and two modes of refinement were tried. In 'block cycling', a specified number of phases were recalculated, using the starting phases on the right hand side. All of these phases were then substituted together for the old ones before a second cycle was calculated. In 'accelerated cycling', a newly calculated phase was immediately substituted for its initial value, where it would then contribute to all reflections of lesser E value.

As a preliminary experiment in refinement, the fifty acentric reflections of largest E_{obs} values were subjected to repeated block and accelerated cycling, with the results shown in Fig. 4. The change in phase from one cycle to another is diminishing, and the phases are asymptotically approaching a set which is self-consistent but which differs from the correct set by about

twelve degrees. It may be that this difference between self-consistency and correctness is a consequence of the limited amount of data used, and that at high resolution this anomaly would not arise.

Two trials were made in which random errors were introduced into the calculated phase set, and the tangent formula was used to refine back again. In the first trial, errors in acentric reflections ranged from 0° to 90° with a mean of 32°, and the signs of two per cent of the centric reflections were reversed. The distribution of errors introduced into the acentric reflections is shown in Fig. 5. The refinement behavior of the 50 largest acentric reflections and then of the 147 largest acentric reflections when subjected to block and accelerated cycling is shown in Fig. 6. The mean phase error drops dramatically in the first cycle or two, levels off, and then begins to rise again with block cycling. Accelerated cycling reduces the mean error more rapidly, but also brings on the subsequent 'blowup' of phase angles sooner.

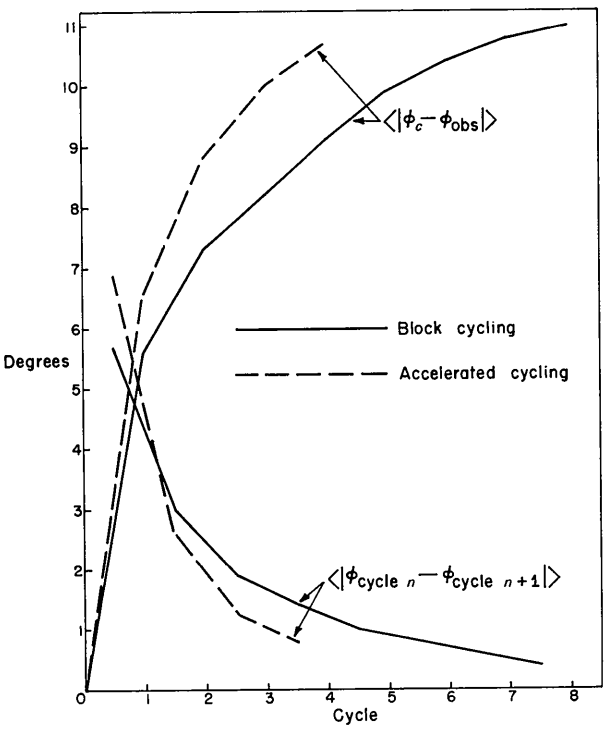

Fig. 4. Refinement behavior of modelglobin phases using the tangent formula, starting with the correct phases. Note that the apparent point of self-consistency differs from the true phase set by a mean of about 12°.

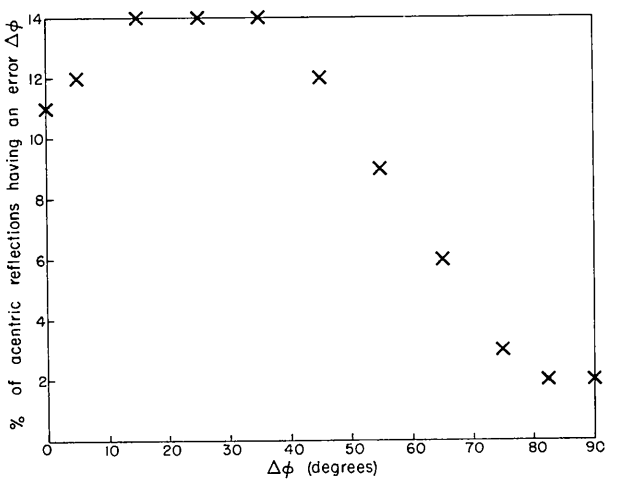

Fig. 5. Distribution of the random errors introduced into the model protein phases as a test tangent formula refinement.

Similar behavior was found in a second set of trials, in which the phases of the most intense 100 reflections were put in error by an average of 56° and the other phases were left as in the 32° trial. Accelerated cycling of the top 72 reflections reduced the mean error from 56° to 31°, then to 27·9° and 28·0° in two more cycles. With the greater initial error in phases, the onset of the drift toward the self-consistent set is delayed compared to the previous run.

The origin of this phase blowup upon continued cycling is not known. One possible explanation might be the one alluded to earlier: with limited data the self-consistent set may differ from the correct one. The rise in phase error after several cycles may be a drive toward self-consistency. This trend may also be present at the beginning, but be masked by improvement in phase angles which were initially badly in error. Only higher resolution data will resolve these questions.

