



**HAL**  
open science

**A translation function combining packing and diffraction  
information: an application to lysozyme  
(high-temperature form)**

Yoshinori Harada, Alain Lifchitz, Jean Berthou, Pierre Jollès

► **To cite this version:**

Yoshinori Harada, Alain Lifchitz, Jean Berthou, Pierre Jollès. A translation function combining packing and diffraction information: an application to lysozyme (high-temperature form). *Acta Crystallographica Section A: Foundations and Advances* [2014-..], 1981, 37, pp.398-406. 10.1107/S0567739481000867. hal-00015162

**HAL Id: hal-00015162**

**<https://hal.science/hal-00015162>**

Submitted on 23 Feb 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## A Translation Function Combining Packing and Diffraction Information: An Application to Lysozyme (High-Temperature Form)

BY YOSHINORI HARADA,\* ALAIN LIFCHITZ† AND JEAN BERTHOU

*Laboratoire de Minéralogie et de Cristallographie, associé au CNRS, Université de Paris VI, 4 place Jussieu,  
F-75230 Paris CEDEX 05, France*

AND PIERRE JOLLES

*Laboratoire des Protéines, Université de Paris V, 45 rue des Saints Pères, F-75270 Paris CEDEX 06, France*

(Received 10 December 1979; accepted 9 December 1980)

### Abstract

A new translation function  $T(\mathbf{t})$  is defined, whose main peak is expected to give directly the absolute position  $\mathbf{t} = \mathbf{t}_0$  of a known molecular configuration (the isostructural model) of known orientation in an unknown crystal structure.  $T(\mathbf{t})$  combines all required information: crystal symmetry, molecular steric properties of the model and experimental data (observed structure factors  $F_o$ ) leading to more reliable results:  $T(\mathbf{t}) = TO(\mathbf{t})/O(\mathbf{t})$ , where  $TO(\mathbf{t})$  measures the degree of agreement between  $F_o$  and  $F_c(\mathbf{t})$ , while  $O(\mathbf{t})$  is a function measuring intermolecular overlap.  $TO$  and  $O$  can both be computed efficiently by the possibility of their expansion in Fourier series, and the use of FFT. A theoretical study followed by application to a simple synthetic structure cast light upon ambiguities and failures [e.g. Bott & Sarma (1976). *J. Mol. Biol.* **106**, 1037–1046] encountered in using translation functions previously described [e.g. Crowther & Blow (1967). *Acta Cryst.* **23**, 544–548]. The crystal structure of the high-temperature form of hen-egg-white lysozyme was determined at 6 Å resolution by the application of this fast translation function and compared with the equivalent structure obtained by the isomorphous replacement method.

### 1. Introduction

Since Rossmann & Blow (1962) formulated the general concepts of molecular replacement, many attempts have been made to apply this method to solve protein structures. Meanwhile, increasing evidence has ap-

peared that protein families exist, and their analysis has demonstrated similarities between the members, in all or part of their structure. This has enlarged the field of applications of this technique.

For an unknown crystal structure, two situations may exist: one where nothing of the molecular configuration is known and a second where part or all of it is already at hand. We are only interested in this last situation.

There are two stages in the application of the molecular replacement technique: one has to determine first the relative or absolute orientation of the known molecular model in the new crystal, and then its position relative to the symmetry elements of the cell, or better, relative to its origin.

The first stage is solved satisfactorily by the rotation function  $R(\theta)$  defined by Rossmann & Blow (1962), in spite of the large amount of computing needed. The numerical calculation of  $R(\theta)$  was substantially improved by Crowther's (1972) fast rotation algorithm using FFT.  $R(\theta)$  proved to be successful, in many cases, such as insulin (Dodson, Harding, Hodgkin & Rossmann, 1966), lactate dehydrogenase (Hackert, Ford & Rossmann, 1973), lysozymes (Joynson, North, Sarma, Dickerson & Steinrauf, 1970; Bott & Sarma, 1976; Nixon & North, 1976), etc.

One of the main obstacles to the application of the molecular replacement method lies in the determination of the translational parameters. Many authors have proposed their own translation function, some of them relevant to macromolecules with their particular characteristics: a large number of atoms and low resolution [Vand & Pepinsky, 1956; Tollin, 1966; Crowther & Blow, 1967 (hereafter CB); Langs, 1973; Hendrickson & Ward, 1976; etc.].

All these functions have been applied more or less satisfactorily to particular cases, but none of them has been generalized. Whatever the correlation functions defined and used, either in reciprocal or real space, the

\* Present address: c/o Professor Y. Iitaka, Faculty of Pharmaceutical Sciences, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113, Japan.

† To whom all communications should be sent.

main drawbacks turned out to be the amount of computing and a general lack of reliability. In order to overcome the former difficulty, many of the authors have proposed to handle only a reduced quantity of data, and in doing so, *i.e.* in limiting the information, have obtained poor correlation or, in other cases, the disappearance of the expected peaks. On the other hand, some authors have tried to improve the reliability by an *a posteriori* checking of a reasonable packing (Bott & Sarma, 1976).

The multiplicity of the translation functions has imposed upon us a necessity to compare the efficiency of some of them and their specific advantages and, finally, to develop, test and apply two new functions: an overlap and a translation function.

## 2. Derivation of the two functions

Let

$$\mathbf{x}_2 = C(\theta) \mathbf{x}_1 + \mathbf{t}, \quad (1)$$

where  $C(\theta)$  is the rotation matrix and  $\mathbf{t}$  the absolute translation vector which relates the equivalent points  $\mathbf{x}_1$  of the isostructural model and  $\mathbf{x}_2$  in the unknown crystal.

Suppose that absolute orientation,  $C(\theta_0)$ , has been already determined by a rotation function such as Rossmann & Blow's. In this situation, we have only to determine  $\mathbf{t}$ . How can we do so with only the above-mentioned information?

In order to determine the value of the translational parameter  $\mathbf{t} = \mathbf{t}_0$ , which will allow the molecular model to build up a reasonable crystal structure in the unknown cell, various translation functions have been defined in direct or in reciprocal space. They can be divided into three classes: (i) product-type, (ii) sum-type and (iii) difference-type functions, applied to the two sets of quantities, observed and calculated, to be brought into agreement. The present investigation is essentially an exploration of the first type.

