vibration in Å of atoms i and j, and B_i and B_j are the isotropic thermal parameters. The values used in the calculation of ErOOH in Fig. 6 are those derived from the work of Christensen (1965). The density distribution of $Er(OH)_3$ was calculated from the thermal parameters as

$$U_{eq} = (U_{11}U_{22}U_{33})^{1/3} \quad \text{and} \quad B_{eq} = 8\pi^2 U_{eq}.$$

These thermal parameters were made available by Beall *et al.* (1977). A second term is added to b_{ij} to adjust for loss of pair correlation with increasing distance. This term is of the form

$$b_{ij},\ r = (r_{ij}/Kr_c)^2,$$

where K is a constant and r_c is 1 Å larger than the maximum r for the RDF calculation. The mean Er–O distance as shown in Fig. 6 is similar in the three systems amorphous hydrous erbium oxide, $Er(OH)_3$ and ErOOH. There is, however, a significant broadening of the first peak in the amorphous material when compared to that of the crystalline $Er(OH)_3$. This could imply a greater bond-length distribution in the amorphous sample or larger thermal parameters. Calculations of interference functions, RDF's and X-ray scattering curves for various nine-coordinated geometries are presently being performed on crystalline model systems in this laboratory. Further, it is believed by conjecture only that the studied erbium amorphous material could be a precursor to another form (termed in this lab as DR) which in turn transforms into $Er(OH)_3$ with ambient aging or with thermal aging. This DR form (Mullica, Milligan & Dillin, 1979) has not been structurally studied, but much effort is being spent to obtain meaningful structural data.

The authors thank the Robert A. Welch Foundation for financial support (grant No. AA-668). Further, the authors acknowledge appreciation for the help given by the facilities of Baylor University, Tarleton University and Arco Corp.

References

AZAROFF, L. V. (1955). *Acta Cryst.* **8**, 701–704.

BEALL, G. W., MILLIGAN, W. O. & WOLCOTT, H. A. (1977). *J. Inorg. Nucl. Chem.* **39**, 65–70.

CHRISTENSEN, A. N. (1965). *Acta Chem. Scand.* **19**, 1391–1396.

CROMER, D. T. (1969). *J. Chem. Phys.* **50**, 11, 4857–4859.

CROMER, D. T. & LIBERMAN, S. (1970). *J. Chem. Phys.* **53**, 5, 1891–1898.

DAMIENS, M. A. (1918). *Ann. Chem. (Warsaw)*, **10**, 137–183.

HAIRE, R. G., LLOYD, M. H., MILLIGAN, W. O. & BEASLEY, M. L. (1977). *J. Inorg. Nucl. Chem.* **39**, 843–847.

International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham: Kynoch Press.

JOYE, P. & GARNIER, C. (1912). *C. R. Acad. Sci.* **154**, 510–522.

KONNERT, J. H. & KARLE, J. (1973). *Acta Cryst.* A**29**, 702–710.

LEVY, H. A., AGRON, P. A. & DANFORD, M. D. (1959). *J. Appl. Phys.* **30**, 2012–2013.

LORCH, E. (1969). *J. Phys. C*, **2**, 229–237.

MILLIGAN, W. O., BEALL, G. W. & MULLICA, D. F. (1983). *Crystallography in North America*, Sect. F, Ch. 9, pp. 362–369. American Crystallographic Association.

MILLIGAN, W. O., MULLICA, D. F. & OLIVER, J. D. (1979). *J. Appl. Cryst.* **12**, 411–412.

MULLICA, D. F., BEALL, G. W., MILLIGAN, W. O. & OLIVER, J. D. (1979). *J. Appl. Cryst.* **12**, 263–266.

MULLICA, D. F., MILLIGAN, W. O. & DILLIN, D. R. (1979). *J. Cryst. Growth*, **47**, 635–638.

TAYLOR, M. (1979). *J. Appl. Cryst.* **12**, 442–449.

WAGNER, C. N. J., LEE, D., TAI, S. & KELLER, L. (1980). *Advances in X-ray Analysis*, pp. 245–252. New York: Plenum Press.

WARREN, B. E. (1969). *X-ray Diffraction.* Reading, MA: Addison-Wesley.

WARREN, B. E. (1975). *J. Appl. Cryst.* **8**, 674–677.

WARREN, B. E. & MOZZI, R. L. (1966). *Acta Cryst.* **21**, 459–461.

WYCKOFF, R. W. R. (1960). *Crystal Structures*, Vol. 2. New York: Interscience.

Acta Cryst. (1984). A**40**, 269–277

Improvement of Protein Phases by Coarse Model Modification

BY V. YU. LUNIN AND A. G. URZHUMTSEV

Research Computer Centre, USSR Academy of Sciences, 142 292 Pushchino, Moscow Region, USSR

(*Received* 25 *April* 1983; *accepted* 2 *December* 1983)

Abstract

A procedure is suggested for the refinement of a set of protein phases and for its extension to a higher resolution, which is a development of the approach of Agarwal & Isaacs [*Proc. Natl Acad. Sci. USA*, (1977), **74**, 2835–2839]. A new set of phases is obtained by combining the starting phases with those calculated from a stereochemically non-conditioned coarse 'atomic' model which is automatically constructed and subjected to a least-squares refinement in reciprocal space. The method has been tested with actinidin data generated from atomic coordinates. Starting from the phases calculated to 3 Å resolution

0108-7673/84/030269-09$01.50

and the amplitudes calculated to 2 Å resolution a new set of phases was obtained with a mean error of 31° for 12 713 non-centrosymmetric reflections in the range to 2 Å. The refinement of the phases to 3 Å resolution for γ-crystallin IIIb from calf lens and its extension to 2·7 Å resolution resulted in a noticeable improvement in the electron density map.

