Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Combining image-seeking functions and a subtraction strategy: a vector-space procedure to improve many-body searches in molecular replacement

Carmen Álvarez-Rúa,* Javier Borge and Santiago García-Granda

Departamento de Química Física y Analítica, Facultad de Química, Universidad de Oviedo, C/Julián Clavería 8, 33006 Oviedo, Spain

Correspondence e-mail: mcara@sauron.quimica.uniovi.es

Received 18 July 2001
Accepted 15 November 2001

Many-body searches in molecular replacement are usually carried out sequentially and each step benefits from the structural information obtained in previous rotational and translational stages. In this context, the incorporation of known structural information has proved to enhance the discrimination of a rotation function in Patterson space when many independent molecules have to be located in the asymmetric unit of the crystal cell. This improvement is achieved by subtraction of the contributions of already positioned molecules from the observed Patterson map, which makes the determination of the correct orientation of the remaining molecules easier. The quality of the resultant difference Patterson map is greatly influenced by the application of a bulk-solvent correction to the structure-factor amplitudes of the molecules that are being subtracted. The results obtained in the rotation search benefit both from the availability of high-resolution data and from the combination of the subtraction strategy and the refinement of a great number of the peaks of the rotation function.

1. Introduction

The greater the number of solved structures is then the greater the increase in the significance of the molecular-replacement method (Rossmann, 1972) in macromolecular crystal structure determination. This method, which was initially conceived as a tool for *ab initio* phase determination by imposing non-crystallographic symmetry restraints (Rossmann & Blow, 1963), is mostly used nowadays for locating a three-dimensional model in the crystal unit cell of the macromolecule concerned. At present, the number of solved structures is very large, as proved by the great number of entries in the Protein Data Bank (Berman *et al.*, 2000). A great effort is being made to classify all the information derived from the structural analysis of macromolecules into databases where structures are arranged according to different sources of information, as fold and function characteristics or protein–protein interactions (Gerstein, 2000). These classifications provide molecular replacement with a set of representative folds that can be used as search models. This approach is based on the idea that there is a list of structural fragments from which all macromolecules can be assembled. Molecular replacement benefits highly from this source of information.

Furthermore, this method is undergoing constant evolution which allows better results to be achieved with poorer data or less homologous models obtained from other experimental techniques such as NMR (Chen *et al.*, 2000) or electron microscopy (Ban *et al.*, 1998).

However, in spite of the significant improvements introduced in the method in the past few years, there are still some

situations that cause molecular replacement to fail. One of these is the presence of various molecules in the asymmetric unit of the crystal. A number of methods have been introduced to cope with this problem.

Some of these procedures aim at locating many bodies through simultaneous *n*-dimensional searches (Chang & Lewis, 1997; Kissinger *et al.*, 1999; Glykos & Kokkinidis, 2000*a*). A different approach consists of incorporating some previous structural knowledge of the crystal in the different steps of the molecular-replacement method, such as the presence of non-crystallographic symmetry elements relating monomers in the asymmetric unit on the locked rotation function (Tong & Rossmann, 1990), the availability of any existing phase information on the phased translation function (Bentley, 1997) or the 'addition' of the Patterson function of an already positioned model to the Patterson function of the search model in each trial orientation and/or position (Zhang & Matthews, 1994). The most commonly used procedure consists of fixing the contributions of already positioned models and searching for the rest of the bodies in the asymmetric unit (Navaza, 1994; Navaza & Saludjian, 1997). Another approach has been presented that reduces the effective number of independent molecules to be located by exploiting the information provided by non-crystallographic symmetries (Navaza *et al.*, 1998). Recently, a new method has been developed that allows the formation of multi-copy search models from correctly oriented monomers to be used in the translation search (Vagin & Teplyakov, 2000).

This kind of incorporation of structural information usually does not affect the observed data but the mathematical model that is being fitted to the experiment. However, a number of procedures have also been introduced that directly modify the experimental data. In these cases, the elimination of the contribution of already positioned models can be carried out either in reciprocal or in direct space. Examples of elimination in reciprocal space can be found in the evaluation of phased translation functions (Bentley, 1997) or in the use of observed structure factors that correspond to the still uninterpreted regions of the unit cell, which are obtained by FFT inversion of a specially calculated electron-density difference map (Vitali *et al.*, 1996). In direct space, this is accomplished by direct subtraction of the Patterson map of the known molecules from the observed Patterson map. Our focus will be on this last case.

Direct space offers a natural way for eliminating known structural information from experimental data. The removal of peaks from the observed Patterson function that correspond to pairs of atoms in special positions which could be hiding peaks arising from the unknown structure has previously been discussed by Patterson (1935) and Buerger (1959). Zhang & Matthews (1994) introduced the same idea to the molecular-replacement method to reduce noise in the evaluation of the rotation and translation functions when many bodies were present in the asymmetric unit of the crystal structure and named this the 'subtraction' strategy. Their approach was to apply this strategy to the correlation form of the rotation function (Rossmann & Blow, 1962). Nevertheless, this methodology can also be applicable in rotation functions that make an explicit use of the values of the Patterson function of the crystal and the model, such as vector-search rotation functions.

These vector-search rotation procedures are based on the comparison between an observed Patterson map, which can be considered as a representation of all the interatomic vectors in the unit cell translated to the origin, and a vector set obtained from the search model. The degree of fit between them is obtained through the evaluation of a function that attains an extremum at the correct orientation of the search model. Image-seeking functions can be used for this purpose. In general, these functions require the evaluation of the Patterson map of the crystal at the end-points of the vectors from the search model, which are systematically rotated through the asymmetric unit of the angular space that corresponds to the symmetry of the crystal. If different molecules are present in the asymmetric unit of the crystal, the observed Patterson map will contain contributions from all of them, making the fitting of the vector set from the model difficult. If the location of one or more of these molecules can be determined by any means, such as a previous molecular-replacement study, the exclusion of this information from the observed Patterson map can be helpful in the location of the rest of the components of the asymmetric unit.

In this context, we have studied the application of the subtraction method to a recently developed procedure using an image-seeking function as a vector-search rotation function in macromolecular crystallography (Borge *et al.*, 2000), which has been implemented in a computer program called *OVIONE* (Álvarez-Rúa *et al.*, 2000*a*). The aim of this new procedure is to improve the discrimination of the rotation function by elimination of the extremes of this function that correspond to already positioned models and therefore reducing noise in the map of the rotation function.

This method also tries to complement the most commonly used strategies for many-body translation searches. These procedures require that the correct orientations of the search models are included in the set of rotation-function solutions that are tested in the translation search. This is not always feasible because some correct orientations appear far down in the rotation list. The new procedure can make these orientations appear in higher positions in this list by subtraction of the contribution to the observed Patterson map of known positioned molecules.