In general, the linear correlation coefficient  $\mathcal{C}$  measuring the degree of agreement between two sets  $\{x_i\}$  and  $\{y_i\}$  of non-negative values is defined as

$$\mathcal{C} = \frac{\sum_i x_i y_i}{\left( \sum_i x_i^2 \sum_i y_i^2 \right)^{1/2}}, \quad (2)$$

$\{x_i\}$  and  $\{y_i\}$  being the coordinates of two vectors  $\mathbf{x}$  and  $\mathbf{y}$ , in the sense of vector space.

$\mathcal{C}$  has two useful properties:

(i) it is independent of the scale of  $\{x_i\}$  and  $\{y_i\}$  (by the way, the scale does not have a very clear meaning, particularly in the case of a partial model and when the two sets are not exactly proportional);

(ii)  $0 \leq \mathcal{C} \leq 1$  because it is intrinsically a cosine; so  $\mathcal{C}$  is a normalized measure of agreement,  $\mathcal{C} = 1$

corresponding to a total agreement between  $\{x_i\}$  and  $\{y_i\}$ ,  $y_i = kx_i$  ( $k = \text{constant}$ ).

It seemed interesting to apply such a concept to the translation problem.  $\mathcal{C}(\mathbf{t})$  will measure the degree of agreement between the observed  $\{F_o^2(\mathbf{h})\}$  and calculated  $\{|F_c(\mathbf{h}, \mathbf{t})|^2\}$  intensities depending on  $\mathbf{t}$ :

$$\mathcal{C}(\mathbf{t}) = \frac{\sum_{\mathbf{h}} F_o^2(\mathbf{h}) |F_c(\mathbf{h}, \mathbf{t})|^2}{\left\{ \sum_{\mathbf{h}} F_o^4(\mathbf{h}) \sum_{\mathbf{h}} |F_c(\mathbf{h}, \mathbf{t})|^4 \right\}^{1/2}}, \quad (3)$$

where  $S$  is a reciprocal sphere whose radius is equal to the resolution of the  $F_o$ 's. The highest  $\mathcal{C}(\mathbf{t})$ , corresponding to  $\mathbf{t} = \mathbf{t}_0$ , is expected to determine directly the position of the model in the unknown crystal and, moreover, it gives an absolute measure of agreement likelihood.

The sole information needed to evaluate (3) is: (i)  $F_o$ 's; (ii) the space group of the observed structure; and (iii)  $F_m(\mathbf{h})$ , the molecular structure factors, *i.e.* the Fourier transform of the electron density of the properly oriented model.\* Such a translation function definition has, *a priori*, several advantages: (i) no explicit information concerning atom occupancies and coordinates have to be known and (ii) the result (*i.e.*  $\mathbf{t}_0$ ) is, as a rule, independent of the resolution (except for the precision).

The first term of the denominator of (3) is obviously constant because it does not depend on  $\mathbf{t}$ . But its second term  $\left\{ \sum_{\mathbf{h}} |F_c(\mathbf{h}, \mathbf{t})|^4 \right\}^{1/2}$  has no reason to be constant. We thought that neglecting this denominator's variation with respect to  $\mathbf{t}$  is the reason why translation functions of scalar-product type have not been as efficient as expected (Nixon & North, 1976; Rae, 1977).

*A priori* the numerical calculation of the whole set of  $\mathcal{C}(\mathbf{t})$  values seems heavy. Although the number of operations can be reduced, because of the crystal and molecular symmetry, nevertheless, one has to compute structure factors for the whole set of  $\mathbf{t}$  values (typically  $32 \times 32 \times 32$  samples),<sup>†</sup> and sum up all  $F_o^2(\mathbf{h})|F_c(\mathbf{h}, \mathbf{t})|^2$  and  $|F_c(\mathbf{h}, \mathbf{t})|^4$  for each  $\mathbf{h}$  pertaining to the set of independent reflexions. This problem is crucial and on its resolution depends the practical interest of  $\mathcal{C}$ .

\* In this paper 'model' means the whole atomic content of the asymmetric unit of the unknown crystal cell.

<sup>†</sup> This volume must be, at most, equal to the crystal unit cell, but, depending on the location of the permissible origins in a given space group, may be smaller or greater than the asymmetric unit, as Hendrickson & Ward (1976) pointed out. For example, in  $P2_12_12_1$ , we have a region  $(0 - a/2, 0 - b/2, 0 - c/2)$ , smaller than the asymmetric unit.

Nixon & North (1976) and Rae (1977) have shown that the iterative calculation of  $F_c(\mathbf{h}, \mathbf{t})$  is practicable, when using linear combination of molecular structure factors  $F_m(\mathbf{h})$ . We propose a new algorithm, restricted to an analytic function of  $|F_c(\mathbf{h}, \mathbf{t})|^2$ , which reduces drastically the total amount of computation. Thus,  $\sum_{\mathbf{h}} A(\mathbf{h}) |F_c(\mathbf{h}, \mathbf{t})|^{2n}$ , a periodic function of  $\mathbf{t}$ , can be easily developed in a Fourier series with respect to the reciprocal index  $\mathbf{p}$  of the translation parameter  $\mathbf{t}$ , as is shown, for  $\sum_{\mathbf{h}} F_o^2(\mathbf{h}) |F_c(\mathbf{h}, \mathbf{t})|^2$ , in the Appendix. The method requires two steps: (i) numerical evaluation of the Fourier expansion coefficients and (ii) Fourier transform, by FFT, of these coefficients. This process allows a global exploration of the entire asymmetric unit of translation space, with standard Fourier synthesis programs.

Let the numerator of  $\mathcal{E}(\mathbf{t})$  be  $TO(\mathbf{t})$ . It is what we call a quasi-correlation function for the reason that the vector\* norms are not taken into account. On the other hand, in direct space,  $TO(\mathbf{t})$  represents a scalar product – which is currently and improperly named ‘correlation’ – between the observed  $P_o(\mathbf{u})$  and calculated  $P_c(\mathbf{u}, \mathbf{t})$  Patterson functions, according to Parseval’s theorem. So,

$$TO(\mathbf{t}) = \sum_{\mathbf{h}} F_o^2(\mathbf{h}) |F_c(\mathbf{h}, \mathbf{t})|^2 / \sum_{\mathbf{h}} F_o^4(\mathbf{h}) \quad (4)$$

$$\simeq \int_{\text{unit cell}} P_o(\mathbf{u}) P_c(\mathbf{u}, \mathbf{t}) d\mathbf{u}, \quad (4')$$

where  $\sum_{\mathbf{h}} F_o^4(\mathbf{h})$ , which does not depend upon  $\mathbf{t}$ , was introduced to make  $TO^\dagger$  a normalized quantity independent of the structure scale and volume, with  $TO(\mathbf{t}) \simeq 1$ .