1. Introduction

In structure determinations of large molecules by the method of isomorphous replacement the phase information is often limited in resolution, not extending beyond 3·5–3·0 Å. These phases, for a number of reasons, may be subjected to considerable errors. At the same time, a substantial portion of the structure-factor amplitudes known from an X-ray experiment are usually discarded from electron density computation because the phases remain uncertain. Over the last decade a number of methods have been elaborated to improve the phases available (by the improvement we assume here the refinement of the phases of the starting set and the determination of new phases which were not present in the starting set).

In this article we describe a development of an approach suggested by Agarwal & Isaacs (1977) and some of its applications. It proceeds from the construction of the coarse 'atomic' model which reproduces as closely as possible the electron density in the starting map. Then the parameters of this model are allowed to refine in the reciprocal space and the refined model is used to calculate the phases of the extended structure-factor set. These new phases along with observed amplitudes are applied to calculate an improved electron density map. A similar idea was put into practice for myoglobin-like proteins at a low resolution (Borovikov, Vainstein, Gelfand & Kalinin 1979; Kalinin, 1980).

It should be noted that such a general concept may be realized in various ways, and these different realizations may essentially affect the result of the extension. Agarwal & Isaacs extended a set of phases for insulin from 3·0 to 1·9 Å (with mean discrepancy 65° between the extended phases and those calculated from the refined model).

In our work, we modify their method. Firstly, we propose a new algorithm for a coarse model construction which implies a more accurate evaluation of the contribution of an individual atom to the weighted synthesis of a given resolution, atomic-distance restrictions being excluded. Secondly, the difference synthesis technique is used. The results of a test with actinidin (extension from 3·0 to 2·0 Å with a mean error of 31°) proves our approach.

Some checks were employed on the phase extension which will be described in § 5.

2. Methods

2.1. *Generation of a coarse 'atomic' model*

As a first step for phase improvement, a coarse 'atomic' model is constructed to satisfy the condition that the set of phases calculated from the model should be close enough to the starting set of phases. The condition given above would be satisfied if the electron density represented by the coarse model is close enough to that in the starting map. It should be emphasized that no stereochemical restriction is imposed on the model. The 'atomic' model here is just a way to describe the starting set of phases.

If we consider these 'atoms' as an approximation to the real ones, the absolute deviation from their positions is important rather than their mutual arrangement. We may assume that the model is composed of atoms of one kind ('nitrogen atoms') described by their coordinates $\mathbf{r}_j$ and individual isotropic temperature factors B_j. In the absence of stereochemical restrictions strong peaks of electron density would generate several nearly coinciding atoms. The number of atoms included in the first model is set to exceed the total number of non-hydrogen atoms in the real structure. One reason is the approximation of strong peaks by several atoms, the other is concerned with the fact that in the course of refinement 'poor atoms' would be removed so that the final number of atoms would be close to that in the real structure.

The construction of the model is based on the principle of 'exhausting' the electron density map. The procedure may be described as follows. The first dummy atom is placed at point $\mathbf{r}_1$ in the starting map $\rho_{st}(\mathbf{r})$ with the maximal value of electron density. If the density of the dummy atom at its center is lower than the density of the starting map of this point, the value of B_1 is set to zero. Otherwise, the density of the dummy atom is lowered to $\rho_{st}(\mathbf{r}_1)$ by a larger value of B_1. Then the contribution of the dummy atom is subtracted from the electron density map and the procedure is repeated with a second dummy atom, *etc.* Such a successive incorporation of dummy atoms would be performed until the selected number of atoms is accumulated or the maximal value of the residual electron density, $\rho_{st}(\mathbf{r}) - \sum_j \rho_j(\mathbf{r})$ drops below the given value ρ_{crit}.

It should be noted that here is given the principle scheme of model construction, the practical algorithm being essentially modified to save computer time.

When applying the above procedure two important questions arise: what kind of Fourier terms are to be used in the calculation of a starting electron density map, $\rho_{st}(\mathbf{r})$, and how to estimate the contribution $\rho_j(\mathbf{r})$ to the map for an atom with coordinates $\mathbf{r}_j$ and temperature factor B_j? In the calculation of the starting Fourier synthesis we assume that for each phase

the probability distribution $P_{st}(\varphi, s)$ is known *a priori* (the probability curve may be derived from any available method of phase determination: isomorphous replacement, anomalous scattering, non-crystallographic averaging *etc.*). Then the minimal r.m.s. error would correspond to the following expression:

$$\rho_{st}(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{s} \in S} m(\mathbf{s}) F_0(\mathbf{s}) \exp [i\varphi_{st}(\mathbf{s}) - 2\pi i(\mathbf{s}, \mathbf{r})], \quad (1)$$

where V is the volume of the unit cell, $F_o(s)$ are the observed amplitudes, S is a set of reciprocal-lattice points with the known starting phase probability distribution $P_{st}(\varphi, \mathbf{s})$, $m(\mathbf{s})$ and $\varphi_{st}(\mathbf{s})$ are a figure of merit and the best phase derived from the expression

$$m(\mathbf{s}) \exp [i\varphi_{st}(\mathbf{s})] = \langle \exp (i\varphi) \rangle = \int_0^{2\pi} \exp (i\varphi) P_{st}(\varphi ; \mathbf{s}) \, d\varphi.$$

In the absence of additional phase information, this Fourier synthesis may be used as the starting one (see § 2.5).

