# The Refinement of Heavy-Atom Parameters in the Presence of Non-Crystallographic Symmetry

By Michael G. Rossmann

*Department of Biological Sciences, Purdue University, West Lafayette, Indiana* 47907, *U.S.A.*

The necessary algebra is derived for the least-squares refinement of heavy-atom parameters in a series of isomorphous derivatives where the heavy atoms are related by non-crystallographic symmetry. In addition the simultaneous refinement of the particle (or molecular) orientation and position is considered, permitting the precise adjustment of the system of non-crystallographic rotation axes.

## Introduction

The first method for systematic adjustment of heavy-atom parameters required in isomorphous replacement phase determination was proposed and used by Hart (1961) as part of the structure determination of myoglobin. This method depends on successive small adjustments of each parameter in turn to find the minimum of a suitable residual derived from a consideration of the phase circle diagrams (Harker, 1956). Rossmann (1960) introduced the use of the least-squares technique, but his method required the independent refinement of the parameters of each heavy-atom compound. Dickerson, Kendrew & Strandberg (1961*a*) and Dickerson, Kopka, Varnum & Weinzierl (1967) developed a more general least-squares analysis, requiring alternate cycles of an initial determination of the best or most probable phases (Blow & Crick, 1959; Cullis, Muirhead, Perutz, Rossmann & North, 1961; Dickerson *et al.*, 1961*b*) and subsequent heavy-atom parameter refinement. Variations of this technique have been employed in most structure determinations of biological macromolecules (Blow & Matthews, 1973) using widely distributed programs (*e.g.* those originally written by H. Muirhead, R. E. Dickerson or M. G. Rossmann). In recent years it has become possible to solve oligomeric proteins where the protein subunits, and hence their associated heavy atoms, are related by a closed point-group symmetry. The molecular symmetry is, however, not always incorporated into the crystal symmetry. One example is the tetrameric lobster glyceraldehyde-3-phosphate dehydrogenase (GAPDH) molecule which has 222 non-crystallographic symmetry (Buehner, Ford, Moras, Olsen & Rossmann, 1974). While for lower molecular symmetries each non-crystallographic asymmetric unit can be taken independently, for higher symmetries, such as icosahedral 532 viruses with 60 asymmetric units, this procedure is cumbersome, time wasting and possibly indeterminate.

The external environment of identical protein subunits related by non-crystallographic symmetry will be different. This might cause small surface conformational differences, but the positions of the heavy-atom sites are not likely to be affected greatly, particularly for higher-order aggregates where the external surface is small compared to internal subunit contacts. This has been well demonstrated by Buehner *et al.* (1974) for GAPDH, where the r.m.s. displacement of any atom from its mean is less than 0·4 Å. However, a breakdown of symmetry may have occurred (up to 2·5 Å) in 'dimeric' α-chymotrypsin (Tulinsky, Mani, Morimoto & Vandlen, 1973) although the large number of closely linked parameters they used makes their results suspect. Similar comments are valid for the notoriously inaccurate occupancy determination. When heavy-atom derivatives have been prepared by diffusion rather than by co-crystallization, there may, however, be greater differences caused by the masking of the site of a given subunit by its environment.

This paper considers the treatment of non-crystallographic symmetry in a manner analogous to that used in linking atoms associated by conventional crystallographic symmetry. Such a procedure has been considered in the refinement of tomato bushy stunt virus (Harrison, 1975). Furthermore, the mathematically similar problem of group refinement (Scheringer, 1963; Doedens, 1970) has been applied in a variety of cases (La Placa & Ibers, 1965; Doedens & Dahl, 1966), and relates to the more general problem of constrained refinements (Pawley, 1972).

## General principles

It will be assumed that the best phase $\alpha_P$ of the native structure factor $\mathbf{F}_P$ has been determined for a given reflection. The vector $\mathbf{f}_H$, representing the heavy-atom structure factor, may be calculated with respect to the current approximate heavy-atom parameters. The phase triangle may then be closed (Fig. 1) with the vector $\mathbf{D}$, the calculated structure factor of the heavy-atom compound for that reflection. Thus

$$\mathbf{D} = \mathbf{F}_P + \mathbf{f}_H. \tag{1}$$

We now wish to find a small change $\Delta D$ in order to minimize the squared difference $(F_{PH} - D - \Delta D)^2$ for each heavy-atom compound, $H$, and for each reflec-