As for the second term of the denominator, we have replaced  $\{\sum_{\mathbf{h}} |F_c(\mathbf{h}, \mathbf{t})|^4\}^{1/2}$  by its first approximation in order to simplify the calculation. This leads to a new function:

$$O(\mathbf{t}) = \sum_{\mathbf{h}} |F_c(\mathbf{h}, \mathbf{t})|^2 / N \sum_{\mathbf{h}} |F_m(\mathbf{h})|^2, \quad (5)$$

where  $N$  stands for the number of asymmetric units in the crystal cell.  $N \sum_{\mathbf{h}} |F_m(\mathbf{h})|^2$ , which does not depend upon  $\mathbf{t}$  and is strictly equal to  $\sum_{\mathbf{h}} |F_c(\mathbf{h}, \mathbf{t})|^2$ , when no overlap occurs, as will be shown later on, plays the same role as  $\sum_{\mathbf{h}} F_o^4(\mathbf{h})$  for  $TO$ . Although  $O(\mathbf{t})$  was introduced as an approximation, it has its own physical meaning: it is a steric function, which will be discussed in the next section.

\* With the sense above stated.

† The definition (4) makes  $TO(\mathbf{t})$  identical with the ‘residual’  $R_3$  by Rae (1977).

Finally, the translation function has the following expression:

$$T(\mathbf{t}) = \frac{TO(\mathbf{t})}{O(\mathbf{t})} = N \sum_{\mathbf{h}} |F_m(\mathbf{h})|^2 \frac{\sum_{\mathbf{h}} F_o^2(\mathbf{h}) |F_c(\mathbf{h}, \mathbf{t})|^2}{\sum_{\mathbf{h}} F_o^4(\mathbf{h}) \sum_{\mathbf{h}} |F_c(\mathbf{h}, \mathbf{t})|^2}. \quad (6)$$

As shown in the following two applications with, on the one hand, ideal model structure data and, on the other, experimental data, introduction of  $O(\mathbf{t})$  has been able to attenuate the ‘false’ peaks in  $TO$  maps owing to its character of scalar product.

### 3. Physical meaning of $O(\mathbf{t})$ and interpretation of $T(\mathbf{t})$

Suppose a hypothetical isolated molecule whose electron density distribution  $\rho_m(\mathbf{x})$  is represented by a rectangular function in the one-dimensional lattice of cell  $a$  (Fig. 1). If there are two identical molecules in this cell, related by a mirror plane, the crystal electron density  $\rho_x(\mathbf{x})$  becomes

$$\rho_x(\mathbf{x}) = \rho_m(\mathbf{x}) + \rho_m(\mathbf{x}'), \quad (7)$$

where  $\mathbf{x}' = A(\mathbf{x} + \mathbf{t}) + \mathbf{d}$ ,  $A$  is the rotation of the symmetry operation,  $\mathbf{d}$  is its associated translation – in this particular case  $A = -1$ ,  $\mathbf{d} = \mathbf{0}$  – and  $\mathbf{t}$  the absolute position of the reference molecule.

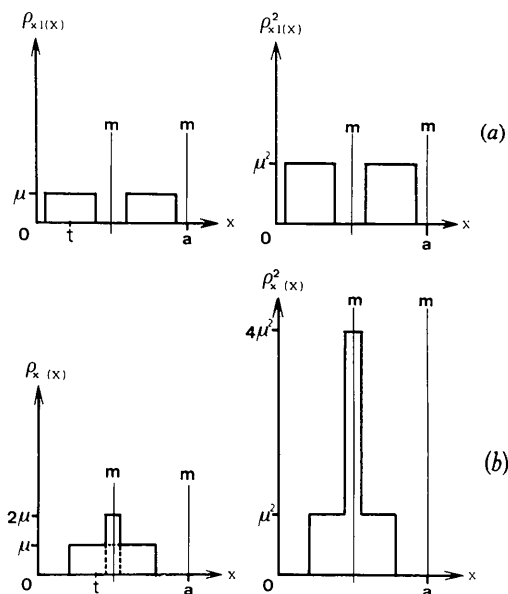


Fig. 1. Schematic interpretation of  $O(\mathbf{t}) = \int_0^a \rho_x^2(x) dx$ , showing how it takes a greater value when molecular electron densities overlap, in spite of conservation of  $\int_0^a \rho_x(x) dx$  in each case. On the left, electron density distribution of a hypothetical crystal structure containing two molecules related by a mirror plane and, on the right, the squared density: (a) when there is no intermolecular overlap; (b) in the case of partial overlap.

Let us square equation (7):

$$\rho_x^2(\mathbf{x}) = \rho_m^2(\mathbf{x}) + \rho_m^2(\mathbf{x}') + 2\rho_m(\mathbf{x})\rho_m(\mathbf{x}'). \quad (8)$$

We will examine the particular value of this expression for each of two possible situations of 'crystal packing', depending on  $\mathbf{t}$ .

In the first situation (Fig. 1a), in which two molecules are completely separated, and no intermolecular electron density overlap occurs, (8) takes the following particular form

$$\rho_{x_1}^2(x) = \rho_m^2(x) + \rho_m^2(x'). \quad (9)$$

By application of Parseval's theorem

$$\int_0^a \rho_m^2(\mathbf{x}) \, d\mathbf{x} = \frac{1}{a} \sum_{\mathbf{h}} |F_m(\mathbf{h})|^2$$

to the preceding equation, we obtain:

$$\int_0^a \rho_{x_1}^2(\mathbf{x}) \, d\mathbf{x} = \frac{2}{a} \sum_{\mathbf{h}} |F_m(\mathbf{h})|^2. \quad (10)$$

The second situation corresponds to partial or entire overlap of the two molecules (Fig. 1b). In the extreme case of total overlap, (8) becomes

$$\rho_{x_2}^2(\mathbf{x}) = \{2\rho_m(\mathbf{x})\}^2 = 4\rho_m^2(\mathbf{x}). \quad (11)$$