This paper presents a description of the methodology proposed for the application of the subtraction strategy to the evaluation of a rotation function in vector space (§2) and a discussion on the conditions of application of the method (§3), along with some results (§4).

2. Methodology

Image-seeking functions (ISFs; Buerger, 1950, 1959) have proved to be useful as rotation functions in macromolecular crystallography, provided that some requirements are fulfilled. One of them is the availability of moderately high resolution data (beyond 2 Å). Below this limit, the performance of ISFs

decreases dramatically. On the contrary, better results are usually obtained as higher resolution data are included in the computation of these functions. Another requirement is the occurrence of some factors (model size, symmetry of the crystal) that cause the ratio between the standard uncertainty of the Patterson maps of both the model and the crystal to have a value above a certain threshold. Details of this methodology and the implementation in a computer program can be found elsewhere (Borge *et al.*, 2000; Álvarez-Rúa *et al.*, 2000*a*,*b*). In brief, the method consists of the following steps.

(i) A self-vector set (SVS) is obtained from the search model by making use of a special SVS selection procedure and an observed Patterson map is calculated from experimental data.

(ii) The SVS is systematically rotated through the angular space. In each angular position an ISF that relates the observed Patterson map and the SVS is evaluated.

(iii) The results of the rotation function undergo a refinement process, by means of a minimization algorithm, known as the 'downhill simplex' method (Nelder & Mead, 1965).

Attention will be focused on one of the ISFs implemented in the program *OVIONE*, the 'weighted minimum-average' function (Nordman, 1966; Nordman & Schilling, 1970). This function is defined as

$$\mathrm{MIN}(M, N) = \sum_{i=1}^{M} P(i) / \sum_{i=1}^{M} W(i), \tag{1}$$

where N is the total number of vectors in the SVS, $P(i)$ is the value of the observed Patterson function at the vector point i, $W(i)$ is the weight associated with vector **i** and the sum includes the M lowest values of $P(i)/W(i)$. The lower this ratio, the worse the accommodation of the vector **i** to the observed Patterson is. Therefore, MIN(M, N) will present a maximum value when a good fit between the worst-fitting vectors is achieved.

If the crystal contains several molecules in the asymmetric unit and the location of some of them has been determined previously, the subtraction method can be used to eliminate the contributions of these molecules from the observed Patterson map. The resulting expression for the ISF would be

$$\mathrm{MIN}(M, N) = \sum_{i=1}^{M} [P(i) - P_{\mathrm{known}}(i)] / \sum_{i=1}^{M} W(i). \tag{2}$$

The term $[P(i) - P_{\mathrm{known}}(i)]$ represents the difference between the Patterson maps of the crystal and the already positioned molecules. This subtraction of Patterson maps is equivalent (through Parseval's theorem) to the evaluation of a difference Patterson map of coefficients $(|F|^2 - |F_{\mathrm{known}}|^2)$, where $|F|^2$ represents the square of the amplitudes of the original crystal and $|F_{\mathrm{known}}|^2$ the square of the amplitudes of the calculated structure factors of the molecules that are being subtracted.

These coefficients can also be expressed as

$$\begin{aligned} |F|^2 - |F_{\mathrm{known}}|^2 &= (F^*_{\mathrm{known}} + F^*_{\mathrm{unknown}})(F_{\mathrm{known}} + F_{\mathrm{unknown}}) \\ &\quad - |F_{\mathrm{known}}|^2 \\ &= F^*_{\mathrm{known}} F_{\mathrm{unknown}} + F^*_{\mathrm{unknown}} F_{\mathrm{known}} + |F_{\mathrm{unknown}}|^2. \end{aligned} \tag{3}$$

The last term on the right ($|F_{\mathrm{unknown}}|^2$) corresponds to the coefficients of the part of the structure still to be solved. In terms of the Patterson function, this term generates all the vectors (self and cross) between the atoms of the molecules that have not yet been located. The first two terms represent the cross vectors between the atoms of the previously located molecules and the rest of the atoms in the unit cell.

Therefore, the difference Patterson map with coefficients $(|F|^2 - |F_{\mathrm{known}}|^2)$ is equivalent to the Patterson map of the remaining contents of the unit cell plus the cross terms that relate the part of the structure to be solved and the known contributions that have been subtracted. These cross terms add noise to the Patterson map of the unknown molecules. In spite of this, the subtraction strategy is still efficient and contributes to the location of new molecules, as the examples given in this paper will show.

Different types of information can be subtracted from the observed Patterson map, depending on the form that the term $|F_{\mathrm{known}}|^2$ adopts.

Let S denote the total number of molecules or fragments in the asymmetric unit that have already been positioned. The expression of $|F|^2$ can then be written as

$$|F|^2 = \left(F^*_R + \sum_{s=1}^{S}\sum_{g=1}^{G} F^*_{sg}\right)\left(F_R + \sum_{s=1}^{S}\sum_{g=1}^{G} F_{sg}\right). \tag{4}$$

In this expression F_{sg} stands for the contribution to the structure factor of the molecule obtained from molecule s by application of the gth symmetry operator. Each term F_{sg} is calculated using $P1$ symmetry. F_R is the contribution to the structure factor of the remainder of the unit-cell contents.

Operating on (4), the following expression results:

$$|F|^2 = \sum_{s=1}^{S}\sum_{s'=1}^{S}\sum_{g=1}^{G}\sum_{g'=1}^{G} F^*_{sg} F_{s'g'} + \sum_{s=1}^{S}\sum_{g=1}^{G}[F^*_{sg} F_R + F^*_R F_{sg}] + F^*_R F_R. \tag{5}$$

In this expression, the first term in the right corresponds to $|F_{\mathrm{known}}|^2$ in (3). The second term in (5) corresponds to the first and second terms on the right in (3). Finally, $F^*_R F_R$ is equivalent to $|F_{\mathrm{unknown}}|^2$ in (3).

Only the first term, which does not involve the component F_R, can be subtracted from $|F|^2$.

The type of positional information available (orientation, orientation plus translation) determines how this term can be subtracted. In all cases, the orientation of the fragments must be known. Two different expressions of $|F_{\mathrm{known}}|^2$ can be obtained from the first term in (5).

(i) $|F_{\mathrm{known}}|^2 = \sum_{s=1}^{S}\sum_{g=1}^{G} |F_{sg}|^2$. After the evaluation of a rotation function, a list of the possible orientations of the search models is available. If one or more of these orientations are clearly distinguishable from the rest and are supposed to represent the correct orientations of the search models, the contribution of the correctly rotated models can be subtracted from the total Patterson map before the positions of these models in the unit cell are determined. Each term $|F_{sg}|$ in this expression is calculated using $P1$ symmetry.

