Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

A new vector-search rotation function: image-seeking functions revisited in macromolecular crystallography

Javier Borge,* Carmen Álvarez-Rúa 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: jjba@sauron.quimica.uniovi.es

Received 24 June 1999
Accepted 7 March 2000

A new rotation function in Patterson space is described. An image-seeking function can be defined as a criterion of fit between the observed Patterson map and a suitable vector set extracted from a specially calculated Patterson map of the search model. The behaviour of image-seeking functions has appeared to be heavily dependent on certain relations between some statistical parameters of both maps. A new algorithm, which carries out the crucial step of selecting the appropriate vector set from the search model, has been established. As a consequence of the combination of these two preceding results, a new vector-search rotation function has been proposed and tested.

1. Introduction

Molecular replacement (Rossmann & Arnold, 1993) is one of the most important techniques for solving macromolecular crystal structures. A high percentage of published structures have been solved by this means. The original method (Hoppe, 1957) was extensively developed (Rossmann & Blow, 1962) in order to establish a mode of locating structural units not related by crystallographic symmetry elements within the unit cell. A further evolution of the method (the most widely used nowadays) was developed by Michael G. Rossmann (Rossmann, 1972). Provided that the three-dimensional structure of the macromolecule or a small fragment of it is roughly known, molecular replacement determines the correct orientation and translation that must be applied to correctly locate the three-dimensional structure (called the search model) in the crystallographic unit cell. Similar structures or fragments from other macromolecules in different crystal lattices may be used as search models. Relevant crystal structures (Rossmann *et al.*, 1985; Kwong *et al.*, 1998) have been determined by molecular replacement.

Nowadays, the number of solved macromolecules is enormous. Since those structures are easily accessible *via* the internet (Sussman *et al.*, 1998), it is very simple to find suitable search models. Moreover, certain structural domains appear frequently in different proteins without a clear sequence homology (Rao & Rossmann, 1973; Rossmann & Liljas, 1974). The above-mentioned reasons guarantee not only the survival but also the expansion of the method in the future.

However, a recently published paper (Brünger, 1997) stated:

> Despite three decades of experience with molecular replacement, there is only a partial understanding of the reasons for success or failure of the method.

We agree with this idea. There are two main reasons which cause molecular-replacement errors: the great number of

variables involved in the process (search-model size and structural similarity, range of observed data, Patterson grid size, Patterson integration radius *etc.*) and the lack of systematization in its use.

Usually, the method is mathematically implemented in two different ways: reciprocal space and Patterson space. The term 'molecular replacement' frequently refers to reciprocal-space algorithms, while 'vector search' is applied to Patterson-space algorithms. In general, reciprocal-space implementations (Navaza & Saludjian, 1997) offer quick solutions in a systematic but not always understandable way without user intervention. On the other hand, Patterson-space procedures (Nordman & Nakatsu, 1963; Huber, 1965) are not so systematically established and user participation is therefore increased. However, these apparent disadvantages can be turned into useful tools. In case of difficult problems, in which crystallographer participation is completely essential, Patterson space methods allow users to interact easily with them.

We will focus on the first step of the molecular-replacement method: the rotation problem. The translation function (the next step in the procedure) depends heavily on the solution of the rotation function. Incorrect orientations will always cause translation-function failures. Otherwise, if the orientation is correct, the translation step does not usually present many complications. Therefore, a great effort must be made to obtain the correct orientation of the search model.

All vector-search procedures are based on the comparison between an observed Patterson map and a vector set obtained from the search model. An observed Patterson map contains interatomic vectors translated to the origin. In fact, it is impossible to distinguish single point vectors in a Patterson map; only broad peaks appear in it. Each peak comes from the contribution of a great number of similar vectors. Overlap between different peaks is always present and this reduces the number of resolved peaks even further.

If an arbitrarily rotated search model is available, an appropriate vector set can be built from it. It is possible to compare this vector set with the observed map. The degree of similarity indicates whether the trial orientation is correct or not. If not, another orientation must be tried. The question is how to measure the fit between the vector set and the Patterson map. Different techniques can be used. Image-seeking functions (ISFs; Buerger, 1959) seem to be very good tools for this purpose. ISFs have proved to be a method for deconvoluting the Patterson maps and solving the crystal structures of small molecules. However, an alternative plan using them as criteria of fit in the case in which we are interested has been suggested (Nordman & Nakatsu, 1963), as ISFs can automatically detect a known vector set included in a Patterson map. The aim of this paper is to establish a vector-search rotation function based on ISFs in a clear, easy and systematic way. The paper is divided into several sections.

The first section describes the statistical analysis required to determine the behaviour of ISFs under common problem conditions, as ISFs do not always work properly under typical macromolecular circumstances (low-resolution data or lack of similarity between the search model and the unknown crystal structure, among others). Numerical simulations (based on computer-generated random numbers) have been performed in order to establish a statistical parameter which determines whether or not the present conditions will cause the failure of ISFs. Previous knowledge of this information avoids wasted effort: other vector-search (or reciprocal-space) techniques should then be used instead of ISFs.

The second section deals with the generation of the vector set. Since ISFs can only find a vector set if it is actually included in the Patterson map, the main task is to generate the appropriate vector set from the search model. The generated vector set must actually appear in the Patterson map; otherwise it will not be a correct input for an ISF, which will try to find something non-existent. An adequate build-up and selection of vectors from the search model are therefore essential.

The third section covers the rotation-search process. A special remark about the significance of a correct scaling of the data is made. The remainder of the paper shows different tests which demonstrate the efficiency of the proposed method. A brief comparison with other existing rotation functions is also included.

Finally, the general conclusions and future developments of the method are presented.

2. Methodology

The relationship between points and vector sets was first established by Dorothy Wrinch (Wrinch, 1939). 11 years later, Martin J. Buerger (Buerger, 1950) asserted that the Patterson function could be considered as the vector set of the crystal electron density and introduced ISFs (Buerger, 1951) to search for electron-density images inside the Patterson map. Similar techniques were developed almost simultaneously (Beevers & Robertson, 1950; Garrido, 1950; Thomas & McLachlan, 1952). They can be grouped and characterized as superposition methods, which are qualitatively equivalent to ISF methods. Some other procedures, such as the use of the accumulation function (Raman & Lipscomb, 1961) and the symmetry-minimum function (Simpson *et al.*, 1965) were proposed later.

The application of ISFs to the rotation-function problem arose some years later. A modified ISF (Nordman & Nakatsu, 1963) was used as a criterion of fit between vector sets and Patterson maps. Since then, many ISFs have been applied in small-molecule and macromolecular crystallography. The 'weighted minimum-average function' (Nordman, 1966; Schilling, 1970) is the basis of the vector-search procedure (for small molecules) successfully implemented in the program *ORIENT* (Beurskens *et al.*, 1987) included in the *DIRDIF*96 system (Beurskens *et al.*, 1996). Nowadays in macromolecular crystallography, vector-search procedures are only implemented, as far as we know, in *X-PLOR* (Brünger, 1992), *PROTEIN* (Steigemann, 1996) and *CNS* (Brunger *et al.*, 1998). All of them perform conventional vector searches based on other sorts of modified ISFs (Huber, 1985). On the other hand,

reciprocal-space implementations dominate the present software: *e.g. MERLOT* (Fitzgerald, 1988), *AMoRe* (Navaza & Saludjian, 1997).

However, ISFs can be very useful in macromolecular crystallography if they are applied correctly. A careful statistical analysis and an appropriate vector-set generation are needed in order to use ISFs in a proper way.

2.1. Statistical analysis

The behaviour of the ISFs in this new field must be carefully studied in order to avoid unpleasant surprises. Working with real data is always complicated, so we have tried to simplify the problem as much as possible. As previously described in the literature (Nordman & Hsu, 1982; Nordman, 1983, 1985), numerical simulations based on random numbers were performed in order to study the dependence of the behaviour of the ISFs on certain statistical parameters. The analysis is split into two parts: (i) random-number generation and (ii) numerical simulation procedure.

2.1.1. Random-number generation. It is known that a deterministic machine like a digital computer cannot generate truly random numbers. Therefore, we must use some sort of algorithm to generate 'pseudorandom' numbers. We can call 'pseudorandom' numbers (random numbers from hereon in this section for clarity) a set of numbers that fulfil some tests of randomness. Usually, compiler developers include random-number generation routines, but all of them suffer from a lack of randomness over short periods because of their simplicity. This fact encourages everybody who needs a set of random numbers to look for better algorithms.