By Parseval's theorem,

$$\int_0^a \rho_{x_2}^2(\mathbf{x}) \, d\mathbf{x} = \frac{4}{a} \sum_{\mathbf{h}} |F_m(\mathbf{h})|^2, \quad (12)$$

where the sum of the squared electron density on an entire cell takes twice the value of that of the preceding situation (no overlap). Therefore, we can deduce the following double inequalities in direct space

$$\int_0^a \rho_{x_1}^2(\mathbf{x}) \, d\mathbf{x} \leq \int_0^a \rho_x^2(\mathbf{x}) \, d\mathbf{x} \leq \int_0^a \rho_{x_2}^2(\mathbf{x}) \, d\mathbf{x} \quad (13)$$

and equivalent double inequalities in reciprocal space

$$2 \sum_{\mathbf{h}} |F_m(\mathbf{h})|^2 \leq \sum_{\mathbf{h}} |F_c(\mathbf{h}, \mathbf{t})|^2 \leq 4 \sum_{\mathbf{h}} |F_m(\mathbf{h})|^2. \quad (14)$$

A term-by-term division of the inequalities (14) by the positive quantity  $2\sum_{\mathbf{h}} |F_m(\mathbf{h})|^2$ , which does not depend upon the variable  $\mathbf{t}$ , gives the following relation

$$1 \leq \frac{\sum_{\mathbf{h}} |F_c(\mathbf{h}, \mathbf{t})|^2}{\sum_{\mathbf{h}} |F_m(\mathbf{h})|^2} \leq 2. \quad (15)$$

This discussion can be extended, without loss of generality, to the three-dimensional case, leading to the overlap function  $O(\mathbf{t})$  (5) with the double inequality

$$1 \leq O(\mathbf{t}) \leq N. \quad (16)$$

$O(\mathbf{t})$  can be considered as the absolute measure of the degree of electron density intermolecular overlap, for

the synthetic crystal built from the reference molecule in position  $\mathbf{t}$ . More precisely,  $O(\mathbf{t})$  is a steric function which demarcates a permissible region, for the model, from crystal packing arrangement considerations.

Therefore,  $T(\mathbf{t})$  (6) can be written formally:

$$T(\mathbf{t}) = \frac{\text{biased measure of agreement between observed and calculated intensities}}{\text{measure of intermolecular overlap in synthetic crystal}}.$$

Being defined like this,  $T(\mathbf{t})$  is expected to have a high amplitude when the agreement between observed and calculated  $F$ 's is important and simultaneously the intermolecular overlap is small. Conversely, the  $T(\mathbf{t})$  amplitude will be lowered when the agreement is weak and/or intermolecular overlap important.

In general, one criterion frequently invoked to establish the validity of a resulting trial crystal structure derived from exploration of a reciprocal quantity is the reasonableness of the packing arrangement in direct space. Expression (6) involves these two kinds of information simultaneously. So  $T(\mathbf{t})$  should be a more reliable translation function than those using only one-sided information and well adapted to practical applications owing to the ease of its computation.

#### 4. Applications

Functions  $O(\mathbf{t})$  and  $T(\mathbf{t})$  have been tested on hypothetical 'observed' data derived from an artificial crystal structure, in order to verify their efficiency and limits; on the other hand, they were used for the low-resolution structure determination of the high-temperature crystal form of hen-egg-white lysozyme.

##### 4.1. Artificial simple crystal structure

Parameters of an artificial simple crystal structure, whose motif consists of two carbon atoms, are summarized in Table 1. This structure is quite peculiar.

Table 1. Parameters of an artificial crystal structure for test

Space group  $P2_12_12_1$

$$\begin{aligned} a &= b = 10 \text{ \AA} \\ c &= 40 \text{ \AA} \\ Z &= 4 \end{aligned}$$

Standard motif

Atom	Coordinates		
	$x$	$y$	$z$
C(1)	$\frac{1}{2}$	$\frac{1}{2}$	0
C(2)	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

Although far from being realistic it has some convenient properties for a test experience:

the most unacceptable packing is geometrically predictable;

the total overlap could be obtained, for this unacceptable packing, owing to molecular cylindrical symmetry;

the artefact peaks will certainly occur in the  $TO$  map, owing to atoms in special positions.

Thus, it is a good candidate to explore and verify the expected behaviour of  $TO(t)$ ,  $O(t)$  and  $T(t)$ . The molecular  $F_m(\mathbf{h})$  and 'observed'  $|F_c(\mathbf{h}, \mathbf{0})|$  structure factors were calculated up to  $3.2 \text{ \AA}$  resolution.

The section  $t_z = 0$  of the function  $TO(t)$  represented in Fig. 2(a) exhibits three distinct peaks  $P1(0,0,0)$ ,  $P3(0, \frac{1}{4}, 0)$  and  $P4(\frac{1}{4}, \frac{1}{4}, 0)$  corresponding to three expected vectorial solutions which are associated with three crystallographically equivalent arrangements, leading finally to a unique crystal structure with different choices of the unit-cell origin.

On the other hand, the peak  $P2(\frac{1}{4}, 0, 0)$  of Fig. 2(a) has a different nature, although it has an amplitude equal to that of the preceding peaks. It corresponds to an unacceptable molecular arrangement from crystal packing considerations. Namely, in this molecular arrangement, built up from a vectorial solution associated with  $P2$ , we observe a total interpenetration of component atoms. This is an example of the kind of drawback which occurs when scalar-product-type translation functions are used: important peaks could occur for sterically impossible solutions.

This fact justifies the division of  $TO(t)$  by the overlap function  $O(t)$  (Fig. 2b) whose principal and unique peak is at  $t_0 = (\frac{1}{4}, 0, 0)$ . As shown in Fig. 2(c), this operation giving  $T(t)$  attenuates considerably the unacceptable peak  $P2$ , leaving only three main solution peaks.

#### 4.2. High-temperature form of hen-egg-white lysozyme

The crystal and molecular structure of lysozyme in its low-temperature crystal form ( $A$  form,  $T < 298 \text{ K}$ ) has been determined at high resolution by Blake, Mair, North, Phillips & Sarma (1967), using the multiple heavy atom isomorphous replacement method. On the other hand, Jolles & Berthou (1972) have described a thermodependent crystal dimorphism of this enzyme,

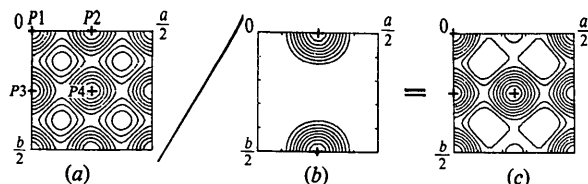


Fig. 2. Sections ( $t_z = 0$ ) through the peak of the highest value for (a)  $TO(t)$ , (b)  $O(t)$  and (c)  $T(t)$ , when using error-free hypothetical data, calculated from a test crystal structure made up of two atoms in the asymmetric unit of  $P2_12_12_1$ . Contours between 10 and 100% of the range of values are shown.

and shown a possibility of crystal and molecular structure determination of its high-temperature form: the  $B$  form (Table 2).