For a dummy atom with parameters $\mathbf{r}_j$ and B_j its contribution to the starting synthesis may be calculated as

$$\rho_j(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{s} \in S} m(\mathbf{s}) f(s) \times \exp (-Bs^2/4) \exp[-2\pi i(\mathbf{s}, \mathbf{r} - \mathbf{r}_j)],$$

where S is the set of phases and $m(\mathbf{s})$ is the weighting factor which coincides with that in (1); $f(s)$ is the atomic form factor for a nitrogen atom. It has been shown (see Appendix 1) that for all the atoms composing the coarse model $\rho_j(r)$ may be well described by a Gaussian function:

$$\rho_j(\mathbf{r}) = C_0[4\pi/(B_0 + B_j)]^{3/2} \times \exp \{-4\pi^2 |\mathbf{r} - \mathbf{r}_j|^2/(B_o + B_j)\},$$

where constants C_0 and B_0 depend only on the set of reciprocal-lattice points presented in $\rho_{st}(\mathbf{r})$ and on the weight function $m(\mathbf{s})$. These constants may be changed when going to a higher resolution or applying other probability distributions for the starting set of phases.

2.2. *Refinement of the model*

The coarse model constructed as described may be modified by the changing of N (the number of atoms in the model), $\mathbf{r}_j$, B_j, $j = 1, \ldots, N$, to provide the least possible discrepancy between structure amplitudes $F_c(\mathbf{s})$ calculated from the model and observed structure amplitudes $F_o(\mathbf{s})$:

$$R_2 = \sum_{\mathbf{s}} [F_c(\mathbf{s}) - F_o(\mathbf{s})]^2 \Rightarrow \min. \quad (2)$$

The minimization may be done by the reciprocal-space refinement (Agarwal, 1978) alternately applied to coordinates $\mathbf{r}_j$ and temperature factors B_j.

The first cycle of the refinement is based only on reflexions present in the starting set of phases, then reflexions at a higher resolution are gradually added and in the final cycles all reflexions for which $F_o(\mathbf{s})$ is available are included except those with $d > 10$ Å strongly affected by the solvent (Phillips, 1980).

In the course of the refinement the list of atoms may be periodically checked for the value of B to remove atoms with an unreasonably great B (evenly 'smeared' over the unit cell). The number of atoms in the model is thus gradually reduced.

2.3. *Combining phase information*

The phases calculated from the model at the final stage of refinement may be improved when correctly combined with the starting-phase information. The principle of combining is provided by the Bayes theorem.

Following an approach of Srinivasan & Parthasarathy (1976) and assuming that the difference between the coarse model and the real structure is entirely due to random independent errors in atomic coordinates, we may derive a conditional distribution of probability for phases $\varphi_c(\mathbf{s})$, providing $F_o(\mathbf{s})$, $F_c(\mathbf{s})$ and $\varphi(\mathbf{s})$ are known:

$$P_c(\varphi_c | F_c, F_o, \varphi ; \mathbf{s}) \sim \exp \left\{ 2 \frac{\alpha}{\beta} F_o F_c \cos (\varphi - \varphi_c) \right\}, \quad (3)$$

where α and β are some parameters describing the 'coarseness' of the model which still should be determined. A normalization factor is omitted. Taking the starting distribution $P_{st}(\varphi, \mathbf{s})$ as a prior distribution of probability for φ, a distribution $P_{comb}(\varphi, \mathbf{s})$, after the random variable φ_c is assumed to have the value calculated from the model, may be derived according to the Bayes theorem (neglecting a normalization factor) as:

$$P_{comb}(\varphi ; \mathbf{s}) \sim P_{st}(\varphi ; \mathbf{s}) P_c(\varphi_c | F_c, F_o, \varphi ; \mathbf{s}).$$

For the lattice points $\mathbf{s}$, not included in the starting set S, *a priori* distribution $P_{st}(\varphi, \mathbf{s})$ may be taken to be uniform:

$$P_{st}(\varphi ; \mathbf{s}) = \frac{1}{2\pi} \quad \text{for } \mathbf{s} \notin S.$$

The main problem in combining the phase information in this way is concerned with the determination of α and β involved in (3). Bricogne (1976) suggested for $t = \alpha/\beta$ the expression

$$t = [\langle |F_o^2 - k^2 \exp(Bs^2/2) F_c^2| \rangle]^{-1} = t(\mathbf{s}), \quad (4)$$

where k and B are derived from the Wilson plot when $F_c(\mathbf{s})$ are set to an absolute scale. Another estimate of α and β based on the maximum of a likelihood

function (Lunin, 1982) is briefly described in Appendix 2.

Special tests showed a good estimate for the distribution parameters from the maximum likelihood approach when the model is not refined in the reciprocal space. Otherwise the method tends to overestimate: a mean error derived from the distribution (3) using the value of t from the maximum likelihood approach proved to be about half a real value. So, when applied to γ-crystallin IIIb data (see § 4), the value of t, calculated from the maximum likelihood function, was artificially lowered to double the mean phase error predicted from $P_c(\varphi)$ (see § 5.4).

2.4. *Extension of a partial model against different maps*

Our tests with actinidin (see § 3) demonstrated a marked phase improvement through the use of the following modification of the method described above. In the initial model after refinement only 'the strongest atoms' (those with a low value of B) are kept (about one half), the rest of them being removed from the model. Then the coarse model is filled with the additional dummy atoms in a way similar to that described in 2.1, by 'exhaustion' of the difference synthesis:

$$\rho_d(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{s} \in S'} m_{\text{comb}}(\mathbf{s}) [F_o(\mathbf{s}) \exp [i\varphi_{\text{comb}}(\mathbf{s})] - F_p(\mathbf{s}) \exp [i\varphi_p(\mathbf{s})] \exp [-2\pi i(\mathbf{s}, \mathbf{r})], \quad (5)$$

where m_{comb} is a figure of merit and φ_{comb} is 'the best' phase calculated from $P_{\text{comb}}(\varphi; \mathbf{s})$; $F_p(\mathbf{s}) \exp [i\varphi_p(\mathbf{s})]$ are structure factors calculated from the strongest atoms in the coarse model with $B < B_{\text{lim}}$. It should be noted that weights $m_{\text{comb}}(\mathbf{s})$ and the set of lattice points here are different from those present in (1), resulting in different parameters B_0 and C_0 describing the contribution of each atom to the $\rho_d(\mathbf{r})$.

