Phase Improvement in Protein Crystallography Using a Mixed Electron Density Model

By V. Yu. Lunin, A. G. Urzhumtsev and E. A. Vernoslova

Research Computer Center, USSR Academy of Sciences, 142292 Pushchino, Moscow Region, USSR

and Yu. N. Chirgadze, N. A. Nevskaya and N. P. Fomenkova

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

Abstract

The proposed technique for phase improvement is based on the refinement of a so-called mixed electron density model. This model consists of two parts. The first is a partial stereochemically correct atomic model of the protein molecule related to the interpreted part of electron density of the unit cell. The second is an artificial atomic model describing the uninterpreted part of the residual electron density of the unit cell. The conventional free-atom crystallographic refinement of such a mixed model results in phase improvement. No attention is paid to the structural sense of the refined atomic positions of the mixed electron density model. This method of phase improvement has been applied for eye lens protein γ-crystallin IIIb at 2·7 Å resolution. A starting partial model of the molecule contained about 56% of the total number of residues. Phase refinement is done in two stages. Each stage leads to a significant increase in the quality of the electron density map, so that the partial atomic model of the protein molecule can be improved and expanded.

1. Introduction

The main aim of X-ray structure determination of proteins is to determine a model of a molecule with a complete set of atoms. However, in many cases the quality of the electron density synthesis allows only part of the model to be constructed. The other part cannot be built without an improved electron density map. In this paper we describe a new method applied in the refinement of γ-crystallin IIIb from calf lens at the stage when a further extension of the obtained partial model was required.

When a partial model is fairly complete, *i.e.* it contains at least 75% of the total number of residues, the synthesis can be improved by crystallographic refinement of the partial model (Rice, 1981; Remington, Wiegand & Huber, 1982). Phases calculated from the refined model combined with those obtained by isomorphous replacement improve the quality of the synthesis. It should be emphasized that such a phase improvement results from a thorough crystallographic refinement, which is a very complex procedure requiring 'hand' operation.

In the case when one has no atomic model at all, Agarwal & Isaacs (1977) proposed that phases could be improved by refining a coarse model. This model is automatically constructed by the program and may not be an accurate stereochemical fit to the real structure. It need only adequately reproduce available electron density maps. This approach is attractive, for it does not require hand construction of the model and implies a relatively simple refinement (without stereochemical restrictions and 'human intervention'). At the same time it is evident that such a simplified approach may not give quite as good results as those obtained by Rice's method; it may be rather a preliminary step in the construction of a partial model. One of the implementations of Agarwal & Isaacs's method was developed by the authors of the present paper. It was tested with actinidin (Lunin & Urzhumtsev, 1981, 1983) and applied to the extension of phases for γ-crystallin IIIb from 3·0 up to 2·7 Å (Lunin & Urzhumtsev, 1984). This application resulted in the improvement of the initial electron density map, which was uninterpretable, to the extent that an incomplete model for about one half of the molecule could be constructed (see § 3).

In this paper we shall consider an intermediate stage in the model construction when the Agarwal & Isaacs procedure, without recourse to a partial model, cannot further improve the synthesis, and the partial model obtained is not quite complete enough to allow Rice's method to be applied.

One of the possible methods in such a situation was proposed by Bhat & Blow (1982). The method is as follows. First, the electron density corresponding to the interpreted part of the molecule is generated. Then the model is expanded to incorporate a part of the map adjacent to the interpreted portion. The calculation of phases from such a model gave an improved synthesis for tyrosyl-tRNA-synthetase. Later (Bhat & Blow, 1983) the procedure also incorporated refinement of a partial model with inclusion of the fixed uninterpreted part of the synthesis.