The most popular way to generate random numbers following a uniform distribution is based on the linear congruential method. If a very large sequence of random numbers is needed, some modifications must be undertaken. The L'Ecuyer algorithm with Bays–Durham shuffle provides very good results. The Box–Muller method must be applied to transform uniform distributions into normal distributions which bear some resemblance to Patterson maps. We have used one of the most famous optimized routines (Press *et al.*, 1992) in order to generate random numbers according to the described methodology. The mathematical details of the described procedure can be found elsewhere (García-Granda *et al.*, 1996).

2.1.2. Numerical simulation procedure. For the sake of simplicity, we have decided to carry out one-dimensional simulations. Extension to two or three dimensions cannot easily be performed because of the possibility of correlation between the random numbers.

A Patterson map (which will be called 'total') may be considered as the sum of two parts: the search group ('signal') in which we are interested and the remainder or background ('noise'). Our numerical simulation procedure is based on the same idea. Let SIGNAL = {SIGNAL(i): $i = 1, n$} and NOISE = {NOISE(i): $i = 1, n$} be normal deviates with means $x_{\rm SIGNAL}$ and $x_{\rm NOISE}$ and standard deviations $\sigma_{\rm SIGNAL}$ and $\sigma_{\rm NOISE}$, respectively. A new set TOTAL = {TOTAL(i): $i = 1, n$} is constructed as SIGNAL plus NOISE: TOTAL(i) = SIGNAL(i) + NOISE(i). Its average and standard deviation can be obtained from $x_{\rm SIGNAL}$, $\sigma_{\rm SIGNAL}$, $x_{\rm NOISE}$ and $\sigma_{\rm NOISE}$,

$$x_{\rm TOTAL} = x_{\rm SIGNAL} + x_{\rm NOISE}, \tag{1}$$

$$\sigma_{\rm TOTAL} = \left\{\sigma^2_{\rm SIGNAL} + \sigma^2_{\rm NOISE} + \frac{2}{n}\left[\sum_{i=1}^{n} {\rm NOISE}(i)\cdot{\rm SIGNAL}(i)\right] - 2x_{\rm SIGNAL}x_{\rm NOISE}\right\}^{1/2}. \tag{2}$$

This is a very important point: the 'signal' set is included in the 'total' set. Once we have our 'total' set, we randomly choose a subset ('fragment') of correlative numbers from SIGNAL: FRAGMENT = {FRAGMENT(i): $i = 1, k$; $k < n$} (k is a parameter fixed by the user). A random number p ($0 \le p \le n - k$) must be chosen to build FRAGMENT: FRAGMENT(i) = SIGNAL($i + p$) ($x_{\rm FRAGMENT} \cong x_{\rm SIGNAL}$, $\sigma_{\rm FRAGMENT} \cong \sigma_{\rm SIGNAL}$). FRAGMENT will be the set of numbers that we will try to locate in the TOTAL set. (In the crystallographic situation, FRAGMENT would be the vector set and TOTAL the observed Patterson map.) There are ($n - k + 1$) possible positions for FRAGMENT. For each position, we calculate the goodness of fit between FRAGMENT and TOTAL (by evaluation of an ISF). The highest goodness of fit indicates the best position of the FRAGMENT in the TOTAL set. If the best position is p, the ISF succeeded. The whole process is then repeated q times (q is a parameter fixed by the user) under equal conditions on different SIGNAL, NOISE and TOTAL sets. Each time the ISF obtains the right solution, a control variable (ISF_success) is increased. If not, ISF_success remains unchanged. The final percentage of successful searches (PSS) is defined as $100 \times$ ISF_success/q.

In all our experiments, we have used SIGNAL and NOISE sets of 600 numbers ($n = 600$) and we have performed 500 experiments ($q = 500$) for each selected condition of $x_{\rm SIGNAL}$, $\sigma_{\rm SIGNAL}$, $x_{\rm NOISE}$, $\sigma_{\rm NOISE}$ and k. Although we have studied several types of ISFs, we are going to focus on the 'weighted minimum-average function'. Originally designed for small molecules, this ISF could not be successfully applied in macromolecular crystallography (Nordman, 1972),

$${\rm MIN}(m, k) = \sum_{i=1}^{m} {\rm TOTAL}(i) / \sum_{i=1}^{m} {\rm FRAGMENT}(i). \tag{3}$$

The new parameter m is the number of 'fragment' points used in the calculation. Not all points need be included in this ISF. Only those having the lowest value of TOTAL(i)/FRAGMENT(i) are useful. Low values of TOTAL(i)/FRAGMENT(i) reveal poor fit between TOTAL and FRAGMENT. The MIN(m, k) function indicates the correct position when a good fit between the worst fitting points is found.

Calculations were performed using SIGNAL and TOTAL average and standard deviation as input, because the search model and the observed Patterson map are always known in the real case. It is obvious that there is a clear relationship between the average and standard deviation from SIGNAL,

NOISE and TOTAL and therefore x_{NOISE} and σ_{NOISE} can be estimated by the formulae

$$x_{NOISE} = x_{TOTAL} - x_{SIGNAL}, \quad (4)$$

$$\sigma_{NOISE} = \frac{1}{n}\Bigg(-\sum_{i=1}^{n} X(i)\cdot \mathrm{SIGNAL}(i) \pm \Bigg\{\Bigg[\sum_{i=1}^{n} X(i)\cdot \mathrm{SIGNAL}(i)\Bigg]^2 + n^2(\sigma^2_{TOTAL} - \sigma^2_{SIGNAL})\Bigg\}^{1/2}\Bigg). \quad (5)$$

$X(i)$ (necessary for the calculation) is any set of n random numbers normally distributed. When this arbitrary set is transformed with the aid of the expression $[X(i)^*\sigma_{NOISE} + x_{NOISE}]$, the required NOISE is obtained. This means that this obtained NOISE plus the input SIGNAL generates the TOTAL first specified.

By changing the size of FRAGMENT, the average and standard deviation of SIGNAL and TOTAL and the parameter m, we will be able to check the behaviour of the studied ISF. All calculations involved in these numerical simulations were performed at the Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Spain on a CRAY T3E computer.

The first result (which may be considered to be obvious) is that increasing the size of FRAGMENT always produces better results, independent of the statistical conditions.

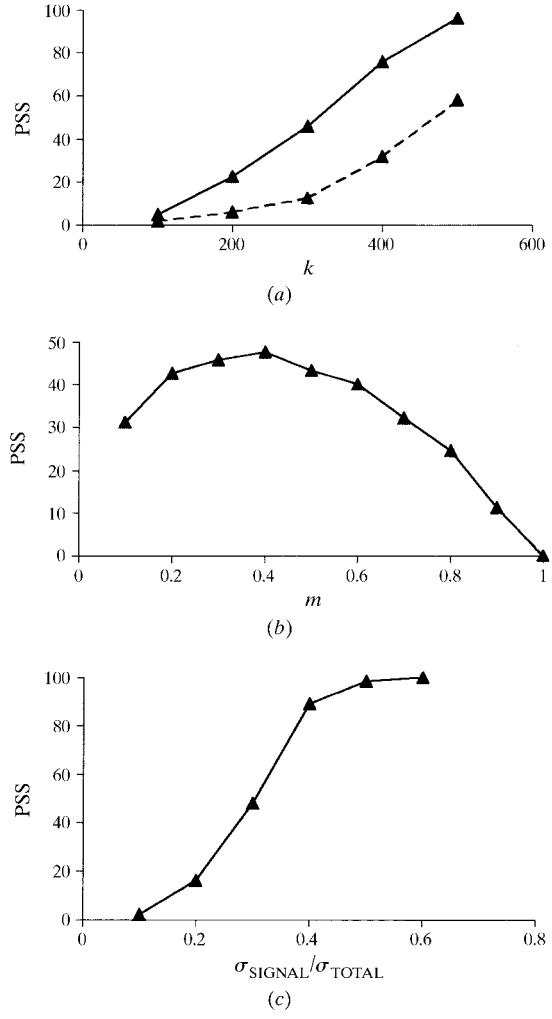

Figure 1
(*a*) Percentage of successful searches (PSS) as a function of the parameter k (size of FRAGMENT). Statistical conditions for the continuous curve are x_{TOTAL} = 1500, σ_{TOTAL} = 428.57, x_{SIGNAL} = 900 and σ_{SIGNAL} = 128.57. Statistical conditions for the dashed curve are x_{TOTAL} = 1500, σ_{TOTAL} = 375.00, x_{SIGNAL} = 900 and σ_{SIGNAL} = 75.00. (*b*) Percentage of successful searches (PSS) as a function of the parameter m (fraction of used points). Statistical conditions are x_{TOTAL} = 1500, σ_{TOTAL} = 428.57, x_{SIGNAL} = 900 and σ_{SIGNAL} = 128.57. (*c*) Percentage of successful searches (PSS) as a function of $\sigma_{SIGNAL}/\sigma_{TOTAL}$. Statistical conditions are x_{TOTAL} = 9000, x_{SIGNAL} = 8100, m = 0.3 and k = 300.