The conclusion from these modelglobin trials is that refinement of MIR phases at low resolution should be carried out for two or three cycles for the strongest reflections and then halted before phase deterioration begins. If this is done, then a certain amount of phase improvement will result.

MIR phase refinement with cytochrome *c*

Unlike many proteins, cytochrome *c* has a prosthetic group, the heme, which is readily recognizable even at 4 Å resolution. The appearance of the heme group was used as a criterion of improvement during tangent formula refinement.

Two cytochrome derivatives were available, Pt and Hg, each with a different single binding site to the protein molecule. Pt was well-substituted, with a mean change in structure factor by the heavy atom of 29% in the centric *hk*0 zone of space group $P4_1$. Hg was weaker, with a mean change of only 15%. Further details about data, derivatives and MIR phase analysis will be found in Dickerson, Kopka, Borders, Varnum, Weinzierl & Margoliash (1967) and Dickerson, Kopka, Weinzierl, Eisenberg & Margoliash (1967).

Preparation of normalized structure factors, E_{obs}, presented a problem at low resolution. With the peaks in the *F* curve at about 10 Å and 5 Å, Wilson plot scaling was impossible. Instead, a temperature factor of

exp ($-0{\cdot}475S^2$) was assumed to be present by analogy with Kendrew's (1967) findings for myoglobin, and was removed by multiplying the data by exp ($0{\cdot}475S^2$). ($S=2\sin\theta$). The data were then scaled so that $\langle E_{obs}^2\rangle = 1{\cdot}0$ for the region from 8 Å to 4 Å. The inner reflections out to a resolution of 13 Å were removed because of their sensitivity to the salt concentration of the crystallizing medium.

Refinement of the MIR phases by accelerated cycling was tried first. After every 100 reflections, a Fourier map of the heme region was calculated. As the refinement proceeded, the iron atom gained in peak intensity in successive Fourier maps and the detail of the surrounding heme was steadily degraded. Block cycling was then tried, with much better results. The strongest 197 reflections (out of 1439 at 3·7 Å resolution) were given one cycle of block refinement. These 197 reflections were used with the unchanged phases of the 1242 reflections of lesser E_{obs} to calculate a much more encouraging map around the heme region. Refinement of the next 50 reflections did not appreciably change the map, nor did a second cycle of the top 100 reflections.

It appeared to make little difference whether or not the MIR figures of merit were used to weight the right hand side of the tangent formula at this point, and in the final calculations they were not used. The SIR work to follow, however, suggested that weighting with figures of merit is the proper thing to do. The refined 197 reflections were given a new figure of merit for the purpose of calculating electron density maps, defined as:

$$m' = \tfrac{1}{2} + \tfrac{1}{2}\cos\pi\left[m - \cos\left(\frac{|\Delta\varphi|}{2}\right)\right], \qquad (7)$$

where m is the old figure of merit and $|\Delta\varphi|$ is $|\varphi_{obs}-\varphi_c|$. The justification for this rather arbitrary function was that the weight should range between 0 and 1, and that a reflection should be given a large weight if it had had a large figure of merit and had not been altered appreciably by tangent refinement, or if it had had a small figure of merit and had been changed considerably. Non-refined reflections retained their old figures of merit. Unweighted maps calculated as controls showed the same general trends as weighted maps.

Of the several maps which were calculated, two are particularly informative: the original unrefined map weighted with figures of merit, and the refined map weighted with new figures of merit. These two maps are shown in Figs. 7 and 8, and 10 and 11, and a theoretical heme at several resolutions is in Fig. 9. Points of identification for the heme maps include the high density at the iron atom, the symmetrical extension of the two propionic acid side chains, the asymmetrical arrangement of the two thioether links to the protein, the polypeptide chain sequence: –Cys14–Ala15 –Gln16–Cys17–His18– running from the upper thioether link to the one at the right, and from there to the fifth iron coordination position on the near side of the heme, and the curved polypeptide chain running be-

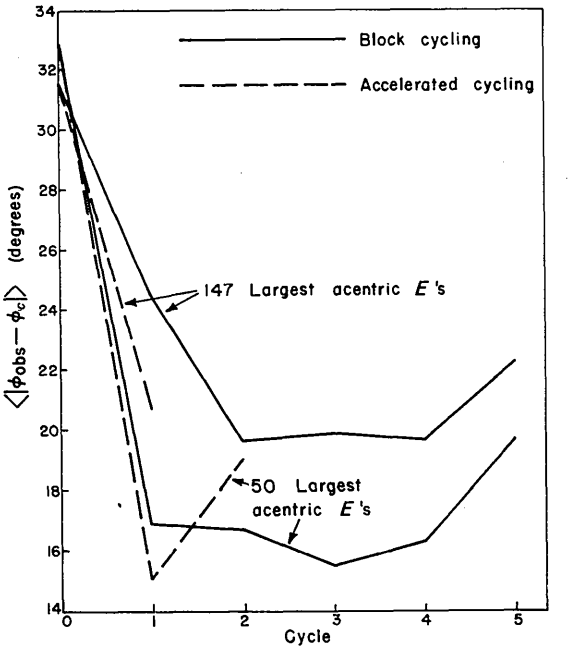

Fig. 6. Refinement behavior of the modelglobin phases after introduction of the errors of Fig. 5. Note the initial rapid return of phases toward the true model set, followed by their slower subsequent divergence.