(ii) $|F_{\rm known}|^2 = \sum_{s=1}^{S}\sum_{s'=1}^{S}\sum_{g=1}^{G}\sum_{g'=1}^{G} F^*_{sg}F_{s'g'}$. When both the orientation and position of the models to be subtracted (one or more) are known, the whole first term in (5) can be eliminated from the squared observed amplitudes $|F|^2$. In practice, the components of this term are not independently calculated but are substituted by the structure factors of the already positioned models altogether, which have to be calculated using the crystal symmetry.

A flow diagram of the application of this methodology is shown in Fig. 1.

This procedure is intended to improve the discrimination of the rotation function by incorporation of information about the orientation or the orientation plus the position of previously located molecules. Two factors contribute to this improvement.

Firstly, the orientations corresponding to the subtracted molecules will not appear in the list of the solutions of the rotation function. Therefore, the determination of the orientation of the remaining molecules benefits from this fact, since some of these orientations could be masked by the solutions corresponding to the known molecules.

Secondly, as has been extensively discussed elsewhere (García-Granda *et al.*, 1996; Borge, 2000), the performance of ISF-based rotation functions critically depends on the value that the ratio between the standard uncertainties of the Patterson map of the model and the crystal ($\sigma_{\rm model}/\sigma_{\rm crystal}$) adopts. If the subtraction of the contribution of the known molecules is correctly carried out, the magnitude of the standard uncertainty of the resulting difference Patterson map is lower than that of the original observed Patterson map. When this last map is substituted by a difference Patterson map in the expression of the rotation function, the value of the ratio ($\sigma_{\rm model}/\sigma_{\rm crystal}$) increases and the performance of the ISF is improved.

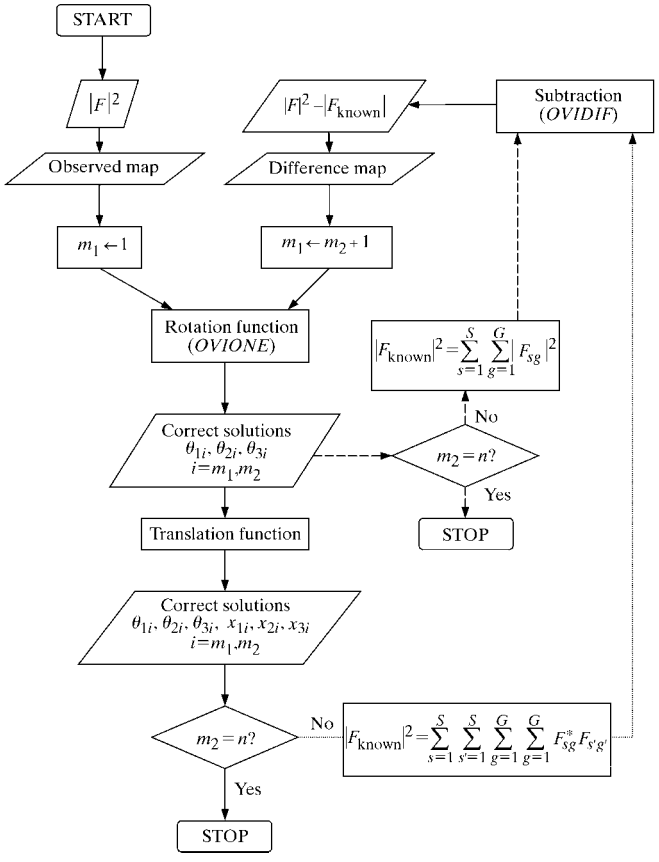

Figure 1
Flowchart of the application of the subtraction strategy in the molecular-replacement process when *n* independent fragments have to be located in the asymmetric unit of the crystal cell. This procedure can be carried out by following two different paths, depending on the type of information used to calculate the term $|F_{\rm known}|^2$: orientation (dashed line) or orientation plus translation (dotted line). In the figure, m_1 and m_2 are counters which refer to the range of molecules located in each run of the rotation (or rotation plus translation) function.

3. Conditions of application

The methodology described above has been implemented in some local programs to test the influence of different types of Patterson maps on the behaviour of the 'weighted minimum-average' function implemented in the program *OVIONE*. A brief discussion on the conditions of application of the method follows.

3.1. Evaluation of the term $|F_{\rm known}|^2$

The two different expressions of $|F_{\rm known}|^2$ are, at least in theory, feasible, although the second one offers a number of advantages over the first one. The great advantage that the first expression of $|F_{\rm known}|^2$ offers is that no information about the position of the known model is required. However, the success of the proposed methodology is directly correlated to the accuracy of the determination of the correct orientation of the model to be subtracted, which is normally better when achieved with a rigid-body refinement of the positional parameters after the translation search. Therefore, if the position of the search model is also known, there is no reason for not introducing information about the symmetry of the crystal into the calculation of the amplitudes of the structure factors of the model. The resulting magnitudes will resemble the contribution of the positioned models to the observed amplitudes better than the amplitudes calculated in space group *P*1.

The best results are obtained when the orientation and position of all the molecules to be subtracted are known and all these independent fragments are considered as a unique body. In this situation, the crystal symmetry is incorporated in the calculation of $|F_{\rm known}|^2$. This is a better representation of the contribution to $|F|^2$ of the set of known molecules. Therefore, the application of this methodology under this condition appears to be mandatory.

3.2. Preparation of the data

The subtraction strategy requires that both observed and calculated data are appropriately prepared. Observed data must be expressed on an absolute scale. This condition is

required both for the calculation of the difference Patterson maps and for the evaluation of the rotation function. A Wilson plot (Wilson, 1942) can be used for this purpose.