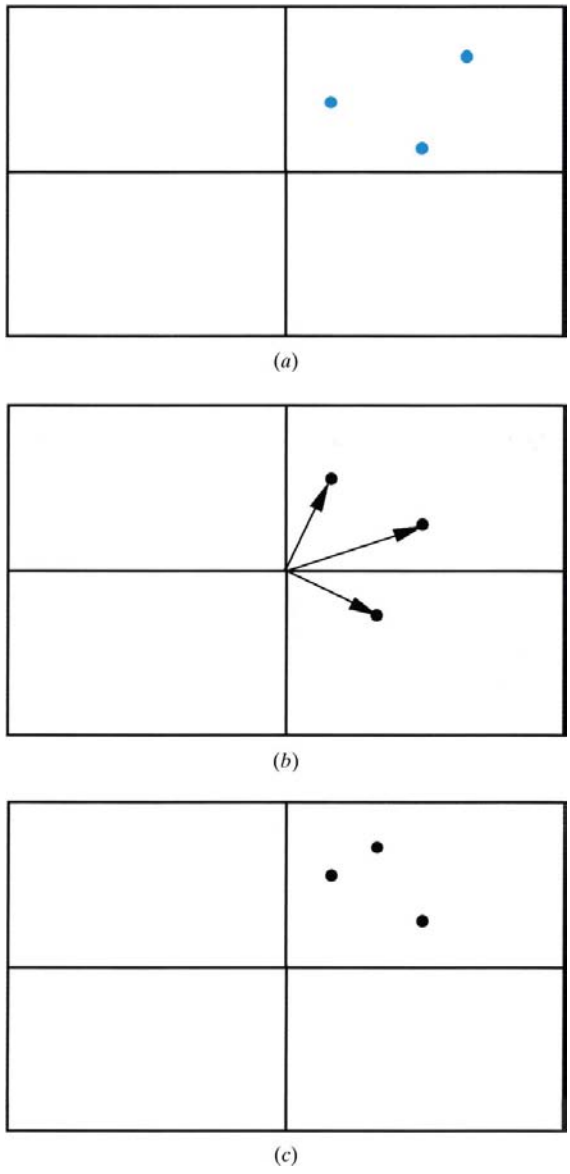

Figure 2
(*a*) Search model. (*b*) Calculated Patterson map of the search model. Peaks related by the inversion centre have been omitted for clarity in all figures. The interpretation of the Patterson map consists of assigning vectors to peaks. (*c*) Calculated Patterson map reduced to the crystallographic unit cell.

Fig. 1(*a*) shows the behaviour of the ISF for different sizes of the FRAGMENT under several statistical situations.

The second result refers to the value of the parameter *m*. After a great number of experiments under completely different conditions, we can conclude that the best value for this parameter is around 0.3*k* or 0.4*k*. Fig. 1(*b*) shows this result.

In a crystallographic problem, the size of the FRAGMENT set is fixed (it is the search model). If the parameter *m* is also fixed, we will try to relate the behaviour of the ISFs to the statistical conditions of the particular problem. The most important result is now presented. There is a high and clear correlation between the ratio $\sigma_{SIGNAL}/\sigma_{TOTAL}$ and the probability of success of the ISFs. Discriminatory capacity of the ISFs increases with $\sigma_{SIGNAL}/\sigma_{TOTAL}$. Failure becomes inevitable when that ratio is very low, as is shown in Fig. 1(*c*).

The represented values must not be taken as absolute criteria as one-dimensional simulations cannot directly be compared with three-dimensional Patterson maps. However, the observed trend is very important and cannot be overlooked in macromolecular cases. Unfortunately, it is not possible to arbitrarily improve the ratio $\sigma_{SIGNAL}/\sigma_{TOTAL}$ without losing the physical sense of the problem in a crystallographic situation. The Patterson function (known) may be considered as the sum of two contributions: the search model (known) and the background (unknown). If the search-model contribution to the total Patterson function is modified so that the values of its characteristic statistical parameters are altered, the background contribution has to change in a similar way. Provided that the same factor is used to scale the search model and the background, the ratio $\sigma_{SIGNAL}/\sigma_{TOTAL}$ remains constant for a given problem. Then, a critical statistical parameter SP1 can be defined as $\sigma_{SIGNAL}/\sigma_{TOTAL}$. It is therefore highly recommended to evaluate SP1 before the calculation starts in order to determine the probability of success of the ISF. In conclusion, SP1 is fixed in a given problem and cannot be altered. The only way to change it consists of increasing the size of the search model or choosing another one.

Besides SP1, other statistical parameters can be defined but they are not so outstanding. They may be altered, but it is not very easy to systematize their behaviour. Therefore, it is not worth calculating them.

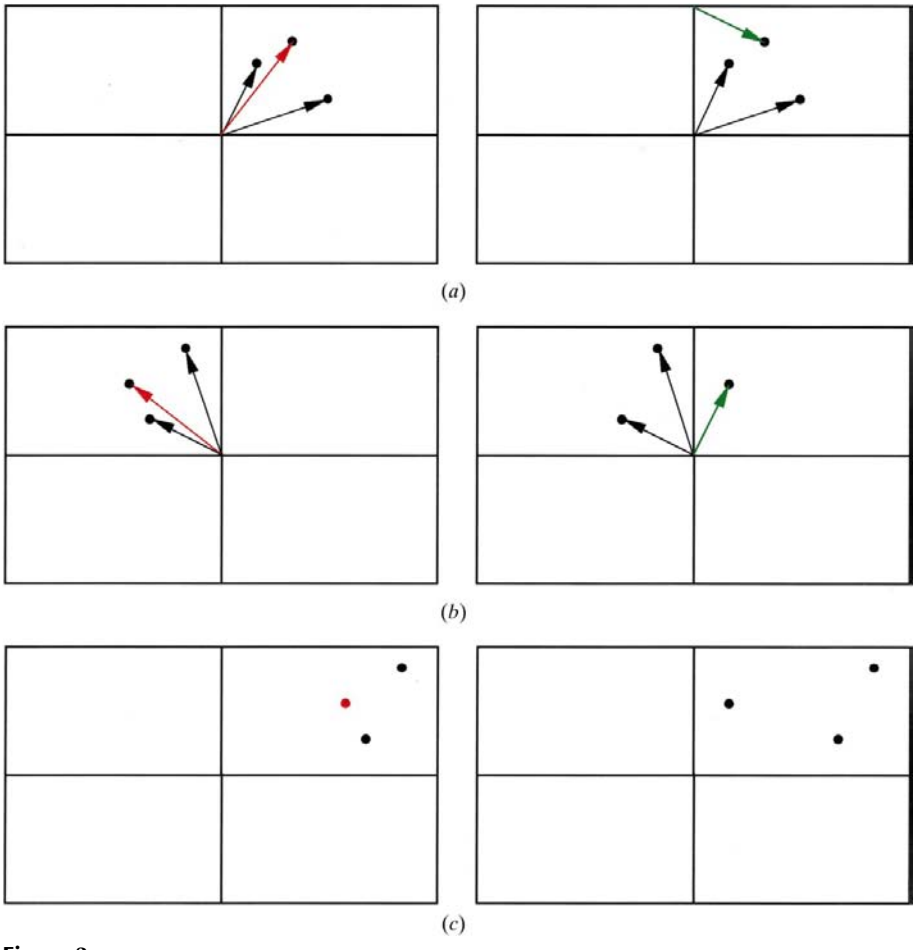

Figure 3
Left column: incorrect interpretation. Right column: correct interpretation. (*a*) The red vector has been incorrectly assigned to a Patterson peak. The green vector represents the correct origin choice (see Fig. 2*b*). (*b*) 90° counterclockwise rotation around an axis through the origin perpendicular to the paper. (*c*) Final rotated peaks inside the crystal unit cell. Since the red and green vectors were different, the final position of the corresponding Patterson peak after rotation is different.

