

- MA, Y. & MARKS, L. D. (1991). *Acta Cryst.* **A47**, 707–715.  
 MAKSYM, P. A. & BEEBY, J. L. (1981). *Surf. Sci.* **110**, 423–436.  
 MENADUE, J. F. (1972). *Acta Cryst.* **A28**, 1–11.  
 MOON, A. R. (1972). *Z. Naturforsch. Teil A*, 390–395.  
 NEAVE, J. H. & JOYCE, B. A. (1983). *Appl. Phys.* **A31**, 1–8.  
 PENG, L. M. & COWLEY, J. M. (1986). *Acta Cryst.* **A42**, 545–552.  
 VAN HOVE, J. M., LENT, C. S., PUKITE, P. R. & COHEN, P. I. (1983). *J. Vac. Sci. Technol.* **B1**(3), 741–746.

*Acta Cryst.* (1991). **A47**, 794–801

## Low-Resolution Phases: Influence on SIR Syntheses and Retrieval with Double-Step Filtration

BY A. G. URZHUMTSEV

*Research Computer Centre, Academy of Sciences of the USSR, Pushchino, Moscow Region, 142292 USSR*

(Received 11 July 1990; accepted 12 June 1991)

### Abstract

Different sources of error in single isomorphous replacement (SIR) maps are analysed. It is shown that SIR maps are good enough in themselves to localize a molecule but the absence of low-resolution reflections can distort the image of the molecule. Even in this case the molecule can be reliably localized using the double-step filtration technique of Urzhumtsev, Lunin & Luzyanina [*Acta Cryst.* (1989), **A45**, 34–39] if parameters are chosen appropriately. On the other hand, such syntheses may be considerably improved by addition of low-resolution structure factors. A very simple procedure is suggested to retrieve unknown phases for these structure factors.

### I. Introduction

At the beginning of a macromolecular structure determination, electron density syntheses may be so noisy that the molecules cannot be localized. Different sources of errors influence synthesis distortions. Some of them have recently been analysed by Fenderson, Herriott & Adman (1990). In our work we have analysed the influence of some other sources of synthesis errors: using isomorphous replacement phases instead of the exact values and lack of very low-resolution reflections in the synthesis (loss of the central zone of reciprocal space).

Earlier, Podjarny, Schevitz & Sigler (1981) and Luzzatti, Mariani & Delacroix (1988) noted the significance of low-resolution phases. Now we have succeeded in both qualitative and quantitative estimation of the effect of these omitted reflections. This omission was shown to be the main cause of molecular image distortion. Unfortunately, determination of the corresponding phases by the isomorphous replacement method is problematic, in particular because of the well known disordered solvent contribution.

When molecule boundaries can be found, the powerful solvent-flattening procedure (Bricogne, 1974) may be used to improve the synthesis. To localize a molecule in a very noisy synthesis, similar procedures were independently developed by Stuart (1988), Urzhumtsev (1985), Wang (1985) and Westbrook (see, Podjarny, Moras, Navaza & Alzari, 1988). The most popular of these approaches is the Wang (1985) procedure. This was successfully inserted into the general density-modification scheme (Podjarny, 1987) which was used from the beginnings of protein crystallography (see e.g. Qurashi, 1953).

There is a widespread point of view that the basis of the Wang procedure is a local averaging of a noisy synthesis. However, as was shown by Urzhumtsev *et al.* (1989), the averaging by itself does not produce molecule boundaries and the key point is the preliminary nonlinear filtration of the synthesis. In the Wang procedure this step is carried out indirectly by calculating the synthesis without  $F_{000}/V$  and removing negative density. Wang has not paid due attention to this step although the results for molecular boundaries depend on its parameters (Urzhumtsev *et al.*, 1989). We have demonstrated that the appropriate choice of parameters for both nonlinear filtration and averaging enables one to localize the molecule even in a very noisy synthesis, in particular, one calculated without low-resolution reflections.

When low-resolution reflections are unphased the synthesis may also be improved by phase extension to the low-resolution region which is the exact opposite to the usual phase extension to high-resolution reflections. Some authors working with viruses structures in particular include these reflections in synthesis calculations, obtaining the phases from a model (e.g. Harrison & Jack, 1975). This situation differs from the one discussed here where a model is absent but more high-resolution phases are known, maybe with some errors. Earlier, Podjarny *et al.*

(1981) used complicated matrix methods (Tsoucaris, 1970) to retrieve phases for low-resolution reflections. We have found that these phases can be retrieved in a much simpler way. The whole procedure consists of nonlinear filtration of the initial synthesis and calculation of Fourier coefficients for a new distribution. Both test calculations and application to the experimental data have shown that the addition of retrieved data to the initial synthesis may show the molecule without any solvent flattening.

## II. Method of testing

This work has arisen from elongation factor  $G$  (EFG) structure determination (Chirgadze, Nikonorov, Brazhnikov, Garber & Reshetnikova, 1983) and the model, the unit cell, the resolution etc. used in the test calculations were taken from this investigation. Such a choice of parameters has allowed us to use the results of the test calculations directly to solve the problem.

### 1. Notation

SIR: single isomorphous replacement method (Blow & Rossmann, 1961).

AMB: automatic molecule bounding method (Wang, 1985).

DSF: double-step filtration method (Urzhumtsev, 1985).

$F_s^H \exp(i\varphi_s^P)$ : structure factors calculated from a protein model.

$F_s^H \exp(i\varphi_s^H)$ : structure factors calculated from a heavy atom.

