Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

RSPS version 4.0: a semi-interactive vector-search program for solving heavy-atom derivatives

Stefan D. Knight

Uppsala Biomedical Center, Department of Molecular Biology, Box 590, S-751 24 Uppsala, Sweden

Correspondence e-mail: stefan@xray.bmc.uu.se

Received 20 April 1999
Accepted 11 October 1999

A program for inspection and interpretation of the Patterson function is described. The program is mainly intended for finding heavy-atom positions from difference Patterson maps, but may also be used to locate molecules with non-crystallographic symmetry when the local axis is nearly parallel to a crystallographic symmetry axis. Options are available for vector-based methods to locate heavy-atom sites, for finding sets from a list of possible heavy-atom positions and for checking of potential solutions. Both crystallographic and non-crystallographic symmetry may be used, either independently or in conjunction.

1. Introduction

Finding heavy-atom positions in at least one heavy-atom derivative is a crucial step in any *de novo* protein structure determination using isomorphous replacement and/or anomalous scattering methods. Traditionally, heavy-atom derivatives have been solved by inspection of the difference Patterson function based on the derivative and native data, and in many cases this is still an efficient method of locating one or a few heavy-atom sites. However, with increasing complexity, following from larger structures and/or high symmetry in the crystal lattice, an automated system for difference Patterson interpretation becomes more and more indispensable. In the case of high symmetry, be it crystallographic (space-group symmetry; SGS) or non-crystallographic symmetry (NCS), the increased complexity arises simply because of the crowding of numerous peaks in the Patterson function. Larger structures give rise to a higher background owing to an increased number of protein–protein and protein–heavy-atom vectors, thus obscuring the already weak signal from the heavy atom(s). Furthermore, with bigger structures the likelihood of multiple heavy-atom binding sites increases. This in turn has two effects. Firstly, the number of peaks in the difference Patterson function increases as $N^2 - N$ (where N is the number of sites in the unit cell), so that crowding of peaks and possible peak overlap becomes more and more serious. Secondly, each additional site adds to the background of the Patterson function, whereas the signal per site remains constant.

Automated methods for solving difference Patterson functions may be loosely divided into two categories: vector-search methods (for some examples, see Tickle, 1983; Terwilliger *et al.*, 1987; Knight, 1989; Knight *et al.*, 1990) and superposition methods (Buerger, 1959; Sheldrick, 1991). *RSPS* is based on vector-search methods, which are essentially trial-and-error approaches. The Patterson function is sampled at points corresponding to the vectors predicted from trial solutions,

either single-atom or multiple-atom, and the values of the Patterson function at these points are somehow combined to generate a 'score' for the trial solution. Trial solutions may for instance be generated by placing either a single atom or a set of atoms with known inter-atomic vectors at different points of a grid covering the asymmetric unit of the crystal.

In order to judge whether a trial solution is likely to be correct or not, some sort of figure-of-merit or score is clearly needed. Buerger (1959) discusses three different scoring functions or image-seeking functions, namely the sum function, the product function and the minimum function. The sum function is the sum of the Patterson 'density' at each of the vectors predicted by a particular trial solution, whereas the product function instead multiplies the values at each of these points. The minimum function is simply the lowest value of the Patterson map at any of the predicted vectors. The sum function is commonly accepted to be the least discriminating of these scoring functions, owing to its insensitivity to missing peaks. This function is also the least sharp and tends to be heavily dominated by contributions from large peaks in the Patterson map. Thus, any position giving one predicted vector at a high-density position in the Patterson map will have a relatively high score when using the sum function. The product function is much sharper and is very sensitive to missing or low peaks, as is the minimum function. However, in the solution of a heavy-atom derivative difference Patterson map this sensitivity to missing or low peaks may sometimes be a disadvantage, since it is not at all uncommon that one or several peaks corresponding to the true solution are missing or are obscured by noise. This problem is avoided in the sum function and also to some extent in the harmonic mean function [hmf = $1/\sum 1/P(u)$]. This latter function tends strongly towards the minimum but also takes some account of the larger values.

In this paper, the Patterson search program *RSPS* version 4.0 is described. This version has many new features, including search options that may be used both to find heavy-atom positions from difference Patterson maps and to locate molecules with NCS.

2. Description of the program

RSPS is a command-driven program that uses vector-search methods to find heavy-atom positions from difference Patterson maps. *RSPS* version 4.0 may also be used to locate molecules with NCS. The input of commands is free format and case insensitive. Commands may be abbreviated and may be freely interspersed by comments. A simple macro facility is available that allows the user to collect sequences of commands that constitute some basic operation such as the picking of peaks on a specific Harker section of the Patterson map. The program is intended to provide a flexible and user-friendly toolkit for inspection and interpretation of the Patterson function, but still requires a basic understanding of vector space and its relation to real space for successful application.