On the other hand, the incorporation of low-resolution data in the computation of Patterson maps is very important, since their amplitudes, which are much larger than those corresponding to the high-resolution terms, dominate the evaluation of the Patterson functions. However, calculated structure-factor amplitudes are systematically larger than observed amplitudes at a resolution below about 5 Å if the contribution from the solvent is not taken into account. This introduces errors in the evaluation of the terms $(|F|^2 - |F_{\rm known}|^2)$ and the resultant difference Patterson maps are meaningless. Therefore, the application of a bulk-solvent correction on the calculated data becomes unavoidable. Two types of solvent models have been tested. The first was the scaling model based on Babinet's principle (Moews & Kretsinger, 1975; Tronrud, 1997), also known as the exponential scaling model, which leads to the following expression for the corrected amplitudes,

$$|F_{\rm bulk\_exp}| = |F_{\rm known}|\{1.0 - k_{\rm sol}\exp[-B_{\rm sol}(\sin\theta/\lambda)^2]\}. \quad (6)$$

The second model was the mask bulk-solvent correction (Jiang & Brünger, 1994). In this case, the expression of the corrected amplitudes is

$$|F_{\rm bulk\_mask}| = |F_{\rm known} + k_{\rm sol}F_{\rm mask}\exp[-B_{\rm sol}(\sin\theta/\lambda)^2]|, \quad (7)$$

where $F_{\rm mask}$, the structure factors of a flat solvent mask, are exponentially scaled and added to the structure factors of the known molecules.

The mask bulk-solvent correction was applied by considering that the region of the unit cell that was not occupied by the molecules to be subtracted was filled with bulk-solvent electron density. This approach is quite unrealistic, particularly in the present situation, where structure factors are calculated for a model which does not fill the unit cell completely. However, when both the orientations and translations of the models to be subtracted are known and the corresponding structure factors are calculated using the symmetry of the crystal, this approach provides a better approximation to the correct distribution of the amplitudes of the calculated structure factors than considering this region to be completely empty. This bulk-solvent correction cannot be applied when only the orientation of the models is known. In this case, each molecule to be subtracted is treated separately and it is not possible to guess a correct mask which describes the solvent region.

Either of these two models substantially improves the performance of the subtraction strategy, as will be shown in §4.

The incorporation of low-resolution data also imposes a limit on the number of molecules that can be subtracted from the observed Patterson map. As this number increases, the calculated structure-factor amplitudes become larger. An appropriate scaling of the observed data and a careful correction for solvent effects can avoid errors in the evaluation of $(|F|^2 - |F_{\rm known}|^2)$ arising from values of $|F_{\rm known}|^2$ incorrectly greater than $|F|^2$. If, in spite of these corrections, the distribution of the calculated structure-factor amplitudes is systematically greater than the distribution of the observed data, the number of molecules to be subtracted has to be reduced so that the errors in the coefficients $(|F|^2 - |F_{\rm known}|^2)$ are minimized.

3.3. Refinement

The many-body search has appeared to be improved by combination of the application of the subtraction strategy and the refinement of the solutions of the rotation function. In many cases, the asymmetric unit of the angular space is too coarsely sampled and the three Euler angles that describe the orientation of the remaining molecules appear far away from the top of the list of possible solutions of the rotation function, even if an improved Patterson map, such as the difference Patterson map obtained with the subtraction strategy, is used. In general, only a subset of the highest solutions of the rotation function is refined and the correct solutions are missed. However, if this number of solutions is substantially increased, the refinement process manages to highlight the correct orientations, which are more clearly distinguishable from the rest when an improved difference Patterson map is used.

4. Results and discussion

Two different protein structures from the PDB are presented here as test examples for this new methodology. Better results were systematically obtained in all the experiments that were carried out to test this methodology when information about the symmetry of the crystal and about the orientation and position of the models to be subtracted was incorporated in the calculation of the structure factors $F_{\rm known}$. Therefore, it seems that the correct application of the subtraction strategy requires that these conditions are fulfilled.

Experimental data were obtained from the PDB and expressed on an absolute scale by means of a Wilson plot. Upper resolution limits were obtained from the information contained in the PDB files. The low-resolution limit was set to 15 Å. Therefore, the use of a bulk-solvent correction to compensate the discrepancies in the distribution of the observed and calculated structure-factor amplitudes becomes mandatory. Two different bulk-solvent models were used in our tests. For the Babinet-based scaling model, the magnitudes of the parameters $k_{\rm sol}$ and $B_{\rm sol}$ were arbitrarily chosen as 0.85 e Å^{-3} and 250 Å^2, respectively. These values are inside the ranges cited in the literature: $k_{\rm sol}$ = 0.75–0.95 e Å^{-3}, $B_{\rm sol}$ = 150–350 Å^2 (Kostrewa, 1997; Badger, 1997). The values of these parameters are usually obtained during macromolecular refinement and experimental evidence exists that these parameters are distributed over a wider range than that quoted above (Glykos & Kokkinidis, 2000*b*, 2001). However, the subtraction strategy benefits from the application of this bulk-solvent correction in spite of the rough approximation that has been introduced in our approach.

The mask bulk-solvent correction was only applied when both the orientations and translations of the subtracted

Table 1
Results of the rotation search with and without application of the subtraction strategy for 1b2s.

$\bar{x}_R$ and σ_R are the mean value and standard uncertainty of the rotation function, respectively. Results (obtained after the refinement process) are expressed in standard uncertainty units above the mean. Numbers in parentheses correspond to the position or range of positions that a particular solution occupies in the list of solutions of the rotation function. The symbol — means that this orientation could not be found in the list of solutions of the rotation search. In all cases, the 100 highest peaks of the rotation function entered the refinement process.

			Subtraction strategy				
			No bulk-solvent correction		Bulk-solvent correction: Babinet-based scaling		Bulk-solvent correction: mask model†
	R.m.s. (Å)	Original results	Orientation	Orientation + position	Orientation	Orientation + position	Orientation + position
Barnase		Subtracted chains: *D*, *E*, *F*					
$\bar{x}_R$		13.276	10.511	5.985	12.429	10.979	11.851
σ_R		6.176	4.415	3.086	4.918	4.502	4.696
Chain *A*	0.395	16.178 (1–4)	18.387 (1–4)	16.892 (1–2)	17.643 (1–3)	20.652 (1–7)	18.875 (1–8, 12)
Chain *B*	0.505	11.561 (5)	11.525 (5)	12.898 (3)	15.554 (4)	11.705 (8–10)	11.834 (9–11)
Chain *C*	0.622	5.786 (59)	8.068 (19)	8.165 (22)	8.432 (12)	9.418 (12)	9.659 (13)
Barstar‡		Subtracted chains: *A*, *B*, *C*					
$\bar{x}_R$		13.909	12.1195	4.791	13.082	11.602	13.329
σ_R		6.724	6.015	3.389	6.258	5.768	6.217
Chain *D*	0.622	11.922 (1)	9.481 (5)	—	10.825 (2)	11.304 (1)	13.281 (1)/*13.637 (1)*
Chain *E*	0.917	—	—	—	—	—	—/*9.491 (12)*
Chain *F*	0.652	7.378 (24)	6.221 (49)	—	7.693 (17)	7.610 (18)	8.278 (14)/*10.243 (9)*
Barstar		Subtracted chains: *A*, *B*, *C*, *D*, *F*					
$\bar{x}_R$		13.909	10.341	2.034	11.432	9.899	11.429
σ_R		6.724	5.873	1.987	6.021	5.035	5.402
Chain *E*	0.917	—	—	8.546 (64)	—	9.961 (3)	8.769 (5)

† Barnase: $k_{sol} = 0.445$ e Å^{-3}, $B_{sol} = 24.310$ Å^2; barstar: $k_{sol} = 0.558$ e Å^{-3}, $B_{sol} = 34.518$ Å^2. ‡ The results in italics were obtained when the 3000 highest peaks of the rotation function were considered for refinement.

molecules were known. Typical values for these scaling parameters are k_{sol} = 0.3–0.4 e Å^{-3}, B_{sol} = 15–40 Å^2 (Kostrewa, 1997).