A program (*OVIEST*) which performs the calculation of statistical parameters in a crystallographic situation has been written in FORTRAN77. The TOTAL set corresponds to the observed Patterson map. The SIGNAL set corresponds to the search-model calculated map from which some vectors (the FRAGMENT set) are extracted. The main steps of the program are as follows.

(i) Calculated and observed Patterson maps are read.

(ii) All negatives values of the maps that appear because of the truncation series error are set to zero in order to avoid the physically meaningless situation.

(iii) Average and standard deviation from both maps and the corresponding statistical parameter SP1 are calculated.

2.2. Vector-set generation

ISFs use as input an appropriate vector set from the search model and try to find the orientation which causes the best fit between that vector set and the observed Patterson map. It is necessary to emphasize again that the generated vector set must be contained inside the calculated Patterson map. The problem now is how to choose the correct vector set.

Geometrical procedures to generate interatomic vectors followed by some modifications to make them look like Patterson peaks (Nordman & Schilling, 1970) are prohibitive. Apart from being excessively CPU time consuming, the main reason for avoiding these procedures is that the generated vector set does not resemble the real set which appears in the observed Patterson map.

The natural way of obtaining the vector set as it actually appears in the observed map is to calculate the search-model Patterson map and to consider the extracted peaks as vectors because Patterson peaks correspond with interatomic vectors. These vectors can be classified in two main groups: self-vectors (between atoms in the same asymmetric unit) and cross-vectors (between atoms in different asymmetric units). Cross-vectors can be divided again in two subgroups: type I (between asymmetric units belonging to the same crystallographic unit cell) and type II (between asymmetric units related by lattice translations). Self-vectors depend only on the orientation of the search model. Type I cross-vectors depend on the orientation and the translation and are therefore useless. Type II cross-vectors and self-vectors are inevitably linked, as they are identical apart from lattice translations. From hereon in this paper, we will refer to type I cross-vectors simply as cross-vectors, since type II cross-vectors are not independent of self-vectors. In conclusion, an appropriate vector set for the evaluation of an ISF would be the self-vector set (SVS) of the search model.

The most popular system for building the SVS is to place the search model in a large orthogonal box without symmetry (cross-vectors do not exist under these conditions) and to perform a structure-factor calculation. The box size is chosen to be approximately equal to the smallest box containing the model plus the integration radius plus the required resolution (Navaza, 1994). A Patterson map is then computed. Patterson peaks extracted from that map constitute the SVS. Obviously, this set depends on the box size and the number of generated reflections in the calculation.

In order to obtain the real SVS, as it appears 'camouflaged' in the observed Patterson map, the search model must be placed in the crystallographic unit cell (excluding symmetry). Only observed reflections (expanded to *P*1) must be used in the structure-factor calculation. A Patterson map is calculated. Patterson peaks appear randomly distributed around the origin of the unit cell. However, its coordinates may be changed so that the peaks are positioned inside the unit-cell limits. Each peak can be situated inside the unit cell by adequately applying lattice translations, as shown in Fig. 2.

At this stage, an important detail must be taken into account. This vector set cannot be correctly rotated unless each peak is expressed with respect to its corresponding origin (Fig. 3).

Final rotated peaks naturally depend on the choice of the origin. The problem is how to estimate the correct origin for the peaks extracted from a calculated map. Calculating the Patterson map in a box around the origin does not solve the problem. Fig. 4 shows this process.

Moreover, in the case of multiple peaks (very common in macromolecular crystallography because of the great amount of peak overlap), only one contribution (correct or not) would be selected. The rest are deleted (Fig. 5).

In order to avoid these problems, some authors have decided to calculate the peaks in each orientation of the search model, instead of applying a rotation matrix to the original set. The generalized Patterson-search technique (Nordman, 1994) and the direct rotation function based on a correlation coefficient (DeLano & Brünger, 1995) are examples of this methodology. The CPU-time effort is enormous because a structure-factor calculation and a Patterson map are needed in each orientation. However, the resolving power of this kind of rotation function is greater because the peaks of the search model are calculated more accurately. Nevertheless, routine and quick applications do not need such resolving power.

A correct deconvolution of the peaks from a Patterson map is not easy. Several attempts for small molecules containing heavy atoms have been made (Goldak, 1969, 1971, 1974). This methodology cannot be easily adapted to macromolecules. Our proposal is to design a rough but quick and useful method that interprets correctly Patterson peaks.

We start from the assumption that the search model and its corresponding SVS rotate in the same way. In other words, the SVS follows a predictable path if the rotation matrix is known. Cross-vectors are not present in the calculated map because all symmetry has been excluded. Therefore, in order to determine whether or not a Patterson peak is correctly expressed we have developed the following procedure.

(i) Select a peak and express its coordinates in each one of the eight translationally related corners of the unit cell.

(ii) Choose an arbitrary rotation and apply the corresponding rotation matrix to each one of the possible coordinates of the peak.

(iii) Apply the same rotation matrix to the original search model, perform a structure-factor calculation, build a new calculated Patterson map and perform a peak search on it.

(iv) Compare the possible rotated peaks (step ii) with these new peaks. A possible rotated peak is accepted if it matches up with some of these new obtained peaks. In order to measure the degree of coincidence of two peaks, we have defined a parameter eps1,

$$\text{eps1} = [(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2]^{1/2}, \quad (6)$$

where (x_1, y_1, z_1) are the Cartesian coordinates of the first peak and (x_2, y_2, z_2) are the Cartesian coordinates of the second peak. Multiple peaks are correctly interpreted because all possible origin choices are taken into account. In this case, more than one possible option will be validated. Fig. 6 shows the whole procedure.

The method just described must be applied simultaneously to all Patterson peaks. If an extremely precise SVS is required, the whole procedure must be repeated with new arbitrary rotations until convergence, because if a peak is correctly

Table 1
Scaling results for 6rxn and 1har.

	6rxn			1har		
	Wilson	Blessing	Cowtan	Wilson	Blessing	Cowtan
k	0.580	0.583	0.441	0.046	0.040	0.055
$B_{\rm overall}$	8.440	8.475	6.203	39.120	43.880	43.600

expressed it must appear in any rotated Patterson map independently of the applied rotation. Usually, two or three rotations are sufficient. The parameter eps2 represents the minimum percentage of rotations in which one possible peak must be valid in order to be included in the SVS. In these cases, the value of the accepted peak will be the minimum obtained from all the rotations. Spurious contributions are then eliminated.

The method, although very rough, has the following advantages over traditional procedures.

(i) Patterson peaks (single and multiple) are correctly interpreted and spurious contributions are purged.

(ii) CPU time required is short because only two or three Patterson maps are calculated. A program, *OVISEL*, written in FORTRAN77 performs the calculations involved in the described method. Input parameters for *OVISEL* are the number of random rotations, eps1 and eps2.

2.3. Rotation-search process

Once the SVS is available, the rotation search can be executed. The program *OVIROT* written in FORTRAN77 calculates the 'weighted minimum-average function', which has proved to be a useful rotation function. A flow diagram of the whole process is shown in Fig. 7.

OVIROT reads its input parameters from a file called ovirot.inp which includes the following data.

(i) Crystallographic symmetry, to avoid unnecessary calculations in the rotation search according to the symmetry of the angular space (Rao *et al.*, 1980). The rotation is performed using Eulerian angles (Rossmann & Blow, 1962) over a grid fixed by the symmetry. Orthogonalization convention keeps X in the crystallographic direction a and Y in the plane defined by a and b. The program *CONVROT* (Urzhumtseva & Urzhumtsev, 1997) can convert this rotation description to any other.

(ii) Angular step size. By default fixed to 10° (although other values are allowed).

