

Correct Calculation of Scale Factor between Diffraction Data Sets of Isomorphous Crystals and Its Application to Protein Crystallography

BY STANISLAV V. NIKONOV

Institute of Protein Research, Academy of Sciences of the USSR, 142292 Pushchino, Moscow Region, USSR

(Received 14 December 1982; accepted 6 April 1983)

Abstract

A correct calculation of the scale factor between the diffraction data of isomorphous crystals is shown on the basis of a statistics criterion. This scale factor requires no additional refinement, improves the signal-to-noise ratio of difference Patterson maps and Wilson's statistics and allows an estimation of the degree of non-isomorphism. The method has been tested for the 3 Å data sets of protein γ -crystallin crystals and a few of its heavy-atom isomorphous derivatives.

Introduction

The solution of protein crystal structures by the multiple isomorphous replacement method requires accurate determination of the parameters of the added heavy atoms. It is well known that the structure factor of heavy atoms can be determined from the expression

$$\mathbf{F}_{Hi} = \mathbf{F}_{pHi} - \mathbf{F}_{pi}, \quad (1)$$

where \mathbf{F}_{pHi} is the structure factor of the i th reflection of the heavy-atom derivative, \mathbf{F}_{pi} , \mathbf{F}_{Hi} are the protein and heavy-atom components which are, strictly speaking, related only to the same single crystal.

In practice, the amplitude F_{pi} is measured from one crystal, F_{pHi} from another and F_{pHi} is usually brought to the scale of the native protein crystal by applying a scale coefficient k_i . Therefore,

$$\mathbf{F}_{Hi} = k_i (\mathbf{F}_{pHi}^D) - \mathbf{F}_{pi}^N = k_i (\mathbf{F}_{pi}^D + \mathbf{F}_{Hi}^D) - \mathbf{F}_{pi}^N,$$

where superscripts N and D apply to the data obtained from the native and derivative crystals, respectively.

The contribution of heavy atoms can be calculated correctly if

$$k_i \mathbf{F}_{pi}^D - \mathbf{F}_{pi}^N = 0.$$

Hence, a correct scale factor can be determined as

$$k_i = F_{pi}^N / F_{pi}^D, \quad (2)$$

assuming that the vectors \mathbf{F}_{pi}^N and \mathbf{F}_{pi}^D have the same directions.

It should be noted that the assumption of perfect isomorphism between the atomic positions in the native crystal and in the native component of the derivative crystal leads to independence of the coefficient k_i of the value of $\sin \theta / \lambda$.

Therefore, the scale factor for scaling up the heavy-atom derivative relative to the parent data can be obtained by averaging (2) over a large number of reflections. Hence, the scale factor is determined as

$$k_{pH} = \bar{k}_i = \overline{(F_{pi}^N / F_{pi}^D)}. \quad (3)$$

In practice it is impossible to calculate k_{pH} from (3) because the values of F_{pi}^D are unknown. A substitution of F_{pi}^D by F_{pHi}^D leads to errors in the value of scale factor and to distortions of the difference Fourier synthesis (Blundell & Johnson, 1976). Various schemes have been used to avoid these problems (Green, Ingram & Perutz, 1954; Blow, 1958; Kraut, Sieker, High & Freer, 1962; Singh & Ramaseshan, 1966; Arnone *et al.*, 1971; Eklund, Samama, Wallen & Branden, 1981) and to adjust the scale factor to compensate for the additional scattering power of the heavy atoms. Some of them required knowledge of the details of heavy-atom substitution (Green *et al.*, 1954). Others (Kraut *et al.*, 1962) did not need these data, but estimated the correct scale factor to within about 4%. An improved value of the scale factor was then obtained using iterative procedures. Therefore the problem of scaling of heavy-atom-derivative data to the native set proved a difficult one and no entirely satisfactory solution was found.

In this paper a method of calculation of a correct value of the scale factor directly from the experimental data is proposed using a definite statistics criterion.