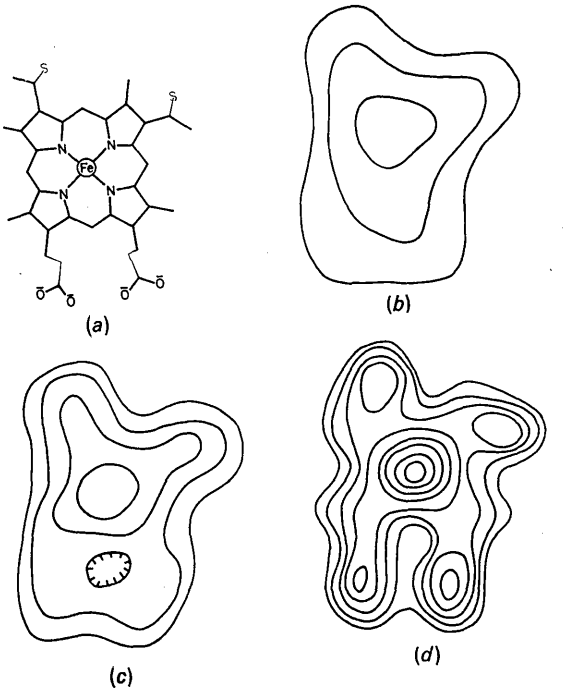

Fig. 9. The heme skeleton of cytochrome *c*, and its expected appearance at several resolutions. For each map, heme structure factors were calculated in the true cytochrome cell, the data were cut off at the desired resolution, and a Fourier map was calculated. The propionic acid side chains at the bottom were assumed here to be in their most extended conformation. (*a*) Heme skeleton. (*b*) 6 Å resoution. (*c*) 5 Å. (*d*) 3·7 Å.

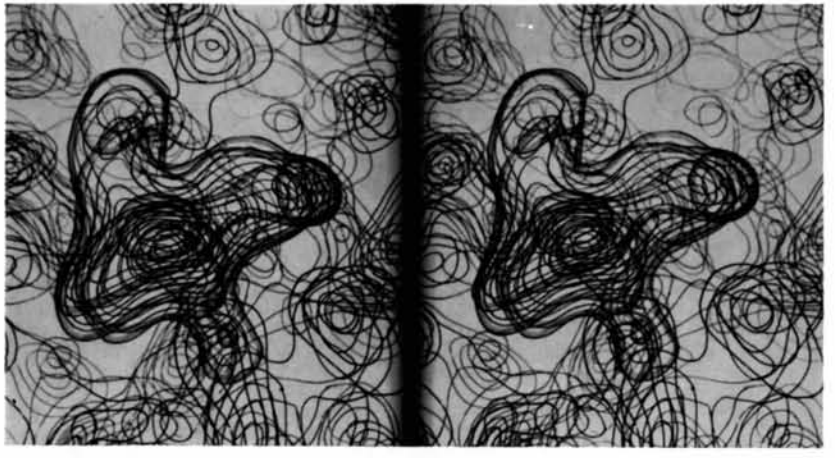

Fig. 7. The heme region of cytochrome *c*. This map is the result of a centroid phase analysis using Pt and Hg derivatives, and is weighted with figures of merit. See text for interpretation.

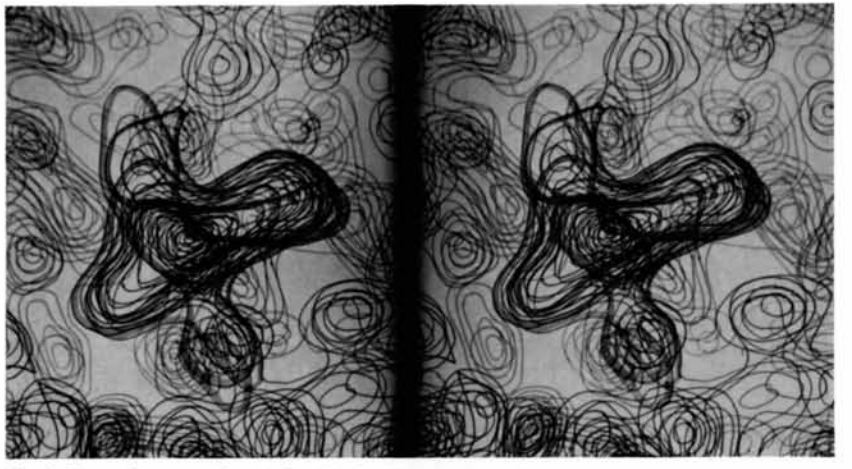

Fig. 8. Same heme region. After Karle–Hauptman refinement. Note the enhancement of the central iron atom.