$F_s^{PH} \exp(i\varphi_s^{PH})$ : structure factors calculated from a heavy-atom derivative.

$\varphi_s^{\text{SIR}}$ ,  $m_s^{\text{SIR}}$ : phase and its figure of merit determined with the SIR.

$\rho^{\text{SIR}}$ : electron density distribution calculated with coefficients  $m_s^{\text{SIR}} F_s^P \exp(i\varphi_s^{\text{SIR}})$ .

$\tau^{\text{SIR}}$ : result of double-step filtration of  $\rho^{\text{SIR}}$ .

$T_s \exp(i\varphi_s)$ : Fourier coefficients of  $\tau^{\text{SIR}}$ .

### 2. Atomic model

The model of the artificial molecule was composed of two atomic models of carboxypeptidase A (Rees, Lewis & Lipscomb, 1983), each of 339 amino acid residues, and was placed without overlapping in the unit cell (space group  $P2_12_12_1$ ,  $a = 76$ ,  $b = 106$ ,  $c = 116$  Å). The model heavy atom was placed in a cavity between the models of carboxypeptidase.

### 3. Structure factors

The sets of structure factors  $\{F_s^P \exp(i\varphi_s^P)\}$  and  $\{F_s^{PH} \exp(i\varphi_s^{PH})\}$  were calculated from the native model and its heavy-atom derivative, respectively, each included about 2500 reflections out to 6 Å resol-

ution. The number of electrons in the heavy atom was chosen such that its contribution  $\{F_s^H \exp(i\varphi_s^H)\}$  to the structure factors was

$$R_H = \sum_s F_s^H / \sum_s F_s^P = 0.1. \quad (1)$$

### 4. SIR simulation

We assumed that values of structure-factor moduli were known for the native protein and for its derivative,  $\{F_s^P\}$  and  $\{F_s^{PH}\}$ , respectively. The exact phases were assumed unknown and were used only for comparison. The heavy atom was supposed to be determined either absolutely accurately or with the coordinate error  $\Delta r$ , depending on the test. The  $F_s^P$  and  $F_s^{PH}$  values and the heavy-atom parameters were used to calculate phases and their figures of merit with the standard SIR technique (Blow & Rossmann, 1961).

### 5. Syntheses

Structure factors with a resolution within  $6\text{\AA} \leq 1/|s| \leq d_{\max}$  were used to calculate the SIR syntheses

$$\rho^{\text{SIR}}(\mathbf{r}) = V^{-1} \left[ F_{000} + \sum_s m_s^{\text{SIR}} F_s^P \times \exp(i\varphi_s^{\text{SIR}}) \exp(-2\pi i s \mathbf{r}) \right] \quad (2)$$

where  $d_{\max}$  was either infinite or 30 Å, depending on the test.

### 6. Molecule region

In spite of widespread use, the term molecule region is not exactly defined. Usually, it means the points of the unit cell which are closest to the atoms of the molecule under investigation. In our case we define the molecule region as a set of points  $\mathbf{r}$  such that  $\rho^{\text{EX}}(\mathbf{r}) \geq \xi$  for them, where  $\rho^{\text{EX}}(\mathbf{r})$  is the ideal electron density map, calculated from the atomic model. Normally,  $\xi$  should be taken in such a way that the corresponding region has a volume equal to the real molecular volume, e.g. 40–45% of the whole unit-cell volume. However, when only the outside molecule boundary is analysed and not the fine details, the  $\xi$  value should be less and the volume of the corresponding region becomes greater. In particular, it may be useful to analyse the regions in which the volume increases to 60–70% of the whole unit-cell volume.

A definition of a molecule region by a more accurate but complicated method has led our test calculations to practically the same results.

### 7. Comparison of the syntheses

We used an approach to synthesis comparison suggested by Lunin (1988).

When an electron density distribution  $\rho(\mathbf{r})$  is analysed the shape and localization of regions bounded by equipotential surfaces are the object of interest rather than the absolute values of the density. For example, we have introduced the molecule region as the set of points  $\mathbf{r}$  such that  $\rho^{\text{EX}}(\mathbf{r}) \geq \xi$  and the more exactly this region may be reproduced with a synthesis approximation  $\rho(\mathbf{r})$  the better this approximation is.

Let  $\Omega_\nu$  be a region which is bounded by some equipotential surface of the synthesis  $\rho^{\text{EX}}$  and occupies a share  $\nu$  of the whole unit-cell volume. If we introduce a uniform grid with the total number  $N_{\text{tot}}$  of grid points in the unit cell, the region  $\Omega_\nu$  normally contains  $\nu N_{\text{tot}}$  grid points.

If the same number  $\nu N_{\text{tot}}$  of grid points are chosen randomly then some of them belong the region  $\Omega_\nu$  and some are outside the region. The mean number of such 'outside of  $\Omega_\nu$  points' is  $\nu(1 - \nu)N_{\text{tot}}$ . Otherwise, the same number  $\nu N_{\text{tot}}$  of grid points may be chosen as points corresponding to the largest values of some function  $\rho(\mathbf{r})$ . In this case the number  $N_{\text{out}}$  of the 'outside of the  $\Omega_\nu$  points' may be less than for a random choice. The value

$$D(\nu) = D(\nu; \rho^{\text{EX}}, \rho) = N_{\text{out}}(\nu) / \nu(1 - \nu)N_{\text{tot}} \quad (3)$$

may be used as a quantitative estimation of the correctness of  $\Omega_\nu$  determination with the use of the function  $\rho(\mathbf{r})$ . The smaller  $D(\nu)$  the better  $\Omega_\nu$  is reproduced.