In this paper we consider another approach mentioned in the previous paper (Lunin & Urzhumtsev, 1984). In the first stage we construct a mixed model of electron density by the addition of the partial model with dummy atoms to reproduce the uninterpreted part of the synthesis. In contrast to Bhat & Blow, we assume that the part of the model related to the uninterpreted synthesis does not have an arbitrary electron density distribution but rather the distribution prescribed by the sum of atomic contributions, i.e. we suggest an additional requirement for the atomicity of the model. At the second stage we refine the mixed model in reciprocal space to calculate phases improving the synthesis. As opposed to Bhat & Blow, we do not only refine the interpreted but also the uninterpreted part of the synthesis.

A principal difference of the indicated approaches from Rice's method is that we expand the model by introducing the uninterpreted part of the synthesis. Such a model allows a better approximation of the electron density distribution in the unit cell. Besides, the additional part of the model can be a stabilizing factor in the refinement, for it prevents the partial model from 'degrading'. For an example of such a 'degradation' observed in the absence of dummy atoms, see § 4 of this paper.

2. Application of a mixed electron-density model for phase improvement

2.1. Initial information

It is assumed that for a set of reflections we know:

(a) The observed structure amplitudes F_o.

(b) The probability distributions $P_o(\varphi)$ for the corresponding phases obtained earlier (e.g. by the method of isomorphous replacement), from which we can determine the 'best' phases φ_o and figures of merit m_o by the equality (Blow & Crick, 1959)

$$m_o \exp(i\varphi_o) = \int_0^{2\pi} \exp(i\varphi) P_o(\varphi) \, d\varphi.$$

(c) The atomic coordinates r_j and isotropic temperature factors B_j of the partial model. The calculated structure factors for a partial molecular model will be denoted by $F_p \exp(i\varphi_p)$.

2.2. Building of a mixed model

A mixed model was made up from the partial molecular model atoms $\{r_j, B_j\}_{j=1}^N$ and a set of dummy atoms with the parameters $\{r'_k, B'_k\}_{k=1}^M$. These atoms were chosen such that the electron density corresponding to their set reproduces the most essential features of the difference synthesis with the coefficients

$$m_o[F_o \exp(i\varphi_o) - F_p \exp(i\varphi_p)]. \qquad (1)$$

The parameters r'_k and B'_k of each dummy atom were determined as follows:

(a) by searching the global maximum r'_k on the difference synthesis;

(b) by determining the parameter B'_k so that the electron density at the center of the atom was about the same as that on the difference synthesis at r'_k;

(c) by subtracting the contribution of the identified dummy atom from the synthesis.

After this the next dummy atom is constructed until the maximal value of the 'residual' electron density drops below a limit value ρ_{crit}. The building of a coarse model of the uninterpreted electron density and the evaluation of an atomic contribution to the weighted synthesis have been described in detail previously (Lunin & Urzhumtsev, 1984).

In the case where one knows the boundary of the molecule in the cell, one could restrict the region of placing dummy atoms so that r'_k are only allowed to be inside the molecule. A further restriction of this region was that dummy atoms were forbidden to be at a distance less than r_{min} from the partial model atoms in the construction of a mixed model. This prevents the partial model atoms from being displaced from their correct position by dummy atoms. The value r_{min} was chosen to approximate to the convergence radius of the applied method of refinement, i.e. at the first stage it was 0·7 Å (Agarwal, 1978) and was increased at later stages. The dummy atom regions can be selected more thoroughly, for instance, these may be introduced as the connected domains adjacent to the interpreted synthesis as in the paper of Bhat & Blow (1982).

2.3. Refinement of a mixed model

For the refinement of a mixed model the Isaacs program was applied, which realized the Agarwal (1978) algorithm of the reciprocal-space refinement without stereochemical restrictions.

In the course of the refinement the mixed model was revised periodically, and the dummy atoms with high B'_k values were omitted from the model.