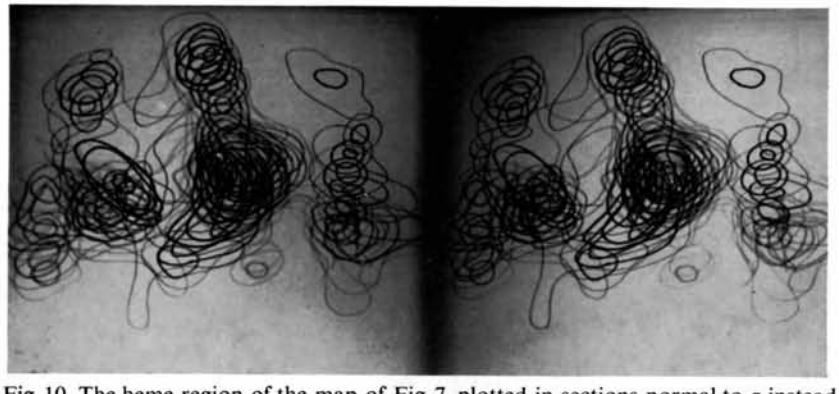

Fig. 10. The heme region of the map of Fig. 7, plotted in sections normal to *z* instead of to *x*, and viewed down from the top right of Fig. 7. The heme is seen nearly edge-on, with the Cys17 bridge coming out from the center and curving to the left. The Cys14 bridge rises in the back and curves forward and to the left through the peptide chain to join Cys17 and then to drop down to His18, which is coordinated to the heme iron from the left. The extended polypeptide chain which carries the sixth ligand curves upward past the heme plane on the right.

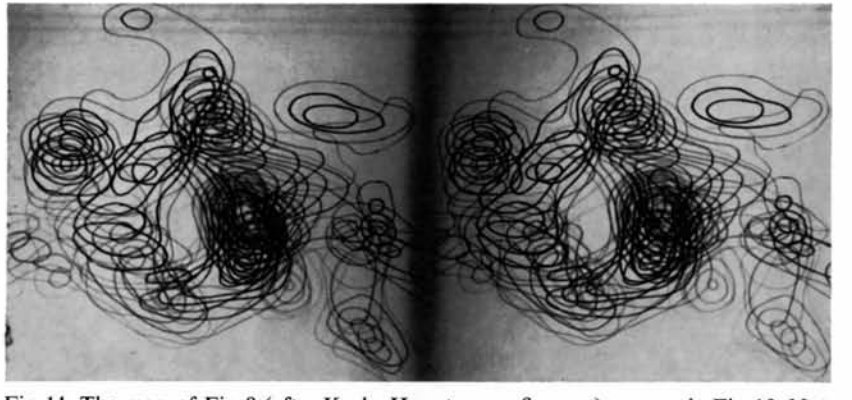

Fig. 11. The map of Fig. 8 (after Karle–Hauptman refinemnt), seen as in Fig. 10. Note the enhancement of density in the regions of the iron, the sulfurs of Cys14 and Cys17, and the sixth iron ligand on the right. Note also the strengthening of the polypeptide chain between Cys14 and Cys17, and the weakening of the extended chain on the right.

To face p. 384

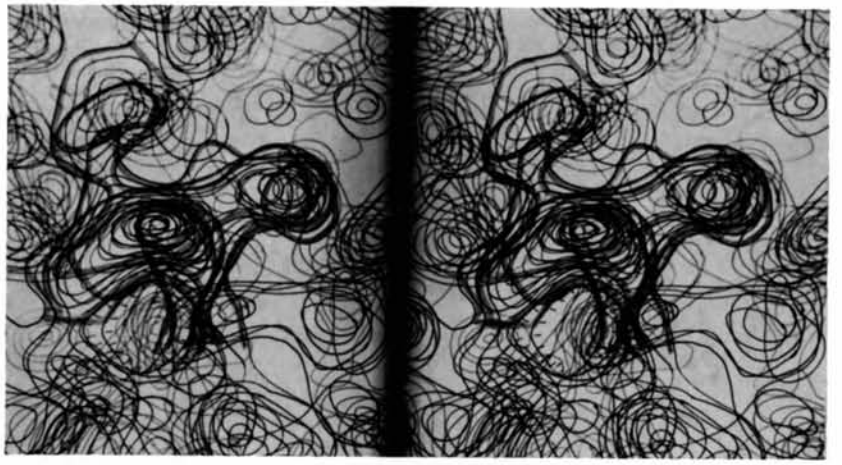

Fig. 14. Heme region of cytochrome *c* using Pt SIR phases and figures of merit.