The crystal structure of both examples had been solved previously by using the molecular-replacement method and the search models cited in the literature were also used in our calculations.

A new program (*OVIDIF*) was written to calculate difference Patterson maps that the rotation-function program *OVIONE* can read in. The program *OVIDIF* reads the calculated structure factors of the models to be subtracted from an external file. In both test examples, the program *SFALL* (Agarwal, 1978) from the *CCP4* program suite (Collaborative Computational Project, Number 4, 1994) was used to compute the calculated structure factors when no bulk-solvent correction or the exponential scaling model were applied. The structure-factor amplitudes, corrected by means of the mask model procedure, were calculated with the program *CNS* (Brunger *et al.*, 1998). In all the examples, *OVIONE* was run using the default values of the parameters that control the performance of the program.

4.1. 1b2s

A barnase–barstar complex was crystallized in space group *C*2, with unit-cell parameters $a = 206.24$, $b = 43.51$, $c = 83.69$ Å, $\beta = 107.42^\circ$ (Vaughan *et al.*, 1999). Data to a high-resolution limit of 1.82 Å were available and were used in the rotation search. The asymmetric unit of the crystal cell contains three barnase–barstar complexes (barnase, chains *A*, *B* and *C*; barstar, chains *D*, *E* and *F*). Two tests were carried out, using as search models the structure of the barnase and the barstar molecules of a previously refined crystal of the same complex (PDB code 1brs; Buckle *et al.*, 1994; chains *A* and *D*, respectively). A summary of the results of the combination of the subtraction strategy and the evaluation of the rotation function are presented in Table 1. A discussion of these results follows.

4.1.1. Barnase. When the rotation function was evaluated under normal conditions, two of the correct orientations of the barnase molecule (chains *A* and *B*) appeared at the top of the list of possible solutions. However, the orientation corresponding to the remaining molecule (chain *C*) was below position 20 in the results obtained both in the rotation and refinement processes.

The following example illustrates how the subtraction procedure could improve this result. We will assume that the three barstar molecules have been previously correctly oriented and located in the unit cell through a standard molecular-replacement protocol. The contribution of these molecules to the observed Patterson map is subtracted by using either only information about the orientation of the three barstar molecules or by assuming that the correct position of these molecules in the crystal unit cell is also known. Three differently calculated sets of structure factors were used: uncorrected and modified by application of a bulk-

solvent correction (Babinet-based and mask model). In all these cases, the orientation of the chain *C* occupied higher positions in the list of solutions of the rotation function compared with the results obtained before the application of the subtraction strategy; a noteworthy improvement was observed when the calculated amplitudes were corrected to take into account the contribution of the solvent.

4.1.2. Barstar. In contrast to the previous example, the orientation of one of the barstar molecules (chain *E*) could not be determined either in a normal run of the program *OVIONE* or when the subtraction strategy (with and without bulk-solvent correction) was used. There are two reasons that could explain the differences between these results and those obtained with the barnase model. Firstly, the overall differences between the search model and the molecules in 1b2s are greater for the barstar case than for the barnase example. This means that the subtraction strategy works more effectively for the barnase case, as the subtracted terms better resemble the contribution of the already positioned models to the observed data. Secondly, chain *E* is the molecule that most differs from the search model (r.m.s. deviation = 0.917 Å) in comparison with the other five molecules in the asymmetric unit. Therefore, the fitting between the SVS obtained from the search model and the observed or the difference Patterson map is worse for the angular position that corresponds to chain *E* than for the rest of molecules in the asymmetric unit and the discrimination of the rotation function decreases for this particular orientation.

In this case, no improvements were obtained when the contribution of the three barnase molecules was subtracted from the experimental Patterson map and the rotation-function program was run under default conditions (see Table 1).

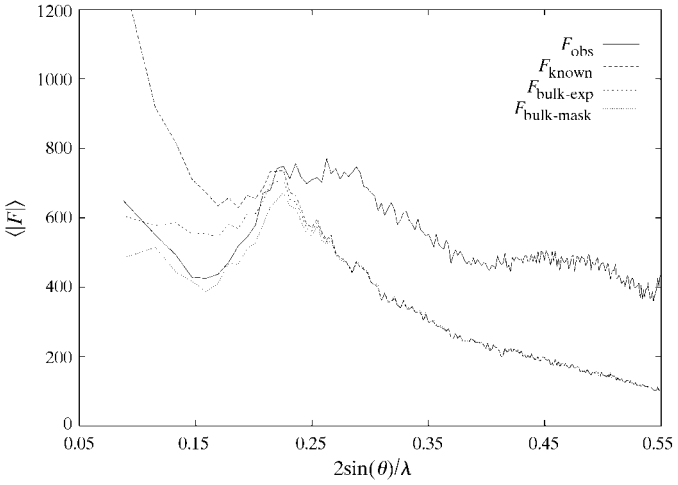

Figure 2
Local average $\langle|F|\rangle$ *versus* $2\sin(\theta)/\lambda$ for 1b2s. F_{obs}, scaled observed data (Wilson plot). F_{known}, calculated structure-factor amplitudes of the three correctly positioned barnase molecules. F_{bulk_exp}, F_{known} corrected with the bulk-solvent model based on Babinet's principle. F_{bulk_mask}, F_{known} corrected with the mask bulk-solvent model. See §4.1.2 for a detailed discussion of the effects of these distributions on the results of the rotation function.

This failure could be circumvented by making use of the refinement procedure implemented in the program *OVIONE*. By default, only a small number (100) of the highest peaks of the rotation function are refined, since the correct orientation of the search model is supposed to be among these peaks. However, when many bodies are present in the unit cell and the angular space is not finely sampled, some of the correct solutions can appear below this boundary and are not included in the refinement process and are therefore missed.

