*Acta Cryst.* (1971). B27, 2402

# The Determination and Refinement of Heavy-Atom Parameters in Protein Heavy-Atom Derivatives. Some Model Calculations Using Acentric Reflexions

BY ELEANOR DODSON AND M. VIJAYAN*

*Laboratory of Molecular Biophysics and Chemical Crystallography Laboratory, South Parks Road, Oxford*

(*Received* 4 *January* 1971)

Some problems associated with the determination and refinement of heavy-atom parameters in protein derivatives using totally acentric data are critically examined with the aid of calculations on model systems. The difference Patterson ($\Delta F^2$ Patterson) synthesis is found to be a good approximation to the heavy-atom vector map even when the substitution is high. A theoretical expansion for the coefficients of the conventional difference Fourier synthesis is given. The quality of the difference map decreases rapidly with increasing heavy-atom substitution. Theoretical considerations and model calculations show that the background noise closely resembles the features of the protein Fourier map. Statistical errors in the X-ray data are likely to lead to a systematic overestimation of the occupancy factors when the heavy-atom parameters are refined by the least-squares method against the heavy-atom contributions estimated from isomorphous and anomalous differences. Refined occupancy factors reasonably close to their true values can, however, be obtained either by using empirically evaluated values of $k$ (the ratio between the real and imaginary parts of the heavy-atom form factor) in estimating the heavy-atom contribution or by applying suitable weighting functions to the terms in the minimization function. Some aspects of the least-squares method based on the minimization of the sum of squares of the lack-of-closure errors are also discussed.

## Introduction

The multiple isomorphous-replacement method has been successfully used, often in combination with anomalous-scattering data, in the solution of protein structures. In the application of this method, the first step in the crystallographic analysis is the determination of the heavy-atom sites in different heavy-atom derivatives. In the absence of any phase information, one most often uses a difference Patterson ($\Delta F^2$ Patterson) synthesis for this purpose. However, once a set of phase angles for the protein structure factors is available from one or more heavy-atom derivatives, the heavy-atom positions in yet another derivative can be obtained more easily from a difference Fourier synthesis. As is well known, difference Patterson and difference Fourier techniques are very powerful in centrosymmetric projections. The nature and the extent of the limitations introduced in these techniques when they are used with acentric reflexions are discussed in the first part of this paper. After determining the approximate locations of the heavy atom sites in various derivatives one refines the positional and thermal parameters and the occupancy factors of these sites. Least-squares methods with different minimization functions have been tried by various workers. In the procedure most commonly used, which we call 'phase refinement', one minimizes the sum of squares of the lack-of-closure errors, *viz.*,

$$\sum w(F_{HP} - |\mathbf{F}_P + \mathbf{F}_H|)^2$$

where $F_{HP}$ is the magnitude of the structure factor of the heavy atom derivative, $\mathbf{F}_P$ is the structure factor of the native protein and $\mathbf{F}_H$ is the contribution from the heavy atom alone (Dickerson, Kendrew & Strandberg, 1961; Muirhead, Cox, Mazzarella & Perutz, 1967; Dickerson, Weinzierl & Palmer, 1968). In practice, one carries through alternate cycles of phase refinement and phase angle calculations until convergence is reached. An alternative procedure, which we call $F_{HLE}$ refinement for reasons discussed later, employed by some workers (Kartha, 1965; Adams, 1968) makes use of the values of the magnitudes of the heavy atom contribution estimated from isomorphous and anomalous differences. This method has the advantage that is does not rely on the phase angles and hence can be carried out for each derivative independent of all others. However, in practice, experimental errors in the anomalous differences result in systematic errors in the estimated values of heavy atom contributions. These give rise to some serious problems, especially

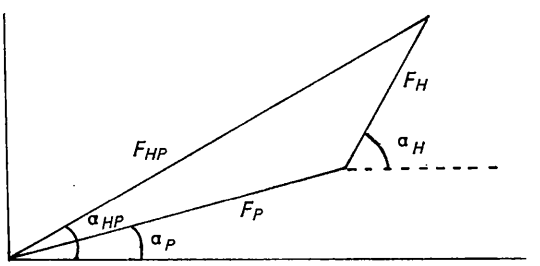


Fig. 1. Vector diagram showing the relation between $\mathbf{F}_P$, $\mathbf{F}_{HP}$ and $\mathbf{F}_H$.

* Present address: Department of Physics, Indian Institute of Science, Bangalore-12, India.

in the refinement of occupancy factors. Such problems are discussed in the latter part of this paper. Some aspects of phase refinement are also discussed.

These investigations resulted mainly from our experience with the X-ray structure analysis of insulin and, hence, all the model calculations were performed using data from rhombohedral 2Zn insulin crystals with space group $R3$ ($a=82{\cdot}5$, $c=34{\cdot}0$ Å; three insulin hexamers in the triply primitive unit cell). However, the results are applicable to other non-centrosymmetric space groups and are of general validity.

## Difference Patterson and difference Fourier syntheses

### *Theory*

Theoretical expansions for the coefficients of the difference Patterson synthesis have been derived by various workers at different levels of approximation. An exact expression has been given by Phillips (1966) following the approach suggested by the methods of Ramachandran and his colleagues (Ramachandran & Raman, 1959). Referring to Fig. 1,

$$F_{HP}=F_P\cos(\alpha_P-\alpha_{HP})+F_H\cos(\alpha_H-\alpha_{HP})$$

and

$$F_{HP}-F_P=-2F_P\sin^2\beta/2+F_H\cos\gamma$$

where

$$\beta=\alpha_P-\alpha_{HP},\ \gamma=\alpha_H-\alpha_{HP}.$$

Thus

$$(F_{HP}-F_P)^2=4F_P^2\sin^4\beta/2+F_H^2\cos^2\gamma-4F_PF_H\sin^2\beta/2\cos\gamma\ . \qquad (1)$$

When $F_H$ is small compared to $F_P$, $\beta$ may be expected to be small in most cases and $\gamma$ may reasonably be assumed to have random values; the first and the third terms in expression (1) would, then, be small and the second term would give rise to heavy-atom vector peaks with half their normal heights. In the general case, the three terms in (1) represent convolutions with no precisely predictable features though the contributions from the first and third terms may be expected to be relatively small even when $F_H$ is large on account of the multiplicative factors $\sin^4\beta/2$ and $\sin^2\beta/2\cos\gamma$ respectively. However, for given values of $F_P$, $|\gamma|$ tends to decrease as $F_H$ increases and consequently the average height of the heavy-atom vector peaks tends to increase with increasing heavy-atom substitution.