We have tried to determine this structure by two independent methods: the isomorphous replacement method and the molecular replacement method which suits this case. There is, of course, no difference in the amino-acid sequence of the two forms  $A$  and  $B$ . Hence, global geometric similarity between the two molecular forms is largely expected. Thus, molecule  $A$  can be used as an isostructural model of molecule  $B$ , although enzymatic activity differences between the two have been detected in solution (Saint-Blancard, Clochard, Cozzone, Berthou & Jolles, 1977) associated with slight structural modifications (Cozzone, Opella, Jardeitzky, Berthou & Jolles, 1975).

The data were collected to a resolution of  $6 \text{ \AA}$  on the heavy-atom derivatives and  $2.5 \text{ \AA}$  for the native crystals, with automatic diffractometers (Philips PW1100 and Hilger & Watts).

The use of the MIR technique proved to be difficult. The heavy atoms, indeed, do not bind easily to the  $B$  form. Only two appeared to be successful,  $K_3UO_2F_5$  and  $UO_2Ac_2$ , both of them giving the same diffraction pattern (Berthou, Laurent, Lifchitz, Jolles, Artymiuk, Blake & Rice, 1977). This difficulty was, by the way, another reason for trying to apply, in this case, the MR method.

$R(\theta)$  has been evaluated from the complete set of independent reflexions  $F_o^2(\mathbf{h})$  of the  $B$  form, in the range  $6\text{--}25 \text{ \AA}$ , and from  $|F_m(\mathbf{h})|^2$  of the isolated model placed in a hypothetical triclinic cell  $a = b = c = 60 \text{ \AA}$ ,  $\alpha = \beta = \gamma = 90^\circ$ .\* In such use of the rotation function, it is currently claimed that the hypothetical unit cell must have, in each direction, dimensions at least twice that of a single molecule (e.g. Joynson *et al.*, 1970). But according to our expectations, we have verified that a smaller one is sufficient, thus halving the computation and leading to the very same results (Lifchitz, 1981).

\* The lysozyme molecule can be circumscribed in a cube of  $40 \times 40 \times 40 \text{ \AA}$ .

Table 2. Parameters of the two crystal forms of hen-egg-white lysozyme

	$A$ form	$B$ form
Crystallization temperature	$< 298 \text{ K}$	$> 298 \text{ K}$
Space group	$P4_32_12$	$P2_12_12_1$
$a$	$79.1 \text{ \AA}$	$56.4 \text{ \AA}$
$b$	$79.1 \text{ \AA}$	$73.8 \text{ \AA}$
$c$	$37.9 \text{ \AA}$	$30.4 \text{ \AA}$
$Z$	8	4
Water content ( $w/w$ )	33.5%	36%
Rotation Eulerian angles of $A$ molecule in $B$ cell (not refined)	$\theta_0 = (240^\circ, 67^\circ, 35^\circ)$	
Translation of $A$ molecule in $B$ cell (not refined)	$t_0 = (0.89, 0.98, 0.53)$	

The three Eulerian angles which specify the orientation (Rossmann & Blow, 1962) of the molecule *A* in the *B*-form cell (Table 2), obtained with an adapted version of the fast rotation program (Crowther, 1972), do correspond to the main peak in the  $R(\theta)$  map: this peak stood 4.3 standard deviations above the mean value.

In the second stage, the translational search was done by the three-dimensional function  $T(\mathbf{t})$  in order to determine the absolute position of the model in the *B* cell, with:

(1) orientational parameters  $\theta_0$  determined in the preceding rotational search;

(2) atomic fractional coordinates of the model *A* for calculating  $F_m(\mathbf{h})$ ;

(3) 380 independent  $F_o(\mathbf{h})$  of the *B* form in the resolution range 6–25 Å.

Computing time of  $TO(\mathbf{t})$  and  $O(\mathbf{t})$  were about 10 s each, on an IBM 370/168 computer for  $\mathbf{t}$  sampling  $\Delta t = (a/64, b/64, c/32)$ , namely  $32 \times 32 \times 16 = 16\,384$  points for region  $(0-a/2, 0-b/2, 0-c/2)$ : an asymmetric unit of translation in  $P2_12_12_1$ .

First,  $TO(\mathbf{t})$  was calculated with normalized structure factors  $E_o(\mathbf{h})$  and  $E_m(\mathbf{h})$  instead of  $F_o(\mathbf{h})$  and  $F_m(\mathbf{h})$  in order to:

(a) avoid the difference of thermal agitation effect between  $\{F_o(\mathbf{h})\}$  and  $\{F_m(\mathbf{h})\}$ ; and

(b) make possible a balanced contribution of translational information, homogeneously distributed throughout the whole reciprocal space, which is improperly called, at low resolution, 'sharpened' data.  $\{F_o(\mathbf{h})\}$  and  $\{F_m(\mathbf{h})\}$  are normalized, by concentric shells, according to the relation

$$E(\mathbf{h}) = F(\mathbf{h}) / \langle |F(\mathbf{h})|^2 \rangle^{1/2}. \quad (17)$$

The particular section of  $TO(\mathbf{t})$  including the solution peak is shown in Fig. 3(a). Fig. 3(b) gives the corresponding section of  $O(\mathbf{t})$ . One region, sterically

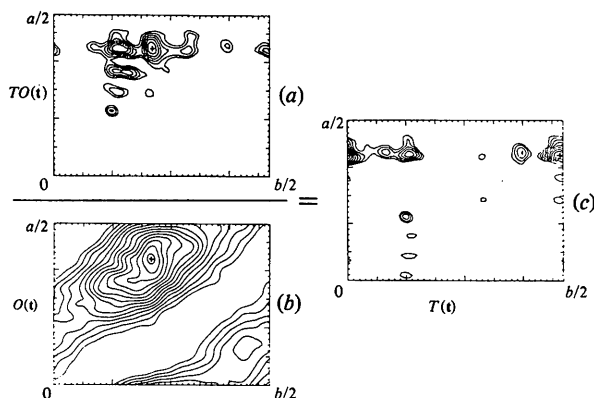


Fig. 3. Section ( $t_z = 1/32$ ) through the highest peak value of  $T(\mathbf{t})$ , and homologous sections of  $TO(\mathbf{t})$  and  $O(\mathbf{t})$ , for 6 Å experimental data of lysozyme *B*. Contours above the mean are shown. + indicates the maximum value for each section.

forbidden, is clearly revealed: comparing Fig. 3(a) to 3(b), we can see that the main peak of  $TO(\mathbf{t})$  is exactly situated in this region. This paradoxical result clearly displays the potential weakness of scalar-product-type translation functions.

