Acta Cryst. (1985). **A40**, 713-716

The Use of Phase Combination in Crystallographic Refinement: the Choice of Amplitude Coefficients in Combined Syntheses

BY DAVID STUART AND PETER ARTYMIUK

*Laboratory of Molecular Biophysics, Zoology Department, South Parks Road, Oxford OX*1 3*PS, England*

(*Received* 17 *April* 1984; *accepted* 12 *July* 1984)

Abstract

When Fourier syntheses are calculated using phase information from both isomorphous replacement and the refined model it is usual to combine the information for each reflection by multiplying the two probability distributions together. It is well known that phase information from a model contains 'bias' towards that model; it is also true that such bias will be diluted by the process of phase combination. A method is proposed of tailoring the amplitude coefficients individually to correct for the likely residual bias in the combined phase.

The refinement of macromolecules by crystallographic techniques remains a time-consuming and often difficult procedure, essentially because the relationship between the X-ray observations and the required atomic parameters is non-linear and the usual least-squares techniques become trapped in false minima. It is therefore important to be able to produce a clear and unbiased representation of the electron density distribution so that the model can be rebuilt by hand. In many cases phase information is available from more than one independent source, most usually from (i) the isomorphous replacement method (independent of any model of the structure) and (ii) calculated from the current model. Similar considerations to those that follow apply to phase information derived from other sources. Phase information is normally represented as a probability distribution in which the phase circle is represented as a ring with variable density; the magnitude of the density corresponds to the probability of that phase angle being the correct one. Blow & Crick (1959) have shown that the minimum root-mean-square error in the final electron density map is obtained by using the distance from the centre of the ring to its centre of gravity as a weighting function in the Fourier synthesis, the phase being taken to be in the direction of the centre of gravity. This quantity is known as the 'figure of merit' (m). Sim first considered the use of weights for phase angles calculated from a partial model of the structure in 1959 and then recast his scheme to follow that of Blow & Crick (Sim, 1960). When calculating maps based on phase information from such partially-known structures a further complication arises from the fact that the heights of atomic peaks depend on whether the atom was included in the phase calculation, so that the resultant map is, in effect, biased towards the starting model even though this may be incomplete and partly wrong. Huber *et al.* (1974), Main (1979) and Vijayan (1980) have investigated this and confirmed that the F_o synthesis may be corrected by the addition of difference terms proportional to $|F_o|-|F_c|$. Thus if

$$\chi^2=\frac{\text{sum of the squares of the scattering factors for correctly placed atoms}}{\text{sum of the squares of the scattering factors of all the atoms included}}$$

then the most truthful Fourier synthesis will be that based on the following amplitude coefficients (Main, 1979):

$$F_o+(2/\chi^2-1)(F_o-F_c) \qquad (1)$$

for acentric terms, and

$$F_o+[2/(1+\chi^2)-1](F_o-F_c) \qquad (2)$$

for centric terms. Let the multiplier for the difference terms be defined as Q_c:

$$F_o+Q_c(F_o-F_c). \qquad (3)$$

Rice (1981) has discussed this for the case of protein refinement and it is clear that while χ^2 cannot easily be identified accurately it is not difficult to estimate approximately (and hence derive Q_c). Rice found that in the early stages of a refinement a sensible value for Q_c would be about three.

To combine the (essentially independent) phase information from isomorphous and calculated probability distributions it is enough simply to multiply the probability distributions together and use the centroid of the resultant distribution as the best 'combined' phase with weight, m_{comb}. However, the distributions must be appropriately weighted. It is by no means obvious what the optimum weights are: as Bricogne (1976) observed 'the only practical solution ... seems to be an inspired relative weighting of the two sources of phase information'. In practice, inspiration has been forthcoming and the procedure is now fairly standard (see *e.g.* Rice, 1981).