tion, h. Here $|\mathbf{F}_{PH}|$ represents the observed structure amplitude of the heavy-atom compound. It can then be shown (cf. Rollett, 1969) that the corresponding normal equations are

$$\sum_j \left\{ \sum_h \sum_H \omega \frac{\partial D}{\partial \xi_i} \frac{\partial D}{\partial \xi_j} \right\} \Delta \xi_j = \sum_h \sum_H \omega (F_{PH} - D) \frac{\partial D}{\partial \xi_i},$$

$$(2)$$

where $\xi$ is a parameter required in the calculation of $D$ and $\omega$ is a suitable weight to be applied to each observational equation. The subscripts $i$ and $j$ run over all parameters.

Now from (1)

$$D^2 = q_H^2 [(a_H + F_P \cos \alpha_P)^2 + (b_H + F_P \sin \alpha_P)^2], \quad (3)$$

where $1/q_H$ is the scale factor to be applied to $\mathbf{F}_{PH}$ relative to $\mathbf{F}_P$, and $a_H$ and $b_H$ are the real and imaginary components of $\mathbf{f}_H$ for compound $H$. Hence

$$\frac{\partial D}{\partial q_H} = \frac{D}{q_H}, \quad (4)$$

and for other parameters

$$\frac{\partial D}{\partial \xi} = \frac{q_H^2}{D} \left[ (a_H + F_P \cos \alpha_P) \frac{\partial a_H}{\partial \xi} \right.$$

$$\left. + (b_H + F_P \sin \alpha_P) \frac{\partial b_H}{\partial \xi} \right]. \quad (5)$$

It is now apparent that in order to determine the shifts $\Delta \xi$ for each refinable parameter it is necessary to set up the normal equations (2). These can be computed by evaluating $D$ from (3) and $\partial D/\partial \xi$ from (4) or (5). However, it is still necessary to derive $a_H, b_H$ and their derivatives in terms of the crystallographic and non-crystallographic constraints.

## Introduction of symmetry into the structure factor expressions

Let $\mathbf{x}$ be the position vector erected from the molecular or particle center with its components expressed in
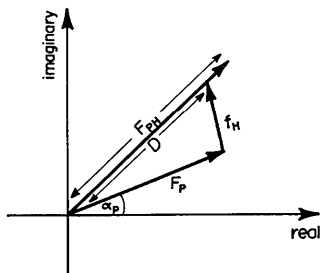


Fig. 1. Vector diagram relating the structure factors of the native protein or virus particle $\mathbf{F}_P$, the heavy-atom component $\mathbf{f}_H$, and the complete heavy-atom derivative $\mathbf{F}_{PH}$. D is the calculated heavy-atom derivative structure factor based on a knowledge of $\mathbf{F}_P$ and $\mathbf{f}_H$.

terms of fractions of the unit-cell lengths. Then

$$\mathbf{x}_n = [C_n]\mathbf{x} \quad (6)$$

where $[C_n]$ is the $n$th non-crystallographic symmetry operator. Let the molecular or particle center be at $\mathbf{S}$, referred to the crystal origin. Thus a heavy atom in the $n$th non-crystallographic asymmetric unit is positioned at

$$\mathbf{p}_n = \mathbf{x}_n + \mathbf{S}$$

$$= [C_n]\mathbf{x} + \mathbf{S} \quad (7)$$

with respect to the cell origin.

Let $[T_m]$ and $\mathbf{d}_m$ be the crystallographic rotational and translational symmetry operators. Thus, an atom in the $n$th non-crystallographic asymmetric unit after operation with the $m$th crystallographic operator will be related to the standard atom, $\mathbf{x}$, by

$$\mathbf{x}_{mn} = [T_m]\mathbf{p}_n + \mathbf{d}_m$$

$$= [T_m][C_n]\mathbf{x} + [T_m]\mathbf{S} + \mathbf{d}_m$$

$$= [R_{mn}]\mathbf{x} + [T_m]\mathbf{S} + \mathbf{d}_m \quad (8)$$

where

$$[R_{mn}] = [T_m][C_n]$$

$$= [T_m][\alpha][\varrho_n][\beta].$$

Here $[\varrho_n]$ is the $n$th non-crystallographic rotation operator with respect to an orthogonal coordinate system, $[\alpha]$ is an orthogonalizing matrix, and $[\beta]$ is a de-orthogonalizing matrix (Rossmann & Blow, 1962).