Fig. 3(c) shows, on a section of the function  $T(\mathbf{t}) = TO(\mathbf{t})/O(\mathbf{t})$ , the main peak which corresponds to a solution vector  $\mathbf{t}'_0 = (25/64, 31/64, 1/32)$ , while its corresponding peak in the homologous section of the map  $TO(\mathbf{t})$  (Fig. 3a) is only the fifth peak in amplitude. So,  $\mathbf{t}_0$  (Table 2) is finally

$$\mathbf{t}_0 = \mathbf{t}'_0 + \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right). \quad (18)$$

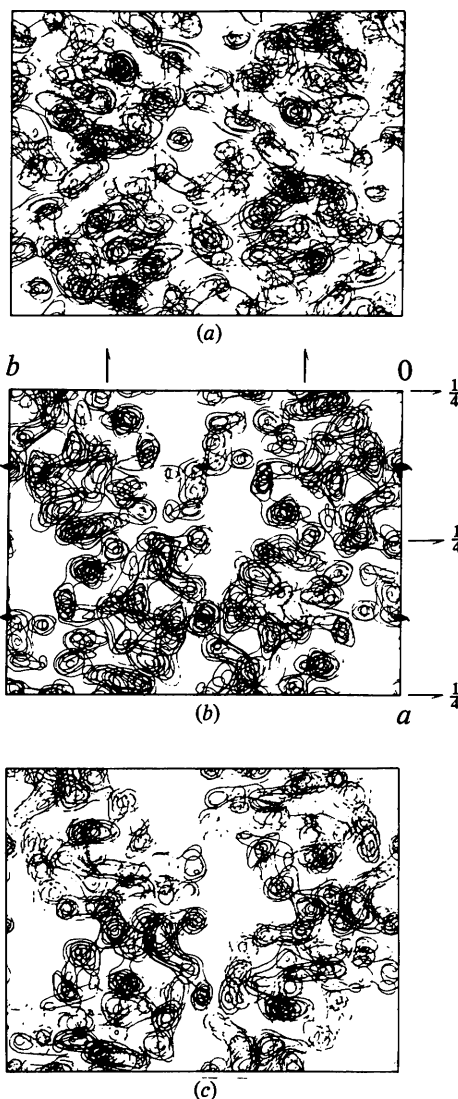


Fig. 4. Composite electron density maps of the whole cell of lysozyme *B* form, at 6 Å resolution: (a) obtained by the double isomorphous replacement method with  $\langle m \rangle = 0.66$ ; (b) calculated by employing phases determined from  $\theta_0$  and  $\mathbf{t}_0$  with  $\{F_o\}$  of the *B* form; (c) calculated from *A* molecule oriented and positioned by  $\theta_0$  and  $\mathbf{t}_0$  in the *B* cell. Contours are at arbitrary equal intervals.

The last vector selects a particular origin among the eight crystallographic equivalent positions, for a convenient comparison with the MIR map.

At this stage, three density maps at 6 Å, shown in Fig. 4, have been calculated:

(a) one obtained by the double isomorphous replacement method, with a mean figure of merit  $\langle m \rangle = 0.66$ ;

(b) one calculated by the molecular replacement method employing phases determined from  $\theta_0$  and  $t_0$  with  $F_0$  of the *B*-form crystal:

(c) one calculated from  $\{F_c(\mathbf{h}, \theta_0, t_0)\}$ .

First of all, the MIR electron density map clearly revealed the lysozyme molecule. Despite the low figure of merit, in the light of the high-resolution interpretation of molecule *A*, it was even possible to trace the molecular backbone in the map (Blake, 1977).

Secondly, as expected (Ramachandran & Srinivasan, 1961), the (b) and (c) maps show a great similarity.

Now it can be easily noted that the global aspect of (a) and (b) maps is quite similar, namely:

the orientation of the molecule;

its position with regard to the symmetry elements;

the region of weak density in the MIR map which corresponds to the solvent molecules are empty in the MR map;

the highest densities in each molecule are located in the same region and correspond to helical segments of the chain. The paths of the helices are in agreement, particularly the 88–96 segment which binds the two halves of the molecule.

This agreement is meaningful because neither the phases derived from the isomorphous replacement method nor the rotational parameter ( $\theta_0$ ) nor the translational one ( $t_0$ ) have been refined.

But, nevertheless, there exists visible orientational and positional differences. Positional errors are of about 3 Å in magnitude, *i.e.* roughly half of the maximum resolution of observed data. The convergence radius of refinement methods is of this order, hence, it is expected that a refinement by the rigid-body technique first at low resolution, then, by least squares, at higher resolution will converge.

## 5. Comparison with Crowther & Blow's translation function

Among the translation functions previously described, those by CB,  $T_0(t_r)$ ,\*  $T_1(t_r)$  and  $T_2(t_r)$ , have been applied most frequently to protein structures. They give only a relative translation vector, but except when ambiguities and artefacts made the problem difficult, these functions were helpful to solve many of them.

In order to compare their respective properties and advantages, with regard to those of  $T(t)$ , we have applied them to the two data sets: the observed  $\{F_o\}$  of lysozyme and a synthetic one  $\{F'_o(\theta_0, t_0)\}$  for which the *A* molecule was oriented and positioned according to the results of the MR technique.

This 'ideal' situation, as in the test used by CB, without any orientation inaccuracy and in which the model fits absolutely the molecule of the observed structure, enabled us to check the behaviour of the proposed translation functions in nearly realistic conditions, unless there is no bias introduced by the orientation stage.