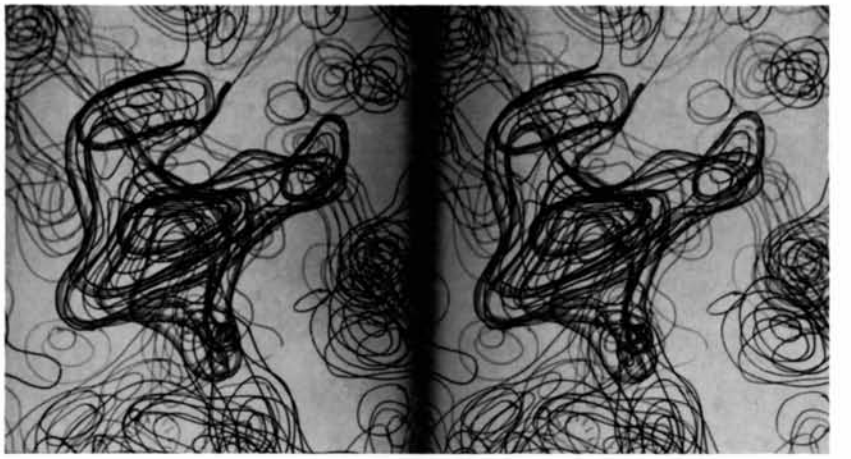

Fig. 15. Heme region using the Hg phase possibilities to choose between the two possible Pt SIR phases.

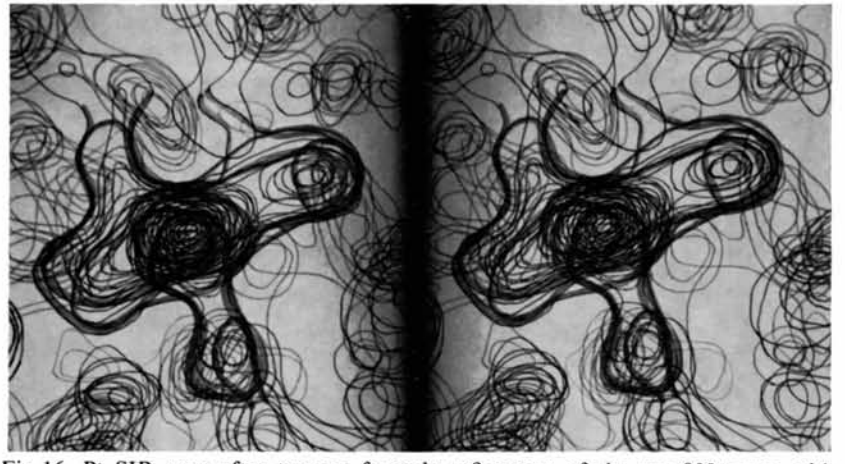

Fig. 16. Pt SIR map after tangent formula refinement of the top 800 terms, with accelerated cycling.

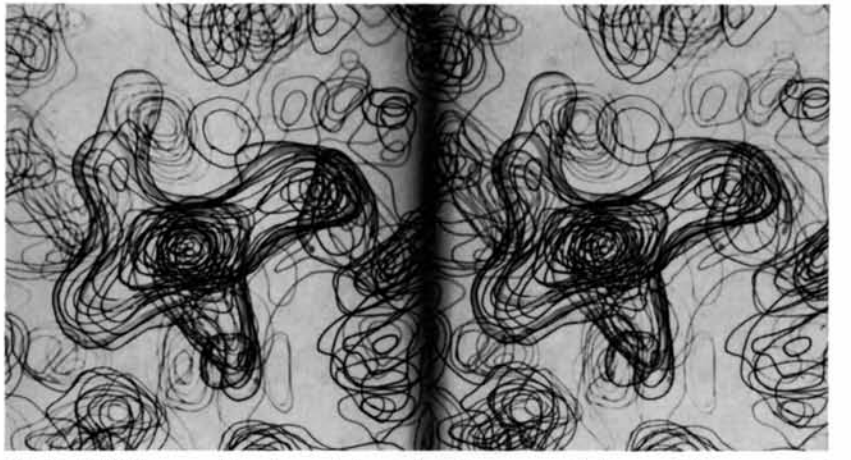

Fig. 17. Pt SIR map after tangent formula refinement of the top 800 terms, with block cycling.

hind the heme and bearing the side group which makes the sixth coordination with the iron.

In plan view, the heme is not drastically altered, other than in having more density at the iron and being generally more skeletal in appearance. But the view from the upper edge (Figs. 10 and 11) shows the differences more clearly. The density and connectedness of the polypeptide chain sequence between thioether links is greatly increased in the refined map. The two maps were put on the same scale by Wilson plots so that relative densities would be comparable. As a result of refinement, the peak iron density rose from 500 to 610 on an arbitrary scale, and the two sulfur sites became more consistent, that of Cys14 rising from 196 to 282 and that of Cys17 falling from 338 to 288. The density at the sixth coordination site, which has been thought on chemical grounds to be methionine, rose from 130 to 205. The only undesirable feature of the refined map is that the extended chain with the sixth ligand was weakened.