(iii) Peak-selection criteria, because not all the peaks are used. Peaks included in the calculation are selected accordingly to their length and height. In general, the minimum length is 5 Å. Shorter vectors provide little angular discrimination because the sphere around the Patterson origin shows peaks (vector density) everywhere. The maximum length ($L_{\rm MAX}$) is calculated as the diagonal of the minimum orthogonal box containing the search model. Although long vectors may suffer from certain errors in their determination, they are very characteristic of the search model and must be included. Peaks whose height is very low are excluded. An eight-point interpolation scheme (Nordman, 1980) has been used to estimate peak Patterson values at arbitrary points in the unit cell.

(iv) The fraction of peaks used in the ISF calculation (usually 0.3).

Results from *OVIROT* are written in a file called ovirot.out. This file contains the 100 highest peaks of the rotation function and its value is expressed in standard deviations above the mean. All calculations (*OVIEST*, *OVISEL* and *OVIROT*) were made at the Universidad de Oviedo (Spain) on the X-ray group DEC ALPHA AXP 3300, VMS workstations.

2.4. Scaling of the data

Correct scaling of the data is very important and its great significance has been recently emphasized (Vagin & Teplyakov, 1997).

Observed data must be expressed on an absolute scale (electron units). This transformation is absolutely necessary in order to avoid nonsensical physical situations. ISFs look for the search model inside the observed Patterson map. The absence of data scaling can lead to incorrect situations (for instance, search models 'greater' than observed maps). Obviously, in these cases ISFs will not work properly.

Only a rough value of the overall temperature factor ($B_{\rm overall}$) and scale factor (k) are needed. Extremely precise values (although useful for later steps in the structure-solution process) are not required at this point. The most popular scaling procedure, the Wilson plot (Wilson, 1942), is not very accurate when applied to macromolecules. Only high-resolution data (beyond 4 Å), which is not always available, must be used. However, the Wilson plot must not be automatically excluded. Successful results of the Wilson plot in protein crystallography may be found in the literature (Nixon & North, 1976; Derewenda *et al.*, 1982). In order to avoid Wilson-plot errors, new scaling procedures based on the shape and height of the Patterson origin peak (Goldak, 1974) were established. The fitting of Gaussian functions to the Patterson origin peak by means of a least-squares procedure (Rogers, 1980) has recently been successfully implemented (Blessing & Langs, 1988). Nowadays, the enveloped cross-rotation function (Vellieux, 1995) makes use of this procedure. Recently, the scaling problem has been solved with the use of an empirical scattering curve (Cowtan & Main, 1998). $B_{\rm overall}$ and k are calculated in Wilson statistics through the fitting of a theoretically generated scattering curve to the observed data. This new experimental curve results from the combination of a great number of scattering curves from different protein structures. The algorithm is implemented in the *DM* program (Cowtan, 1994) included in the *CCP*4 (Collaborative Computational Project, Number 4, 1994) package. Wilson, Blessing and Cowtan algorithms have been checked in our test cases. Some conclusions are presented in §3.

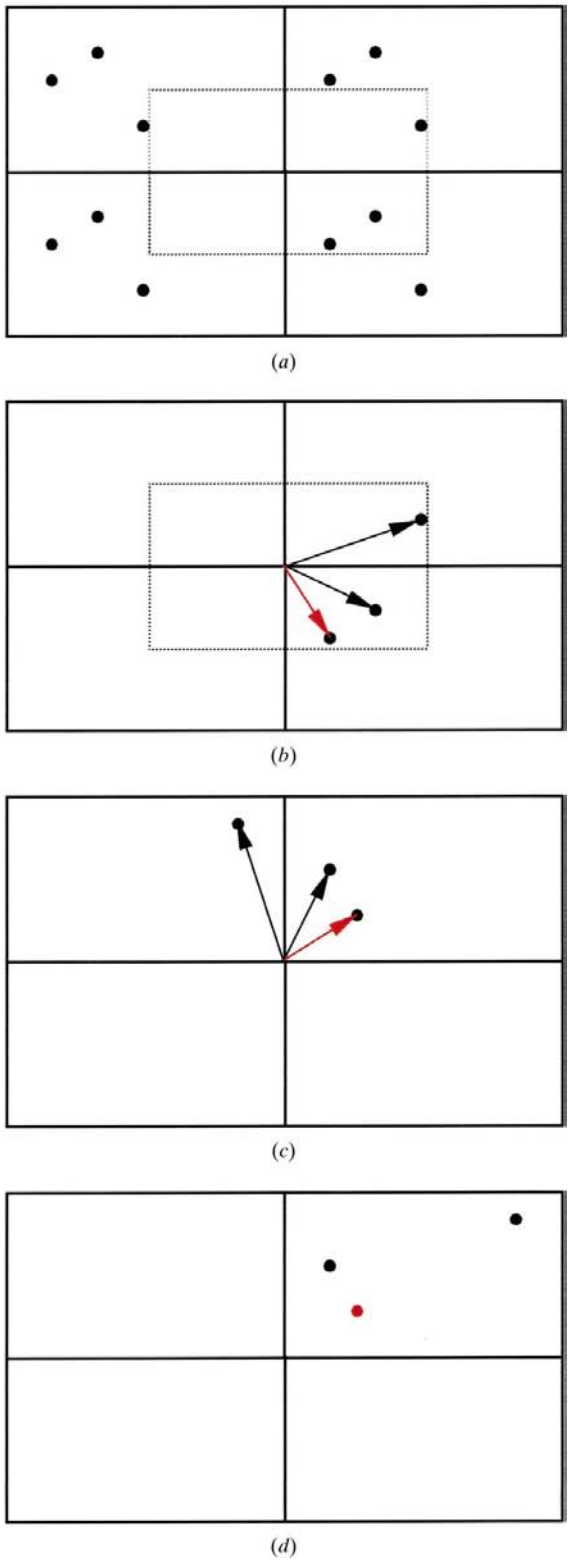

Figure 4
(*a*) Calculated Patterson map. (*b*) Only peaks inside the box are selected. The red vector represents an incorrect interpretation (see Fig. 2*b*). (*c*) 90° counterclockwise rotation around an axis through the origin perpendicular to the paper. (*d*) Final rotated peaks inside the crystallographic unit cell (the box and the vectors have been deleted for clarity). Compare this result with Fig. 3(*c*) (right column): the maps differ because the red vector in the present case is incorrect.

3. Results and discussion

We have used four different protein structures as tests for our new methodology. Data for 6rxn, 1har and 1pdo in the present analysis consisting of $|F|$ and $\sigma(|F|)$ have been obtained from the Protein Data Bank (PDB codes: 6rxn, r6rxnsf; 1har, r1harsf; 1pdo; r1pdosf). The experimental details of data collection, crystal solution and refinement have been reported in the literature. Professor Lennart Sjölin (University of Göteborg, Sweden) provided us with the coordinates and structure factors of the last protein (LSAZ in our internal code).

It is assumed that the scale factor and overall temperature factor for the observed data are available. Table 1 shows the

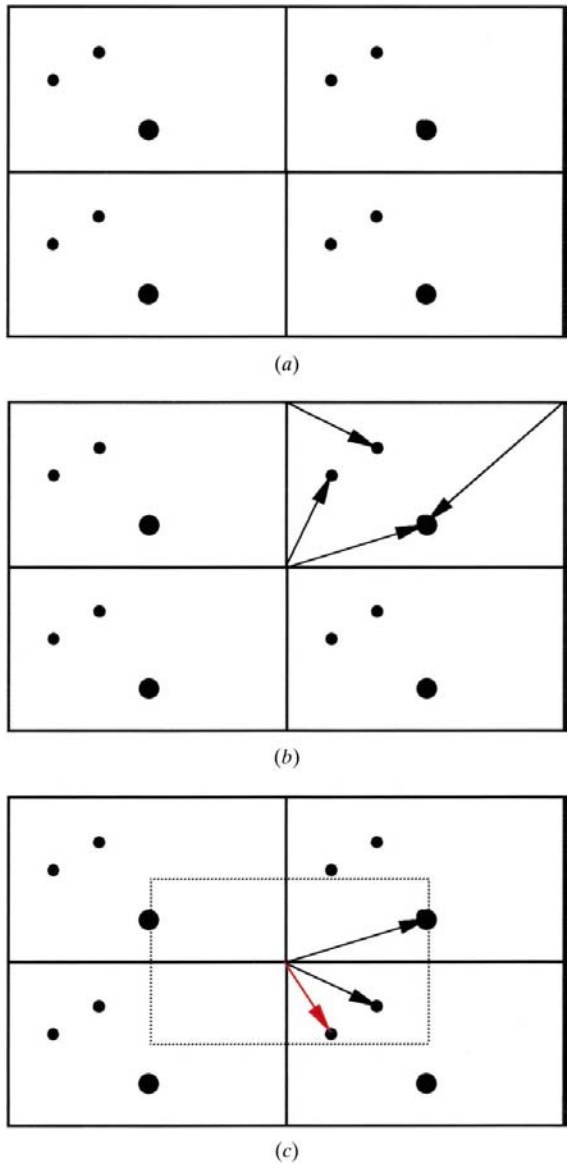

Figure 5
(*a*) Calculated Patterson map. The heavy dot is supposed to be a multiple (double) peak. (*b*) Patterson peaks correctly expressed. Two single vectors contribute to the heavy dot. (*c*) Peaks inside the box are selected. One contribution to the multiple peak is eliminated; that vector is lost. In addition, a single vector is incorrectly assigned (see Fig. 2*b*).