After the insertion of additional atoms generated from $\rho_d(\mathbf{r})$, the model is subjected to the same refinement procedure, described in § 2.2. Such a substitution of weak atoms of the coarse model by atoms extracted from the difference map may be repeated several times. Test calculations showed that this procedure is more efficient than the construction of a new coarse model using the improved map, calculated from $m_{\text{comb}} F_o \exp (i\varphi_{\text{comb}})$, as was proposed by Agarwal & Isaacs.

2.5. *Accounting for the known structural information*

Up to this point we have assumed that nothing is known about the structure which is under study apart from the starting probability distribution $P_{\text{st}}(\varphi; \mathbf{s})$ for $\mathbf{s} \in S$ and the set of structure amplitudes $F_o(\mathbf{s})$ for $s \in S'$. However, if some structural information is available, it may be taken into account in the construction of a coarse model. For instance, if the molecular envelope is known, it is reasonable to include in the model only atoms inside the envelope (*i.e.* the 'exhaustion' of the starting map should be limited only by the region inside the envelope). In a similar way the coarse model may cover only some large continuous regions of high electron density determined as indicated by Bhat & Blow (1982).

If we have already the stereochemically consistent model accounting for a part of the structure (the electron density map is partially interpreted) we may improve the set of phases by the modification of a mixed model that is the partial model complemented with dummy atoms in the non-interpreted regions fitted in the difference map (5). As the fraction of stereochemically 'correct' atoms increases, the phase improvement by the mixed-model modification is transformed into the usual crystallographic refinement of the atomic model.

2.6. *Geometrical interpretation of the method*

The procedure described here allows a simple geometrical interpretation both in direct and reciprocal space.

In direct space the starting coarse model of the crystalline structure may be considered just as a way to describe the initial electron density map by a relatively small number of parameters $\mathbf{r}_j$, B_j, $j = 1, \ldots, N$ (to compare with the number of grid points taken in computation of the map). Further steps may be interpreted as a modification of the map through a variance of these parameters in such a way that the structure amplitudes derived after Fourier inversion of the map are close enough to the experimental values. It may be noted that without minimization of the model according to (2), when the phases are obtained from the 'mixed' model (a partial model plus dummy atoms in non-interpreted regions of the map), the phases would be close to those obtained by the method of Bhat & Blow (1982). A minimization procedure applied to the non-interpreted regions would provide a further gain in accuracy of phases.

A more profound interpretation of the method in reciprocal space draws our attention to the fact that for improvement of the set of phases we impose strong additional limitations restricting ourselves only to those structure factors which correspond to a structure consisting of discrete atoms. This requirement for 'atomicity' leaves us with a much more restricted class of allowed structure factors. A modification of the coarse model implies a search of 'atomic' structures and a selection of that one which gives the closest agreement with the set of observed structure amplitudes $F_o(\mathbf{s})$.

Table 1. *Distribution of an average phase error* $\langle|\varphi-\varphi_c|\rangle$ *(in ° for non-centrosymmetric reflexions) on the test phase improvement for actinidin*

	Limits, s^2										Total
Model	0·000–0025	0·025–0050	0·050–0·075	0·075–0·100	0·100–0·125	0·125–0·150	0·150–0·175	0·175–0·200	0·200–0·225	0·225–0·250	0·000–0·250
M_0	21	20	19	27	47	63	68	70	72	76	57
M_7	17	14	14	22	31	43	51	53	53	58	42
M_{12}	21	18	18	24	32	41	49	51	55	59	42
M_{20}	17	14	17	22	27	35	40	43	43	50	35
M_{24}	17	14	13	19	24	29	34	37	39	44	31
M_{100}	24	28	36	40	47	—	—	—	—	—	—
M_{101}	17	15	20	25	31	—	—	—	—	—	—
Number of reflexions	320	681	957	1078	1278	1475	1502	1691	1783	1948	12 713

3. Test phase improvement for actinidin

3.1. *Extension of the set of phases*

The technique presented above was tested using data for actinidin (Lunin & Urzhumtsev, 1981, 1983). The structure factors $F(s)\exp[i\varphi(\mathbf{s})]$ to 2 Å resolution were calculated from atomic coordinates deposited with the National Bureau of Standards Protein Data Bank (Bernstein *et al.*, 1977). The magnitudes $F(\mathbf{s})$ were considered further as observed structure amplitudes $F_o(\mathbf{s})$ and phases $\varphi(\mathbf{s})$ simulated a starting set of phases $\varphi_{st}(\mathbf{s})$. The phases in the resolution range $2{\cdot}0 \le d < 3{\cdot}0$ Å were eliminated from the refinement process and were used only to check the result.

The electron density map was calculated at 3·0 Å resolution using 4641 independent Fourier terms, and the coarse model M_0 was automatically built with 2168 atoms (to compare with 1653 non-hydrogen atoms in the actinidin structure). The distribution of the mean phase error over the ranges of s^2 for structure factors calculated from the model M_0 is given in Table 1. Then the model M_0 was incorporated into the refinement process (18 cycles for the refinement of $\mathbf{r}_j$ and 12 cycles for the refinement of B_j, with a gradual extension of the resolution from 3·0 to 2·0 Å). The phases reproduced from the resulting model M_7 showed significantly less error (Table 1). When a new coarse model was generated from the electron density map at 2·5 Å resolution calculated with the model M_7 phases, but after its refinement no substantial improvement was obtained (model M_{12} in Table 1). On the other hand, when the technique of completion of a partial model by new atoms fitting the difference map was applied at 2·5 Å resolution (addition of 1793 atoms to the strongest 1140 atoms of model M_7), the refinement of the resulting model led to model M_{20} reproducing phases with a reduced error (Table 1). The same procedure using the difference map at 2·0 Å resolution with M_{20} phases resulted in a further reduction of the mean phase error (model M_{24} in Table 1).