2.4. Improvement of the synthesis

To improve the synthesis we used the coefficients

$$m_{MR} F_o \exp(i\varphi_{MR}), \qquad (2)$$

where

$$m_{MR} \exp(i\varphi_{MR}) = \int_0^{2\pi} \exp(i\varphi) P_{MR}(\varphi) \, d\varphi \qquad (3)$$

and (Srinivasan & Parthasarathy, 1976)

$$P_{MR}(\varphi) \sim \exp[2t F_o F_{MR} \cos(\varphi - \varphi_{MR})]. \qquad (4)$$

Here $F_{MR} \exp(i\varphi_{MR})$ are the structure factors calculated from the refined mixed electron density model,

t is a parameter defining to what extent the model is incomplete and coarse. It was determined from the maximum of a likelihood function (Lunin & Urzhumtsev, 1984). The mixed model was refined within the range $d_{min} \leq d \leq d_{max}$. For the reflections with $d > d_{max}$ the initial phase probability distribution $P_0(\varphi)$ was used.

3. Phase improvement for γ-crystallin IIIb at 2·7 Å resolution

Crystallins are the main structural proteins from the eye lens of vertebrates. Crystals of γ-crystallin IIIb from calf lens with a molecular mass of about 20 000 daltons belong to space group $P2_12_12_1$ with cell dimensions $58·7 \times 69·5 \times 116·9$ Å (Chirgadze, Sergeev, Fomenkova & Oreshin, 1981). The asymmetric unit contains two molecules with a total mass of 40 000 daltons. The polypeptide chain of the molecule comprises about 170 residues with partially known amino-acid sequence (Croft & Slingsby, 1973).

The native protein crystal of γ-crystallin IIIb had a diffraction limit of about 2·7 Å. Earlier, an electron density map at 3·0 Å resolution had been obtained using the method of multiple isomorphous replacement. Five heavy-atom derivatives were used with anomalous scattering for three of them, the total number of reflections was 6777, and the mean figure of merit was 0·72 (Chirgadze *et al.*, 1981). The electron density of only one molecule was interpreted because of the similarity of the density distribution in the two molecules. The polypeptide chain is clearly traced in the elongated segments of β sheets (γ-crystallin does not contain α helices) as well as in the region of the peptide joining two domains. Several segments of the loops and irregular conformations of the polypeptide chain were poorly defined. The interpretation of side chains also presented a problem. Hence, the polypeptide chain might be traced over C_α atoms well enough for about 80% of the length. The supposed total number of residues was 165.

In the first stage the phases were refined and extended in the range 10 to 2·7 Å (Lunin & Urzhumtsev, 1984) by the method based on the approach of Agarwal & Isaacs (1977). As a starting model for refinement we have constructed a coarse model from dummy atoms reproducing the 'isomorphous' synthesis. No structural data were used. The features of a starting coarse model and the course of the initial refinement are specified in Table 1. The refinement gave phases for about 3800 additional reflections, mainly in the range 3·0-2·7 Å. The resulting weighted synthesis was obtained with the combined phases. A new synthesis at 2·7 Å resolution (Fig. 1) had a higher noise level. Nevertheless, the synthesis was of higher resolution, and a partial model of one protein molecule could be built. The C_α tracing was improved with the total of 167 residues. The map was inter-preted on the basis of the amino-acid sequence. Most of the Cys, Met, Phe and Tyr residues commonly used as markers were well localized. But the precise site of the two Trp residues could not be determined at this stage. In all, 93 side chains were localized, *i.e.* about 56% of the total. The partial molecular model I contained 967 atoms, *i.e.* about 69% of the total (Table 2).

The partial molecular model was fitted to the electron density using the interactive molecular graphics program *ISOCUB* (Nevskaya, Kurochkina & Chirgadze, 1985). The model of each peptide unit is built by an operator on the basis of the crude coordinates of the C_α and C_β atoms. The overall geometry was almost ideal except for the angle τ (NC$_\alpha$C) between two peptide groups and the angle β(C$_\beta$C$_\alpha$C) between the side and main chains.