A greater number of solutions (3000) were then included in the refinement process. However, the orientation of chain *E* did not appear except when the mask bulk-solvent correction was applied to the structure factors corresponding to the three barnase molecules.

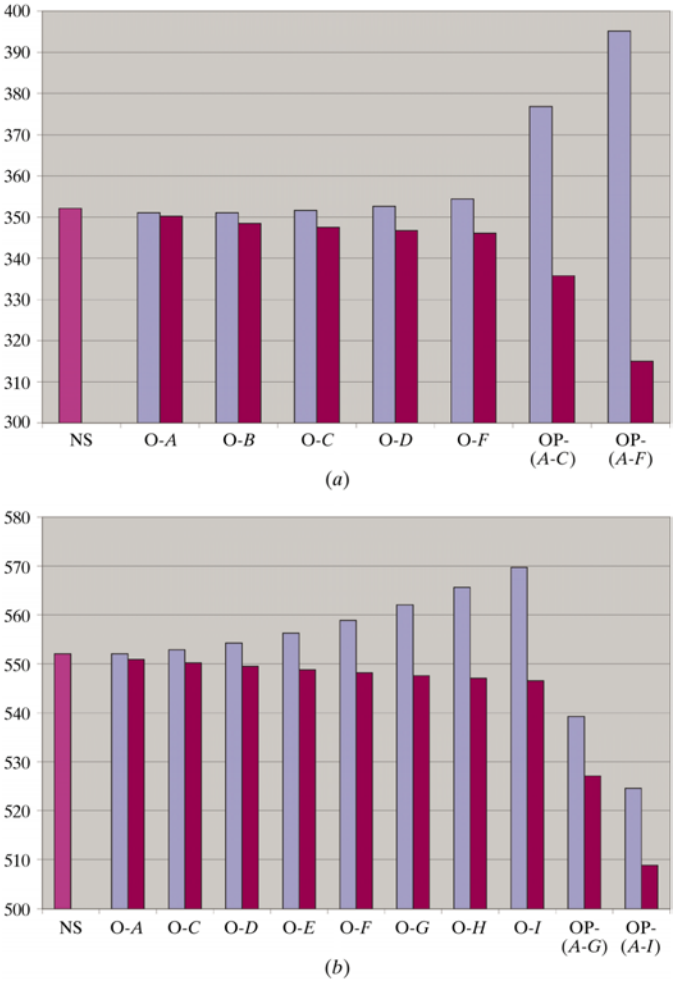

Figure 3
Standard uncertainties (σ_{map}) of the difference Patterson maps as a function of the number of subtracted molecules for (*a*) 1b2s and (*b*) 1dwk. Blue, data not corrected. Red, bulk-solvent correction (Babinet-based model) applied to the calculated structure factor of the subtracted molecules. Codes used in the figure: NS, no subtraction (total Patterson map); O-*X*, orientation of the subtracted molecule known; OP-(*X*-*X*′), orientation and position known. *X*, subtracted molecule. (*X*-*X*′), range of subtracted molecules. When only the orientation of the molecules is known, the values in the figure correspond to the standard uncertainty of the difference Patterson map obtained when the Patterson map of chain *X* was subtracted from the contiguous difference Patterson map on the left.

Table 2
Results of the rotation search with and without application of the subtraction strategy for 1dwk.

The format of this table conforms to the specifications given in Table 1. Results are expressed in standard uncertainty units above the mean before (Rot.) and after (Ref.) the refinement process. In addition, the results obtained when a different number of orientations (NSOL = 100, 3000) were refined are also presented.

(*a*) Original results. $\bar{x}_R = 14.800$, $\sigma_R = 2.300$.

	NSOL = 100		NSOL = 3000	
Chain	Rot.	Ref.	Rot.	Ref.
A	5.915 (7)	10.454 (2)	5.915 (7)	10.500 (2–5)
B	—	—	2.844 (1027)	5.449 (82)
C	5.079 (10)	9.736 (6)	5.079 (10)	10.053 (8)
D	6.772 (3)	10.941 (1)	6.772 (3)	10.941 (1)
E	6.101 (5)	9.910 (5)	6.101 (5)	9.910 (10)
F	7.135 (2)	8.758 (8)	7.135 (2)	8.758 (15)
G	8.685 (1)	10.196 (4)	8.685 (1)	10.196 (6–7)
H	—	—	3.407 (299)	7.030 (21)
I	—	—	2.462 (2247)	7.350 (18)
J	—	—	—	—

(*b*) Subtracted chains: *A*, *C*, *D*, *E*, *F*, *G*.

	No bulk-solvent correction $\bar{x}_R = 13.253$, $\sigma_R = 2.089$				Bulk-solvent correction: Babinet-based scaling† $\bar{x}_R = 14.473$, $\sigma_R = 2.145$			
	NSOL = 100		NSOL = 3000		NSOL = 100		NSOL = 3000	
Chain	Rot.	Ref.	Rot.	Ref.	Rot.	Ref.	Rot.	Ref.
B	—	—	2.442 (2379)	7.122 (3)	—	—	2.828 (1015)/ *2.348 (2967)*	7.446 (4)/ *6.981 (2)*
H	—	—	3.304 (369)	7.497 (2)	5.187 (3)	7.847 (1)	5.187 (3)/ *3.622 (159)*	7.847 (3)/ *7.492 (1)*
I	—	—	—	—	—	—	2.388 (2604)	8.143 (2)
J	—	—	2.444 (2368)	9.941 (1)	—	—	2.807 (1067)	10.278 (1)

† Numbers in italics correspond to the results of the subtraction strategy when only information about the orientation of the subtracted molecules was included in the calculations ($\bar{x}_R = 14.167$, $\sigma_R = 2.208$).

An interpretation of this result can be found in Fig. 2. When no bulk-solvent correction is applied to the calculated data, the magnitudes of the structure factors of the subtracted model are much larger than the observed data at resolutions below 5 Å. This causes the subtraction method to fail and no correct solution (even those that were perfectly recognized without subtraction) is found. When a bulk-solvent correction was used, a better fit between the magnitudes of the calculated and observed data is observed at low resolution. In this case, the choice of the parameters that were used in the scaling model based on Babinet's principle was not adequate to fit the distribution of the amplitudes of the observed data in every point of the resolution range used in the calculations. However, the magnitudes of the amplitudes of the calculated structure factors corrected with the mask bulk-solvent model remain below the values of the corresponding observed amplitudes in the whole resolution interval and no meaningless physical situation occurs. This result could suggest that the mask model should be the method of choice for systematically correcting the calculated structure factors of the models to be subtracted. However, experimental tests (not shown here) reveal that not very different results are obtained in the subtraction strategy with both bulk-solvent models, provided that the distribution of the amplitudes of the calculated structure factors does not pass over the observed data. This last requirement seems to be more important for the subtraction strategy than the type of bulk-solvent correction that is being applied.