0108-7673/84/060713-04$01.50

What does not appear to have been adequately considered is the effect that such phase combination has on the choice of best amplitude coefficients for the Fourier synthesis. For those reflections well phased by the isomorphous replacement method the combination with the calculated phase information will introduce rather little bias; in other cases the bias will be essentially the same as in the partial structure phase (Q_c). In other words, the bias in the combined phase will vary from one reflection to another between zero and Q_c. Thus, we need to replace Q_c in (3) with $Q_{\rm comb}$, $Q_{\rm comb}$ being different for each reflection. Then the coefficients used in the final Fourier will be

$$m_{\rm comb}[F_o+Q_{\rm comb}(F_o-F_c)]\exp i\,\alpha_{\rm comb}. \qquad (4)$$

Now, if the phase distributions are normal we can associate an error, σ, with each, and the bias, $Q_{\rm comb}$, in the combined distribution is simply

$$Q_{\rm comb}=\left[\frac{Q_i}{\sigma_i^2}+\frac{Q_c}{\sigma_c^2}\right]\left[\frac{1}{\sigma_i^2}+\frac{1}{\sigma_c^2}\right]^{-1}, \qquad (5)$$

where the subscripts i and c refer to the isomorphous and partial structure distributions, respectively.

Now, since the isomorphous replacement phase is independent of the model we may assume that Q_i is zero, hence

$$Q_{\rm comb}=\left[\frac{Q_c}{\sigma_c^2}\right]\left[\frac{1}{\sigma_i^2}+\frac{1}{\sigma_c^2}\right]^{-1}. \qquad (6)$$

Where the phase is uncertain and there is a Gaussian probability distribution of phase errors, the probability that a phase α is correct is given by

$$P(\alpha)=[1/(2\pi)^{1/2}\sigma]\exp[-\alpha^2/2\sigma^2], \qquad (7)$$

where α is the phase angle and σ is the standard deviation of the distribution. Blow & Crick (1959) show that the best weighting function to minimize errors is given by the centre of gravity of this distribution, which for a phase circle with unit radius is at

$$m=\exp[-\tfrac{1}{2}\sigma^2], \qquad (8)$$

where m is the usual figure of merit. In the special case of combining two unimodal normal distributions we would expect the resultant distribution to be weighted by $1/\sigma^2$ as in (5) above. Then from (8) it is clear that σ^2 is proportional to $-\log m$ and (6) becomes

$$Q_{\rm comb}=\left[\frac{Q_c}{-\log m_c}\right]\left[\frac{1}{-\log m_i}+\frac{1}{-\log m_c}\right]^{-1} \qquad (9)$$

or, more simply,

$$Q_{\rm comb}=Q_c[\log m_i/(\log m_c+\log m_i)]. \qquad (10)$$

The probability distribution for the partial structure phase is well represented by this simple model; unfortunately, unless the isomorphous phasing is very good, some isomorphous distributions will be bimodal and thus m_i will lead to an underestimation of the amount of phase information in the distribution so that the estimate of $Q_{\rm comb}$ will be unduly pessimistic. In practice, we have found that even simpler schemes of determining $Q_{\rm comb}$ are capable of producing marked improvements in the clarity of maps while still avoiding bias.

We will briefly discuss the effect of the choice of Fourier coefficients for two of the enzymes we are currently refining.

Phosphorylase b

The first example is from the refinement of rabbit muscle phosphorylase *b* at 2·0 Å resolution (Sansom, Stuart, Babu & Johnson, unpublished work). The phosphorylase molecule is large (M = 100 000 for the single subunit in the crystallographic asymmetric unit) and the initial model was built to an isomorphous replacement map of only moderate quality so that refinement has been a time-consuming business. Fig. 1 shows identical portions of two maps calculated at the stage where the R factor for the 52 300 data in the range 5 to 2 Å was 0·37. The refinement at that stage did not include any model for the solvent and therefore calculated phases were not used below 5 Å resolution. Combined phases were used between 5 and 3 Å resolution and calculated phases from 3 to 2 Å. Fig. 1(*a*) shows the map calculated with $m_{\rm comb}(2F_o-F_c)\exp i\alpha_{\rm comb}$ coefficients whereas the map in Fig. 1(*b*) is calculated with the coefficients of (4) and (10) with Q_c chosen as 1. Clearly, the new coefficients produce an enormous increase in clarity, the contrast between the protein and the solvent is much greater than in Fig. 1(*a*) (owing in part to the inappropriate choice of very low resolution Fourier coefficients in Fig. 1*a*) and closer inspection reveals that the protein density itself is greatly improved. Note in particular the much improved density for Trp 824 in the centre of the diagram and for the helix running along the top of the figure.