Should it, however, be necessary to slightly rotate the system of non-crystallographic axes by $[\Delta]$, then

$$\mathbf{x}' = [\alpha][\Delta][\beta]\mathbf{x}$$

and

$$\mathbf{x}_n = [C_n]\mathbf{x}$$

$$= [C_n][\alpha][\Delta][\beta]\mathbf{x}.$$

It follows that

$$[R_{mn}] = [T_m][\alpha][\varrho_n][\Delta][\beta] \quad \text{(since } [\beta][\alpha] = \mathbf{1}). \quad (9)$$

Each rotation matrix, $[R_{mn}]$, can be evaluated since the crystallographic operations $[T_m]$ are completely known, and the absolute directions of the non-crystallographic operators $[\varrho_n]$ are at least roughly known while their relative directions are completely known. The elements of $[\varrho_n]$ are expressed most easily in terms of the rotation, $\kappa$, about an axis defined by its latitude and longitude, $\psi$ and $\varphi$, respectively. The non-crystallographic adjustment matrix $[\Delta]$ can be expressed in terms of Eulerian angles $\psi_1, \psi_2, \psi_3$ which are chosen to give an identity matrix before refinement (Rossmann & Blow, 1962). Instability of this matrix can be avoided by fixing $\psi_1$ equal to zero under these special conditions.

The structure factor expression of reflection $\mathbf{h}$ for the $J$ heavy atoms in the compound $H$ can then be written as

$$\mathbf{f}_H = \sum_{j=1}^{J} Z_j \exp(-B_j \sigma^2) \left\{ \sum_{m=1}^{M} \sum_{n=1}^{N} \exp 2\pi i \mathbf{h} \cdot \mathbf{x}_{mn} \right\}$$

where $\sigma^2 = (\sin \theta/\lambda)^2$ (using the usual notation for Bragg angle, wavelength and temperature factor).

Therefore

$$\mathbf{f}_H = \sum_{j=1}^{J} Z_j \exp(-B_j\sigma^2)$$

$$\times \left\{ \sum_{m=1}^{M} \sum_{n=1}^{N} \exp 2\pi i \mathbf{h} \cdot ([R_{mn}]\mathbf{x}_j + [T_m]\mathbf{S} + \mathbf{d}_m) \right\}$$

$$= \sum_{j=1}^{J} Z_j \exp(-B_j\sigma^2)$$

$$\times \left\{ \sum_{m=1}^{M} \exp 2\pi \mathbf{h} \cdot ([T_m]\mathbf{S} + \mathbf{d}_m) \right.$$

$$\times \left. \sum_{n=1}^{N} \exp 2\pi i \mathbf{h} \cdot [R_{mn}]\mathbf{x}_j \right\}.$$

Now set

$$\theta_m = 2\pi \mathbf{h} \cdot ([T_m]\mathbf{S} + \mathbf{d}_m)$$

and

$$\varphi_{mn} = 2\pi \mathbf{h} \cdot [R_{mn}]\mathbf{x}_j .$$

These definitions can be rewritten as

$$\theta_m = 2\pi \mathbf{h}'_m \cdot \mathbf{S} + \Phi_m$$

and

$$\varphi_{mn} = 2\pi \mathbf{h}'_{mn} \cdot \mathbf{x}_j$$

where

$$\mathbf{h}'_m = [T_m]^T \mathbf{h} ,$$
$$\mathbf{h}'_{mn} = [R_{mn}]^T \mathbf{h} ,$$

and

$$\Phi_m = \mathbf{h} \cdot \mathbf{d}_m .$$

$\left. \right\}$ (10)

Hence

$$a_H = \sum_j Z_j \exp(-B_j\sigma^2)\left\{ \sum_m \cos\theta_m \sum_n \cos\varphi_{mn} \right.$$

$$\left. - \sum_m \sin\theta_m \sum_n \sin\varphi_{mn} \right\}$$

$$b_H = \sum_j Z_j \exp(-B_j\sigma^2)\left\{ \sum_m \sin\theta_m \sum_n \cos\varphi_{mn} \right.$$

$$\left. - \sum_m \cos\theta_m \sum_n \sin\varphi_{mn} \right\} .$$

$\left. \right\}$ (11)