The results obtained when the mask bulk-solvent correction was applied were used as input for a translational search with the program *AMoRe* (Navaza, 1994, 2001). This program produced an unambiguous result: the correct orientation and position of the three barstar molecules appeared at the top of the list of solutions and were clearly distinguishable from the rest. Therefore, the subtraction strategy managed to produce the appropriate input for the translation function, which clearly solved the many-body search.

The performance of the rotation function improves as the number of subtracted molecules increases, but this also makes the calculated structure-factor amplitudes have larger values. As a consequence, greater errors are introduced in the evaluation of the coefficients ($|F|^2 - |F_{\rm known}|^2$) and results are usually unpredictable. The following test illustrates this last statement. In order to determine the correct orientation of chain *E*, we considered that the correct orientation and positions of the remaining protein molecules in the unit cell were known. A difference Patterson map was computed by subtracting the contributions of these five molecules from the original Patterson map. As shown in Table 1, chain *E* could be located in all cases when information about the position of the subtracted molecules was used. The best results were attained when a bulk-solvent correction was applied. In this example, the subtraction of a large fraction of the unit-cell contents seems to have compensated the errors introduced in the coefficients of the difference Patterson map and the identification of the orientation of chain *E* was possible. However, this is not a general result, as will be shown in §4.2.

4.2. 1dwk

This PDB entry corresponds to the structure of cyanase with bound oxalate anions at the active sites (Walsh *et al.*, 2000). Cyanase crystals are triclinic (unit-cell parameters $a = 76.33$, $b = 80.93$, $c = 82.13$ Å, $\alpha = 70.10$, $\beta = 71.95$, $\gamma = 66.42^\circ$) and contain ten monomers (chains *A*–*J*) in the asymmetric unit, which are assembled into a decamer to form the active sites of this enzyme. The refined coordinates of the seleno-

methionine-labelled cyanase (PDB code 1dw9), which was solved using the MAD method, were used as a starting model for solving the structure of the enzyme in complex with the dianion oxalate. Data to 1.65 Å were available.

In this example, only one of the monomers (chain *A*) of the uncomplexed form of cyanase was used as search model. The r.m.s. deviation between the search model and the ten monomers in the test crystal ranges from 0.26 to 0.56 Å. A normal run of the rotation function yielded the orientation of six of the ten monomers in the asymmetric unit among the ten highest peaks of the list of solutions, both in the rotation and after the refinement process. The program was re-run under the same conditions but refining a greater number of solutions of the rotation function. This time, the orientation of nine monomers could be determined. However, the ranks of the three monomers that were not previously located in the list of solutions of the rotation function were very low (positions 18, 21 and 82 after the refinement process). The orientation of one of the monomers (chain *J*) could not be determined.

These results suggested that the contribution to the observed Patterson map of the six easily located chains (*A*, *C*, *D*, *E*, *F* and *G*) could be subtracted to try to highlight the orientations of the four chains that were not clearly distinguishable under normal conditions. We will focus first in the most favourable case for the subtraction strategy; that is, when both the correct orientations and positions of the molecules to be subtracted are known.

Although the crystal space group is *P*1, it was necessary to determine the relative position of the six molecules to be subtracted with regard to one of them in order to compute $|F_{\rm known}|^2$. Chain *D* was chosen as the reference model and a translation search was carried out on the following 20 highest peaks in the list of solutions of the rotation function. The phased translation function implemented in *AMoRe* was used for this purpose. As a result, the orientation and position of the other five molecules of interest appeared on positions 1 to 5 in the list of solutions of the translation function and were clearly distinguishable from the rest.

Two types of difference Patterson maps were computed with and without correction of the calculated amplitudes to take into account the bulk-solvent contribution. When no correction was applied, the solutions corresponding to the six subtracted monomers did not appear in the list of possible solutions, as expected. However, no new correct orientations could be identified. The rotation-function program was re-run and a greater number of orientations were included in the refinement process. This time, the second and third highest peaks of the rotation search could be assigned to the orientation of two of the chains (*B* and *H*) that were also identified under the same conditions when the crystal Patterson map was used. An outstanding improvement was achieved in relation to the positions that these solutions occupied in the list of refined orientations (chain *B*, from position 82 to 3; chain *H*, from position 21 to 2). Besides, the orientation of chain *J*, which could not be determined in the normal run of the rotation-function program, appeared in the first position in the list of refined solutions. Unfortunately, the orientation of chain *I* could not be found under these application conditions of the method. This result could be attributed to the use of calculated structure-factor amplitudes which had not been corrected to take into account the contribution of the solvent.

On the other hand, the application of the Babinet-based scaling bulk-solvent correction to the calculated data produced a further improvement of the previous results. Chain *H* was easily identified even when a small number of orientations were refined. All the solutions that could not be identified without the subtraction of the known structural information clearly appeared at the top of the list of refined solutions (positions 1– 4).

For comparison, the results obtained when only the contribution to $|F_{\rm known}|^2$ corresponding to each molecule (treated independently from the rest) was subtracted from the observed data are, as expected, worse. The orientation of chains *B*, *H*, *I* and *J* could not be determined when no bulk-solvent correction was applied to the calculated data. Nevertheless, the orientation of two chains (*B* and *H*) could be clearly identified when a bulk-solvent correction was applied and a large number of orientations were considered for refinement (NSOL = 3000).

These examples show that the combination of the refinement procedure and the subtraction method can substantially improve the discrimination of the rotation function.

A final experiment was carried out that illustrates that the number of known molecules to be subtracted cannot be arbitrarily large. In this case, the 25 highest peaks in the list of solutions of the rotation function (NSOL = 3000; see Table 2) were used as input for the phased translation function in *AMoRe*, chain *D* again being the reference model. As a result, the correct positional parameters corresponding to chains *A*, *C*, *E*, *F*, *G*, *H* and *I* were clearly determined. The solution corresponding to chain *B* was far down in the list of orientations and was not considered for translation. The contribution of the eight correctly located monomers was eliminated from the observed Patterson map using the program *OVIDIF*. When no bulk-solvent correction was applied to the calculated data, *OVIONE* always failed to determine the orientations of chains *B* and *J*. When the exponential scaling bulk-solvent model was used to correct the calculated amplitudes, the orientation of chain *J* appeared in the first position in the list of refined solutions (NSOL = 3000). However, the orientation of chain *B* could not be determined. In this situation, the calculated structure-factor amplitudes were larger than the observed amplitudes at low resolution even when the bulk-solvent correction was applied. This introduces errors in the evaluation of the difference Patterson map and the discrimination of the rotation function is worse. This result suggests that a more careful application of a bulk-solvent correction and/or an appropriate selection of the low-resolution limit of the data included in the calculations have to be performed as the number of molecules to be subtracted increases.