Resolution of the SIR phase ambiguity with the tangent formula

The results of the refinement of MIR phases were encouraging but not conclusive. It appeared that some improvement in the appearance of the heme in cytochrome had been produced. It was also apparent that working at low resolution meant working just this side of disaster, and that the method would be considerably more useful at high resolution.

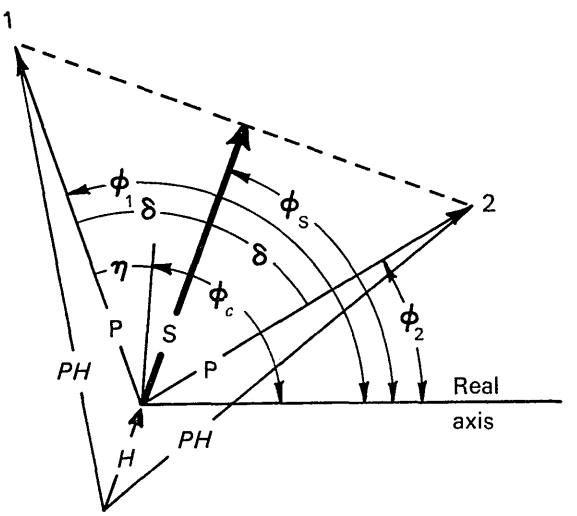

Fig. 12. Single isomorphous replacement phase diagram. Protein, heavy atom and derivative structure factor vectors are labeled *P*, *H* and *PH*, respectively. The two possible phase angles are φ_1 and φ_2. The SIR centroid vector (in the limit of perfectly sharp probability peaks) is labeled S and the SIR phase is φ_s. The phase calculated from the tangent formula is φ_c. (All φ angles are measured from the real axis.) δ measures the half angle of the SIR phase ambiguity, and η is the angle between φ_c and the nearer of the two SIR phase possibilities.

The Karle–Hauptman method could be more powerful at low resolution, paradoxically, if it were asked to do less. Instead of being used to calculate a new phase angle, could it usefully be employed only to indicate *which* of the two possible phases in a SIR analysis was the correct one?

The principles of SIR phasing are illustrated in Fig. 12. The use of the centroid vector S is equivalent to choosing half of each of the phase alternatives. The map will then be half correct density and half incoherent noise. Any means of choosing phase alternates would greatly improve the map. If the bimodal probability curve around φ were perfectly sharp, then the centroid figure of merit would be equal to $|\cos \delta|$. A traditional heavy atom Fourier synthesis of the protein would be built up from coefficients: $F_P \exp(i \varphi_H)$. The SIR map coefficients can be thought of either as: $mF_P \exp(i \varphi_H)$, or as: $|m|F_P \exp(i \varphi_S)$, where $m = \cos \delta$ and is positive if δ is less than 90° or negative if it is greater. In this way of thinking, the function of the addition of parent compound data to an isomorphous heavy atom derivative is to provide the proper weighting factors for the heavy atom Fourier synthesis.

The tangent formula can be used to calculate phase angles, φ_c, using SIR phases weighted with figures of merit on the right hand side of the equation. If it is assumed that the SIR method is accurate but ambiguous, and the Karle–Hauptman method is definite but inaccurate, then the best strategy is to choose the SIR phase which lies nearest to φ_c.

The problem of weighting factors to replace the SIR figures of merit remains. But if δ is small, then it matters little whether the Karle–Hauptman phase is indicative or not. On the other hand, if φ_c falls very near to one phase choice and η is small, then it is irrelevant how large the SIR phase ambiguity initially was. The new figure of merit in the cytochrome experiments was therefore chosen to be either $\cos \delta$ or $\cos \eta$, whichever was greater.

Test of the phase choice method with the model protein

In the tests of the method with modelglobin, SIR phases were generated for a hypothetical two-site derivative having mercuric ions at (0·275, 0·250, 0·142) and (0·400, 0·412, 0·485) and the symmetry related positions of the myoglobin unit cell.* These are actually the PCMBS and Au positions from sperm whale myoglobin (Bodo, Dintzis, Kendrew & Wyckoff, 1959). The SIR phases and $\cos \delta$ weights were used in the tangent formula to calculate a new set of phases, φ_c, in order of decreasing E_{obs}. Each φ_c was used immediately upon calculation to choose between the SIR

* It should be noted that this method of SIR phase choice cannot be used if the heavy atom cluster is centrosymmetric, for then there is no way of driving φ_c away from the real axis. This situation arises, for example, in space group $P2_1$ with a single site derivative, but not in $P4_1$. A method of circumventing this difficulty has been suggested by Karle (1966).

phase alternatives, and this chosen phase was substituted for the initial SIR value in a variation of accelerated cycling.