A useful expansion for the coefficients of the difference Fourier synthesis may be derived as follows. Again referring to Fig. 1,

$$F_{HP}=[F_P^2+F_H^2+2F_PF_H\cos(\alpha_H-\alpha_P)]^{1/2}$$

$$=(F_P+F_H)\left[1+\frac{2F_PF_H}{(F_P+F_H)^2}(\cos(\alpha_H-\alpha_P)-1)\right]^{1/2}.$$

Now

$$\left|\frac{2F_PF_H}{(F_P+F_H)^2}(\cos(\alpha_H-\alpha_P)-1)\right|\leq 1$$

and using $(1+x)^{1/2}\simeq 1+\frac{1}{2}x$ when $x\leq 1$,

$$F_{HP}\simeq F_P+F_H^2/(F_H+F_P)+\frac{F_PF_H}{F_P+F_H}\cos(\alpha_H-\alpha_P)\ . \qquad (2)$$

But

$$\cos(\alpha_H-\alpha_P)=\tfrac{1}{2}[\exp i(\alpha_H-\alpha_P)+\exp\{-i(\alpha_H-\alpha_P)\}]$$

and therefore

$$(F_{HP}-F_P)\ .\ \exp(i\alpha_P)\simeq \underbrace{\frac{F_H^2\ .\ \exp(i\alpha_P)}{F_H+F_P}}_{\text{(I)}}$$

$$+\underbrace{\frac{F_P\ .\ F_H\ .\ \exp(i\alpha_H)}{2(F_H+F_P)}}_{\text{(II)}}+\underbrace{\frac{F_P\ .\ F_H\ .\ \exp i(2\alpha_P-\alpha_H)}{2(F_H+F_P)}}_{\text{(III)}}\ . \qquad (3)$$

The validity of the approximation involved in (2) has been tested by computing Fourier maps using the sum of the terms (I), (II), (III), and comparing these with the corresponding difference Fourier maps (see later). In expression (3), term (II) will give rise to a weighted heavy atom map and term (III) background noise as $(2\alpha_P-\alpha_H)$ has random values. Although the amplitude of (I) is not related to $F_P$, this term will give rise to some features of the protein map since the phase factor corresponds to that of the structure factor of the native protein crystal (see Srinivasan, 1961). When $F_P$ is much greater than $F_H$, the contribution of the first term to the synthesis is very small and the heavy atom peaks appear approximately at half their normal heights. However, for given values of $F_P$, the first term increases and the second term decreases, relative to $F_H$ as $F_H$ increases. This means that the difference Fourier map more and more tends to resemble a highly distorted protein map and the height of the heavy-atom peaks progressively decreases. The problem is more serious in two-dimensional projections on account of the overlap of the distorted protein map. However, whether in projection or in three dimensions, for acentric reflexions, conventional difference Fourier maps are poor approximations to heavy-atom maps for highly substituted derivatives.

### *Model calculations*

Some model calculations were carried out to verify the theoretical deductions described above. Some model derivatives for rhombohedral 2Zn insulin were constructed and the $F_{HP}$ values for each of them were calculated using the relation

$$F_{HP}=|F_P\ .\ \exp(i\alpha_P)+F_H\ .\ \exp(i\alpha_H)|$$

where $\alpha_P$'s are the protein phases derived from isomorphous and anomalous data from Zn free 0·01 *M* lead acetate derivative, and $F_H$'s and $\alpha_H$'s are the heavy-atom amplitudes and phases respectively calculated for mercury atoms placed at appropriate positions. Care was taken to see that the heavy atoms were placed at positions with zero density in the protein map. Altogether six model derivatives were constructed: model (I) has one mercury atom per asymmetric

part, model (II) has two mercury atoms *etc.* The positional coordinates of the heavy atoms in the six models are given in Table 1. The temperature factor assigned to the heavy atoms, 12 Å$^2$, was the same as that of the protein. It should be mentioned that from a theoretical point of view the protein phases can be chosen arbitrarily. The phases used were those available at the time when the calculations were actually carried out.

Table 1. *Heavy-atom positions in model derivatives used in difference Fourier and difference Patterson calculations*

| Site | $x$ | $y$ | $z$ |
|---|---|---|---|
| *A* | 0·0667 | 0·1833 | 0·3000 |
| *B* | 0·1667 | 0·2500 | 0·4400 |
| *C* | 0·1667 | 0·1667 | 0 |
| *D* | 0·2667 | 0·2500 | 0·2400 |
| *E* | 0·3000 | 0·0333 | 0·7200 |
| *F* | 0·0667 | 0·2833 | 0·9000 |

Model (I): Site *A*.
Model (II): Sites *A* and *B*.
Model (III): Sites *A*, *B* and *C*.
Model (IV): Sites *A*, *B*, *C* and *D*.
Model (V): Sites *A*, *B*, *C*, *D* and *E*.
Model (VI): Sites *A*, *B*, *C*, *D*, *E* and *F*.

Difference Patterson projections at 2·8 Å resolution using *hki*0 zonal data and three-dimensional syntheses at 4·5 Å resolution were computed with the usual coefficients $(F_{HP}-F_P)^2$ for all the models. The projection maps for models (I) to (IV) are given in Fig. 2(*a*) to (*d*). The expected heavy atom vector peaks are also indicated. In all four maps the peaks clearly stand out. There is no systematic diminution of the heights of these peaks with increasing heavy-atom substitution; on the contrary, the average peak height tends to go up as substitution increases. The same conclusions were arrived at on the examination of the three-dimensional maps. It should, however, be emphasized that the probable errors in data and scaling and the effects of non-isomorphism are not taken into account in these calculations. Further, if the heavy-atom substitution pattern consists of a combination of major and minor sites, it is very probable that the vector peaks involving the latter are obscured by the background resulting from the first and the third terms in equation (1). In general, and especially in situations like the one mentioned above, a Patterson map with the heavy-atom contributions estimated from isomorphous and anomalous data as coefficients would, obviously, be superior to a difference Patterson synthesis in defining the heavy-atom vector peaks (Kartha & Parthasarthy, 1965; Matthews, 1966).