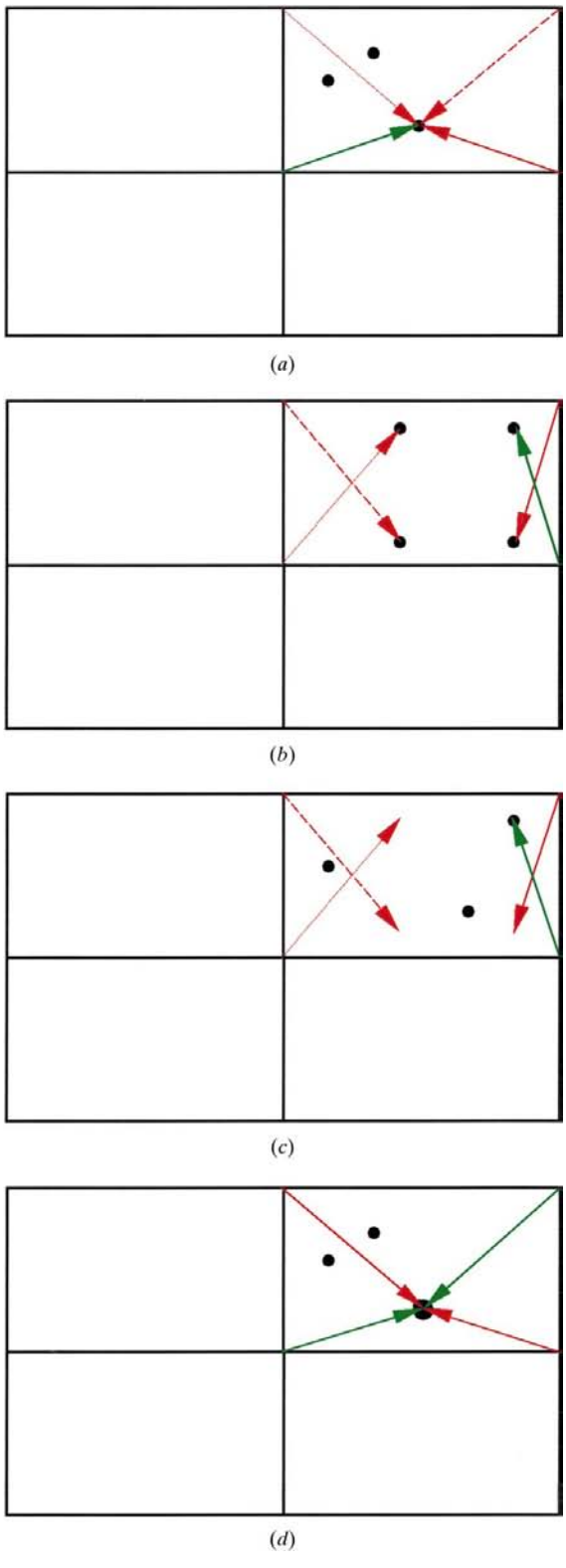

Figure 6
(*a*) All possible vectors for a peak are considered. Only one of them is correct: the green vector (see Fig. 2*b*). (*b*) Final possible positions for the peak after a 90° counterclockwise rotation around an axis through the origin perpendicular to the paper. Four different results depending on the origin choice are obtained. (*c*) Correct rotated Patterson map (Fig. 3*c*, right column) superimposed on the rotated vectors reveals the correct origin for the peak. The red vectors do not match any Patterson peak. (*d*) In our procedure, multiple peaks (see Fig. 5*b*) are correctly interpreted.

scaling results that were obtained when the three techniques mentioned in the previous section were applied to some of our protein tests. No relevant differences were found between the three scaling procedures when Patterson maps and statistical descriptors were calculated. Therefore, we have decided to apply the Wilson-plot procedure for scaling all our observed data sets. *X-PLOR* (Brünger, 1992) was used to produce our Wilson-plot calculations.

Observed and calculated Patterson maps must also be available before the rotation procedure starts. We have used *X-PLOR* (Brünger, 1992) to compute these maps. Reflections from 15 Å to the highest resolution limit available were used in the rotation searches. The same reflections were used in both maps. In the case of the calculated map, the original observed set was expanded to *P*1. Reflections with $F_{\rm obs} < 2\sigma(F_{\rm obs})$ were excluded. The origin peak was subtracted from the Patterson maps. Structure factors obtained from measured reflections and modified with $B_{\rm overall}$ and k were used in the computation of the observed map,

$$\left|F_{\rm absolute}\right|^2 = k^2\left|F_{\rm obs}\right|^2 \exp\left(2B_{\rm overall}\frac{\sin^2\theta}{\lambda^2}\right). \qquad (7)$$

The calculated map was computed from structure factors of the atomic model. These structure factors were calculated by FFT inversion of the electron density generated by the search model in the crystallographic unit cell (symmetry excluded) on a grid whose size was one third of the high-resolution limit (Ten Eyck, 1977). An additional sharpening parameter (usually B_s = 5 Å^2) was also applied in both maps,

$$\left|F_{\rm sharp}\right| = |F| \exp\left(B_s\frac{\sin^2\theta}{\lambda^2}\right). \qquad (8)$$

Three arbitrary rotations were used in the SVS-generation procedure. In all cases, the values of the parameters eps1 and eps2 were chosen to be 0.9 and 60%, respectively.

These default parameters were used throughout in all tests. Other possibilities will be noted. The results of the rotation function are presented in Table 2. The parameter rfr (rotation-function ratio) is the ratio between the value of the rotation function for the correct orientation and the value for the highest false orientation (both expressed in standard deviations above the mean to guarantee that rfr is scale- and shift-invariant).

3.1. 6rxn

This protein is a rubredoxin, an electron-transfer protein from the bacterium *Desulfovibrio desulfuricans* (46 amino acids). It may be considered as a small molecule rather than a typical macromolecule. Relatively high-resolution data are available (5031 reflections collected in the resolution range 24.0–1.5 Å). The protein crystallized in space group *P*1, with unit-cell parameters a = 24.92, b = 17.79, c = 19.72 Å, α = 101.00, β = 83.30, γ = 104.50° and one molecule in the asymmetric unit (Sieker *et al.*, 1983). No cross-vectors appear in the observed Patterson map because symmetry is absent in this crystal structure. There are no excessive difficulties for the ISF. This

protein was extensively used to test all our FORTRAN77 codes.

When the intact protein was used as a search model (model 1), the statistical parameter SP1 was very high (0.36) and a great number of vectors were selected by *OVISEL*. The solution was straightforward and an excellent figure of merit (rfr = 4.16) was obtained.

A second search model (model 2) consisting of the first sixteen amino acids (approximately one third of the whole protein) was tried. In this case, the SP1 value decreased and the number of vectors was also lower than before. *OVIROT* found the correct solution but the figure of merit was worse (rfr = 1.92). The main reason for this drastic reduction is the magnitude of the SP1 parameter. The number of selected vectors (although quite low) was sufficient for the ISF because most of them were correct.

Table 2
Rotation-search results.