Calculation formulae for the scale factor

From (1) the structure amplitude of a heavy-atom derivative can be written as

$$(F_{pHi}^D)^2 = (F_{pi}^D)^2 + 2F_{pi}^D F_{Hi}^D \cos \gamma_i + (F_{Hi}^D)^2,$$

where γ_i is the angle between vectors \mathbf{F}_{pi}^D and \mathbf{F}_{Hi}^D .

Using (2) we can write

$$(F_{pHi}^D/F_{pi}^N)^2/(F_{pi}^N)^2 = k_i^{-2} (1 + 2\delta_i \cos \gamma_i + \delta_i^2), \quad (4)$$

where $\delta_i = F_{Hi}^D/F_{pi}^D$ and

$$F_{pHi}^D/F_{pi}^N = k_i^{-1} [(1 + \delta_i \cos \gamma_i)^2 + \delta_i^2 \sin^2 \gamma_i]^{1/2}.$$

Making the approximation that

$$\delta_i^2 \sin^2 \gamma_i / (1 + \delta_i \cos \gamma_i)^2 < 1, \quad (5)$$

one can obtain

$$F_{pHi}^D/F_{pi}^N \simeq k_i^{-1} [1 + \delta_i \cos \gamma_i + \delta_i^2 \sin^2 \gamma_i / 2(1 + \delta_i \cos \gamma_i)]$$

since $(1 + \chi)^{1/2} \simeq 1 + \chi/2$ if $\chi < 1$.

It is easy to show that the right-hand side of this expression is always positive if $\delta_i < 1$, and so

$$F_{pHi}^D/F_{pi}^N = k_i^{-1} [1 + \delta_i \cos \gamma_i + \delta_i^2 \sin^2 \gamma_i / 2(1 + \delta_i \cos \gamma_i)]. \quad (6)$$

For the centric case, $\sin \gamma_i = 0$ and one can use (6) if $\delta_i < 1$, that is

$$F_{Hi}^D < F_{pi}^D. \quad (7)$$

By averaging (6) over a large number of reflections it is easy to obtain

$$(k_{pH}^{-1})_{\text{centric}} = \overline{(F_{pHi}^D/F_{pi}^N)}. \quad (8)$$

For the acentric case, it is easy to show that the condition of (5) is always correct if $\delta_i < 1/\sqrt{2}$, that is

$$F_{Hi}^D < F_{pi}^D/\sqrt{2}. \quad (9)$$

The average value of $\delta_i^2 \sin^2 \gamma_i / 2(1 + \delta_i \cos \gamma_i)$ can be obtained by a series expansion of $1/(1 + \delta_i \cos \gamma_i)$.

For this case we have

$$\overline{[\delta_i^2 \sin^2 \gamma_i / 2(1 + \delta_i \cos \gamma_i)]} \simeq \overline{\delta_i^2} / 4$$

since $\sin^2 \gamma = \cos^2 \gamma = 1/2$ and $\cos^4 \gamma = 3/8$. Hence

$$\overline{(F_{pHi}^D/F_{pi}^N)} \simeq k_i^{-1} (1 + \overline{\delta_i^2} / 4). \quad (10)$$

But from (4)

$$\overline{(F_{pHi}^D)^2 / (F_{pi}^N)^2} = k_i^{-2} (1 + \overline{\delta_i^2}).$$

Therefore, if $\delta_i < 1$, we can write

$$\overline{[(F_{pHi}^D)^2 / (F_{pi}^N)^2]}^{1/2} \simeq k_i^{-1} (1 + \overline{\delta_i^2} / 2). \quad (11)$$

From (10) and (11) we finally obtain the scale factor

$$(k_{pH}^{-1})_{\text{acentric}} \simeq 2 \overline{(F_{pHi}^D/F_{pi}^N)} - \overline{[(F_{pHi}^D)^2 / (F_{pi}^N)^2]}^{1/2}. \quad (12)$$

Therefore, by using (8) and (12) it is possible to determine the value of k_{pH} from measured data, if the contribution of the native component in the used reflections exceeds the contribution of the heavy-atom component as determined by (7) and (9).