Subsequently, difference Fourier maps were computed at 2·8 Å resolution in the *hki*0 projection for each model with the coefficients $(F_{HP}-F_P)\,.\,\exp(i\alpha_P)$. The protein phases, $\alpha_P$'s, were the same as those used for calculating the $F_{HP}$ values, which implies the assumption that the phases contain no errors. The maps

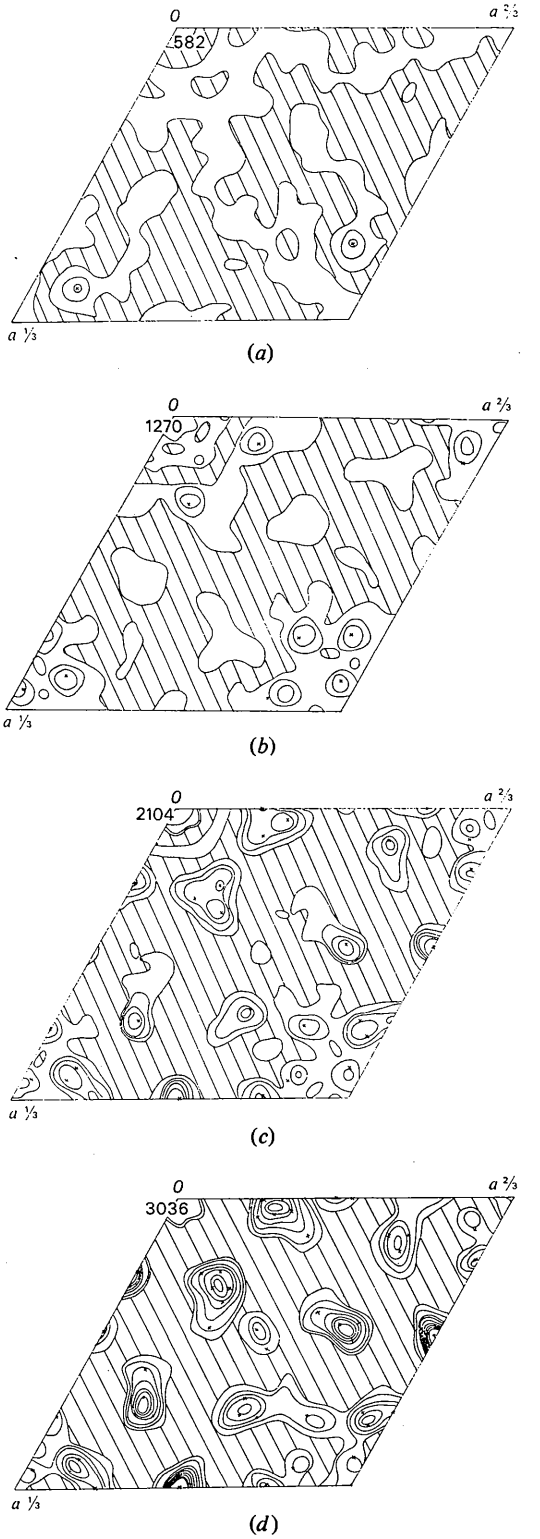


Fig. 2. *hki*0 difference Patterson projections corresponding to (*a*) model (I), (*b*) model (II), (*c*) model (III) and (*d*) model (IV). Contours are drawn at arbitrary intervals of 0, 100, 200 *etc.* The heights of the origin peaks are given. The crosses indicate the positions of the expected vector peaks.

corresponding to the first four models are shown in Fig. 3(*a*) to (*d*). In the map corresponding to model (I), with one heavy atom, the heavy-atom peak clearly stands out. The map corresponding to model (II) is interpretable though some of the background peaks are nearly half as high as the heavy-atom peaks. Without prior knowledge of the atomic positions these might be erroneously interpreted as minor heavy-atom sites. When the number of heavy atoms is more than two, as in models (III) and (IV), the difference Fourier maps clearly become misleading. As can be seen from Fig. 3(*c*) and (*d*), some of the spurious peaks are higher than the heavy-atom peaks.

The protein Fourier map is shown in Fig. 3(*e*). As expected from the theory, some of the spurious features that develop in the difference Fourier maps, as the number of heavy atoms increases, bear a close resemblance to the features in the protein map. Finally, the sum of the Fourier maps of terms (I), (II) and (III) in equation (3) is shown in Fig. 3(*f*), corresponding

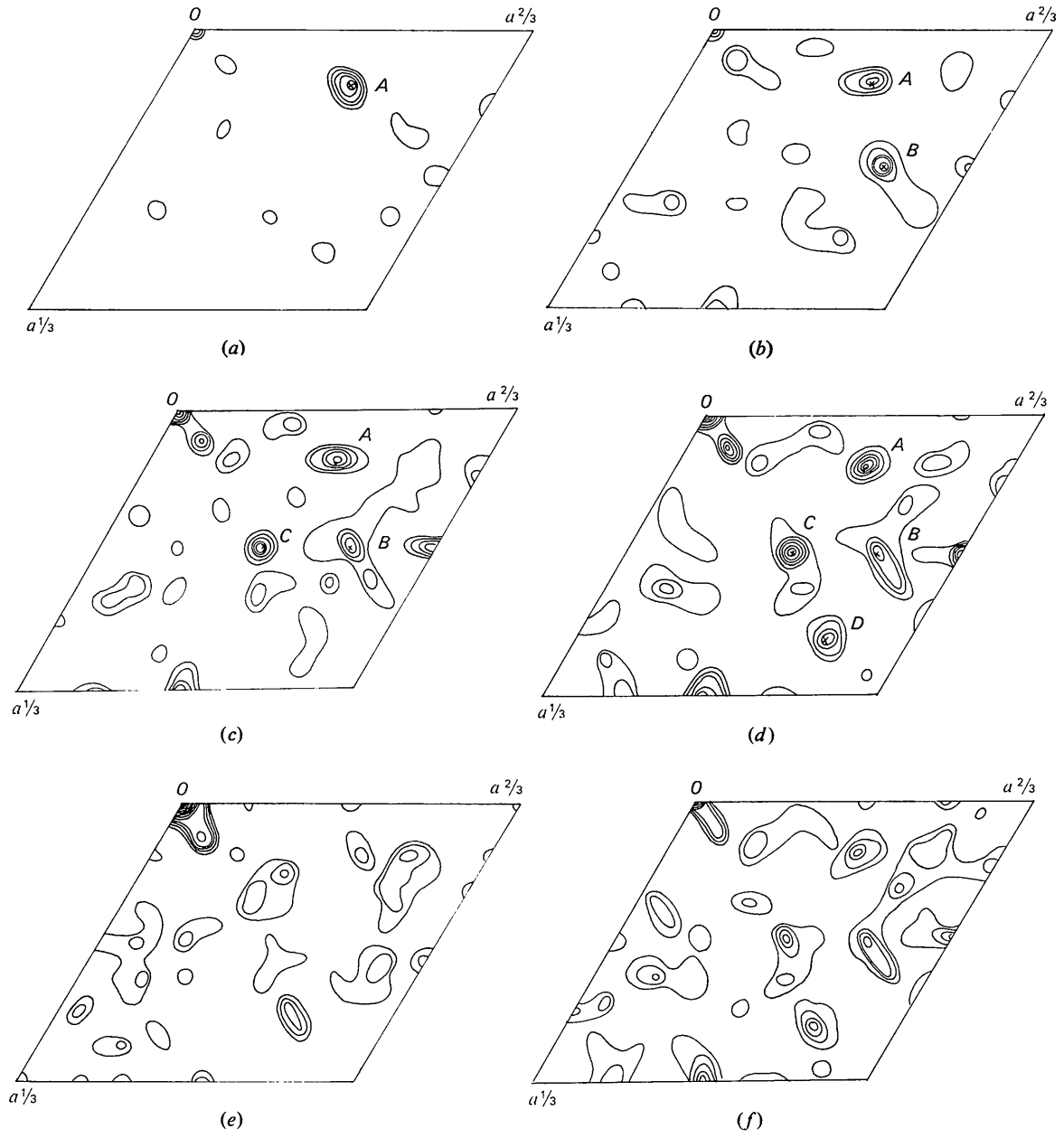


Fig. 3. *hki*0 difference Fourier projections corresponding to (*a*) model (I), (*b*) model (II), (*c*) model (III) and (*d*) model (VI); (*e*) *hki*0 Fourier projection; (*f*) *hki*0 Fourier projection with the sum of (I), (II) and (III) in equation (3) as coefficients. The contours are at arbitrary intervals of 100, 200 *etc.* The heavy-atom positions are indicated.