3.2. *Phase refinement*

This test was the same as the previous one, but a random error with the mean value of 36° was introduced into the starting set of phases at 3·0 Å resolution. The refinement was limited to the 3·0 Å resolution data. The comparison of the starting (M_{100}) and refined (M_{111}) phases with phases of the original set is also given in Table 1.

4. Phase improvement for γ-crystallin IIIb

4.1. *Description of the data*

γ-crystallin IIIb is a protein from calf lens of molecular weight about 20 000 daltons. Crystals of the protein used for X-ray work belong to space group $P2_12_12_1$, with unit-cell dimensions $a = 58{\cdot}7$ Å, $b = 69{\cdot}5$ Å and $c = 116{\cdot}9$ Å (Chirgadze, Sergeev, Fomenkova & Oreshin, 1981). There are two protein molecules in the asymmetric unit related by a non-crystallographic twofold screw axis. It should be noted that the refinement process described below was applied to the whole asymmetric unit taking no advantage from the non-crystallographic symmetry. The structure amplitudes available corresponded to the set to 2·7 Å resolution. The total set of amplitudes comprised about 10 500 reflexions in the resolution range between 20 and 2·7 Å with $F > 2\sigma$, where σ is a standard deviation. Within the range below 3·0 Å resolution the number of reflexions measured was 8533. The starting phases were determined for 6777 reflexions by the multiple isomorphous replacement method (Chirgadze *et al.*, 1981). The electron density map calculated with these data made it possible to trace the course of the polypeptide chain. In the absence of the sequence data the interpretation of the side-chain density encountered many difficulties and required some improvement of the electron density map. For this purpose the procedure described above was applied.

Table 2. *Plan for refinement of the coarse model of crystallin structure (for γ-crystallin* IIIb)

Initial model	Number of atoms	B_{lim}	Number of atoms with $B \le B_{lim}$	d_{min} (Å)	Number of reflexions	*R* factor Before refinement	*R* factor After refinement	Model obtained
M_0	3104		3104	3·0	8 533	0·353	0·166	M_2
M_2	3104	80	3052	3·0	8 533	0·166	0·115	M_4
M_4	3052	80	3045	2·7	10 527	0·169	0·119	M_7
M_7	3045	70	2962	2·5	11 727	0·146	0·123	M_9
M_{11}	3632		3632	3·0	8 533	0·248	0·120	M_{13}
M_{13}	3632	80	3628	2·7	10 527	0·152	0·115	M_{15}
M_{15}	3628	70	3614	2·5	11 727	0·136	0·114	M_{17}
M_{17}	3614	40	3229	2·3	12 510	0·172	0·142	M_{20}
M_{20}	3229		3229	2·3	12 510	0·136	0·116	M_{22}

4.2. *Construction and modification of the coarse model*

The starting map was calculated at 3·0 Å resolution using 6777 reflexions weighted with figures of merit. An absolute scale was derived from the Wilson plot. The starting phase probability distribution $P_{st}(\varphi; \mathbf{s})$ was presented in a general form

$$P_{st}(\varphi; \mathbf{s}) \sim \exp\{A(\mathbf{s}) \cos \varphi + B(\mathbf{s}) \sin \varphi + C(\mathbf{s}) \cos 2\varphi + D(\mathbf{s}) \sin 2\varphi\},$$

by the method of Hendrickson & Lattman (1970). The values of m and φ_{st} calculated from this distribution only slightly differed from similar values derived by the method of Ten Eyck & Arnone (1976) (see Table 3 and Fig. 3). For the contribution of a single nitrogen atom the parameters $B_0 = 83{\cdot}9$ Å^2 and $C_0 = 8{\cdot}4$ e were applied (from an estimate based on the structure-factor set and values of m, see Appendix 1). Then the coarse model was automatically built using the procedure described in § 2.1. This model M_0 was composed of 3104 dummy atoms representing the two protein molecules in the asymmetric unit.

The refinement of the model M_0 was done by series composed of two cycles of refinement of B_j and three cycles of refinement of $\mathbf{r}_j$. After each refinement of B_j the atoms with $B_j > B_{lim}$ were excluded from the model. The statistics of the refinement are given in Table 2. After calculation of the model M_9 only 2602 atoms were kept with $B_j \le 50$ Å^2 and 1030 additional atoms were inserted from the difference map calculated at 3·0 Å resolution (see § 2.4). A contribution from a nitrogen atom to the difference map was estimated as $B_0 = 64{\cdot}2$ Å^2 and $C_0 = 10{\cdot}3$ e, *i.e.* sharpening of peaks was observed compared with the starting map. The model M_{11} obtained in this way was further refined as shown in Table 2.

4.3. *The final map*

The final improved electron density map was calculated using phases generated from the last model M_{22} and combined with phases of the starting set as described in § 2.3, except for the phases with $d >$ 10·0 Å. Owing to the strong influence of the solvent on structure factors in this region the latter were omitted from the refinement of the coarse model. Test calculations showed that phases of reflexions in this range of resolution are determined with a considerable error. Therefore, for reflexions with $d > 10{\cdot}0$ Å the starting phases were applied in the calculation of the final map. Fig. 1 shows the same portion of the electron density map for the starting map, the improved map at 3·0 Å resolution and the final map at 2·7 Å resolution.