Triose phosphate isomerase (TIM)

Chicken TIM, a dimer of identical subunits of 247 residues each, is being refined at 2.5 Å resolution (Artymiuk, Phillips & Taylor, unpublished work) using the restrained structure-factor least-squares program of Hendrickson & Konnert (1980). Between groups of cycles of refinement, rebuilding of the model into $(3F_o-2F_c)\exp i\alpha_{\rm comb}$ maps has been performed on the laboratory's Evans & Sutherland PS2 graphics system.

The R factor at the time of calculating the maps was 0·24 from 10 to 2·5 Å, with good chemical geometry. Fig. 2 shows density for representative residues calculated with identical combined phases and $m_{\rm comb}$ but different coefficients, namely:

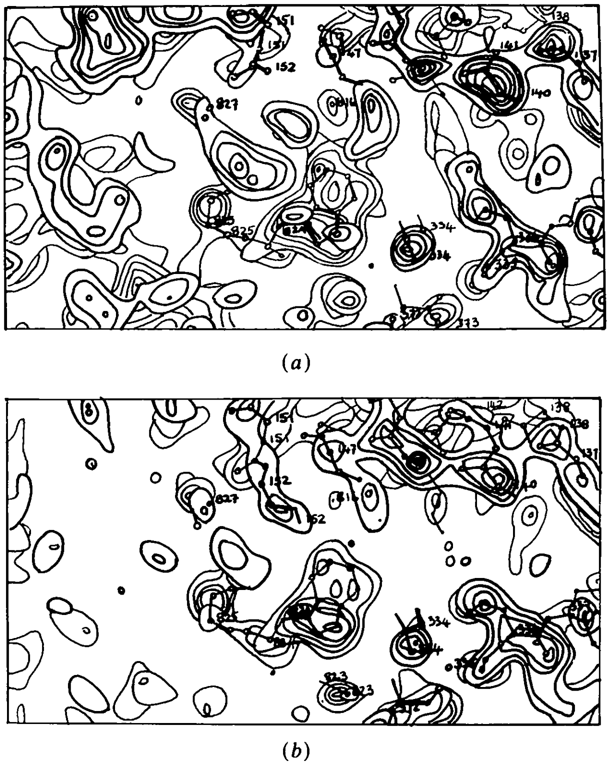
(i) $3F_o - 2F_c \equiv F_o + Q_c(F_o - F_c)$,

where $Q_c = 2$;

(ii) $F_o + Q_{\rm comb}(F_o - F_c)$,

where $Q_{\rm comb}$ is evaluated using (10).