to model (IV). This map is clearly very similar to the difference Fourier synthesis for model (IV), thus justifying the approximation made in equation (2).

Some of these calculations were repeated using three-dimensional data with $d > 4{\cdot}5$ Å. Superimposed sections of the difference Fourier syntheses for models (II), (IV) and (VI) are shown in Fig. 4(*a*), (*b*) and (*c*). Superimposed sections of the protein Fourier synthesis are given in Fig. 4(*d*). The progressive deterioration of the quality of the map with increasing heavy atom substitution can clearly be seen from the comparison of the maps. The heights of the heavy atom peaks and the number of spurious peaks in the maps are given in Table 2. The heights of the heavy atom peaks decrease and the number and heights of the spurious peaks increase as the number of heavy atoms in the derivative increases. Also one notices that the positions of many spurious peaks correspond to the regions of high electron density in the protein map. However, the number of mercury atoms that could be introduced before the spurious peaks become as high as, or higher than, the heavy atom peaks in the 2·8 Å resolution projection difference Fourier maps is two, whereas the corresponding number in the three-dimensional maps at 4·5 Å resolution is four. To that extent it is advantageous to work in three dimensions compared to two-dimensional projections.

Table 2. *The peak heights, on arbitrary scale, of heavy atom sites in three-dimensional difference Fourier syntheses*

| Site | Model (II) | Model (IV) | Model (VI) |
|---|---|---|---|
| *A* | 406 | 301 | 234 |
| *B* | 411 | 404 | 342 |
| *C* | | 443 | 416 |
| *D* | | 374 | 291 |
| *E* | | | 220 |
| *F* | | | 143 |
| Number of spurious peaks greater than | | | |
| 200 | 1 | 6 | 22 |
| 300 | 0 | 1 | 5 |
| 400 | 0 | 0 | 3 |
| 500 | 0 | 0 | 1 |

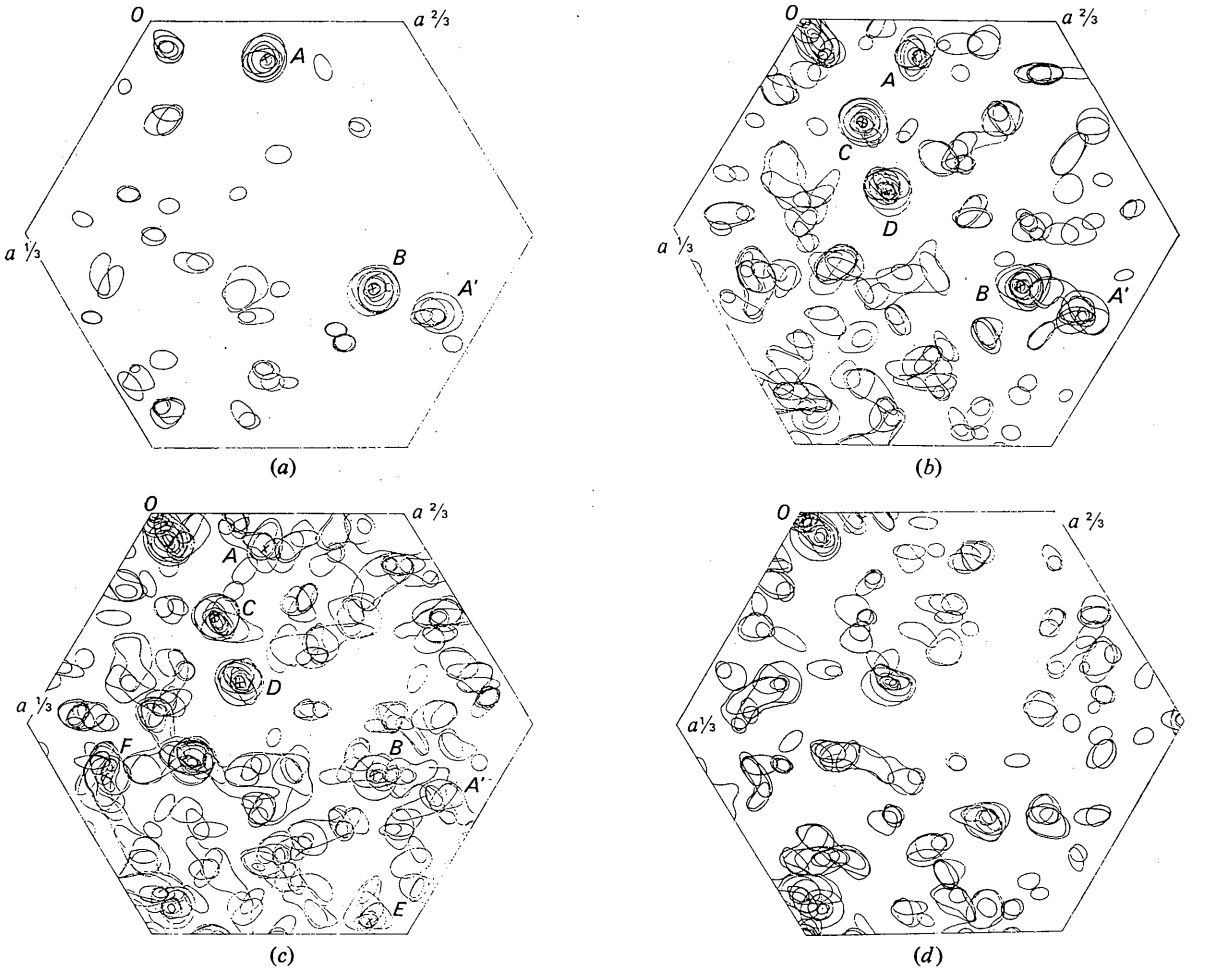


Fig. 4. Superimposed sections of the three-dimensional difference Fourier syntheses corresponding to (*a*) model (II), (*b*) model (IV) and (*c*) model (VI); (*d*) superimposed sections of the three-dimensional Fourier syntheses. Contours are at arbitrary intervals of 100, 200 *etc.* The heavy atom positions are indicated.