In the second stage the proposed procedure was applied to a mixed model. The model consisted of the 967 atoms of the partial model I and of a set of 1444 dummy atoms simulating the uninterpreted region of half of the asymmetric unit (Table 1). The mixed model was expanded to generate the asymmetric unit by use of non-crystallographic symmetry. The dummy atoms were placed inside the region equal to a set of spheres with a 5·5 Å radius centering on the C_α atoms. At this stage the dummy atoms were at a minimal distance of 0·7 Å from the atoms of model I. In constructing the synthesis, a grid along the cell axes was chosen as $100 \times 120 \times 200$, which corresponds to a step of ~0·6 Å along each of the crystallographic axes. The dummy atoms were generated until the electron density was exhausted with $\rho > \rho_{crit}$ for a synthesis with the coefficients of (1). The contribution of each dummy atom to the weighted
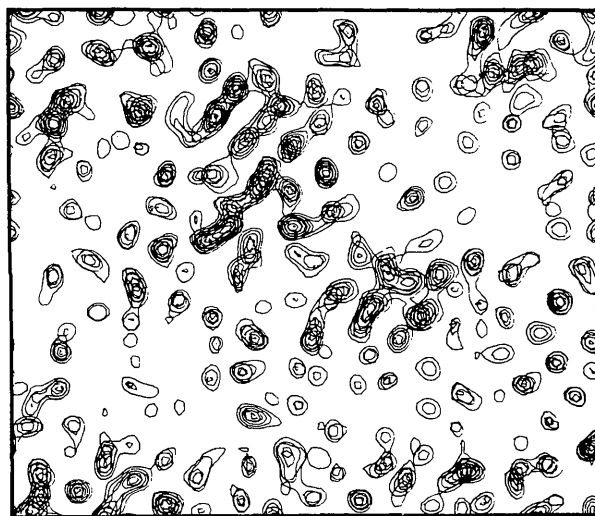

Fig. 1. Synthesis of γ-crystallin IIIb at 2·7 Å resolution after the first stage of phase improvement (the partial model is not used). The xy projection with $z = (27-30)/144$ is shown.

Table 1. *Phase improvement for γ-crystallin IIIb at 2·7 Å resolution*

	Stages of phase improvement		
	Stage I Use of uninterpreted electron density model	Stage II Use of mixed density model	Stage III Use of mixed density model
Starting model			
Number of atoms in the partial model of molecule	no	967	1130
Number of dummy atoms per half of asymmetric unit	1552	1444	873
Parameters for constructing dummy atoms:			
$B_0(\text{Å}^2)$, $C_0(e)$, $\rho_{crit}(e\,\text{Å}^{-3})$	83·9, 8·40, 0·28	72·2, 9·11, 0·40	47·4, 7·69, 0·65
Accounting for molecular boundary	No	Yes	Yes
Refinement procedure			
10·0–3·0 Å			
Number of cycles	10 (6xyz, 4B)	18 (12xyz, 6B)	31 (21xyz, 10B)
$R_{init} \rightarrow R_{final}$	0·35→0·13	0·38→0·14	0·46→0·16
10·0–2·7 Å			
Number of cycles	5(3xyz, 2B)	18 (12xyz, 6B)	18 (12xyz, 6B)
$R_{init} \rightarrow R_{final}$	0·18→0·12	0·19→0·13	0·21→0·15
Quality of synthesis			
Number of reflections	10 527	10 804	10 804
Coefficients of Fourier series	$\begin{cases} m_0F_0\exp(i\varphi_0^{best}) & d>10\,\text{Å} \\ m_{comb}F_0\exp(i\varphi_{comp}^{best}) & 10\ge d\ge2\cdot7\,\text{Å}\end{cases}$	$\begin{cases} m_0F_0\exp(i\varphi_0^{best}) & d>10\,\text{Å} \\ m_{MR}F_0\exp(i\varphi_{MR}) & 10\ge d\ge2\cdot7\,\text{Å}\end{cases}$	Same as stage II
Electron density parameters:			
ρ_{max}^{abs} (e Å^{-3})	2·06	2·83	3·07
ρ_{min}^{map} (e Å^{-3})*	0·56	0·55	0·46

* The minimum level that can be traced on the maps.