Protein	6rxn	6rxn	1har	1har	1pdo	LSAZ
Model	1	2	1	2	1	1
Previous						
L_{MAX}	43.11	31.37	82.37	62.32	57.26	64.09
k	0.58	0.58	0.046	0.046	1.14	14.89
$B_{overall}$	8.44	8.44	39.12	39.12	15.91	18.10
Vectors	204	139	678	245	2220	2895
CPU (s)	74.55	72.61	495.38	469.51	1243.63	356.74
OVIEST						
SP1	0.36	0.21	0.13	0.07	0.02	0.07
CPU (s)	5.98	5.72	46.58	48.31	167.47	38.66
OVISEL						
Vectors	333	172	188	52	343	1124
CPU (s)	149.88	145.68	935.36	932.8	2198.73	1110.27
OVIROT						
Solution	Yes	Yes	Yes	No	Yes	Yes
rfr	4.16	1.92	1.43		1.13	4.84
CPU(s)	76.89	41.15	60.81	56.49	173.06	112.16

3.2. 1har

This protein (Unge *et al.*, 1994) is the HIV-1 (human immunodeficiency virus type 1) reverse transcriptase (216 amino acids). It represents a typical situation in macromolecular crystallography: relatively high symmetry (space group $P4_3$, with one molecule in the asymmetric unit), large unit cell ($a = b = 98.20$, $c = 31.70$ Å) and not very high resolution data (15 150 reflections collected in the resolution range 19.64–2.2 Å). Two different search models were again used: the whole protein and a small fragment.

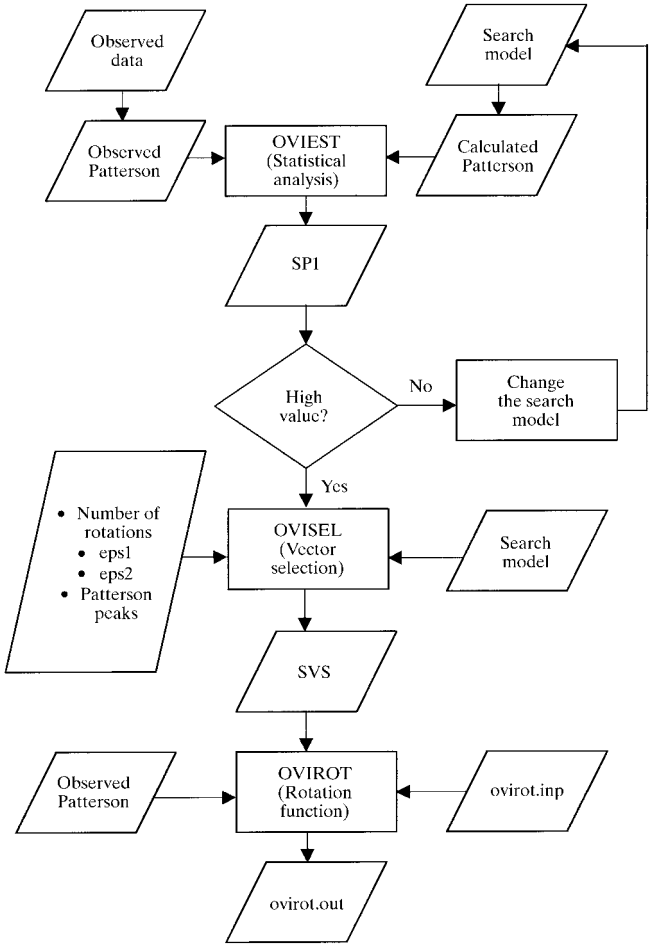

Figure 7
Flowchart of the Patterson space rotation-function procedure.

When the complete protein was used as search model, SP1 was 0.13 and 188 vectors were extracted from the calculated map. The number of vectors is moderately low. However, *OVIROT* again found the correct solution, but the figure of merit was poor (rfr = 1.43).

When the first 56 amino acids of the protein sequence were chosen as the search model (one fourth of the protein), SP1 decreased to 0.07. Only 52 vectors were determined by *OVISEL*. *OVIROT* could not find the correct solution as was expected owing to the low value of the SP1 parameter. A new and more restrictive selection of the vectors was needed. *OVISEL* was run again. This time, six arbitrary rotations (instead of three) were applied. The number of selected vectors was 41, but in this new situation the correct solution was achieved. The figure of merit (rfr = 1.08) was very low.

This example shows the great importance of an appropriate selection of the vectors. ISF does not need a great number of vectors, only good vectors. Introducing wrong peaks always causes ISF errors (the small 6rxn could be oriented using only 20 good vectors from model 1, although the rfr was very low – close to 1.00). All efforts must concentrate on obtaining correct vectors from the search model. Other possibilities are not recommended. For instance, increasing B_s allows more vectors to be obtained, but causes a decrease in SP1; eliminating high-resolution reflections increases SP1, but reduces the number of selected vectors. In conclusion, these tricks must be avoided.

3.3. 1pdo

A mannose permease from *Escherichia coli* (135 amino acids) was crystallized in a very high symmetry space group ($P6_122$), with unit-cell parameters $a = b = 76.36$, $c = 88.73$ Å and one molecule in the asymmetric unit (Nunn *et al.*, 1996). A

great number of cross-vectors caused enormous problems for the ISF. The search model was approximately two thirds of the total protein structure. Although relatively high-resolution data was available (16 836 reflections in the resolution range 19.8–1.7 Å), the SP1 parameter was very poor (0.02) because of the high symmetry. However, the number of vectors was good enough. These effects may cancel each other out. *OVIROT* found the correct solution, although the figure of merit was poor (rfr = 1.13). This example clearly shows the essence of this methodology: we must direct all our efforts towards reaching a great number of good vectors and high SP1 values or, at least, one of these conditions.

3.4. LSAZ

This final example summarizes the last-mentioned idea. LSAZ is the apo form of azurin, a protein involved in the respiratory chain of the bacterium *Pseudomonas aeruginosa* (128 amino acids). The protein crystallized in space group $P2_12_12_1$, with unit-cell parameters a = 57.40, b = 87.40, c = 110.31 Å (local data; 26 578 reflections in the resolution range 22.2–2.79 Å). When the four protein molecules in the asymmetric unit were used as a search model, the SP1 value was very high (0.24) and a great number of vectors were available (1382). *OVIROT* then produced an unambiguous result: the figure of merit (rfr) was 6.50. If just one molecule was used as search model (model 1) the SP1 obviously decreased (0.07). This value was good enough for the rotation function, which found the correct solution; the obtained figure of merit was still quite high (rfr = 4.84).

3.5. Comparison with other molecular-replacement programs

We have studied the performance of our procedure and of two other widely used rotation functions: the direct-rotation function implemented in *X-PLOR* version 3.851 (Delano & Brünger, 1995) and the fast rotation function (Crowther, 1972) as implemented in the program *AMoRe* (Navaza & Saludjian, 1997).

The direct-rotation function solved correctly all the studied test cases, even the most difficult ones {[correct rotation function (RF) value/highest wrong RF value] for some selected cases: 6rxn model 2, 2.41; 1har model 2, 1.51}. As a disadvantage, this function took an excessive CPU time (several hours) to find the same solution as our program (a few minutes).

On the other hand, *AMoRe* was faster than our procedure in all cases. However, when running the program with the default specifications, it could not solve one of the most difficult tests [(correct RF value/highest wrong RF value) for some selected cases: 6rxn model 2, 1.34; 1har model 2, failure]. An appropriate selection of parameters such as resolution range, integration radius or box size could surely have yielded better results. Nevertheless, it is not always straightforward to know how all these parameters affect the performance of this reciprocal-space rotation function.

In conclusion, the new vector-search rotation function can be considered as comparable to other existing functions in speed (similar to *AMoRe*) and accuracy (intermediate between *X-PLOR* and *AMoRe*).

4. Concluding remarks

Indiscriminate usage of ISFs with macromolecular data is risky. Several tests based on computer-generated random numbers show how the behaviour of the ISFs depends largely on the $\sigma_{\rm SIGNAL}/\sigma_{\rm TOTAL}$ relation between the calculated and observed Patterson maps ($\sigma_{\rm CALCULATED}/\sigma_{\rm OBSERVED}$).

Correct scaling of the observed data and appropriate selection of the vectors of the search model are also crucial steps in the rotation search. A new algorithm for the correct selection has been proposed and tested.

ISFs have been recovered for macromolecular crystallography. The proposed methodology works in a systematic way and the CPU time required for the calculations is relatively short.

Future developments of the methodology include the use of data extrapolation (Langs, 1998) beyond the high-resolution limit measured in order to obtain good-quality vectors and the establishment of a powerful algorithm to carry out a correct interpretation of the search model Patterson peaks with the purpose of obtaining an accurate SVS (Álvarez-Rúa *et al.*, 2000).