Usually, the value  $D(\nu)$  varies with  $\nu$  showing the accuracy with which a function  $\rho(\mathbf{r})$  can reproduce different regions of  $\Omega_\nu$ .

### III. The influence of different errors on SIR syntheses

#### 1. SIR phase errors

The 6 Å resolution sets of the native structure-factor moduli and SIR phases and merits were simulated using the atomic model as described above. The SIR synthesis (2) was calculated with a full 6 Å resolution set of structure factors for the exactly determined heavy atom. A comparison of the  $\varphi_s^{\text{SIR}}$  with the exact phases has given an average error of about 45° (Fig. 1). The quantitative comparison (Fig. 2) of the synthesis with the exact one for 6 Å resolution (Fig. 3a) has shown that the substitution of SIR phases for the exact phase values does not markedly influence middle-density regions of the synthesis. The latter allows a fairly easy localization of the molecule.

#### 2. Errors influenced by heavy-atom displacement

In a more complicated situation with the heavy-atom coordinates containing an error  $\Delta r$  one observes more substantial phase errors (Fig. 1) and a poorer synthesis quality. A comparison of the syntheses with the exact 6 Å resolution synthesis (Fig. 2) shows that the regions of middle values of electron density

remain the most error free and that reasonable errors in the atom position (here up to 3–4 Å) do not essentially influence molecule localization (Fig. 3c).

#### 3. Omitting very low-resolution reflections

The effect of omitting very low-resolution reflections may be clearly seen for the synthesis with the exact phase values. When only 29 reflections with a resolution over 30 Å (1% of the total 6 Å resolution

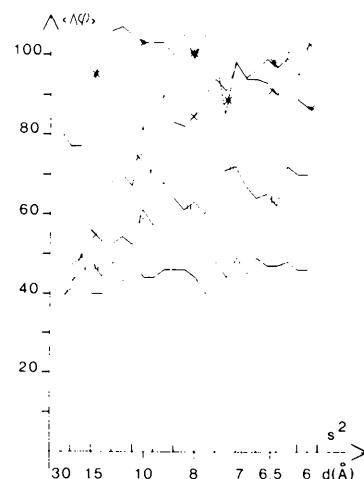


Fig. 1. Mean phase error *versus* resolution  $d = 1/s$  for SIR phases when the heavy-atom position error is  $\Delta r$ . —○—  $\Delta r = 0$ ; —||—  $\Delta r = 2 \text{ \AA}$ ; —×—  $\Delta r = 4 \text{ \AA}$ ; —\*—  $\Delta r = 7 \text{ \AA}$ .

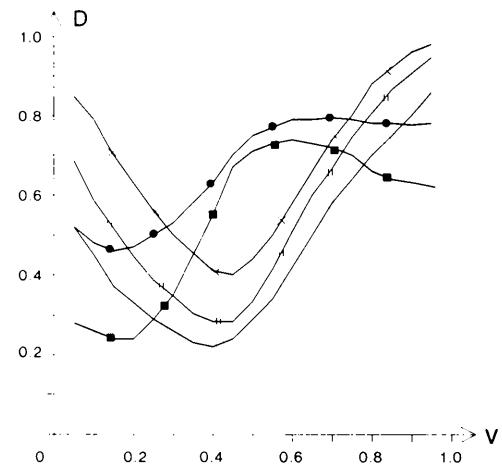


Fig. 2. Comparison of 6 Å resolution syntheses with the exact synthesis of the same resolution.  $D(\nu)$  is the relative discrepancy of the syntheses plotted *versus* relative volume (see II.6). —○— SIR synthesis, correct heavy-atom position; —||— SIR synthesis, heavy-atom position error is 2 Å; —×— SIR synthesis, heavy-atom position error is 4 Å; —■— exact phases, reflections with a resolution lower than 30 Å are excluded from the calculation of the synthesis; —●— SIR synthesis, reflections with a resolution lower than 30 Å are excluded from the calculation of the synthesis.

reflection set) were excluded from the synthesis calculation the compact molecule region has disappeared (Fig. 3d). A quantitative comparison with the exact synthesis (Fig. 2) has shown that corresponding distortions differ from those in the previous test. Exclusion of the reflections of the reciprocal-space central zone strongly affects middle and low-density regions, slightly distorting the regions of the peaks. A visual analysis shows that in this synthesis (Fig. 3d), as well as in the similar synthesis with SIR phases shown in Fig. 3(e), the localization of the molecule is very problematic. It is clear that any additional error  $\Delta r$  introduced in heavy-atom parameters can only redouble the problem of molecule localization. However, high-density values are still more reliable than middle-density ones.

#### IV. Automatic molecule localization

##### 1. Double-step filtration method

The DSF was elaborated (Urzhumtsev, 1985) for automatic molecule localization in a noisy synthesis. We use this term to emphasize the importance of each of two successive steps in the initial synthesis processing.

The first step (non-linear filtration) is to determine the measure of reliability,  $p(\mathbf{r})$ , that the grid point  $\mathbf{r}$

enter the molecule region. Different functions (Urzhumtsev, Lunin & Luzyanina, 1986, 1989) may be used to do this. In this work we used the simplest function  $p(\mathbf{r}) = \chi[\rho(\mathbf{r})]$ :

$$\chi(\rho; \mu) = \begin{cases} (\rho - \mu)/(\rho_{\max} - \mu) & \text{if } \rho \geq \mu \\ 0 & \text{if } \rho < \mu \end{cases} \quad (4)$$

with the prescribed value  $\mu$ .