Table 2. *Partial models of the protein molecule of γ-crystallin IIIb used for phase improvement*

Molecular model parameters	Model I	Model II
Complete molecular model		
Number of residues*	167	172
Total number of atoms*	1411	1476
Partial molecular model		
Peptide groups (% of the total)	137 (82)	148 (86)
Side chains (% of the total)	93 (56)	118 (69)
Atoms (% of the total)	967 (69)	1130 (77)

* The exact number of residues in the molecule of γ-crystallin is unknown.

synthesis was in the form (Lunin & Urzhumtsev, 1983)

$$C_0\left(\frac{4\pi}{B_0}\right)^{3/2} \exp\left(-\frac{4\pi^2|\mathbf{r}|^2}{B_0}\right), \tag{5}$$

and the values of ρ_{crit}, C_0 and B_0 are listed in Table 1.

The mixed model was refined in two series (Table 1). In the first series of eighteen cycles the refinement was done for the range 10–3·0 Å comprising ~8600 structure factors. The coordinates (x, y, z) and temperature factor B were refined alternately. The crystallographic factor

$$R = \frac{\sum |F_o - F_M|}{\sum F_M},$$

equal to 0·38 for the initial model, decreased to 0·14 during the refinement. In the second series, refinement was carried out in the range 10–2·7 Å with about 10 800 structure factors, *i.e.* one and half times as many as in the first series. The R factor was 0·13. It should be noted that this value is much lower in comparison with R-factor values for the conventional crystallographic free-atom refinement of protein structure. Such a phenomenon appears to be explained by the larger number of degrees of freedom for the mixed model of electron density.

The second improvement made the synthesis more readily interpretable compared with the first one. The map became markedly sharper (Fig. 2). The sharpness of the synthesis became better resulting in a clearer tracing of the main chain. Several 'bulges' indicating the peptide carbonyl locations appeared on the density of the main chain. The density clouds of the side chains also became much more interpretable. This stage demonstrates the efficiency of the applied method of phase improvement showing that it is possible to build an initial model of the protein molecule at a relatively low resolution even in the absence of reliable sequence data.

The obtained synthesis led to a further extension of the protein molecule model. The course of the main chain was improved, and the total number of residues increased to 172, 118 side chains, *i.e.* ~69% of the total, were placed in the electron density. The partial model II then contained 1130 atoms, which was ~77% of the total. The statistics of geometric distortions from the angles τ and β deformed in the model building were also better.

At the third stage the mixed model consisted of 1130 atoms of the partial model II and 873 dummy

atoms simulating the electron density in the uninterpreted region of half of the asymmetric unit (Table 1). The second molecule in the asymmetric part of the unit cell was generated by the use of non-crystallographic symmetry relations. The refinement was carried out in 49 cycles within two resolution ranges. The final R value was 0·15. The synthesis (Fig. 3) was substantially improved at this stage compared with the previous one.

4. The role of dummy atoms in a mixed model

To check how important dummy atoms are for a mixed model, we carried out the following test. The partial model I (967 atoms) used at the second stage was refined in reciprocal space by the method applied earlier to the mixed model (35 refinement cycles to 2·7 Å). As a result, the crystallographic R factor dropped from 0·53 for the 3·0 Å resolution range to 0·27 for the 2·7 Å resolution range. However, the Fourier synthesis with the phases calculated from the model refined in such a way was considerably poorer than the initial one. This shows that the uninterpreted synthesis part stabilizes refinement of a model when it is not complete.

5. Discussion

The method has proved to be successful in phase refinement for γ-crystallin IIIb from calf lens at 2·7 Å resolution. The synthesis of the electron density is greatly improved at both stages of refinement. The extension of the partial atomic model from 69 to 77%

at one stage illustrates the progress achieved. The refinement of a stereochemically restrained protein molecule, considered commonly as a final step of the analysis, becomes now practically possible.