It may be noted that in the above calculations the protein phases are assumed to contain no errors. Also, the protein and the heavy-atom derivatives are assumed to be perfectly isomorphous. Errors in phase angles and non-isomorphism would obviously render the difference-Fourier technique still less effective when the substitution is heavy.

### Refinement of heavy atom parameters

*$F_{HLE}$ refinement*

As mentioned in the Introduction, the heavy atom parameters of each derivative can be refined independently of all other derivatives using the values of $F_H$ estimated from isomorphous and anomalous differences. It has been shown that, in the presence of anomalous scattering, the magnitudes of the heavy atom contribution can be calculated using the relation,

$$F_H^2 = F_{HP}^2 + F_P^2 - 2F_{HP}F_P \cos(\alpha_P - \alpha_{HP}) = F_{HP}^2 + F_P^2 \pm 2F_{HP}F_P[1 - (k[F_{HP}^{(+)} - F_{HP}^{(-)}])/2F_P)^2]^{1/2} \quad (4)$$

(Kartha & Parthasarathy, 1965; Matthews, 1966; Singh & Ramaseshan, 1966). Here $k = f_H/f''_H$ where $f_H$ and $f''_H$ are the real and imaginary components of the heavy-atom form factors, and $F_{HP}^{(+)}$ and $F_{HP}^{(-)}$ are the structure factors of the heavy-atom derivative for reflexions $hkl$ and $\bar{h}\bar{k}\bar{l}$ respectively. $F_{HP}$ can be approximated to $[F_{HP}^{(+)} + F_{HP}(-)]/2$. In the above expression, the lower estimate is relevant when $|\alpha_P - \alpha_{HP}| < 90°$ and the upper estimate is relevant when $|\alpha_P - \alpha_{HP}| > 90°$. We refer to the lower and upper estimates as $F_{HLE}$ and $F_{HUE}$ respectively. In practice, in most cases, $F_{HLE}$ corresponds to the true value of $F_H$ and can be treated as the observed value of the heavy-atom contribution and used as such in refinement calculations.

In the refinement using the estimated values of $F_H$, there are two aspects that merit special consideration. First, one should consider how often $F_{HUE}$ is likely to correspond to the true value of $F_H$. As stated earlier, the upper estimate is relevant only when $|\alpha_P - \alpha_{HP}| > 90°$.

Referring to Fig. 1, $|\alpha_P - \alpha_{HP}|$ can be greater than 90° only when $|F_P| < |F_H \cos(\alpha_H - \alpha_P)|$ and $|\alpha_H - \alpha_P| > 90°$. Normally the average value of $F_H$ is much smaller than that of $F_P$, and only half of those reflexions for which $F_P < F_H$ will have $|\alpha_H - \alpha_P| > 90°$, so even in a highly substituted derivative these conditions are satisfied in very few cases. Hence one is justified in taking $F_{HLE}$ as a reasonable estimate of $F_H$ provided sufficient checks are made to reject these reflexions for which $F_{HVE}$ is likely to be appropriate. In practice this can be achieved by rejecting all reflexions from consideration for which $F_{HVE}$ is less than the maximum expected value of $F_H$.

The second aspect that deserves detailed consideration is the effect of experimental errors on the refined parameters. Especially the anomalous differences $[F_{PH}^{(+)} - F_{HP}^{(-)}]$ are likely to be in serious error as they are small differences between large quantities. Generally the average value of $[F_{HP}^{(+)} - F_{HP}^{(-)}]$ is similar in magnitude to the average error present in $F_{HP}^{(+)}$ and $F_{HP}^{(-)}$. Assuming the errors in $F_{HP}^{(+)}$ and $F_{HP}^{(-)}$ to be independent and random, therefore, it is readily seen that on the average the errors in the data tend to increase the $[F_{HP}^{(+)} - F_{HP}^{(-)}]$ values. In expression (4), $[F_{HP}^{(+)} - F_{HP}^{(-)}]$ is multiplied by $k$, a number much greater than unity, and hence, in the presence of experimental errors, the $F_{HLE}$ values tend to get overestimated thus leading to an overestimation of the occupancy factors.

Calculations carried out on some model derivatives were in agreement with these conclusions. Structure factors for reflexions from three model derivatives were constructed using the relation

$$F_{HP}^{(\pm)} = |\mathbf{F}_P + \mathbf{F}_H \pm i\mathbf{F}''_H| .$$

The protein phases used in these calculations were those obtained for 2Zn insulin using five heavy-atom derivatives. $\mathbf{F}''_H$ is the imaginary part of the structure factor of the heavy atoms. The positional and thermal parameters, and the occupancy factors of the heavy-atom sites in the three model derivatives are given in Tables 3, 4 and 5. The form factors and dispersion corrections used were those for mercury atoms (Cromer, 1965; Cromer & Waber, 1965). In order to simulate the real situation, errors of the type $\pm p \times F_{HP}/200$, where $p$ is a random number ranging from 0 to 16, were added to $F_{HP}^{(+)}$ and $F_{HP}^{(-)}$. Thus the errors introduced to $F_{HP}^{(+)}$ and $F_{HP}^{(-)}$ were assured to be random and independent. The random numbers were multiplied by $F_{HP}/200$ to make the errors dependent on the magnitude of the structure factors of the derivatives.

Table 3. *Model* (1). *$F_{HLE}$ refinement*

The e.s.d.'s of refined occupancy factors are given in parentheses.

| Site | $x$ | $y$ | $z$ | $B$ | True occupancy | Refined occupancy $k_{\text{theor}}$ unweighted | Refined occupancy $k_{\text{theor}}$ weighted | $k_{\text{emp}}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 0·031 | 0·031 | 0 | 12 | 0·333 | 0·461 (30) | 0·367 (20) | 0·343 (17) |
| 2 | 0 | 0 | 0·320 | 12 | 0·111 | 0·136 (17) | 0·103 (12) | 0·086 (10) |
| 3 | 0 | 0 | −0·320 | 12 | 0·111 | 0·080 (17) | 0·096 (11) | 0·075 (10) |
| 4 | −0·340 | −0·160 | −0·430 | 12 | 0·333 | 0·500 (26) | 0·407 (17) | 0·367 (15) |
| $\langle F_{HLE} \rangle$ | | | | | | 77·7 | 77·7 | 56·9 |
| $\langle F_H \rangle$ | | | | | 55·2 | 73·2 | 59·4 | 53·9 |
| $R = \sum\|F_{HLE} - F_H\|/\sum F_{HLE}$ | | | | | | 0·457 | 0·446 | 0·377 |

The magnitudes of the errors applied are similar to those found in the derivatives of rhombohedral 2*Zn* insulin.