Now we can introduce a combined value

$$k_{pH} = [w_1(k_{pH})_{\text{centric}} + w_2(k_{pH})_{\text{acentric}}] / (w_1 + w_2), \quad (13)$$

where w_1 and w_2 are numbers of centric and acentric reflections, respectively.

Statistical criterion for selection of scaling reflections

Now let us assume that F_{pi} and F_{Hi} are non-correlated and $F_{pi} > F_{Hi}$. It is clear that an infringement of conditions (7) and (9) is less probable if F_{pi} is large enough. In order to estimate the number of reflections of the data set which may be used in the calculation of the scale factor we shall obtain the probability distributions of the structure amplitudes.

Let the probability distributions of the structure amplitudes of the native and heavy-atom compounds be $P_p(F)$ and $P_H(F)$. Denote by F_0 the minimum value of a structure amplitude of the native crystal which may be used in calculations.

Suppose that the structure amplitude F_{pHi} of the derivative crystal contains $F < F_{Hi} < F + dF$ and $F_0 < F_{pi} < F$ for the centric case or $F_0 < F_{pi} < \sqrt{2}F$ for the acentric case. It is clear that this structure amplitude is not in accordance with (7) or (9). The probability of such reflections can be determined as the probability of the compound event

$$r(F) dF = q(F) P_H(F) dF,$$

where

$$q(F) = \int_{F_0}^{F_1} P_p(F) dF$$

is the probability that the structure amplitude of the native component lies between F_0 and F_1 ($F_1 = F$ for the centric case and $F_1 = \sqrt{2}F$ for the acentric case). Therefore, for the interval between F_0 and ∞ , we can write

$$R(F) = \int_{F_0}^{\infty} q(F) P_H(F) dF.$$

The probability that the structure amplitude of the native crystal lies between F_0 and ∞ can be defined by

$$Q(F) = \int_{F_0}^{\infty} P_p(F) dF.$$

Hence the part S of the total amount of used reflections which are not in accordance with (7) or (9) if $F_{pi}^N > F_0$ is determined by

$$S(F_0) = \int_{F_0}^{\infty} q(F) P_H(F) dF / \int_{F_0}^{\infty} P_p(F) dF. \quad (14)$$

Assuming that Wilson's statistics (Wilson, 1949) are obeyed, we can write for the centric case

$$P_p(F) = 1/(2\pi\overline{F_p^2})^{1/2} \exp(-F^2/2\overline{F_p^2}),$$

$$P_H(F) = 1/(2\pi\overline{F_p^2} t^2)^{1/2} \exp[-F^2/(2t^2\overline{F_p^2})];$$

and for the acentric case

$$P_p(F) = 2(F/\overline{F_p^2}) \exp(-F^2/\overline{F_p^2}),$$

$$P_H(F) = 2[F/(t^2 \overline{F_p^2})] \exp[-F^2/(t^2 \overline{F_p^2})];$$

where $t^2 = \overline{F_H^2}/\overline{F_p^2}$. Substituting these expressions into (14) and integrating the latter it is easy to show that for the acentric case

$$S(F_0)_{\text{acentric}} = \exp[-F_0^2/(t^2 \overline{F_p^2})] - [1/(t^2 \sqrt{2} + 1)] \\ \times \exp[-(F_0^2/\overline{F_p^2})(\sqrt{2} - 1 + 1/t^2)]. \quad (15)$$