It was expected, for  $P2_12_12_1$ , to find an intermolecular vector  $(t_x, t_y, \frac{1}{2}) = (2X, 2Y, \frac{1}{2})$  derived from pairs of molecules related by the  $2_1$  operator parallel to *c*, where (*X, Y, Z*) are coordinates of the local origin of the model with respect to the crystal origin. Therefore, to determine only *X* and *Y* from a relative translation function, a section  $t_z = \frac{1}{2}$  or even a projection along *c* would suffice.

First, we tried  $T_0(t_r)$  on  $F'_o$ . But the section  $t_z = \frac{1}{2}$  did not exhibit the expected peak, and only when we have used normalized  $F'_o: E'_o$  has it appeared in the map.

CB have introduced  $T_1(t_r)$  as an improved version of  $T_0(t_r)$ , where intramolecular vectors are removed from the observed Patterson function, increasing the signal-to-noise ratio. The previous experiment on  $T_0(t_r)$  was reproduced on  $T_1(t_r)$ . The expected peak was found, when using both normalized and unnormalized structure factors, but in the unnormalized map, which has poor contrast, this peak is slightly shifted.

$T_2(t_r)$  is a sum of  $T_0$ -like expressions, so it was not surprising not to find a solution with a peak-to-background ratio better than that of  $T_0(t_r)$ .

These comparative tests on error-free data clearly showed that (a)  $T_1(t_r)$  was superior to  $T_0(t_r)$  and  $T_2(t_r)$ , as CB predicted, and, particularly, (b) the signal-to-noise ratio of  $T_1(t_r)$ , and its accuracy too, can be improved by using  $E_m$  and  $E'_o$  instead of  $F_m$  and  $F'_o$  (necessity of normalization).

It is worthwhile to point out that if the peak solution was clear in the case of *c* projection, it was not so easily detected in the *a* and *b* projections, even when using  $T_1(t_r)$ , the most powerful function, on normalized factors  $E_m$  and  $E'_o$ .

In a second series of experiments,  $T_1(t_x, t_y, \frac{1}{2})$  was calculated with  $E_m$ , which necessarily contains experimental orientation inaccuracy, with  $E_o$  for the *B* form. The global aspect of this map resembled considerably the one calculated with error-free data. However, no peak corresponding to the expected intermolecular vector was observed. This failure can be understood from the nature of  $T_1$ : if the rotation is not known accurately enough, the intramolecular vectors

\* To avoid confusion with  $T(t)$  of this paper, the function  $T(t)$  of CB is written  $T_0(t_r)$ , where  $t_r$  indicates a relative translation vector.



will be incompletely removed and, possibly, will introduce artefacts.

So, it seems that  $T_1$  is sensitive to orientation error of the model, while  $T$  is intrinsically more robust against inaccuracy in orientation of the model (as verified on lysozyme):  $T$  is a ratio of scalar product and overlap functions which are both not very sensitive to small disorientations.

However,  $T_0(t_r)$  and  $T_1(t_r)$  have an advantage over  $T(t)$ : in general, two specific 2D maps are expected to be sufficient to solve the 3D translation problem. Conversely,  $T(t)$  must be evaluated in 3D, which, of course, assumes a larger volume to be explored. But, the risk of intermolecular vector superposition, for  $T_0(t_r)$  and  $T_1(t_r)$ , is important. In this case we might have a blurred map in which the solution peak is not easy to detect, as was the case for **a** and **b** projections mentioned above. This situation does not exist with  $T(t)$ , because the search is performed in three dimensions and, mainly, by the fact that  $O(t)$  enhances the signal-to-noise ratio.

## 6. Conclusions

We have presented here a quite new approach of the translation problem, known as the most difficult part of the molecular replacement technique. Referring only to the notion of linear correlation coefficient, new functions have been defined independently of the concepts of Patterson vectors and residual factor. This particularity permits prediction of their behaviour independent of the spatial resolution. Moreover, being analytic functions, they provide computational facility. We can sum up our results:

(1) Three functions, defined as  $T(t)$ ,  $O(t)$ ,  $TO(t)$ , have been tested and they behaved as expected. However, separate utilization of functions  $TO$  and  $O$  was not sufficient to solve the translation problem though they provide their own information. Only the combination  $T = TO/O$  was really efficacious. It means that information obtained in direct space and information obtained in reciprocal space must be combined.

(2) Crystal packing considerations intervening directly, owing to the overlap function  $O(t)$ , in the search for the solution, play a crucial role in the reliability of the translation function  $T(t)$ . Instead of taking into account, *a posteriori*, the steric conditions, we thought it was better to include them in the function definition itself.

(3) Two structures of very different size, 2 and  $10^3$  atoms, have been worked out and verified, once more, the property of MR to be independent of the atom number.

(4) The 3D maps of the three functions were produced with a reduced volume of computing, because it is possible to use the FFT algorithm.

(5) The application to the experimental data of the lysozyme *B* form led to acceptable results, prior to any refinement, as has been checked by comparison with those of the isomorphous replacement method.

(6)  $E_o$  and  $E_m$  (normalized data) are indispensable for  $TO$  and related functions to fulfil their expected role.

(7) The robustness of  $T$  against inaccurate orientation of the model, in an unknown crystal, is quite good and seems to be an advantage over Crowther & Blow's best translation function  $T_1$ .

The consistency of the results of MR with those of MIR confirms that both are essentially correct and emphasizes the efficiency of  $T(t)$ . However, the high-temperature form of lysozyme being an ideal case,  $T$  must be applied to other structures, particularly when the molecular model is only partial or inaccurate.

On such an occasion, it would be interesting to put the centre of mass of the model at the origin before translation, for that should make easier the comparison and the interpretation of the maps and, later on, the refinement process. Furthermore, using  $O'(t) \simeq \{\sum_h |F_c(\mathbf{h}, t)|^4\}^{1/2}$  instead of  $O(t)$  might improve  $T(t)$ .  $O'(t)$  could be numerically evaluated by an algorithm very similar to the one used for  $O(t)$ , but at the cost of a greater amount of computing.

We wish to thank Dr R. A. Crowther for kindly supplying a copy of his fast rotation program, Professor D. C. Phillips for providing atomic coordinates of HEW lysozyme, Professor H. Curien for encouragement, A. Laurent for friendly assistance and V. Bhide for reading the manuscript. We are grateful to referees for their interesting and helpful criticisms. One of us (YH) has been supported by a grant (no. 76A/6263) from Ministère des Affaires Etrangères and, partly, by Fondation de France (Fonds Cuttoli).