Accelerated cycling was used in spite of our experience with MIR refinement because of the way in which the tangent formula phases were used. It was hoped that the additional factor, that the phases actually used were from isomorphous replacement, would keep the refinement out of the self-consistency trap and would make phasing meaningful down to much lower E_{obs} values than before.

Fig. 13 shows the results of accelerated SIR refinement of the 700 reflections of largest E_{obs} in the 4 Å modelglobin data set. Although they include only half of the reflections, they include about 85% of the total E_{obs} sum. The angle between a SIR phase and one of its possible phase choices can vary from 0° to 90°, so the SIR set before refinement will be something like an average of 45° in error. After one accelerated pass, this error was reduced to a mean of 14·5° for the top 600 reflections and 35° for the next 100. The second accelerated cycle dropped the mean error by another 3·2°.

Expectations that the SIR choice procedure would be valid for a much greater range of E than the MIR refinement seem to be amply borne out. The modelglobin trials, of course, tell nothing about the effect of experimental errors on the refinement process; this aspect is best studied with real data.

Application of SIR/KH phasing to cytochrome *c*

Since the electron density in the heme region of cytochrome *c* was to be used as the quality criterion of tangent formula phases, several auxiliary maps of the heme region were first computed as controls. In Fig. 14 is shown the Pt SIR map, which should be compared with the unrefined MIR map of Fig. 7. The heme in Fig. 14 is clearly visible, but is badly distorted. The center of the iron (peak density 450 rather than 500) is no longer the highest point in the map, the heme is emaciated in appearance, the Cys connections are erratic and the propionic acid side chains at the bottom are both continuous with a mass of density below the heme. The extended chain on the far side of the heme is wiped out. The equivalent Hg SIR map reduces the heme to little more than a flattened blob, although the extended chain in back is visible. Fig. 15 shows a map produced not by incorporating Hg data into the centroid phase analysis as in Fig. 7, but by merely using the Hg phase intersections to choose between the two Pt possibilities. The heme is very little improved over the Pt SIR map except for better definition of the propionic acid groups. The extended chain is broken but is present.

The success of accelerated cycling in resolving the SIR ambiguity in modelglobin led us to apply the same method to the Pt SIR results for cytochrome *c*. Reflections 1–450 were given two cycles of refinement, then reflections 451–800 were given one cycle, and finally reflections 1–100 were given a third cycle. In computing the electron density map, the $\cos\delta - \cos\eta$ figures of merit were used, and the unrefined reflections of low E and resolution lower than 13 Å were included with their SIR phases and figures of merit. The results, shown in Fig. 16, were disappointing. The heme was distorted, its upper left corner was nearly wiped out, and the density of the iron position was increased from 500 on an arbitrary scale in the MIR map of Fig. 7 to 750.

Block cycling of the top 800 reflections was then tried instead of accelerated cycling, and as in the MIR

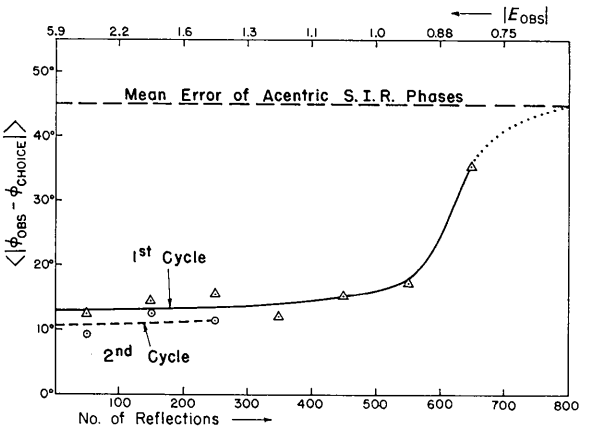

Fig. 13. Refinement behavior of the modelglobin acentric SIR phases. The first accelerated cycle removes nearly 70% of the error contained in the acentric phases associated with the 600 largest E's; the second cycle makes a small additional improvement.

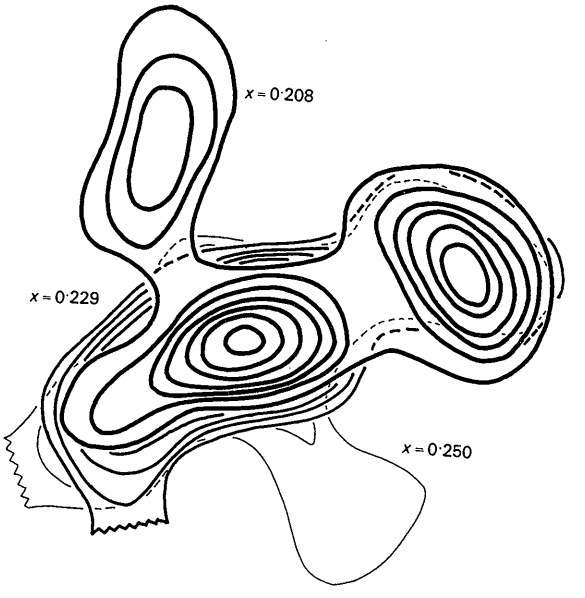

Fig. 18. Map produced by using the joint probability distribution obtained from Pt SIR and tangent formula information.