5. Checking the process of phase improvement

5.1. *Change in R factor*

The coarse-model modification was carried out to minimize (2). A similar estimate of the quality of the coarse map may be obtained from the value of the *R* factor:

$$R = \frac{\sum_{\mathbf{s}} |F_o(\mathbf{s}) - F_c(\mathbf{s})|}{\sum_{\mathbf{s}} F_0(\mathbf{s})}.$$

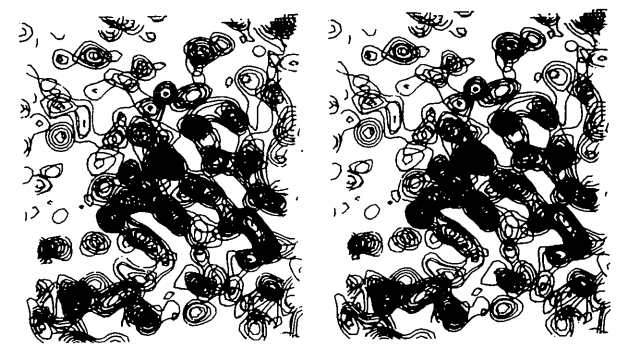
The changes in the *R* factor in the course of refinement are given in Table 2. It should be stressed that these values of *R* factor bear no correlation with those obtained in the true crystallographic refinement of protein structures. Owing to an excessive number of atoms and the absence of any stereochemical restrictions, the coarse model possesses many more degrees of freedom in comparison with a stereochemically reasonable model of the protein structure, allowing for lower values of R_2 and *R*. The values of *R* may therefore be used only as indicators of the coarse-model refinement and for comparison of coarse models with each other.

5.2. *Increase in the maximal value of electron density*

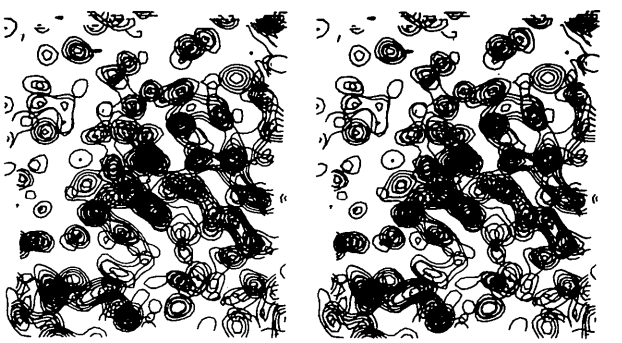
The electron density map based on the more accurate phases usually shows more sharpened peaks. However, a reasonable quantitative estimate criterion is difficult to find. An increase in the maximal value of electron density in the map may be achieved giving more weight to high-angle Fourier terms without change in phases (a sharpening of the map), so this

increase may bear no sign of phase improvement. The increase of ρ_{max} when unweighted terms are altered is a more reliable criterion of phase improvement, however even in that case the systematic errors in phases may lead to strong false peaks, and no guarantee for phase improvement may be given. In the refinement of γ-crystallin IIIb the value of ρ_{max} for the unweighted starting map, the improved 3·0 Å map and the final 2·7 Å map was 2·29, 2·40 and 2·64 e Å^{-2}, respectively.

(*a*)



(*b*)



(*c*)

Fig. 1. Stereopairs of a fragment of the distribution of electron density of γ-crystallin IIIb designed with different sets of phases: (*a*) starting set of phases, 3·0 Å; (*b*) improved set of phases, 3·0 Å; (*c*) improved set of phases, 2·7 Å.

5.3. *Deviation of phases from starting values*

One reasonable check for the quality of phases generated from a model may be extracted from an analysis of variance of $\delta(m)$, where δ is the mean difference $|\varphi_{st}-\varphi_c|$ between starting and calculated phases for reflexions having a figure of merit close to m. To illustrate what $\delta(m)$ looks like in different cases let us consider three 'idealized' situations. If φ_c are close enough to the true values, a deviation from the starting phases would be less as the figure of merit m increases, *i.e.* $\delta(m)$ is a decreasing function, with $\delta(1)=0$ and $\delta(0)$ close to 90°. If the set of φ_c is represented by random values $\delta(m)$ would be equal to 90° for any value of m. If the set of φ_c just reproduces the set of $\varphi_{st}(s)$, *i.e.* $\varphi_c=\varphi_{st}$ for all reflexions, $\delta(m)=0$ for all values of m. In Fig. 2 function $\delta(m)$ is plotted for phases generated from the model M_{22}. Although a plot of this kind could not be a guarantee of the correctness of the set of phases $\varphi_c(s)$, any other type of function would imply significant errors in $\varphi_c(s)$ or some unrealiable estimates of the starting set of phases. Therefore, this test does not raise any doubts on the improvement of the phases for γ-crystalline IIIb.

5.4. *Estimate of a mean phase error*

An *a posteriori* probability distribution $P_{comb}(\varphi)$ calculated from the extended set of phases allow for each phase the determination of its figure of merit and the mean error:

$$\Delta=\langle|\varphi-\varphi^{best}|\rangle=\int_0^{2\pi}|\varphi-\varphi^{best}|P_{comb}(\varphi)\,d\varphi.$$

Estimated mean errors in phases in ranges of s^2 for the improved set of phases are given in Table 3. In Fig. 3 the figure of merit is plotted *versus* s for different sets of phases. It may be noted that for reflexions with $d>10$ Å, omitted in the last stages of refinement, figures of merit estimated from $P_{comb}(\varphi)$ are lower than the starting values. For this reason in calculation of the final map the starting values of the figure of merit and phases were used for reflexions in this range.

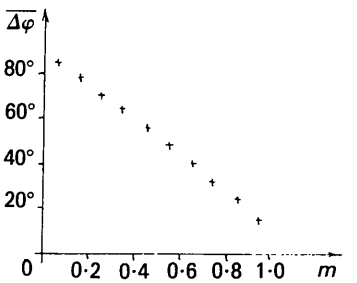


Fig. 2. Dependence of $\langle|\Delta\varphi|\rangle=\langle|\varphi_{st}-\varphi_{M_{22}}|\rangle$ on the values of the figure of merit m of φ_{st} phases for γ-crystallin IIIb.