## APPENDIX

### Fourier expansions of

$$\sum_{\mathbf{h}} |F_c(\mathbf{h}, t)|^2 \text{ and } \sum_{\mathbf{h}} F_o^2(\mathbf{h}) |F_c(\mathbf{h}, t)|^2$$

$$\begin{aligned} F_c(\mathbf{h}, t) &= \sum_{j=1}^N F_m(\mathbf{h}A_j) \exp \{-2\pi i \mathbf{h}(\mathbf{d}_j + A_j \mathbf{t})\} \\ &= \sum_{j=1}^N F_m(\mathbf{h}A_j) \exp(-2\pi i \mathbf{h} \mathbf{d}_j) \exp(-2\pi i \mathbf{h} A_j \mathbf{t}), \end{aligned}$$

where  $A_j$  is the  $j$ th rotation matrix of crystal symmetry and  $\mathbf{d}_j$  its associated translation vector.

$$\text{As } |F_c(\mathbf{h}, \mathbf{t})|^2 = F_c(\mathbf{h}, \mathbf{t})F_c^*(\mathbf{h}, \mathbf{t}),$$

$$\sum_{\mathbf{h}} |F_c(\mathbf{h}, \mathbf{t})|^2 = \sum_{\mathbf{h}} \sum_{j=1}^N \sum_{k=1}^N F_m(\mathbf{h}A_j)F_m^*(\mathbf{h}A_k) \\ \times \exp\{-2\pi i \mathbf{h}(\mathbf{d}_j - \mathbf{d}_k)\} \\ \times \exp\{-2\pi i \mathbf{h}(A_j - A_k)\mathbf{t}\},$$

which results in

$$\begin{cases} Q_{jk}(\mathbf{p}) = F_m(\mathbf{h}A_j)F_m^*(\mathbf{h}A_k) \exp\{-2\pi i \mathbf{h}(\mathbf{d}_j - \mathbf{d}_k)\}, \\ \text{if } \mathbf{p} = \mathbf{h}(A_j - A_k) \end{cases}$$

and

$$\begin{cases} Q_{jk}(\mathbf{p}) \equiv 0 \\ \text{if } \mathbf{p} \neq \mathbf{h}(A_j - A_k). \end{cases}$$

From this, two properties for  $Q_{jk}(\mathbf{p})$  are seen:

$$Q_{jk}(-\mathbf{p}) = Q_{jk}^*(\mathbf{p}) = Q_{kj}(\mathbf{p}).$$

Therefore,

$$\sum_{\mathbf{h}} |F_c(\mathbf{h}, \mathbf{t})|^2 = \sum_{\mathbf{h}} \sum_{j=1}^N |F_m(\mathbf{h}A_j)|^2 \\ + 2\text{Re} \left[ \sum_{\mathbf{p}} \sum_{j>k}^{N-1} Q_{jk}(\mathbf{p}) \right. \\ \left. \times \exp(-2\pi i \mathbf{p}\mathbf{t}) \right].$$

The first summation term of the second member does not depend upon  $\mathbf{t}$ .  $S'$  indicates a reciprocal sphere whose radius is at most twice that of  $S$ .

In the same way, by substituting  $F'_m(\mathbf{h}) = F_o(\mathbf{h})F_m(\mathbf{h})$  for  $F_m(\mathbf{h})$ :

$$\sum_{\mathbf{h}} F_o^2(\mathbf{h}) |F_c(\mathbf{h}, \mathbf{t})|^2 = \sum_{\mathbf{h}} \sum_{j=1}^N |F'_m(\mathbf{h}A_j)|^2 \\ + 2\text{Re} \left[ \sum_{\mathbf{p}} \sum_{j>k}^{N-1} Q'_{jk}(\mathbf{p}) \right. \\ \left. \times \exp(-2\pi i \mathbf{p}\mathbf{t}) \right].$$

## References

- BERTHOU, J., LAURENT, A., LIFCHITZ, A., JOLLES, P., ARTYMIUK, P., BLAKE, C. C. F. & RICE, D. W. (1977). *Abstracts of the 4th European Crystallographic Meeting*, Vol. A, p. 357.
- BLAKE, C. C. F. (1977). Private communication.
- BLAKE, C. C. F., MAIR, G. A., NORTH, A. C. T., PHILLIPS, D. C. & SARMA, R. (1967). *Proc. R. Soc. London Ser. B*, **167**, 365-377.
- BOTT, R. & SARMA, R. (1976). *J. Mol. Biol.* **106**, 1037-1046.
- COZZONE, P., OPELLA, S. J., JARDETZKY, O., BERTHOU, J. & JOLLES, P. (1975). *Proc. Natl. Acad. Sci. USA*, **72**, 2095.
- CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 174-177. New York: Gordon and Breach.
- CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544-548.
- DODSON, E., HARDING, M. M., HODGKIN, D. C. & ROSSMANN, M. G. (1966). *J. Mol. Biol.* **16**, 227-241.
- HACKERT, M. L., FORD, G. C. & ROSSMANN, M. G. (1973). *J. Mol. Biol.* **78**, 665-673.
- HENDRICKSON, W. A. & WARD, K. B. (1976). *Acta Cryst.* **A32**, 778-780.
- JOLLES, P. & BERTHOU, J. (1972). *FEBS Lett.* **23**, 31.
- JOYNSON, M. A., NORTH, A. C. T., SARMA, V. R., DICKERSON, R. E. & STEINRAUF, L. K. (1970). *J. Mol. Biol.* **50**, 137-142.
- LANGS, D. A. (1973). Abstr. ACA meeting, Storrs, Connecticut.
- LIFCHITZ, A. (1981). To be published.
- NIXON, P. E. & NORTH, A. C. T. (1976). *Acta Cryst.* **A32**, 320-325.
- RAE, A. D. (1977). *Acta Cryst.* **A33**, 423-425.
- RAMACHANDRAN, G. N. & SRINIVASAN, R. (1961). *Nature (London)*, **190**, 159-161.
- ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.
- SAINT-BLANCARD, J., CLOCHARD, A., COZZONE, P., BERTHOU, J. & JOLLES, P. (1977). *Biochim. Biophys. Acta*, **491**, 354.
- TOLLIN, P. (1966). *Acta Cryst.* **21**, 613-614.
- VAND, V. & PEPINSKY, R. (1956). *Z. Kristallogr.* **108**, 1-14.