refinements, it was much more successful. The resulting map is shown in Fig. 17; in contrast to Fig. 16, it is far superior to the Pt SIR map and is almost as good as the complete MIR map. The electron density at the iron has increased only to 600, and the features around the periphery of the heme are clear and well-formed. A final pass of block cycling through the top 100 reflections made no further improvement in the heme, and retained a peak density of 600 at the iron.

One final use was made of the Karle–Hauptman method. The angle φ_c and the coefficient α were calculated from the tangent formula and then used in equation (4) with the pre-exponential Bessel function omitted to calculate the relative Karle–Hauptman probabilities of phase angle φ' around the phase circle. This probability curve and the normal Blow-Crick SIR phase probability curve were then multiplied to form the joint probability, and the centroid of this probability function found in the normal way. After block refinement of the top 800 terms, φ_c and α were fed into a modified phase program and centroid phases and figures of merit were produced. It was hoped that this would be the best way to merge isomorphous and statistical phase information, but in fact the map produced was the unfavorable one shown in Fig. 18. The density at the iron increased to 950, there was a termination-of-series ripple around the iron, and the map was closer to Fig. 16 than to Fig. 17.

Conclusions

These experiments show that the Karle–Hauptman tangent formula is useful in several ways in refining protein phases. Once phases have been obtained from a low resolution MIR study, those for the strongest reflections (which are usually among the least well-determined phases) can be improved by one or two block cycles with the tangent formula. Further block cycling or accelerated cycling tends to degrade the map, and to produce a self-consistent but wrong phase set which piles additional density on the strongest features of the map. This difficulty may arise from the constraints imposed by the tangent formula itself, such as the assumptions of point atoms and non-negative electron density, and may also be related to the relatively small amount of data at low resolution.

The use of the tangent formula to refine high resolution protein data has not been explored yet, but a comparison of results at 5 Å and at 4 Å suggests that a much greater fraction of the E's can be refined at higher resolution. At high resolution, the coefficient α for a reflection may be more strongly correlated with the correctness of its phase angle, and it may then be more appropriate to incorporate the Karle–Hauptman results directly into the phase program as outlined in the previous section.

The tangent formula is more effective in the SIR application because it is asked to choose between two isomorphous phases and not to predict a numerical value. Our trials with modelglobin show that under favorable conditions this method can break the SIR ambiguity for the majority of reflections, even at low resolution. In a situation where there is one excellent single-site derivative of good isomorphism out to high resolution, and a collection of other poorer multiple-site derivatives, the wisest strategy may be to throw out the poor derivatives, recollect parent and single-site data again and again to eliminate as much experimental error as possible, and then to use Karle–Hauptman methods to resolve the phase ambiguity.

We should like to express our appreciation to Dr Richard H. Stanford for the seminar which initially triggered our interest in Karle–Hauptman methods, and for discussions at all stages of the work. We should also like to thank Drs J. Karle and I. Karle for their helpful comments and suggestions during the A.C.A. meeting in Tucson.

This work was performed under the auspices of National Science Foundation research grant GB-3053, the help of which is greatly appreciated. One of us (JEW) is also the holder of a National Institutes of Health Predoctoral Traineeship (GM-1262).

References

BLOW, D. M. & ROSSMANN, M. G. (1961). *Acta Cryst.* **14**, 1195.

BODO, G., DINTZIS, H. M., KENDREW, J. C. & WYCKOFF, H. W. (1959). *Proc. Roy. Soc.* A**253**, 70.

COULTER, C. L. (1965). *J. Mol. Biol.* **12**, 292.

DICKERSON, R. E., KOPKA, M. L., BORDERS, C. L., JR, VARNUM, J. C., WEINZIERL, J. E. & MARGOLIASH, E. (1967). *J. Mol. Biol.* **29**, 77.

DICKERSON, R. E., KOPKA, M. L., WEINZIERL, J. E., EISENBERG, D. & MARGOLIASH, E. (1967). *J. Biol. Chem.* **242**, 3015.

KARLE, J. (1964). In *Advances in Structure Research by Diffraction Methods*, Vol.I, R.BRILL, editor. New York: Interscience.

KARLE, J. (1966). *Acta Cryst.* **21**, 273.

KARLE, J. & KARLE, I. (1966). *Acta Cryst.* **21**, 849.

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

KENDREW, J. C. (1967). Private communication.