$F_{HLE}$'s and $F_{HUE}$'s were calculated in two different ways for planes with $d>4{\cdot}5$ Å for the three models. In the first case, the ratio $k$ in equation (4) was calculated from the theoretical values of $f_H$, $f'_H$ and $f''_H$ ($k_{\text{theor}}$) and in the second case, empirical values of $k$ were evaluated as a function of sin $\theta/\lambda$, as suggested by Kartha (1965) and Matthews (1966), using the relation

$$k=2\sum|F_{HP}-F_P|/\sum|F_{HP}^{(+)}-F_{HP}^{(-)}|\,.$$

Three sets of least-squares calculations were carried out for each model to refine the positional coordinates and the occupancy factors of the heavy atom sites. The refined positional coordinates in each case were not significantly different from the true ones and hence are not represented here. The refined occupancy factors in each set of calculations for models (1), (2) and (3) are given in Tables 3, 4 and 5. Other relevant statistics are presented Table 6.

In the first set of calculations, $F_{HLE}$'s estimated with $k_{\text{theor}}$ were used for refinement. The refined occupancy factors and, therefore, the average values of $F_H$ have gone up substantially. This is to be expected as the effect of the errors present in the data has been to increase the average value of $|F_{HP}^{(+)}-F_{HP}^{(-)}|$ and consequently to make the average values of $F_{HLE}$ considerably greater than the average values of true $F_H$. However, it is interesting to note that the percentage increase in the average value of $F_H$ is the greatest for model (1) and the least for model (3). This is clearly a consequence of the fact that the ratio of the average anomalous contribution to the average error is the greatest for model (3) and the least for model (1).

In the second set of calculations, the same set of $F_{HLE}$'s were used, but weighting factors of the form

$$W=1\Big/\left[1+\left(\frac{F_{HLE}-b}{a}\right)^2\right]^{1/2}$$

were introduced, as is normally done in the refinement of small structures, to the individual terms in the weighting function. The parameters $a$ and $b$ were chosen from the agreement analysis for $F_{HLE}$'s. It is clearly seen from the Tables that the introduction of the weighting scheme has the effect of bringing, on an average, the refined occupancy factors and $F_H$'s close to their true values. The standard deviations in their estimation cannot be related to the $F_{HLE}$'s in any simple way and hence the theoretical justification for applying an empirical weighting scheme would appear to be slender. However, the large $F_{HLE}$ coefficients have a higher probability of being less reliable as they are more likely to be estimated from spuriously high anomalous differences. The effect of the weighting scheme applied is mainly to decrease the contribution of the large $F_{HLE}$ terms to the minimisation function. This probably explains the success of the weighting scheme in these calculations.

$F_{HLE}$ coefficients estimated using empirical $k$ values were used in the third set of calculations. The empirical $k$'s for the three models are shown in Fig. 5. The theoretical $k$'s are also given for comparison. As is to be expected, the empirical values are substantially lower than the theoretical ones. The average values of

Table 4. *Model* (2). $F_{HLE}$ *refinement*

The e.s.d.'s of refined occupancy factors are given in parentheses.

| Site | $x$ | $y$ | $z$ | $B$ | True occupancy | Refined occupancy $k_{\text{theor}}$ unweighted | Refined occupancy $k_{\text{theor}}$ weighted | Refined occupancy $k_{\text{emp}}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 0·031 | 0·031 | 0 | 12 | 0·333 | 0·476 (29) | 0·390 (20) | 0·375 (17) |
| 2 | 0·280 | 0·260 | 0·020 | 12 | 0·333 | 0·434 (28) | 0·331 (19) | 0·332 (17) |
| 3 | 0·100 | 0·049 | 0·493 | 12 | 0·333 | 0·401 (28) | 0·351 (18) | 0·311 (17) |
| $\langle F_{HLE}\rangle$ | | | | | | 82·5 | 82·5 | 62·7 |
| $\langle F_H\rangle$ | | | | | 59·2 | 76·5 | 62·2 | 59·5 |
| $R=\sum\|F_{HLE}-F_H\|/\sum F_{HLE}$ | | | | | | 0·440 | 0·425 | 0·351 |

Table 5. *Model* (3). $F_{HLE}$ *refinement*

The e.s.d.'s of refined occupancy factors are given in parentheses.

| Site | $x$ | $y$ | $z$ | $B$ | True occupancy | Refined occupancy $k_{\text{theor}}$ unweighted | Refined occupancy $k_{\text{theor}}$ weighted | Refined occupancy $k_{\text{emp}}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 0·031 | 0·031 | 0 | 12 | 0·333 | 0·480 (33) | 0·370 (22) | 0·422 (25) |
| 2 | −0·340 | −0·160 | −0·430 | 12 | 0·333 | 0·417 (31) | 0·334 (21) | 0·371 (24) |
| 3 | 0·092 | 0·095 | 0·389 | 12 | 0·333 | 0·431 (32) | 0·370 (21) | 0·382 (24) |
| 4 | 0·031 | 0·296 | 0·233 | 12 | 0·333 | 0·366 (32) | 0·348 (20) | 0·312 (23) |
| 5 | 0·190 | 0·271 | 0·783 | 12 | 0·333 | 0·357 (32) | 0·298 (21) | 0·316 (24) |
| $\langle F_{HLE}\rangle$ | | | | | | 96·0 | 96·0 | 83·9 |
| $\langle F_H\rangle$ | | | | | 75·8 | 91·0 | 75·3 | 80·5 |
| $R=\sum\|F_{HLE}-F_H\|/\sum F_{HLE}$ | | | | | | 0·390 | 0·384 | 0·348 |

Table 6. *Some statistics relevant to the $F_{HLE}$ refinement*

| | $\langle F_P \rangle$ | $\langle F_H \rangle$ | $\langle$Error$\rangle$ in $F_{HP}^{(+)}$ and $F_{HP}^{(-)}$ | $\langle F_{HP} \rangle$ | $\langle F_{HP} \rangle$ with error | $\langle \lvert F_{HP} - F_P \rvert \rangle$ | $\langle \lvert F_{HP}^{(+)} - F_{HP}^{(-)} \rvert \rangle$ | $\langle \lvert F_{HP}^{(+)} - F_{HP}^{(-)} \rvert \rangle$ with error |
|---|---|---|---|---|---|---|---|---|
| Model 1 | 165 | 55·2 | 6·85 | 175 | 175·4 | 33·4 | 6·6 | 11·6 |
| Model 2 | 165 | 59·2 | 7·30 | 176·4 | 176·7 | 37·6 | 6·9 | 11·5 |
| Model 3 | 165 | 75·8 | 7·15 | 181·4 | 182·2 | 47·0 | 8·7 | 13·6 |

$F_{HLE}$ estimated using $k_{emp}$, though greater than the average values of true $F_H$, are much lower than the average values of those using $k_{theor}$ suggesting thereby that the errors in data have essentially been taken care of by introducing empirically evaluated $k$'s. The occupancy factors and the average values of $F_H$ that resulted from the third set of refinement are again comparable to the true values.