The authors wish to thank Professor P. T. Beurskens and Professor G. Beurskens (University of Nijmegen, The Netherlands) for their contributions to the early stages of this project, Professor L. Sjölin (University of Göteborg, Sweden) for providing us with the coordinates of LSAZ and for his interest and suggestions, Professor A. T. Brunger (University of Yale, USA) and Professor K. Cowtan (University of York, England) for their assistance in running *X-PLOR* and *DM*, respectively, Professor D. A. Langs (Hauptman–Woodward Medical Research Institute, USA) for providing us with a copy of his software for scaling the observed data and Dr J. Rodríguez-Calonge (Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Spain) for assistance in performing the computations on the CRAY T3E computer. We thank a referee for valuable comments on the manuscript. This work was partially supported by DGICYT (Project PB96-0556). CA-R thanks the Ministerio de Educación y Cultura (Spain) for a fellowship.

References

Álvarez-Rúa, C., Borge, J. & García-Granda, S. (2000). In preparation.

Beevers, C. A. & Robertson, J. H. (1950). *Acta Cryst.* **3**, 164.

Beurskens, P. T., Beurskens, G., Bosman, W. P., de Gelder, R., García-Granda, S., Gould, R. O., Israël, R. & Smits, J. M. M. (1996). *The DIRDIF96 Program System.* University of Nijmegen, The Netherlands.

Beurskens, P. T., Beurskens, G., Strumpel, M. & Nordman, C. E. (1987). *Patterson and Pattersons. Fifty Years of the Patterson Function*, edited by J. P. Glusker, B. K. Patterson & M. Rossy, pp. 356–367. Oxford University Press.
Blessing, R. H. & Langs, D. A. (1988). *Acta Cryst.* A**44**, 729–735.
Brünger, A. T. (1992). *X-PLOR, Version* 3.1. *A System for X-ray Crystallography and NMR.* New Haven: Yale University Press.
Brünger, A. T. (1997). *Methods Enzymol.* **276**, 558–580.
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.
Buerger, M. J. (1950). *Acta Cryst.* **3**, 87–97.
Buerger, M. J. (1951). *Acta Cryst.* **4**, 531–544.
Buerger, M. J. (1959). *Vector Space and its Application in Crystal Structure Investigation.* New York: John Wiley & Sons, Inc.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994). *Jnt CCP4 ESF–EACBM Newslett. Protein Crystallogr.* **31**, 34–38.
Cowtan, K. & Main, P. (1998). *Acta Cryst.* D**54**, 487–493.
Crowther, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 173–178. New York: Gordon & Breach.
DeLano, W. L. & Brünger, A. T. (1995). *Acta Cryst.* D**51**, 740–748.
Derewenda, Z. S., Brzozowski, A. M., Stepien, A. & Grabowski, M. J. (1982). *Acta Cryst.* A**38**, 432–438.
Fitzgerald, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.
García-Granda, S., Borge, J. & Gutiérrez-Rodríguez, A. (1996). *An. Quim. Int. Ed.* **92**(5), 294–298.
Garrido, J. (1950). *Compt. Rendu.* **230**, 1878–1879.
Goldak, G. R. (1969). *Acta Cryst.* A**25**, 367–369.
Goldak, G. R. (1971). *Acta Cryst.* A**27**, 211–216.
Goldak, G. R. (1974). *Acta Cryst.* A**30**, 153–160.
Hoppe, W. (1957). *Acta Cryst.* **10**, 750–751.
Huber, R. (1965). *Acta Cryst.* **19**, 353–356.
Huber, R. (1985). *Proceedings of the Daresbury Study Weekend*, edited by P. A. Machin, pp. 58–61. Warrington: Daresbury Laboratory.
Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1998). *Nature (London)*, **393**, 648–659.
Langs, D. A. (1998). *Acta Cryst.* A**54**, 44–48.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Navaza, J. & Saludjian, P. (1997). *Methods Enzymol.* **276**, 581–594.
Nixon, P. E. & North, A. C. T. (1976). *Acta Cryst.* A**32**, 320–325.
Nordman, C. E. (1966). *Trans. Amer. Crystallogr. Assoc.* **2**, 29–38.
Nordman, C. E. (1972). *Acta Cryst.* A**28**, 134–143.
Nordman, C. E. (1980). *Acta Cryst.* A**36**, 747–754.
Nordman, C. E. (1983). *Proc. Indian Acad. Sci. (Chem. Sci.)*, **92**, 329–334.
Nordman, C. E. (1985). *Crystallographic Computing 3: Data Collection, Structure Determination, Proteins and Databases*, edited by G. M. Sheldrick, C. Krüger & R. Goddard, pp. 232–239. Oxford: Clarendon Press.
Nordman, C. E. (1994). *Acta Cryst.* A**50**, 68–72.
Nordman, C. E. & Hsu, L.-Y. R. (1982). *Computational Crystallography*, edited by D. Sayre, pp. 141–149. Oxford: Clarendon Press.
Nordman, C. E. & Nakatsu, K. (1963). *J. Am. Chem. Soc.* **85**, 353–354.
Nordman, C. E. & Schilling, J. W. (1970). *Crystallographic Computing*, edited by F. R. Ahmed, pp. 110–114. Copenhagen: Munksgaard.
Nunn, R. S., Markovic Housley, Z., Genovesio-Taverne, J. C., Flukiger, K., Rizkallah, P. J., Jansonius, J. N., Schirmer, T. & Erni, B. (1996). *J. Mol. Biol.* **259**, 502–511.
Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1992). *Numerical Recipes. The Art of Scientific Computing (FORTRAN Version)*. Cambridge University Press.
Raman, S. & Lipscomb, W. N. (1961). *Z. Kristallogr.* **116**, 314–327.
Rao, S. N., Jih, J.-H. & Hartsuck, J. A. (1980). *Acta Cryst.* A**36**, 878–884.
Rao, S. T. & Rossmann, M. G. (1973). *J. Mol. Biol.* **76**, 241–256.
Rogers, D. (1980). *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. Ladd & R. A. Palmer, pp. 82–92. London: Plenum Press.
Rossmann, M. G. (1972). Editor. *The Molecular Replacement Method.* New York: Gordon & Breach.
Rossmann, M. G. & Arnold, E. (1993). *International Tables for Crystallography. Volume B. Reciprocal Space*, edited by U. Shmueli, pp. 230–263. Dordrecht: Kluwer Academic Publishers.
Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H.-J., Johnson, J. E., Kamer, G., Luo, M., Mosser, A. G., Rueckert, R. R., Sherry, B. & Vriend, G. (1985). *Nature (London)*, **317**, 145–153.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Rossmann, M. G. & Liljas, A. (1974). *J. Mol. Biol.* **85**, 177–181.
Schilling, J. W. (1970). *Crystallographic Computing*, edited by F. R. Ahmed, pp. 115–123. Copenhagen: Munksgaard.
Sieker, L. C., Jensen, L. H., Pickril, B. C. & Le Gall, J. (1983). *J. Mol. Biol.* **171**, 101–103.
Simpson, P. G., Dobrott, R. D. & Lipscomb, W. N. (1965). *Acta Cryst.* **18**, 169–179.
Steigemann, W. (1996). *Proceedings of the Macromolecular Crystallography Computing School*, edited by P. Bourne & K. Watenpaugh, pp. 204–219. Bellingham: Western Washington University.
Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O. & Abola, E. A. (1998). *Acta Cryst.* D**54**, 1078–1084.
Ten Eyck, L. F. (1977). *Acta Cryst.* A**33**, 486–492.
Thomas, I. D. & McLachlan, D. Jr (1952). *Acta Cryst.* **5**, 301–306.
Unge, T., Knight, S., Bhikhabhai, R., Lovgren, S., Dauter, Z., Wilson, K. & Strandberg, B. (1994). *Structure*, **2**, 953–961.
Urzhumtseva, L. M. & Urzhumtsev, A. G. (1997). *J. Appl. Cryst.* **30**, 402–410.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Vellieux, F. M. D. A. P. (1995). *J. Appl. Cryst.* **28**, 834–836.
Wilson, A. J. C. (1942). *Nature (London)*, **150**, 151–152.
Wrinch, D. (1939). *Philos. Mag.* **27**, 98–122.