Fig. 3 illustrates the effect of the bulk-solvent correction on the standard uncertainties ($\sigma_{\rm map}$) of some of the difference Patterson maps commented on in §§4.1 and 4.2. When no correction is applied, the greater the number of subtracted

molecules, the greater the magnitudes of these standard uncertainties. A correct subtraction of the contribution of the known molecules leads to lower values of $\sigma_{\rm diff}$. As a consequence of this, the ratio between the standard uncertainties of the Patterson map of the model and the crystal increases. As commented on in §2, this result has a positive effect on the performance of the rotation function, as demonstrated in the discussed examples.

5. Conclusions

The methodology presented in this paper has proved to enhance the discrimination of a rotation function evaluated in vector space when many bodies are present in the asymmetric unit of the crystal cell. The subtraction strategy that Zhang & Matthews (1994) proposed should be applied to the correlation form of the rotation function has also appeared to improve the performance of this vector-search rotation function, provided that some requirements are met.

(i) It is advisable that the correct orientation and position of the molecules to be subtracted are known, even though the method has been designed to be applied either when only the orientation or both the orientation and translation of the already positioned models are known.

(ii) The calculated structure factors of the molecules to be subtracted must be corrected in order to downweight the low-resolution terms, which are systematically larger than the observed amplitudes if the contribution of the solvent is not considered.

(iii) The incorporation of a great number of orientations to the refinement process substantially improves the determination of the correct solutions of the rotation function.

The programs *OVIDIF* and *OVIONE* are available free of charge to non-profit-making institutions. These programs, together with a manual and some test examples, are available from http://www11.uniovi.es/~mr/ovione/ovione_main.html or directly from the authors upon request.

The authors wish to thank Dr Jorge Navaza, whose advice and helpful discussions have been a source of inspiration in the development of this method. This work was partially supported by CICYT (BQU2000-0219).

References

Agarwal, R. C. (1978). *Acta Cryst.* A**34**, 791–809.
Álvarez-Rúa, C., Borge, J. & García-Granda, S. (2000*a*). *J. Appl. Cryst.* **33**, 1436–1444.
Álvarez-Rúa, C., Borge, J. & García-Granda, S. (2000*b*). *OVIONE* 1.0. Technical Report, Universidad de Oviedo, Spain. http://www11.uniovi.es/~mr/ovione/ovione_main.html.
Badger, J. (1997). *Methods Enzymol.* **277**, 344–352.
Ban, N., Freeborn, B., Nissen, P., Penczek, P., Grassucci, R. A., Sweet, R., Frank, J., Moore, P. B. & Steitz, T. A. (1998). *Cell*, **93**, 1105–1115.
Bentley, G. A. (1997). *Methods Enzymol.* **276**, 611–619.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Borge, J. (2000). PhD thesis, Universidad de Oviedo, Spain.
Borge, J., Álvarez-Rúa, C. & García-Granda, S. (2000). *Acta Cryst.* D**56**, 735–746.
Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Buckle, A. M., Schreiber, G. & Fersht, A. R. (1994). *Biochemistry*, **33**, 8879–8889.
Buerger, M. J. (1950). *Proc. Natl Acad. Sci. USA*, **36**, 376–382.
Buerger, M. J. (1959). *Vector Space and its Application in Crystal Structure Investigation.* New York: John Wiley & Sons.
Chang, G. & Lewis, M. (1997). *Acta Cryst.* D**53**, 279–289.
Chen, Y. W., Dodson, E. J. & Kleywegt, G. J. (2000). *Structure*, **8**, 213–220.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
García-Granda, S., Borge, J. & Gutiérrez-Rodríguez, A. (1996). *Anal. Quím. Int. Ed.* **92**, 294–298.
Gerstein, M. (2000). *Nature Struct. Biol.* **7**, 960–963.
Glykos, N. M. & Kokkinidis, M. (2000*a*). *Acta Cryst.* D**56**, 169–174.
Glykos, N. M. & Kokkinidis, M. (2000*b*). *Acta Cryst.* D**56**, 1070–1072.
Glykos, N. M. & Kokkinidis, M. (2001). *Acta Cryst.* D**57**, 1462–1473.
Jiang, J.-S. & Brünger, A. T. (1994). *J. Mol. Biol.* **243**, 100–115.
Kissinger, C. R., Gehlaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Kostrewa, D. (1997). *CCP4 Newsl.* **34**, 9–22.
Moews, P. C. & Kretsinger, R. H. (1975). *J. Mol. Biol.* **91**, 201–228.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.
Navaza, J., Panepucci, E. H. & Martin, C. (1998). *Acta Cryst.* D**54**, 817–821.
Navaza, J. & Saludjian, P. (1997). *Methods Enzymol.* **276**, 581–594.
Nelder, J. A. & Mead, R. (1965). *Comput. J.* **7**, 308–313.
Nordman, C. E. (1966). *Trans. Am. Crystallogr. Assoc.* **2**, 29–38.
Nordman, C. E. & Schilling, J. W. (1970). *Crystallographic Computing*, edited by F. R. Ahmed, pp. 110–114. Copenhagen: Munksgaard.
Patterson, A. L. (1935). *Z. Kristallogr.* **90**, 517–542.
Rossmann, M. G. (1972). Editor. *The Molecular Replacement Method.* New York: Gordon & Breach.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Rossmann, M. G. & Blow, D. M. (1963). *Acta Cryst.* **16**, 39–45.
Tong, L. & Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 783–792.
Tronrud, D. E. (1997). *Methods Enzymol.* **277**, 306–319.
Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.
Vaughan, C. K., Buckle, A. M. & Fersht, A. R. (1999). *J. Mol. Biol.* **286**, 1487–1506.
Vitali, J., Martin, P. D., Malkowski, M. G., Olsen, C. M., Johnson, P. H. & Edwards, B. F. P. (1996). *Acta Cryst.* D**52**, 453–464.
Walsh, M. A., Otwinowski, Z., Perrakis, A., Anderson, P. M. & Joachimiak, A. (2000). *Structure*, **8**, 505–514.
Wilson, A. J. C. (1942). *Nature (London)*, **150**, 151–152.
Zhang, X.-J. & Matthews, B. W. (1994). *Acta Cryst.* D**50**, 675–686.