Table 3. *Evaluation of an average error for phase design with respect to coarse models of γ-crystallin* IIIb *(in °, for non-centrosymmetric reflections)*

	Limit, s^2									Total
Model	0·00–0·02	0·02–0·04	0·04–0·06	0·06–0·08	0·08–0·10	0·10–0·12	0·12–0·14	0·14–0·16	0·16–0·18	0·00–0·18
M_0	38	36	31	38	41	52	90	84	75	51
M_2	54	26	20	22	24	27	90	90	80	41
M_4	50	20	16	16	16	24	90	85	77	37
M_7	48	18	14	16	16	20	42	90	77	30
M_9	50	18	14	14	16	20	26	34	76	23
M_{11}	30	24	17	19	20	26	37	43	81	27
M_{13}	50	22	16	16	16	16	40	56	77	27
M_{15}	48	18	14	16	16	18	28	61	90	26
M_{17}	46	18	14	14	14	18	20	30	90	22
M_{20}	48	22	16	16	16	18	20	26	36	22
M_{22}	48	20	14	14	16	16	18	22	28	18
Number of reflexions	504	1068	1442	1759	1667	1634	1360	872	417	10723

5.5. *A visual check*

The main criterion for the correctness of the improved set of phases requires that the map appears to be more interpretable. As an example, Fig. 1 shows a fragment of the map in the region $0{\cdot}17 \le x \le 0{\cdot}67$, $0{\cdot}15 \le y \le 0{\cdot}65$, $0{\cdot}16 \le z \le 0{\cdot}25$ (in fractional coordinates). A comparison of the starting map and the improved one at 3·0 Å resolution may illustrate the progress achieved.

Going now to the map at 2·7 Å resolution (about 2000 more reflexions are added to those taken for the map at 3·0 Å resolution) (Fig. 1) the extent of improvement by visual estimates seems to be smaller relative to that within the 3·0 Å zone. We may notice more fine details in the map, but the density along the chain is broken at several places at high resolution.

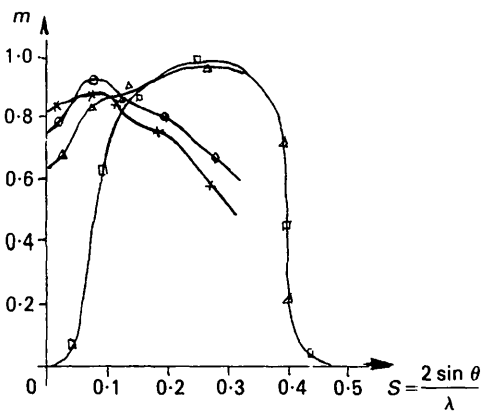


Fig. 3. Plot showing the average value of the figure of merit m against $s = 2 \sin \theta/\lambda$ for different sets of phases for γ-crystallin IIIb: × φ_{best} phases, technique of Ten Eyck & Arnone (1976); ○ φ_{st} phases, technique of Hendrickson & Lattman (1970); □ phases designed according to M_{22} model; △ φ_{comb} phases designed with respect to distribution $P_{comb}(\varphi) \simeq P_{st} P_{M_{22}}(\varphi)$.

5.6. *Use of non-crystallographic symmetry*

It has been mentioned that the presence of a non-crystallographic twofold screw axis was not used for the phase refinement. The same concerns the starting-phase probability distribution $P_{st}(\varphi)$. Therefore, similar features seen in symmetrically related portions of the improved map may be interpreted in favour of the real existence of these features in the protein structure rather being due to phase errors.

6. Conclusions

(*a*) It has been shown by tests and structural investigations on γ-crystallin IIIb that the modification of a coarse model may serve as an efficient instrument for improvement of electron density maps and extension of resolution.

(*b*) An algorithm of the coarse-model construction described above allows a model to be built which well reproduces the starting set of phases and is suitable for its modification leading to improvement of the phases.

(*c*) The procedure is more efficient when weak atoms are periodically eliminated from the model and new atoms are added in the difference maps.

(*d*) An improvement of the electron density map for γ-crystallin IIIb has been achieved by phase refinement at 3·0 Å resolution and an extension of the phases to 2·7 Å resolution.

The authors thank Drs Yu. N. Chirgadze, Yu. V. Sergeyev and N. P. Fomenkova for their interest in the development of the technique described and active promotion of its application in the X-ray study of γ-crystallin IIIb. The authors are indebted to Dr E. E. Shnol for his constant attention and support.

APPENDIX 1

Approximation of a single-atom contribution to the weighted synthesis of limited resolution

If the formula

$$\rho(r)=\frac{1}{V}\sum_{\mathbf{s}\in S} m(\mathbf{s})F(\mathbf{s})\exp\left[i\varphi(\mathbf{s})\exp\left[-2\pi i(\mathbf{s},\mathbf{r})\right]\right.$$

was applied to calculate the electron density synthesis, then a separate atom with the temperature factor B_j^t, $\mathbf{r}_j$ centre coordinates and $f_j(s)$ as an atomic scattering function may contribute to the extent of $\rho_j(\mathbf{r}-\mathbf{r}_j)$, where

$$\rho_j(\mathbf{r})=\frac{1}{V}\sum_{\mathbf{s}\in S} m(\mathbf{s})f_j(\mathbf{s})\exp\left[-B_j^t s^2/4\right]\exp\left[-2\pi i(\mathbf{s},\mathbf{r})\right]$$

is the contribution of the atom removed to the origin. Our aim is to find an approximation for $\rho_j(\mathbf{r})$ allowing the calculation of the $\rho_j(\mathbf{r})$ function with different values of B_j^t without resorting each time to summation of the three-dimensional Fourier series.