In the second step the mean number,  $\tau(\mathbf{r})$ , of points belonging to the molecule region is calculated for a neighbourhood of each point  $\mathbf{r}$ . This value may be calculated as a weighted sum of  $p(\mathbf{u})$  with a weighting function  $a(|\mathbf{u} - \mathbf{r}|)$ . In this work we used the function  $a(\mathbf{r})$  introduced by Wang (1985):

$$a_w(r; R) = \begin{cases} r/R & \text{if } 0 \leq r < R \\ 0 & \text{if } r > R \end{cases} \quad (5)$$

where  $R$  is a prescribed radius of averaging. This function was used to make the comparison of the approaches easier although  $a(\mathbf{r})$  may vary (Urzhumtsev *et al.*, 1989).

The direct  $p(\mathbf{u})$  summation (Wang, 1985) may be replaced by the equivalent method of corresponding structure-factor multiplication. This method of linear filtration in reciprocal space which was proposed by Lunin (see Urzhumtsev, 1985) and Leslie (1987) is much less time consuming.

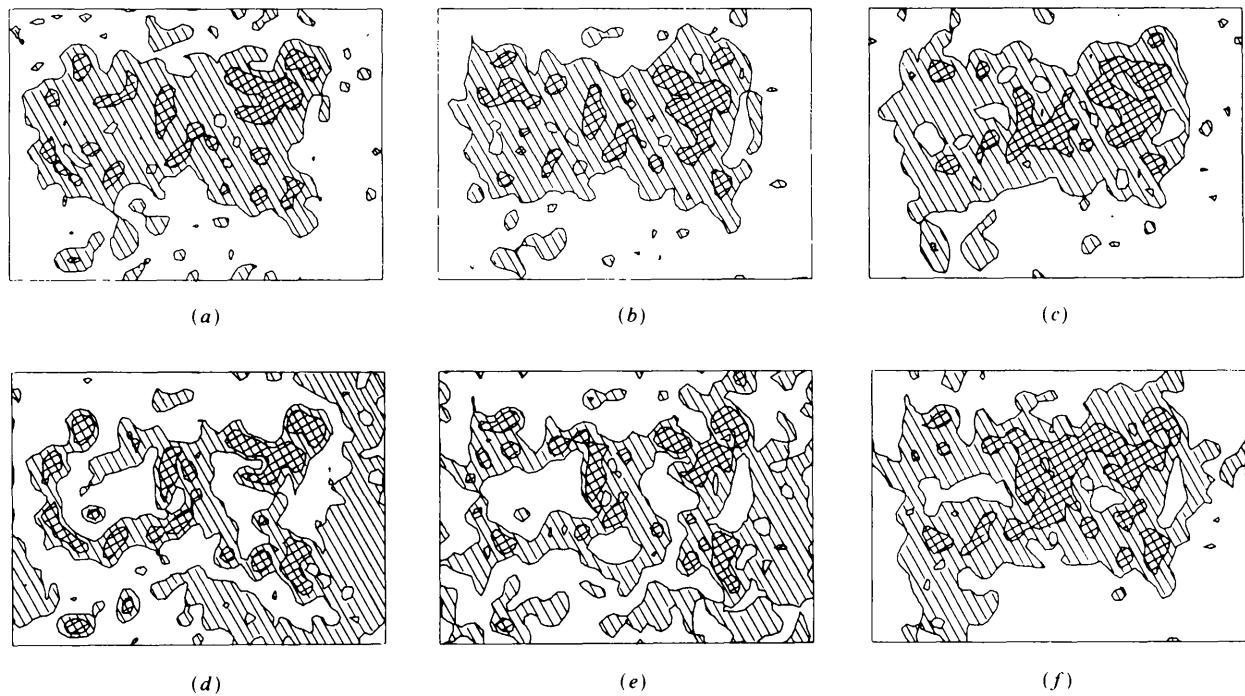


Fig. 3. Model protein electron density maps at 6 Å resolution, section  $z = 10/60$ . Equipotential surfaces bound 10 and 50% of the unit-cell volume. (a) Synthesis with the exact phases; (b) SIR synthesis, correct heavy-atom position; (c) SIR synthesis, heavy-atom-position error is 4 Å; (d) synthesis with the exact phases, reflections with a resolution lower than 30 Å are excluded; (e) SIR synthesis, correct heavy-atom position, reflections with a resolution lower than 30 Å are excluded; (f) SIR synthesis, low-resolution structure-factor phases are retrieved with the DSF technique.

The full scheme of the DSF may be represented as

$$\begin{array}{c} \rho(\mathbf{r}) \xrightarrow{\chi(\rho)} p(\mathbf{r}) \xrightarrow{*a(\mathbf{r})} \tau(\mathbf{r}) \longrightarrow \\ \downarrow \mathcal{F} \qquad \qquad \qquad \uparrow \mathcal{F}^{-1} \\ P_s \exp(i\varphi_s) \xrightarrow{\times A(s)} T_s \exp(i\varphi_s) \end{array}$$

Here  $*a(\mathbf{r})$  means a convolution with spherically symmetric  $a(\mathbf{r})$  and  $\times A(s)$  means multiplication by its Fourier coefficients which are real in this case. This approach is rather different from the Wang (1985) method of automatic molecule bounding where the first step is hidden. The necessity for non-linear filtration has been shown previously by Urzhumtsev *et al.* (1989). Note that the result of this step,  $p(\mathbf{r})$ , describes features of the point  $\mathbf{r}$  only while the final function  $\tau(\mathbf{r})$  describes features of the neighbourhood of the point  $\mathbf{r}$ .