Hence

$$\frac{\partial a_H}{\partial S_1} = \sum_j Z_j \exp(-B_j\sigma^2)\left\{ \sum_m (-2\pi h_m) \sin\theta_m \right.$$

$$\times \sum_n \cos\varphi_{mn} - \sum_m (2\pi h_m) \cos\theta_m \sum_n \sin\varphi_{mn} \right\},$$

and similarly for $S_2$ and $S_3$, which are the $x,y,z$ components of $\mathbf{S}$;

$$\frac{\partial a_H}{\partial x_j} = \sum_j Z_j \exp(-B_j\sigma^2)\left\{ \sum_m \cos\theta_m \sum_n (-2\pi h'_{mn}) \right.$$

$$\times \sin\varphi_{mn} - \sum_m \sin\theta_m \sum_n (2\pi h'_{mn}) \cos\varphi_{mn} \right\},$$

and similarly for $y_j$ and $z_j$, which are the $x,y,z$ components of $\mathbf{x}_j$;

$$\frac{\partial a_H}{\partial \psi_1} = \sum_j Z_j \exp(-B_j\sigma^2)\left\{ \sum_m \cos\theta_m \sum_n \frac{\partial\varphi_{mn}}{\partial\psi_1} \right.$$

$$\times \sin\varphi_{mn} - \sum_m \sin\theta_m \sum_n \frac{\partial\varphi_{mn}}{\partial\psi_1} \cos\varphi_{mn} \right\},$$

and similarly for $\psi_2$ and $\psi_3$, which are the Eulerian adjustment angles. Similar expressions can be written with respect to $b_H$. The derivatives with respect to $Z_j$ and $B_j$ are straightforward.

Finally, from (10)

$$\frac{\partial\varphi_{mn}}{\partial\psi} = 2\pi \mathbf{h} \cdot \frac{\partial}{\partial\psi}[R_{mn}]$$

$$= [T_m][\alpha][\varrho_n]\left\{\frac{\partial}{\partial\psi}[A]\right\}[\beta] \quad \text{[from (9)]}$$

Derivatives of the elements of the Eulerian matrix $[A]$ can be easily evaluated (Rossmann & Blow, 1962; Rossmann & Argos, 1975).

The Eulerian variables $\psi_1, \psi_2, \psi_3$ and the translational parameters $S_1, S_2, S_3$ ($\mathbf{S}$) affect the structure factors, $\mathbf{f}_H$, of every heavy-atom derivative, whereas the other parameters $q_H$, $Z_j$, $B_j$ and $\mathbf{x}_j$ affect only the structure factors of a particular compound. Care must thus be taken in setting up the elements of the normal equations. Those elements relating to parameters governing the orientation and position of the molecule are formed from sums over all heavy-atom compounds, where the other elements contain terms only from one compound.

## Results

This method has been satisfactorily applied to three problems. The program was set up to read the independent molecular or particle symmetry operators from which the program then developed the $[C_n]$ rotation matrices. Thus for 222 symmetry it is necessary to read the $\kappa, \psi, \varphi$ values defining two perpendicular twofold axes which generate four rotation matrices. Similarly for an icosahedral virus it is only necessary to read in the $\kappa, \psi, \varphi$ values defining *in order* the three, two and fivefold axes in a plane and finally a perpendicular twofold axis (Argos, Rossmann & Ford, 1975), which generate 60 rotation matrices.

Pig $H_4$ LDH crystals contain one molecule per asymmetric unit in space group $C2$ (Eventoff, Hackert & Rossmann, 1975). The molecule has 222 symmetry and there was one heavy-atom site per polypeptide chain (non-crystallographic asymmetric unit). The temperature factor was kept constant at 10 Å$^2$. Hence there were four parameters to be refined for the one independent heavy atom. In addition it was necessary to refine the relative scale factor of the heavy-atom compound structure amplitudes, two origin translational parameters of the molecular center (the third

component defined the $y$ axis origin) and three rotational parameters of the molecular orientation; which is a total of 10 parameters. The refinement was a single isomorphous refinement in the absence of any other derivative. After three cycles the parameters had essentially refined. As expected (see Buehner *et al.*, 1974) the positional parameters agreed well with those found by independent refinement of the heavy-atom parameters. However, the occupancy refined to $Z = 58.6 \pm 0.4$ electrons while on independent refinement the occupancies were found to be 53, 42, 66 and 69 electrons, giving a mean of 57·5 electrons. Final residuals did not differ significantly from those found when refining the atoms independently of non-crystallographic symmetry. The amount of computer time needed per independent parameter per reflection was $2.8 \times 10^{-2}$ seconds using the non-crystallographic constraints as opposed to $2.0 \times 10^{-2}$ seconds without constraints.