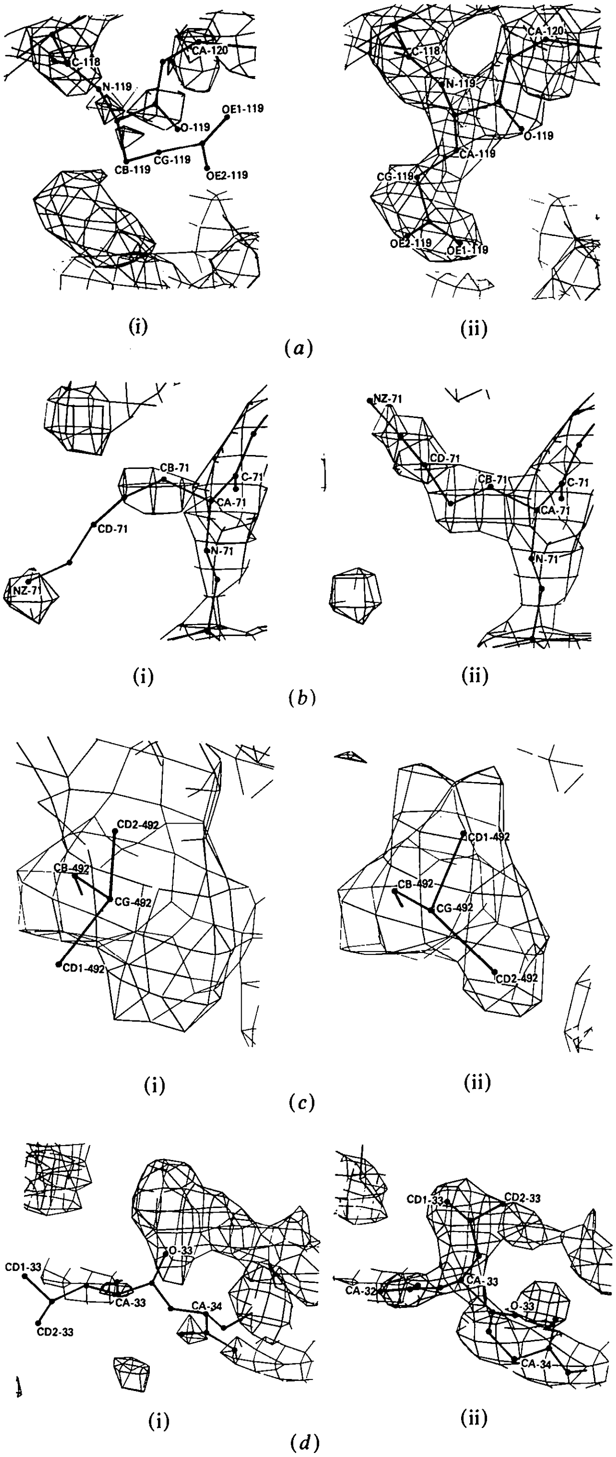
The maps were examined on the graphics: there were no cases where the density was subjectively degraded using coefficients (ii) and numerous instances where it was improved (see Fig. 2); spurious density disappeared or was diminished (Fig. 2*c*) and new connections appeared to isolated strong density (Figs. 2*a*, *b*, *d*). Contouring map (i) at a lower level allowed some correct connections to be seen, but at the cost of still more spurious density features. Where major rebuilding was indicated in map (ii) but not map (i) the indicated modifications led to improved and chemically reasonable contacts with the rest of the molecule and to better agreement between the two molecules in the asymmetric unit (related by a non-crystallographic twofold axis, which had not been used as a restraint in the refinement).



(*a*)

(*b*)

Fig. 1. Electron density distributions for a portion of the phosphorylase *b* unit cell. The two maps (*a*) and (*b*) are calculated as described in the text, with identical phases but different amplitude coefficients. The contour levels in both cases are at equal but arbitrary levels of positive density with the zero level omitted. Coordinates are shown superimposed on the density, these are those used in the structure-factor calculation.



(i) (*a*) (ii)

(i) (*b*) (ii)

(i) (*c*) (ii)

(i) (*d*) (ii)

Fig. 2. Examples of electron densities for selected residues in TIM calculated using the same combined phases (and all terms 10–2.5 Å) but with: (i) coefficients $3F_o - 2F_c$; (ii) coefficients modified as described in the text. The contour level in all cases is 0·32 e Å^3 and for any residue the identical volume of map is shown in both cases. The coordinates shown in maps (i) are those used to calculate α_c and F_c for both syntheses; the rebuilt coordinates are shown in maps (ii). (*a*) Glu 119; (*b*) Lys 71; (*c*) Leu 492 (*i.e.* Leu 192 of molecule 2 in the asymmetric unit); (*d*) Residues 33 and 34.

References

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

Bricogne, G. (1976). *Acta Cryst.* **A32**, 832–847.