It is seen from the above discussion that it is highly desirable to use empirically evaluated $k$ values to estimate the coefficients or to employ suitable weighting functions when the coefficients are estimated with $k_{theor}$. In these circumstances, the refinement yields occupancy factors which are, on an average, reasonably close to the true ones. However, it should be noted that individual occupancy factors may differ from their true values by amounts not explained by their estimated standard deviations. Therefore, once a set of protein phases is available, it is desirable to refine the heavy atom parameters concurrently using the more widely used 'phase refinement' procedure. Once the protein phase angles are accurate, this method would obviously yield satisfactory results. However, it is of interest to investigate the effect of the inevitable errors in phase angles on the refined parameters, especially, the occupancy factors. A related point of interest is the behaviour of the occupancy

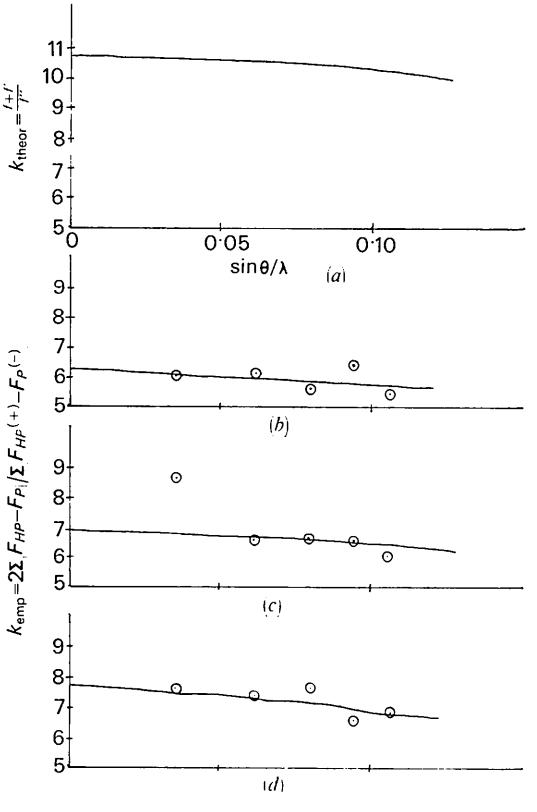


Fig. 5. (*a*) $k_{theor}$; $k_{emp}$ corresponding to (*b*) model (1), (*c*) model (2) and (*d*) model (3).

factors of the heavy-atom sites common to two or more derivatives. Hence, it was thought worth while to carry out phase refinement on the heavy-atom parameters in the model derivatives. It may be noted that site 1 is common to all three model derivatives. Also site 4 in model (1) is the same as site 2 in model (3).

*Phase refinement*

Two sets of protein phases were calculated, to be used in refinement procedures, using both isomorphous and anomalous differences according to the well known procedures described in the literature (Blow & Crick, 1959; North, 1965). In phase set (1), model (1) was used to derive phase angles whereas models (1) and (2) were used to derive phase set (2). The root-mean square values of the lack-of-closure – the conventional $E$'s – were estimated by comparing the $F_{HLE}$ coefficients with the $F_H$ values obtained from $F_{HLE}$ refinement; $E'$ values were taken as a thrid of the $E$ values as suggested by North.

Four sets of least-squares calculations were carried out to refine the heavy atom parameters in models (2) and (3) using phase set (1). In the first set of calculations, all reflections, irrespective of their figures of merit, were included. In the second set, only those reflexions with figure of merit higher than 0·5 were included. In the third set, the cut off value for figure of merit was raised to 0·8. Calculations were repeated to refine the heavy atom parameters in model (3) using phase set (2). No weighting functions were employed in these refinement cycles. The refined occupancy factors that resulted from these calculations are presented in Tables 7, 8 and 9. The average errors in phase angles, obtained by comparing the phases of phase sets (1) and (2) to the true phases, for each set of calculations are also given in the Tables. Even though, in general, the average phase error decreases as the minimum figure of merit of the reflexions included increases, the quantitative relationship between the two values is not the same for phase sets (1) and (2).

Table 7. *Model* (2). *Occupancy factors after refinement using phase set* (1)

| Minimum figure of merit | 0 | 0·5 | 0·8 |
|---|---|---|---|
| ⟨Phase error⟩ | 46·3° | 31·2° | 23·2° |
| Number of terms | 518 | 275 | 110 |
| Occupancy | | | |
| Site (1) (common site) | 0·320 (12) | 0·319 (20) | 0·355 (20) |
| Site (2) | 0·315 (13) | 0·324 (16) | 0·284 (22) |
| Site (3) | 0·293 (13) | 0·306 (16) | 0·292 (21) |
| Average of non-common sites | 0·304 | 0·315 | 0·288 |

One feature that may be noticed from the results is that the occupancy factors tend to be underestimated when the phase angles contain large errors. However, in general satisfactory results can be obtained by using only reflexions with high figures of merit for refinement. The second feature that may be noted is concerned with the behaviour of the heavy atom sites common to the derivative or derivatives used in phasing calculations, and the derivative the heavy atom parameters of which are refined. The mean value of the common sites is consistently higher than that of the non-common sites in model (2) after refinement using phase set (1) and in model (3) after refinement using phase set (2). However, no such marked difference is observed in model (3) after refinement using phase set (1). Thus, the evidence from the present series of calculations alone does not enable us to arrive at any firm conclusion regarding this particular point. However, on the whole, there appears to be a tendency for the occupancy factors of the common sites to be overestimated compared to the other sites. Again it should be emphasized that the effects of possible systematic errors such as those resulting from non-isomorphism are not taken into account in these calculations.