For the centric case it is convenient to use the expression

$$S(F_0)_{\text{centric}} = \sum_{F_0}^{\infty} q'(F) P_H(F) \Delta F / \sum_{F_0}^{\infty} P_p(F) \Delta F, \quad (16)$$

where

$$q'(F) = \sum_{F_0}^F P_p(F) \Delta F.$$

The value of t can be determined from the expression (Crick & Magdoff, 1956)

$$|[(F_{pH}^2 - F_p^2)^{1/2}] / \overline{F_p^2} = g(\overline{F_H^2}/\overline{F_p^2})^{1/2} = gt, \quad (17)$$

where $g = 2$ for the centric case and $g = \sqrt{2}$ for the acentric case. Since usually the left part of (17) is less than 1, we can suppose for all practical cases that $t^2 \leq 0.25$.

Substituting $t^2 = 0.25$ into (15) and (16) one can calculate values $S(F_0)_{\text{centric}}$ and $S(F_0)_{\text{acentric}}$ as functions of $F_0/(\overline{F_p^2})^{1/2}$ for different F_0 (Fig. 1).

It is evident from Fig. 1 that the values of $S(F_0)_{\text{centric}}$ and $S(F_0)_{\text{acentric}}$ are small enough if $F_0/(\overline{F_p^2})^{1/2} \geq 1$. [$S(F_0)_{\text{centric}} \sim S(F_0)_{\text{acentric}} \sim 1\%$ if $F_0 = (\overline{F_p^2})^{1/2}$]. Therefore, using (8) and (12), we can obtain the correct value of scale factor for all practical cases if

$$F_0 = (\overline{F_p^2})^{1/2}. \quad (18)$$

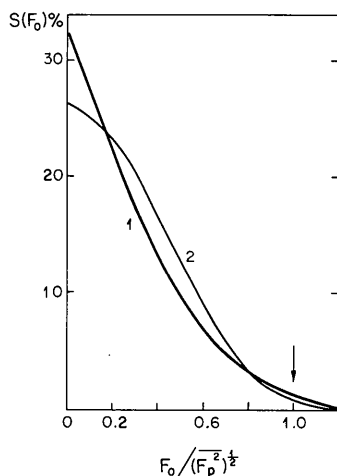


Fig. 1. Relationship between the relative number of reflections leading to errors in calculation of scale factor and value of $F_0/[(\overline{F_p^2})^{1/2}]$. Curve (1) centric case; curve (2) acentric case.

Results

The efficiency of this method of scaling has been tested in calculations of scale factors of different derivatives of γ -crystallin IIIb from calf lens at 3 Å resolution. Data were collected on screened precession photographs. Three or five film packs were taken per crystal. To bring all the planes of native or derivative data to a common scale, the inter-layer scaling factors were determined by the method of Hamilton, Rollett & Sparks (1965). The scale coefficients calculated using (13) have been compared with relations of $\sum F_p^2 / \sum F_{pH}^2$. The Wilson plots for these cases are shown in Fig. 2. In the first case we have omitted all reflections with $F_{pi}^N \leq (\overline{F_p^2})^{1/2}$, in the second case we have omitted all reflections with $F_{pi}^N \leq 2 \sigma_p$ or $F_{pHi} \leq 2 \sigma_{pH}$, where σ_p and σ_{pH} are mean values of standard deviations. It is evident from Fig. 2 that the scale factor calculated via the relations of $\sum F_p^2 / \sum F_{pH}^2$ considerably depends on $(\sin \theta/\lambda)^2$.

The same effect of seeming imperfect isomorphism in this case is shown in Fig. 3, where the relations $2|F_{pH} - F_p| / (F_{pH} + F_p) = \delta F$ are plotted as functions of $(\sin \theta/\lambda)^2$. We suppose that the source of these seeming distortions of isomorphism is a larger contribution of the heavy atom to the value of $\sum F_{pH}^2$ in the second case. For the first case the isomorphism appears to be more perfect for all derivatives.