The other two problems were tests on model data for viruses. The first test was a simulation of satellite tobacco necrosis virus (STNV) where one whole particle is in the crystallographic asymmetric unit of the $C2$ cell (Lentz & Strandberg, 1974). The CPU time per independent parameter per reflection was $1.4 \times 10^{-2}$ seconds. The second test simulated the type II southern bean mosaic virus (Akimoto, Wagner, Johnson & Rossmann, 1975) where the virus sits on the intersection of the crystallographic threefold axis and the three perpendicular twofold axes in space group $R32$. Thus in this case there was no refinement of the particle orientation or position. The CPU time per independent parameter per reflection was $0.1 \times 10^{-2}$ seconds.

Further experience will be required with real problems to determine the range of convergence with poorly placed atoms. The program was, however, subsequently successfully employed in the double isomorphous replacement refinement of real STNV data (Strandberg & Lentz, 1976).

All calculations were performed on Purdue University's CDC 6500 computer.

## References

Akimoto, T., Wagner, M. A., Johnson, J. E. & Rossmann, M. G. (1975). *J. Ultrastruct. Res.* **53**, 306–318.

Argos, P., Rossmann, M. G. & Ford, G. C. (1975). *Proceedings of the Summer School on Crystallographic Computing.* Prague: *Crystallographic Computing* II. Edited by F. R. Ahmed. Copenhagen: Munksgaard. In the press.

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.

Blow, D. M. & Matthews, B. W. (1973). *Acta Cryst.* A**29**, 56–62.

Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. (1974). *J. Mol. Biol.* **82**, 563–585.

Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1961). *Proc. Roy. Soc.* A**265**, 15–38.

Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961a). *Computing Methods and the Phase Problem in X-ray Crystal Analysis.* Edited by R. Pepinsky, J. M. Robertson & J. C. Speakman, pp. 236–251. Oxford: Pergamon Press.

Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961b). *Acta Cryst.* **14**, 1188–1194.

Dickerson, R. E., Kopka, M. L., Varnum, J. C. & Weinzierl, J. E. (1967). *Acta Cryst.* **23**, 511–522.

Doedens, R. J. (1970). *Crystallographic Computing.* Edited by F. R. Ahmed, S. R. Hall & C. P. Huber, pp. 198–200. Copenhagen: Munksgaard.

Doedens, R. J. & Dahl, L. F. (1966). *J. Amer. Chem. Soc.* **88**, 4847–4855.

Eventoff, W., Hackert, M. L. & Rossmann, M. G. (1975). *J. Mol. Biol.* **98**, 249–258.

Harker, D. (1956). *Acta Cryst.* **9**, 1–9.

Harrison, S. C. (1975). Private communication.

Hart, R. G. (1961). *Acta Cryst.* **14**, 1194–1195.

La Placa, S. J. & Ibers, J. A. (1965). *Acta Cryst.* **18**, 511–519.

Lentz, P. J. Jr & Strandberg, B. E. (1974). *Acta Cryst.* A**30**, 552–559.

Pawley, G. S. (1972). *Advances in Structure Research by Diffraction Methods.* Edited by W. Hoppe & R. Mason, pp. 1–64. Oxford: Pergamon Press.

Rollett, J. S. (1969). *Crystallographic Computing.* Edited by F. R. Ahmed, S. R. Hall, & C. P. Huber, pp. 167–181. Copenhagen: Munksgaard.

Rossmann, M. G. (1960). *Acta Cryst.* **13**, 221–226.

Rossmann, M. G. & Argos, P. (1975). *J. Biol. Chem.* **250**, 7525–7532.

Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.

Scheringer, C. (1963). *Acta Cryst.* **16**, 546–550.

Strandberg, B. E. & Lentz, P. J. Jr (1976). Private communication.

Tulinsky, A., Mani, N. V., Morimoto, C. N. & Vandlen, R. L. (1973). *Acta Cryst.* B**29**, 1309–1322.