## 2. Use of Wang's approach

Molecule bounding in the exact synthesis does not cause problems. For the synthesis  $\rho^{\text{SIR}}(\mathbf{r})$  calculated with true SIR phases the quality of the molecule localization was worse (Fig. 4). Using the filtration functions  $\chi$  (4) and  $a_w$  (5) and parameter values  $\mu = F_{000}/V$  and  $R = 10 \text{ \AA}$  suitable for AMB, the distribution  $\tau^{\text{SIR}}(\mathbf{r})$  was obtained from  $\rho^{\text{SIR}}(\mathbf{r})$ . Bounding of the molecule region in  $\tau^{\text{SIR}}(\mathbf{r})$  has demonstrated its greater closeness to the molecule region than the  $\rho^{\text{SIR}}$  synthesis (Fig. 4). Besides, this distribution showed greater stability against phase errors.

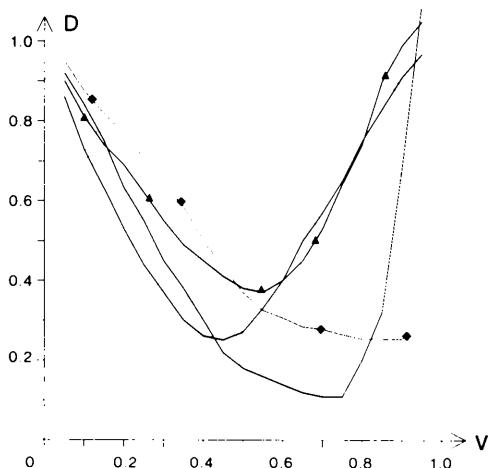
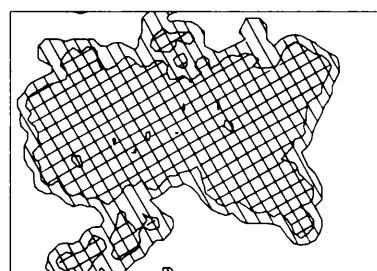


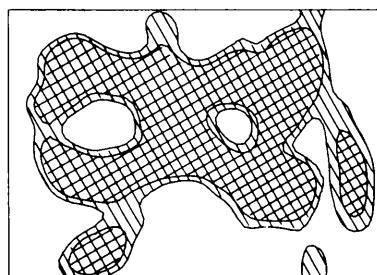
Fig. 4. Accuracy of molecule-region bounding in SIR synthesis and in filtered distributions.  $D(v)$  is the synthesis relative discrepancy plotted versus relative volume (see II.6). —○— SIR synthesis of  $6 \text{ \AA}$  resolution, correct heavy-atom position; —△— result of the filtration of the SIR synthesis, standard filtration parameters; —▲— result of the filtration of the SIR synthesis without low-resolution reflections, standard filtration parameters; —◆— result of the filtration of the same synthesis, filtration parameters are optimal.

Introduction of a heavy-atom position error of  $3-4 \text{ \AA}$  practically did not affect the molecule localization.

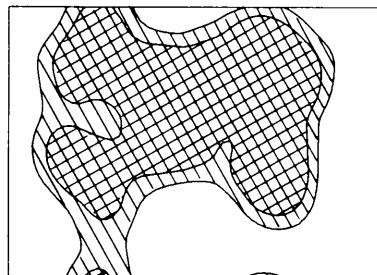
It has already been shown that for a synthesis without central-zone reflections visual molecule localization is very problematic (Fig. 3e). The Wang procedure coped, in general, with this problem too; however, the result was essentially worse (Fig. 4). When the region  $\tau^{\text{SIR}}(\mathbf{r}) \geq \xi$  occupied the whole molecule volume (for the test it was 40–45% of the unit-cell volume), a perceptible portion of the molecule was lacking (Figs. 5a, b). Increasing the volume bounded to 60–65% did not radically improve the situation. Such a missing portion is dangerous because of the solvent flattening which usually follows the procedure in the density modification



(a)



(b)



(c)

Fig. 5. Molecule localization in the ideal synthesis and in the filtered distributions for the SIR synthesis without low-resolution reflections; section  $z = 10/60$ . Equipotential surfaces bound 40 and 60% of the unit-cell volume. (a) The ideal synthesis calculated from the atomic parameters; (b)  $\rho(\mathbf{r}) = \mu$  bounds 50% of the unit-cell volume,  $R = 10 \text{ \AA}$  (Wang); (c)  $\rho(\mathbf{r}) = \mu$  bounds 10% of the unit-cell volume,  $R = 20 \text{ \AA}$ .

method. It may erase part of the synthesis containing important molecular information that is very difficult to restore.