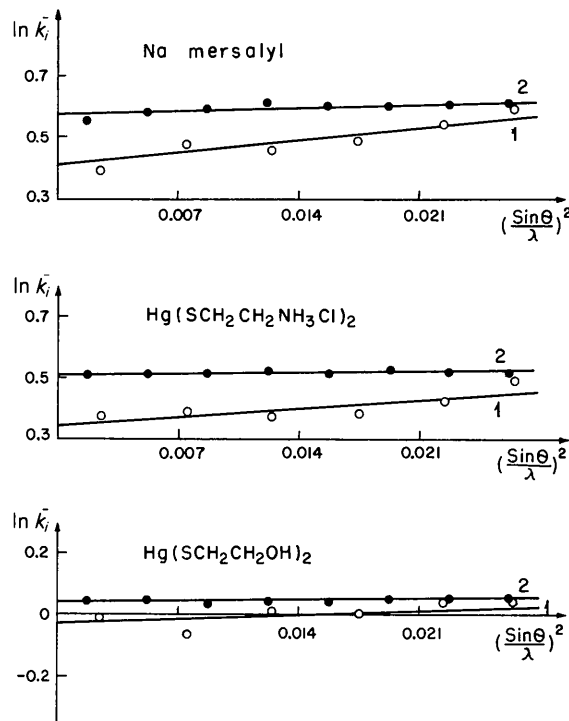


Fig. 2. Wilson plots for calculation of scale factors of different derivatives of γ -crystallin IIIb. Curves (1) are calculated using the relation $\sum (F_p^N)^2 / \sum (F_{pH}^D)^2$; curves (2) are calculated using formula (13) of this paper.

Table 1. Comparison of scale factors for the isomorphous heavy-atom derivatives of γ -crystallin IIIb protein at 3 Å resolution calculated by different methods

Heavy-atom derivative	Method of calculation				
	Proposed method: see formulas (8), (12), (13) Limit for structure amplitudes: $F_{pi}^N \leq 2\sigma_p; F_{pHi}^D \leq 2\sigma_{pH}$ (1)	Proposed method: see formulas (8), (12), (13) Limit for structure amplitudes: $F_{pi}^N \leq [(F_p^N)^2]^{1/2}$ (2)	$k' = \sum (F_p^N)^2 / \sum (F_{pH}^D)^2$ Limit for structure amplitudes: $F_{pi}^N \leq [(F_p^N)^2]^{1/2}$ (3)	Phase refinement program (Ten Eyck & Arnone, 1976) Limit for structure amplitudes: $F_{pi}^N \leq 2\sigma_p; F_{pHi}^D \leq 2\sigma_{pH}$ (4)	
Na mersalyl	k_0	1.519	1.786	1.721	1.804
	ΔB	-6.890	-1.630	-3.050	—
	N	4464	2576	2576	4464
Hg(SCH ₂ COONa) ₂	k_0	1.615	1.874	1.817	1.888
	ΔB	-5.120	+0.810	-0.121	—
	N	5590	2821	2821	5590
Hg(SCH ₂ CH ₂ NH ₃ Cl) ₂	k_0	1.443	1.671	1.611	1.652
	ΔB	-3.890	-0.190	-0.970	—
	N	5298	2722	2722	5298
KAu(CN) ₂	k_0	1.560	1.602	1.565	1.605
	ΔB	-1.270	-1.010	-2.785	—
	N	5478	2836	2836	5478
Hg(SCH ₂ CH ₂ OH) ₂	k_0	1.004	1.030	1.025	1.018
	ΔB	-1.910	-0.870	-0.880	—
	N	6199	2885	2885	6199

Designations: k'_i scale coefficient in the i th interval of $(\sin \theta/\lambda)^2$; $k_{pH} = k_0 \exp[-\Delta B(\sin \theta/\lambda)^2]$; N number of reflections.