The suggested technique allows the use of information from the interpreted part of the synthesis to improve the overall density. In the case when the molecular model contains a very small number of atoms the mixed model is composed mostly of the dummy atoms. In the final stages the relative number of dummy atoms drops and the mixed model is subjected to a free-atom crystallographic refinement of protein structure.

In the intermediate stages, when the partial model contains about half of the atoms, the application of dummy atoms may reduce errors in the calculated phases owing to the absence of part of the molecule. If the mixed model is refined without stereochemical restrictions, the dummy atoms also present a stabilizing factor preventing deviations of the partial model atoms from their positions. It should be emphasized that further progress could be achieved if stereochemical restrictions were imposed on the partial model in the final phase refinement.

We have utilized the coarse and the mixed density model for phase improvement over the range 3·5–2·0 Å. At higher resolution one deals essentially with a fairly complete atomic model requiring no dummy atoms. For lower resolution a similar approach has been applied to proteins of the globin family at 5 Å resolution (Borovikov, Vainstein, Gelfand & Kalinin, 1979; Kalinin, 1980). The model here was built from 'quasicylinders' imitating α helices with refinable parameters. The application of the technique of this
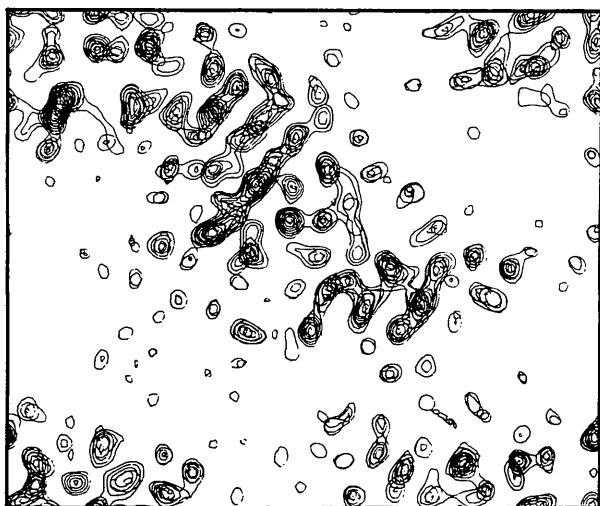

Fig. 2. Synthesis of γ-crystallin IIIb at 2·7 Å resolution after the second stage of phase improvement obtained by the method proposed in this paper. The projection is as in Fig. 1. For details see Table 2.
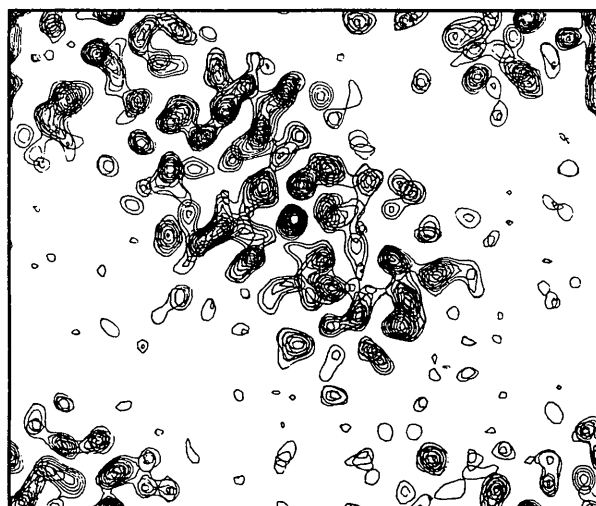

Fig. 3. Synthesis of γ-crystallin IIIb at 2·7 Å resolution after the third stage of phase improvement obtained by the proposed method. The projection is as in Fig. 1. For details see Table 2.

paper for the 5–3·5 Å range has not been studied but it is obvious that a coarse model should be built of larger blocks rather than from a set of atoms to obtain a reasonable relation between the number of parameters describing the model and those of the structure factors used for the refinement.

In modern protein crystallography the following main stages can be distinguished as necessary.