### 3. Optimal parameters for the filtration functions

A different choice of parameter  $\mu$  determines a share of the whole unit-cell volume bounded by the surface  $\rho(\mathbf{r}) = \mu$ . As was shown by Urzhumtsev *et al.* (1989), parameter  $\mu$  in (4) is the key parameter providing the molecule region or its negative, or nothing. Since the high-density synthesis points can be referred to the molecule region with greater certainty than the middle- or low-density ones (Fig. 2), increasing  $\mu$  essentially improves the quality of the molecule localization. For the SIR synthesis without central-zone reflections the  $\mu$  value corresponding to the bounding of only 10% of the unit-cell volume has provided the best result. Further increase of  $\mu$  did not improve the situation, possibly because very little information was preserved. Parameter  $R$  (the radius of averaging) of the function  $a_w(\mathbf{r})$  also influences the filtration. For  $\mu = F_{000}/V$  the best result was achieved at  $R = 20 \text{ \AA}$ . An optimum for the molecule localization seems a value of  $R$  close to half of the characteristic molecule radius. Using the best parameter values of  $\mu$  and  $R$  at the same time gave a filtered synthesis (Fig. 5c) where molecular localization was essentially better (Fig. 4).

It should be noted that the region bounded in the filtered synthesis is very stable against heavy-atom displacements. However, when the atom position is very wrong (for the given model the heavy atom must be shifted by about 15  $\text{\AA}$ ) the true molecule region disappears, giving rise to a false globule at another site. Thus, the presence of a globule in the filtered synthesis does not guarantee a correct solution of the problem.

It is natural that the image of the molecule in the calculated distribution  $\tau(\mathbf{r})$  should lose fine detail if  $R$  increases. To protect the molecule, one should bound not 40–45% but 60–70% of the unit-cell volume. This, of course, gives less additional information in the iterative process of phase improvement which usually follows this procedure, but the probability that the molecule remains inside the bounded region should surely increase. We think that the error may be smaller if our approach is combined with the technique developed by Bhat & Blow (1982), in which connected regions with large density values are added to the basic region (here to the bounded region with a relative volume of 40–45%).

### V. Retrieval of the low-resolution structure factors

Since the main difficulty in localizing the molecule (manually or automatically) in SIR density maps is caused by the absence of reciprocal-space central-

zone reflections, the maps may be improved by retrieving these reflections. Previously, Podjarny *et al.* (1981) successfully retrieved low-resolution structure-factor phases for tRNA using matrix methods (Tsoucaris, 1970). Also, Zelwer (1988) tried to retrieve these phases with an isomorphous pseudo-derivative technique. Here we suggest another procedure to retrieve phases (and the moduli if absent) which is extremely simple.

Since the distribution  $\tau(\mathbf{r})$  calculated with the DSF resembles a low-resolution electron density distribution the question arises whether the Fourier coefficients of  $\tau(\mathbf{r})$  are close to the corresponding low-resolution structure factors. When the answer is positive, the calculated phases  $\varphi(s)$  may be used as an approximation for the unknown values.

We have carried out two tests. In the first one the initial synthesis was calculated with the SIR structure factors (2) at a resolution of 6  $\text{\AA}$  where very low-resolution reflections (with resolution lower than 30  $\text{\AA}$ ) were excluded (Fig. 3e). The distribution  $\tau(\mathbf{r})$  was calculated with the DSF using the functions (4), (5) and the optimal values of  $\mu$  and  $R$  determined above. The phase of its Fourier coefficients,  $T_s \exp(i\varphi_s)$ , were used for the synthesis improvement. The Fourier coefficients  $F_s^P \exp(i\varphi_s)$  of low resolution were added to the new synthesis calculation while structure factors for the resolution 6–30  $\text{\AA}$  were unchanged. The section of the improved synthesis is shown in Fig. 3(f) where molecule localization is quite clear and is close to that in the full SIR synthesis (Fig. 6). The average phase error for the retrieved

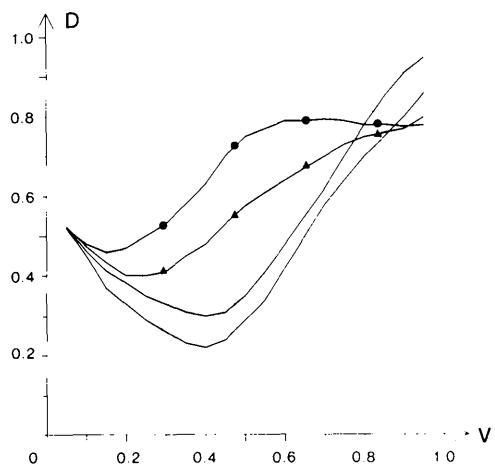


Fig. 6. Comparison of 6  $\text{\AA}$  resolution syntheses with the exact 6  $\text{\AA}$  resolution synthesis. Middle-resolution ( $30 \geq 1/|s| \geq 6 \text{ \AA}^{-1}$ ) Fourier coefficients of the syntheses compared are the same,  $m_s^{\text{SIR}} F_s^P \exp(i\varphi_s^{\text{SIR}})$  and low-resolution ( $1/|s| > 30 \text{ \AA}^{-1}$ ) Fourier coefficients of the syntheses are  $\square - F_s^P \exp(i\varphi_s^{\text{SIR}})$ ;  $\bullet - m_s^{\text{SIR}}$  absent;  $\triangle - F_s^P \exp(i\varphi_s)$ ;  $\blacktriangle - \kappa T_s \exp(i\varphi_s)$ . Here  $T_s \exp(i\varphi_s)$  are structure factors after filtration of 30–6  $\text{\AA}$  resolution  $\rho^{\text{SIR}}$  and  $\kappa$  is the best scaling coefficient of  $T_s$  to  $F_s^P$ .  $D(v)$  is the synthesis relative discrepancy plotted versus relative volume (see II.6).

phases  $\varphi_s$  was about  $50^\circ$ . The standard correlation coefficient calculated with respect to structure-factor moduli was 0.86.