Table 8. *Model* (3). *Occupancy factors after refinement using phase set* (1)

| Minimum figure of merit | 0 | 0·5 | 0·8 |
|---|---|---|---|
| ⟨Phase error⟩ | 46·3° | 31·2° | 23·2° |
| Number of terms | 518 | 275 | 110 |
| Occupancy | | | |
| Site (1) | 0·318 (14) | 0·314 (18) | 0·333 (26) |
| Site (2) | 0·327 (13) | 0·334 (17) | 0·341 (23) |
| Site (3) | 0·317 (14) | 0·326 (19) | 0·326 (27) |
| Site (4) | 0·337 (14) | 0·349 (18) | 0·373 (23) |
| Site (5) | 0·303 (13) | 0·307 (16) | 0·303 (22) |
| Average of common sites | 0·323 | 0·324 | 0·337 |
| Average of non-common sites | 0·319 | 0·327 | 0·335 |

Table 9. *Model* (3). *Occupancy factors after refinement using phase set* (2)

| Minimum figure of merit | 0 | 0·5 | 0·8 |
|---|---|---|---|
| ⟨Phase error⟩ | 31° | 25° | 18·5° |
| Number of terms | 518 | 423 | 254 |
| Occupancy | | | |
| Site (1) | 0·326 (12) | 0·325 (12) | 0·320 (15) |
| Site (2) | 0·344 (12) | 0·348 (12) | 0·340 (15) |
| Site (3) | 0·308 (12) | 0·312 (13) | 0·315 (16) |
| Site (4) | 0·316 (11) | 0·322 (12) | 0·327 (15) |
| Site (5) | 0·304 (11) | 0·307 (12) | 0·309 (14) |
| Average of common sites | 0·337 | 0·337 | 0·330 |
| Average of non-common sites | 0·309 | 0·314 | 0·317 |

We wish to thank Professor Dorothy Hodgkin and Drs Tom Blundell and Guy Dodson for the many stimulating and useful discussions we had with them. Our thanks are also due to Mrs L. Blundell and S. Cutfield for help with the preparation of the diagrams.

### References

ADAMS, M. J. (1968). Ph.D. Thesis, Oxford University.
BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794.
CROMER, D. T. (1965). *Acta Cryst.* **18**, 17.
CROMER, D. T. & WABER, J. T. (1965). *Acta Cryst.* **18**, 104.
DICKERSON, R. E., KENDREW, J. C. & STRANDBERG, B. E. (1961). *Acta Cryst.* **14**, 1188.
DICKERSON, R. E., WEINZIERL, J. E. & PALMER, R. A. (1968). *Acta Cryst.* B**24**, 997.
KARTHA, G. (1965). *Acta Cryst.* **19**, 883.
KARTHA, G. & PARTHASARATHY, R. (1965). *Acta Cryst.* **18**, 745.
MATTHEWS, B. W. (1966). *Acta Cryst.* **20**, 230.
MUIRHEAD, H., COX, J. M., MAZZARELLA, L. & PERUTZ, M. F. (1967). *J. Mol. Biol.* **28**, 117.
NORTH, A. C. T. (1965). *Acta Cryst.* **18**, 212.
PHILLIPS, D. C. (1966). *Advances in Structure Research by Diffraction Methods.* Edited by R. BRILL and R. MASON. New York and London: Interscience Publishers.
RAMACHANDRAN, G. N. & RAMAN, S. (1959). *Acta Cryst.* **12**, 957.
SINGH, A. K. & RAMASESHAN, S. (1966). *Acta Cryst.* **21**, 279.
SRINIVASAN, R. (1961). *Proc. Indian Acad. Sci.* A**53**, 252.


*Acta Cryst.* (1971). B**27**, 2411

# The Crystal Structure of Bis-(L-phenylalaninato)copper(II)*

BY DICK VAN DER HELM,† M. B. LAWSON‡ AND E. L. ENWALL

*Chemistry Department, University of Oklahoma, Norman, Oklahoma* 73069, *U.S.A.*

(*Received* 30 *December* 1970)

The crystal structure of bis-(L-phenylalaninato)copper(II), $CuC_{18}H_{20}O_4N_2$, has been determined, and refined by three-dimensional least-squares techniques. The crystals are monoclinic, space group $P2_1$, with $a=16{\cdot}710$, $b=5{\cdot}317$, $c=9{\cdot}509$ Å and $\beta=98{\cdot}40°$. The final $R$ value for 1745 reflections is 0·041. The standard deviations are between 0·004 and 0·010 Å for the C, N and O atom positions. The structure closely resembles that of bis-(L-alaninato)Cu(II). The copper coordination is best described as a tetragonally distorted octahedron. The conformations of both phenylalanine molecules are similar, although there are distinct differences in detail, and such that the aromatic rings are pointed away from the metal coordination.

## Introduction

The structure of the Cu chelate of L-phenylalanine was determined as one of a series of chelates of amino acids and peptides with transition metal ions. This project is designed to elucidate the factors which determine the bonding geometry of metal ions in biological systems. In particular we were interested to see if in the present structure an interaction occurred between the aromatic ring and the $Cu^{2+}$ ion. This kind of interaction was previously observed between the $Cu^{2+}$ ion and the aromatic ring of a tyrosine residue in the copper chelates of L-tyrosine (Tatsch & van der Helm, 1969) and glycyl-L-leucyl-L-tyrosine (Franks & van der Helm, 1971).

## Experimental

The bis-(L-phenylalaninato)copper(II) complex was prepared by slow diffusion of aqueous solutions of L-phenylalanine and copper(II) acetate ($p$H 5 to 6). The light-blue crystals grew as thin plates with the plate face being the (100) plane and all crystals examined showed a rather high mosaic spread ($\geq 1°$). This preparation yields the *trans* form of the complex which has also been prepared in a different way by Laurie (1967). The *cis* form has recently been prepared by Herlinger, Wenhold & Long (1970).

X-ray investigation of the crystals showed them to be monoclinic and the space group to be $P2_1$ (systematic absences: $0k0$, $k=2n+1$. Because L-phenylalanine was the ligand, the space group $P2_1/m$ could be excluded.). The cell dimensions, determined by a least-squares fit to the $2\theta$ values of 43 reflections measured at 22°C, are: $a=16{\cdot}710\pm0{\cdot}014$, $b=5{\cdot}217\pm0{\cdot}009$, $c=9{\cdot}509\pm0{\cdot}007$ Å, and $\beta=98{\cdot}40\pm0{\cdot}06°$ [$\lambda$(Cu $K\alpha$)= 1·5418 Å]. The F.W. for Cu(L-phenylalanine)$_2$ is 391·91, yielding a $\varrho_c$ of 1·587 g.cm$^{-3}$ with $Z=2$. A density of 1·575 g.cm$^{-3}$ was measured by the flotation method, using a mixture of carbon tetrachloride and hexane.

The integrated intensities were taken on a General Electric XRD-5 diffraction unit using the $\theta$–$2\theta$ scan technique and nickel-filtered Cu $K\alpha$ radiation. The diffraction unit was equipped with a SPG single-crystal orienter, a scintillation counter, and pulse-height analyzer.

The data crystal had the dimensions of 0·36 × 0·10 × 0·02 mm and a mosaic spread of 1°. All unique reflec-

* Work supported in part by Grant GM-10514 from the National Institutes of Health.
† Supported by N.I.H. Development Award 1-K4-GM-42,572.
‡ National Science Foundation Predoctoral Trainee. Present address: Chemistry Department, University of California at Irvine. This work was presented in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Oklahoma.