Hendrickson, W. A. & Konnert, J. H. (1980). In *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan. Bangalore: Indian Academy of Sciences.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. & Steigemann, W. (1974). *J. Mol. Biol.* **89**, 73–101.

Main, P. (1979). *Acta Cryst.* **A35**, 779–785.

Rice, D. W. (1981). *Acta Cryst.* **A37**, 491–500.

Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.

Sim, G. A. (1960). *Acta Cryst.* **13**, 511–512.

Vijayan, M. (1980). *Acta Cryst.* **A36**, 295–298.

Acta Cryst. (1984). **A40**, 716–721

Modification of X-ray Atomic Scattering Factors in Electric Fields–Validation of the Shell Model

By J. A. D. Matthew and S. Y. Yousif

Physics Department, University of York, Heslington, York YO1 5DD, England

(*Received* 11 *November* 1983; *accepted* 27 *July* 1984)

Abstract

Quantum-mechanical calculations of the modification of the X-ray scattering factor of an atom/ion in an electric field are compared with predictions of the semi-classical shell model. If the shell parameters are fitted to the dipole polarizability of the ion, the shell model is a very good representation of the scattering factor of the deformed ion, giving support to its use in estimating the effect of ionic deformation on X-ray diffuse scattering and X-ray Debye–Waller factors.

1. Introduction

Born (1942) was the first to consider the effect of modification of atomic scattering factors due to deformation of vibrating atoms and ions. More recently, Buyers & Smith (1966), Melvin, Pirie & Smith (1968) and Reid (1974*a*, *b*) have applied the shell model to investigate the effects of ionic deformation on X-ray diffuse scattering from alkali halides. In its simplest form the shell model, originally developed by Dick & Overhauser (1958) to account for dielectric properties of alkali halides, separates the ion into a core and a shell coupled by a spring and has, with various modifications, proved a very useful classical parametrization of inter-ionic force constants. However, its ability to describe modifications of atomic scattering due to ionic deformation is less clear. Melvin, Pirie & Smith (1968) and Reid (1974*a*, *b*) plausibly assign to the core and the shell quantum-mechanically calculated charge distributions characteristic of the inner and outer electrons in the ion and, using the core–shell displacements calculated from the lattice-dynamical shell-model calculations, estimate the resultant changes in diffuse X-ray intensity due to the modified scattering factors. Robertson & Reid (1979) have made similar applications of the shell model to X-ray scattering from Si, while Reid & Pirie (1980) and Reid (1983) have used the model to estimate the effect of ionic deformation on Debye–Waller factors. Although March & Wilkins (1978) have developed a method of calculating elastic X-ray scattering from solids in terms of non-rigid pseudoatoms, this quantum-mechanical approach cannot be readily compared with semi-classical shell-model specifications of charge deformation.

In this paper we attempt an evaluation of the effectiveness of the simple shell model in specifying charge deformation of atoms and ions by comparing quantum-mechanical calculations of perturbed X-ray atomic scattering factors with shell-model predictions for isolated atoms or ions in electric fields. § 2 compares the shell-model theory with the corresponding quantum-mechanical theory. § 3 discusses the Kirkwood–Pople–Schofield method of calculating perturbed wave functions, while § 4 analyses the results of the calculations. Finally, § 5 discusses the validation of the shell model in simulating dynamical deformation of atomic/ionic scattering factors, and potential application of such mechanical models to describe static modifications of scattering factors in low-symmetry environments.

2. Theory

The X-ray atomic scattering factor for an unperturbed atom with N electrons is given by

$$f_0=\sum_{j=1}^{N}\int \psi_0^*\exp\left(i\chi\mathbf{S}\,.\,\mathbf{r}_j\right)\psi_0\,\mathrm{d}\tau, \qquad (1)$$

where $\psi_0(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_j, \ldots, \mathbf{r}_N)$ is the many-electron wave function of the atom, $\mathbf{S}=\mathbf{s}-\mathbf{s}_0$ is the difference

0108-7673/84/060716-06$01.50