It is important that when structure-factor phases are retrieved the second step of the DSF, averaging of density, is unnecessary because the linear filtration with real  $A(s)$  does not change the phases  $\varphi_s$  of low-resolution structure factors,  $P_s \exp(i\varphi_s) \rightarrow T_s \exp(i\varphi_s)$ . So only simple synthesis modification (4) with appropriate  $\mu$  gives phases whose addition to the initial synthesis shows a molecule with no artificial solvent flattening (Fig. 3f).

In the second test the situation was simulated when both low-resolution structure-factor phases and moduli are unknown. The initial synthesis and the procedure for its filtration were the same. The value of the scale factor  $\kappa$  between the exact moduli,  $F_s^P$ , and the calculated values,  $T_s$ , was determined from the minimum of  $\sum_s (\kappa T_s - F_s^P)^2$  where the summation was carried out over structure factors with resolution  $30 \text{ \AA} \geq 1/|s| \geq 25 \text{ \AA}$ . The standard crystallographic  $R$  factor has reached 40% for the optimal scale factor. Then the contribution of the Fourier coefficients  $\kappa T_s \exp(i\varphi_s)$  of the resolution lower than  $30 \text{ \AA}$  was added to the initial synthesis. This new synthesis was a bit worse than a similar synthesis from the first test but was better than the initial one.

## VI. Application to EFG structure investigation

The three-dimensional structure of the elongation factor  $G$  (EFG) from *Thermus thermophilus* has been investigated in the Protein Research Institute and Research Computing Centre of the USSR Academy of Sciences. The molecular weight of EFG is about 77 kDa, crystals belong to space group  $P2_12_12_1$ ,  $a = 75.9$ ,  $b = 105.6$ ,  $c = 115.9 \text{ \AA}$ , the solvent volume is 55–60% (Chirgadze *et al.*, 1983). The observed structure-factor moduli with a resolution up to  $6 \text{ \AA}$  were measured for the native protein and for one heavy-atom derivative. The standard SIR gave a practically uninterpretable  $8 \text{ \AA}$  resolution synthesis (Fig. 7a) calculated with about 1700 reflections. However, 29 reflections with resolution below  $30 \text{ \AA}$  were excluded from the calculation because of the absence of the phase values for them. The application of the standard AMB approach (Wang, 1985) failed to give a compact region, whereas the DSF technique with a  $\mu$  value corresponding to bounding of 10–30% of the unit-cell volume and  $R = 20 \text{ \AA}$  gave a clear molecule package. The similarity of the results with those of an independent direct-phase determination (Lunin, Urzhumtsev & Skovoroda, 1990) and other approaches (Chirgadze *et al.*, 1991) make us think the results are reliable.

The addition of 29 low-resolution structure factors after phases have been retrieved with the described technique has led to a considerable improvement of the synthesis (Fig. 7b). The molecule region can

clearly be seen in the electron density map without any artificial solvent flattening.

## VII. Concluding remarks

1. The problem of localizing the molecule in a noisy SIR synthesis of low resolution is due to the omission of low-resolution structure factors from the calculation of the synthesis rather than to the substitution of SIR phases for the exact phases and the usual errors in the heavy-atom parameters.

2. Molecule localization in a noisy synthesis may be obtained by non-linear filtration leaving only 10–30% of the unit-cell volume, followed by linear filtration with a large averaging radius  $R$ .

3. To retrieve phases of low-resolution reflections, non-linear filtration on the initial synthesis is sufficient and the phases of the Fourier coefficients of the filtered distribution may be taken as an approximation. When corresponding moduli are also absent, the addition of calculated Fourier coefficients instead

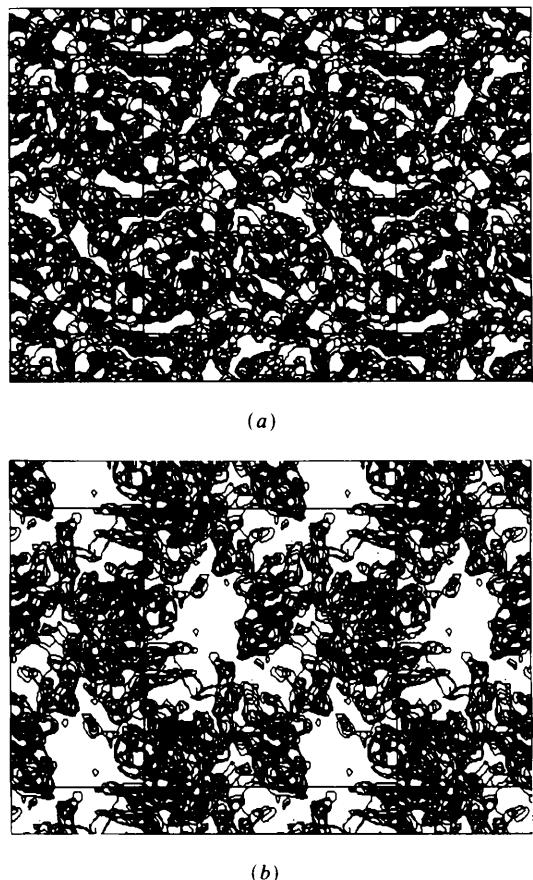


Fig. 7. Electron density maps of the elongation factor  $G$  at  $8 \text{ \AA}$  resolution, sections  $x = 0/36-18/36$ . (a) Initial SIR synthesis, reflections with a resolution lower than  $30 \text{ \AA}$  are absent. (b) Combined synthesis; 29 low-resolution reflections with phases of the filtered synthesis are added to the initial synthesis.

of missing low-resolution structure factors may improve the synthesis.