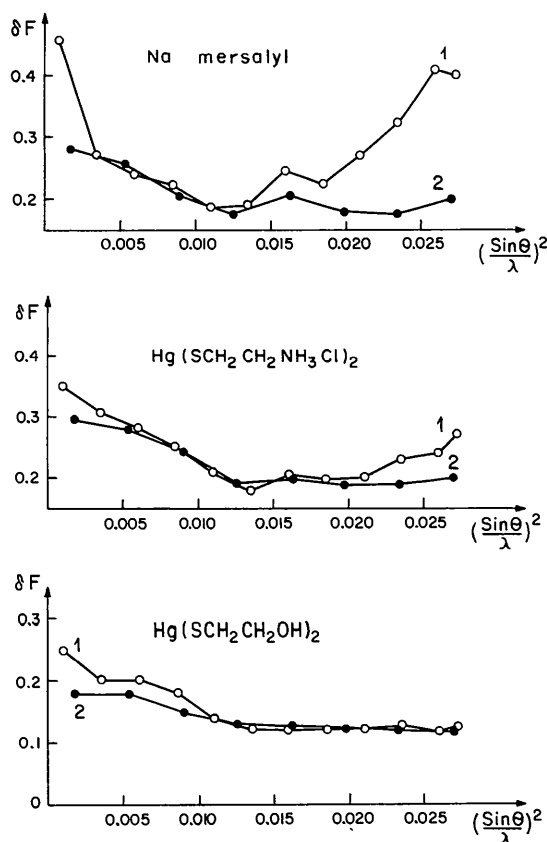


Fig. 3. Plots of $\Delta F = 2(F_{pH} - F_p)/(F_{pH} + F_p)$ against $(\sin \theta/\lambda)^2$.
(1) Scale factor is calculated using the relation $\sum (F_p^N)^2 / \sum (F_{pH}^D)^2$;
(2) scale factor is calculated using formula (13) of this paper.

The overall scale and temperature coefficients determined in different ways are shown in Table 1. The temperature coefficients determined from (13) are small enough and overall scale coefficients are close to that determined by the phase refinement program (Chirgadze, Sergeev, Fomenkova & Oreshin, 1981). Therefore, ignoring the reflections with $F_{pi}^N < (F_p^N)^{1/2}$ and using (8), (12) and (13), we can obtain a correct value of scale factors without post-refinement even for such strong derivatives as Na mersalyl.

The overall scale coefficients derived from the relation $\sum F_p^2 / \sum F_{pH}^2$ on the same basis as the scale coefficients derived from (13) differ from those determined by the phase refinement program to within about 5% for strong derivatives and must be refined using an iterative procedure.

Fig. 4 presents the UV 1/2 projections of the difference Patterson synthesis at 5 Å resolution. The scale and temperature coefficients for Figs. 4(a) and (b) have been selected from columns 1 and 2 of Table 1, respectively. It is clear from Fig. 4 that the use of correctly calculated scale factors improves the signal-to-background ratio (about 20%) and eliminates some false peaks, whereas the principal peaks remain unchanged.

The tests described here demonstrate that in a variety of cases the direct calculation of the scale factors improves the accuracy of isomorphous differences and allows one to evaluate the degree of isomorphism between the native structure and the native component of the derivative. However, it must