It is well known that the Fourier transform is used to convert the product of functions into convolution of their images. Hence,

$$\rho_j(\mathbf{r})=\left\{\frac{1}{V}\sum_{\mathbf{s}\in S} m(\mathbf{s})f_j(\mathbf{s})\exp\left[-2\pi i(\mathbf{s},\mathbf{r})\right]\right\}$$
$$*\left\{\frac{1}{V}\sum_{\mathbf{s}\in R'}\exp\left[-B_j^t s^2/4\right]\exp\left[-2\pi i(\mathbf{s},\mathbf{r})\right]\right\},$$

where $*$ denotes convolution. The first component on the right-hand side of the equality does not depend on B_j^t, so it is subjected to approximation by the Gaussian function (after calculating the appropriate series only once)

$$C_0(4\pi/B_0)^{3/2}\exp\left\{-4\pi^2R^2(\mathbf{r})/B_0\right\},$$

where $R(\mathbf{r})$ is the distance from the origin to point $\mathbf{r}_0$. C_0 and B_0 coefficients are, of course, dependent on the weight function $m(\mathbf{s})$ and the set of reflexions s, incorporated in the synthesis. As

$$\frac{1}{V}\sum_{\mathbf{s}\in R'}\exp\left[-B_j^t s^2/4\right]\exp\left[-2\pi i(\mathbf{s},\mathbf{r})\right]$$
$$=\sum_{\mathbf{u}\in R}(4\pi/B_j^t)^{3/2}\exp\left\{-4\pi^2R^2(\mathbf{r}+\mathbf{u})/B_j^t\right\}$$

(where R is the direct-space lattice) and at every $\mathbf{r}$ point the contributions of the Gauss functions, shifted from the origin by the vector of lattice translation, are negligible, hence

$$\frac{1}{V}\sum_{\mathbf{s}\in R'}\exp\left[-B_j^t s^2/4\right]\exp\left[-2\pi i(\mathbf{s},\mathbf{r})\right]$$
$$\simeq(4\pi/B_j^t)^{3/2}\exp\left\{-4\pi^2R^2(\mathbf{r})/B_j^t\right\}.$$

That is, the convolution of two Gaussian functions may be represented as

$$\rho_j(\mathbf{r})\simeq C_0\left(\frac{4\pi}{B_0+B_j^t}\right)^{3/2}\exp\left\{\frac{-4\pi^2R^2(\mathbf{r})}{B_0+B_j^t}\right\}. \quad (6)$$

APPENDIX 2

Evaluation of distribution parameters $P_c(\varphi_c|F_c, F_o, \varphi)$ out of the likelihood function maximum

With the assumption that a coarse model differs from a true structure in having random independent errors in the coordinates of model atoms, it is possible to demonstrate (Srinivasan & Parthasarathy, 1976) that the φ_c phase, which was calculated according to the model, corresponds to a probability distribution

$$P_c(\varphi_c|F_c, F_o, \varphi; \mathbf{s})\simeq\exp\left\{2\frac{\alpha}{\beta}F_oF_c\cos(\varphi-\varphi_c)\right\}$$

while modulus F_c complies with

$$P_{c,F}(F_c|F_o; \mathbf{s})=\frac{2F_c}{\mu}\exp\left\{-\frac{F_c^2+\alpha F_o^2}{\beta}\right\}I_o\left(2\frac{\alpha}{\beta}F_cF_o\right), \quad (7)$$

where α and β are identical and denote the crudeness of the model.

We consider α and β to be constant in the narrow spherical layer in s^2. We also regard $\tilde{F}_c$ (calculated according to the model) as realization of an independent random value F_c with distribution (7).

This suggests that the values of α and β should be chosen with a view to maximizing the likelihood function

$$\Psi(\alpha,\beta)=\prod_{\mathbf{s}}2\tilde{F}_c(s)/\beta\exp\left\{-\left[\tilde{F}_c^2(\mathbf{s})+\alpha F_o^2(\mathbf{s})\right]/\beta\right\}$$
$$\times I_o\left[2\frac{\alpha}{\beta}F_o(\mathbf{s})\tilde{F}_c(\mathbf{s})\right]$$

(Lunin, 1982).

References

AGARWAL, R. C. (1978). *Acta Cryst.* **A34**, 791–809.

AGARWAL, R. C. & ISAACS, N. W. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 2835–2839.

BHAT, T. N. & BLOW, D. N. (1982). *Acta Cryst.* **A38**, 21–29.

BERNSTEIN, F. C. *et al.* (1977). *J. Mol. Biol.* **112**, 535–542.

BOROVIKOV, V. A., VAINSTEIN, B. K., GELFAND, I. M. & KALININ, D. I. (1979). *Kristallografiya*, **24**, No. 2, 227–238.

BRICOGNE, G. (1976). *Acta Cryst.* **A32**, 823–847.

CHIRGADZE, YU. N., SERGEEV, YU. V., FOMENKOVA, N. P. & ORESHIN, V. P. (1981). *FEBS Lett.* **131**, 81–84.

HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* **B26**, 136–143.

KALININ, D. I. (1980). *Kristallografiya*, **25**, No. 3, 535–544.

LUNIN, V. YU. (1982). *The Use of a Maximum Likelihood Method for the Evaluation of Errors when Estimating Phases in Protein Crystallography.* Preprint, USSR Academy of Sciences, Pushchino.

LUNIN, V. YU. & URZHUMTSEV, A. G. (1981). *The Extension of Resolution of Protein Electron Density Maps by Refinement of Model Structure.* 1. *The Method's Description.* Preprint. USSR Academy of Sciences, Pushchino.

LUNIN, V. YU & URZHUMTSEV, A. G. (1983). *Kristallografiya*, **28**, 310–315.

PHILLIPS, S. E. V. (1980). *J. Mol. Biol.* **142**, 531–554.

SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography.* Oxford: Pergamon Press.

TEN EYCK, L. F. & ARNONE, A. (1976). *J. Mol. Biol.* **100**, 3–11.