1. Initial phase determination, for example, by the method of isomorphous replacement.

2. Intermediate phase improvement resulting in a preliminary model for the protein molecule.

3. Final refinement of the molecular model.

Attempts to obtain reliable phases at the first stage of phase determination for most of the proteins with a medium and large molecular mass have failed. Thus, an elaboration of a powerful method for the second stage is very desirable, with the application of additional information for phase improvement by different methods. The method proposed in this paper could be one of them.

The authors wish to thank Dr N. W. Isaacs for the text of the refinement program.

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. M. (1982). *Acta Cryst.* A38, 21–29.
BHAT, T. N. & BLOW, D. M. (1983). *Acta Cryst.* A39, 166–170.
BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* 12, 794–802.
BOROVIKOV, B. A., VAINSTEIN, B. K., GELFAND, I. M. & KALININ, D. I. (1979). *Kristallografiya*, 24, 227–238.
CHIRGADZE, YU. N., SERGEEV, YU. V., FOMENKOVA, N. P. & ORESHIN, V. D. (1981). *FEBS Lett.* 131, 81–84.
CROFT, L. R. & SLINGSBY, G. (1973). In *The Human Lens–In Relation to Cataract. Ciba Found. Symp.* 19, 212–213. Amsterdam, New York: Elsevier/Excerpta Medica.
KALININ, D. I. (1980). *Kristallografiya*, 25, 535–544.
LUNIN, V. YU. & URZHUMTSEV, A. G. (1981). *The Extension of Resolution of Protein Electron Density Maps by Refinement of Model Structure. I. The Method's Description.* Preprint. USSR Academy of Sciences, Pushchino.
LUNIN, V. YU. & URZHUMTSEV, A. G. (1983). *Kristallografiya*, 28, 310–315.
LUNIN, V. YU. & URZHUMTSEV, A. G. (1984). *Acta Cryst.* A40, 269–277.
NEVSKAYA, N. A., KUROCHKINA, N. A. & CHIRGADZE, YU. N. (1985). *Kristallografiya.* In the press.
REMINGTON, S., WIEGAND, G. & HUBER, R. (1982). *J. Mol. Biol.* 158, 111–152.
RICE, D. W. (1981). *Acta Cryst.* A37, 491–500.
SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography.* Oxford: Pergamon Press.

Electron Density Studies of KCl and LiF by γ-ray Diffractometry

BY M. C. SCHMIDT,* R. COLELLA AND D. R. YODER-SHORT

Department of Physics, Purdue University, West Lafayette, Indiana 47907, USA

Abstract

Use of hard electromagnetic radiation ($\lambda = 0·12$ Å) in diffraction experiments, in conjunction with thin crystals in transmission geometry (Laue case), minimizes multiple scattering, often referred to as the 'extinction problem'. A large crystal slab, completely intercepting the incident beam, was used in this experiment. Several structure factors have been measured in LiF and KCl, with greater accuracy than ever before, using the γ-ray spectrometer of the University of Missouri at Columbia. In the case of LiF the results indicates some compression of the valence electron densities, similar to what has been found earlier for NaF. In the case of KCl all reflections agree well with the values calculated from free-atom Hartree–Fock wave functions except for the 004 and 006, which are appreciably weaker than corresponding reflections with similar or identical $\sin \theta / \lambda$ values. This result indicates asphericity of the valence electron density. The 111 is also much weaker than the calculated value, indicating that the negative ions are slightly compressed and the positive ions slightly expanded with respect to the free-ion charge densities.

1. Introduction

Using high-energy γ-rays ($E \geq 100$ keV) for diffraction experiments, in conjunction with thin crystal slabs set for Laue-case diffraction, greatly minimizes the occurrence of extinction, which makes it possible to extract structure factors from intensity measurements with accuracies by far superior to those hitherto obtained in traditional X-ray experiments ($E = 6$–20 keV) (Schneider, 1974, 1976). The physical rea-

---

* Work based in part on the thesis by M. C. Schmidt, submitted in partial fulfilment of the requirements for the PhD degree to Purdue University.