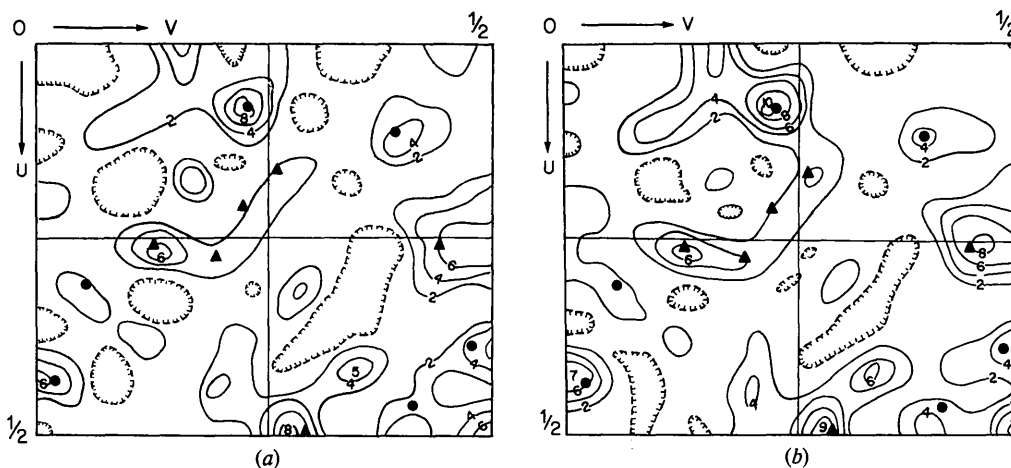


Fig. 4. Projections UV $1/2$ of the difference Patterson synthesis of the $\text{Hg}(\text{SCH}_2\text{CH}_2\text{NH}_3\text{Cl})_2$ derivative of γ -crystallin IIIb. Overall scale and temperature coefficients are calculated (a) using the relation $\sum (F_p^N)^2 / \sum (F_{pH}^D)^2$ and (b) using formula (13) of this paper.

be added that the use of this principle changes significantly only the scale factors of the strong derivatives. For the weak derivatives these changes are small. Finally, it may be noted that the use of the direct method of calculation of scale factors does not require essential changes in the computer program.

I am indebted to Dr Yu. N. Chirgadze for constant interest and help throughout this work.

References

- ARNONE, A., BIER, C. J., COTTON, F. A., DAY, V. W., HAZEN, E. E. JR, RICHARDSON, D. C., RICHARDSON, J. S. & YONATH, A. (1971). *J. Biol. Chem.* **246**, 2302–2316.
- BLOW, D. M. (1958). *Proc. R. Soc. London Ser. A*, **247**, 302–335.
- BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*, pp. 333–336. New York, London, San Francisco: Academic Press.
- CHIRGADZE, YU. N., SERGEEV, YU. V., FOMENKOVA, N. P. & ORESHIN, V. D. (1981). *FEBS Lett.* **131**, 81–84.
- CRICK, F. H. C. & MAGDOFF, B. S. (1956). *Acta Cryst.* **9**, 901–908.
- EKLUND, H., SAMAMA, J.-P., WALLEN, L. & BRANDEN, C.-I. (1981). *J. Mol. Biol.* **146**, 561–587.
- GREEN, D. W., INGRAM, V. M. & PERUTZ, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 287–307.
- HAMILTON, W. C., ROLLETT, I. S. & SPARKS, R. A. (1965). *Acta Cryst.* **18**, 129–130.
- KRAUT, J., SIEKER, L. C., HIGH, D. F. & FREER, S. T. (1962). *Proc. Natl Acad. Sci. USA*, **48**, 1417–1430.
- SINGH, A. K. & RAMASESHAN, S. (1966). *Acta Cryst.* **21**, 279–280.
- TEN EYCK, L. F. & ARNONE, A. (1976). *J. Mol. Biol.* **100**, 3–11.
- WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

Acta Cryst. (1983). **A39**, 697–706

The Contrast of Defects in Inelastically Scattered Electrons

BY PETER REZ*

Department of Materials Science, University of California, Berkeley, California 94720, USA

(Received 25 June 1979; accepted 6 April 1983)

Abstract

The theory for the contrast of stacking faults and dislocations in electrons which have been scattered

inelastically is derived. Small-angle plasmon and single-electron scattering show similar contrast to the elastically scattered electrons. Phonon scattering by large angles away from strongly excited Bragg reflections shows reversed contrast and small-angle phonon scattering gives better contrast for defects near the top of the specimen.

* Present address: V. G. Microscopes Ltd, Charlwoods Road, East Grinstead, Sussex RH19 2JQ, England.