The author thanks Dr V. Yu. Lunin for useful discussions during the work and O. M. Liguinchenco for help in the preparation of the paper. The author is indebted to Professor Yu. N. Chirgadze for allowing him to work with the EFG experimental data. Electron density maps were analysed on an IBM PC computer using the graphical program *FAN* by E. A. Vernoslova and V. Yu. Lunin (Research Computing Centre, Pushchino, USSR).

#### References

- BHAT, T. N. & BLOW, D. M. (1982). *Acta Cryst. A* **38**, 21-29.  
 BLOW, D. M. & ROSSMANN, M. G. (1961). *Acta Cryst. A* **14**, 1195-1202.  
 BRICOGNE, G. (1974). *Acta Cryst. A* **30**, 395-405.  
 CHIRGADZE, YU. N., BRAZHNICKOV, E. V., GARBER, M. B., NIKONOV, S. V., FOMENKOVA, N. P., LUNIN, V. YU., URZHUMTSEV, A. G., CHIRGADZE, N. YU. & NEKRASOV, YU. V. (1991). *Dokl. Akad. Nauk SSSR*. In the press.  
 CHIRGADZE, YU. N., NIKONOV, S. V., BRAZHNICKOV, E. V., GARBER, M. B. & RESHETNIKOVA, L. S. (1983). *J. Mol. Biol.* **168**, 449-450.  
 FENDERSON, F. F., HERRIOTT, J. R. & ADMAN, E. T. (1990). *J. Appl. Cryst.* **23**, 115-131.  
 HARRISON, S. C. & JACK, A. (1975). *J. Mol. Biol.* **97**, 173-191.  
 LESLIE, A. G. V. (1987). *Acta Cryst. A* **43**, 134-136.  
 LUNIN, V. YU. (1988). *Acta Cryst. A* **44**, 144-150.  
 LUNIN, V. YU., URZHUMTSEV, A. G. & SKOVORODA, T. P. (1990). *Acta Cryst. A* **46**, 540-544.  
 LUZZATTI, V., MARIANI, P. & DELACROIX, H. (1988). *Makromol. Chem. Macromol. Symp.* **15**, 1-17.  
 PODJARNY, A. D. (1987). In *Crystallography in Molecular Biology*. NATO Advanced Science Institute Series, Ser. A, Vol. 126, pp. 63-79. New York, London: Plenum Press.  
 PODJARNY, A. D., MORAS, D., NAVAZA, J. & ALZARI, P. M. (1988). *Acta Cryst. A* **44**, 545-551.  
 PODJARNY, A. D., SCHEVITZ, R. W. & SIGLER, P. B. (1981). *Acta Cryst. A* **37**, 662-668.  
 QURASHI, M. M. (1953). *Acta Cryst.* **6**, 103.  
 REES, D. S., LEWIS, M. & LIPSCOMB, W. N. (1983). *J. Mol. Biol.* **168**, 367-387.  
 STUART, D. (1988). Private communication.  
 TSOUCARIS, G. (1970). *Acta Cryst. A* **26**, 492-499.  
 URZHUMTSEV, A. G. (1985). *The Use of Local Averaging in Macromolecular Image Analysis with Electron Density Maps*. Preprint, Pushchino, USSR.  
 URZHUMTSEV, A. G., LUNIN, V. YU. & LUZYANINA, T. B. (1986). Tenth Eur. Crystallogr. Meet., Wroclaw, Poland, 5-9 August 1986. Collected Abstracts pp. 51-52.  
 URZHUMTSEV, A. G., LUNIN, V. YU. & LUZYANINA, T. B. (1989). *Acta Cryst. A* **45**, 34-39.  
 WANG, B. C. (1985). In *Methods Enzymol.* **115**, 90-112.  
 ZELWER, CH. (1988). *Acta Cryst. A* **44**, 485-495.

*Acta Cryst.* (1991). **A47**, 801-803

## Unstressed Superlattices

BY M. E. POLYAKOV

*B. I. Stepanov Institute of Physics, BSSR Academy of Sciences, Leninsky prospekt 70, 220602 Minsk, USSR*

(Received 25 February 1991; accepted 12 June 1991)

#### Abstract

A method for calculating residual stresses in superlattices has been developed in the general form. Allowances have been made for the misfit dislocations, lattice mismatch and difference in the thermal expansion coefficients. The condition for the absence of residual stresses in superlattices has been obtained analytically in the case of both coherent growth and misfit dislocations.

#### Introduction

The main feature of the growth of semiconductor structures consisting of alternating layers of chemically different materials is the mismatch between the bulk lattice constants of different materials at epitaxy

temperature, interfacial misfit dislocations and the difference in linear expansion coefficients.

Recently there has been strong interest in the heteroepitaxial growth of III-V compound\* semiconductors on Si. However, there are serious problems concerning the growth of III-V semiconductors on Si substrates, one of which is the high density of dislocations generated and the other is residual stress in the growing films due to the large lattice mismatch and the difference in expansion coefficients of III-V films and Si substrates.

In this paper a method for calculating residual stresses in superlattices has been developed in the general form. Allowances have been made for the misfit dislocations, lattice mismatch and difference in thermal expansion coefficients of layers. The condition for the absence of residual stresses in superlattices has been obtained analytically in the case of both coherent growth and misfit dislocations.

\* 3-5 compounds in IUPAC (1988) nomenclature.