All options in *RSPS* operate in real or vector space. Search options are divided into six groups depending on the vector set to be used in the search. The available sets are of two main types: atom vector sets, which are used to search for the position of heavy atoms, and molecule vector sets, which are used to find the position of molecules with NCS when at least one NCS axis is closely parallel to a crystallographic symmetry axis. The two main types of vector sets each have three subcategories termed single, more and translate. With single vector sets (selected by VECTORSET SINGLE), the position of one site at a time is determined by considering vectors between symmetry-related positions. This type of search is referred to as a single-site search. When only SGS is used in a single-site search, this corresponds to using only Harker vectors, *i.e.* vectors between SGS-related positions (solid arrows in Fig. 1). When NCS is applied, cross vectors that may be termed pseudo-Harker vectors will be generated between NCS-related positions (dashed arrows in Fig. 1). The combination of SGS and NCS will in addition generate cross vectors between positions on different copies of the NCS protein molecule (dotted arrows in Fig. 1). Once at least one atom position (or NCS molecule) has been found, it may be fixed and used to search for further sites by considering cross

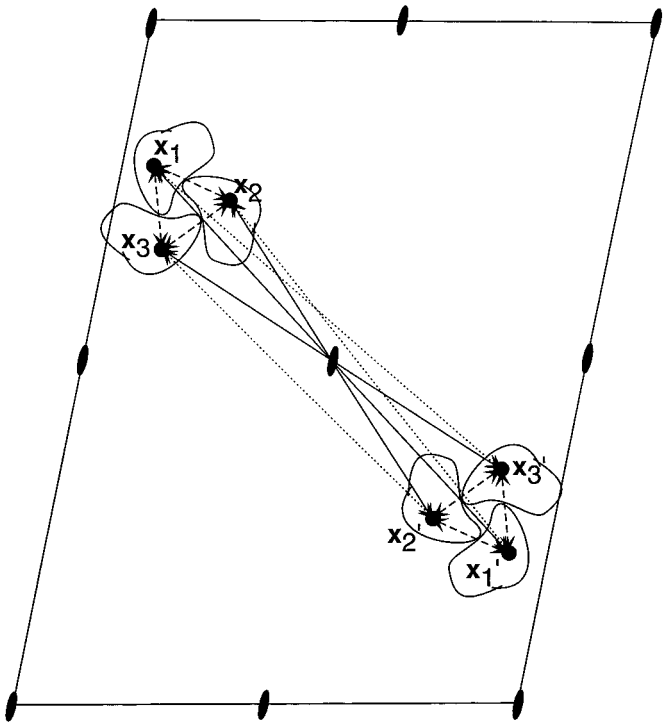

Figure 1
Schematic drawing of a protein trimer with threefold NCS in the space group *P*2, illustrating various types of vectors between heavy atoms. Each protein monomer contains a single heavy-atom site (black circles). Each heavy atom $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ (related by the threefold NCS) generates a Harker vector to its SGS mate $\mathbf{x}'_1$, $\mathbf{x}'_2$ and $\mathbf{x}'_3$, respectively (solid arrows). Pseudo-Harker cross vectors (dashed lines) are generated between heavy atoms within each NCS copy of the trimer in the cell (*e.g.* $\mathbf{x}_1 - \mathbf{x}_2$, $\mathbf{x}_2 - \mathbf{x}_3$, $\mathbf{x}_3 - \mathbf{x}_1$). These vectors only depend on the orientation of the trimer and are independent of the position of the NCS molecule in the cell. Additional cross vectors are generated between heavy atoms on different trimers (dotted arrows). For clarity, only the three unique vectors of this type ($\mathbf{x}_1 - \mathbf{x}'_3$, $\mathbf{x}_2 - \mathbf{x}'_1$, $\mathbf{x}_3 - \mathbf{x}'_2$) are shown.

vectors from trial sites to the fixed site(s) (using VECTORSET MORE). VECTORSET TRANSLATE is used to search simultaneously for two or more atoms provided that their inter-atomic vectors are known. Both SGS and NCS may be independently switched on and off, giving a very flexible means of controlling the type of vectors to be used in the search. A summary of the search options available in *RSPS* is given in Table 1.

Prior to any search in *RSPS*, all unique transformations that map the search fragment, which may consist of a single or several atoms, to vectors in Patterson space are generated. In all cases, the formulations lead to simple representations of the vector sets to be used in terms of 3×3 matrices and 3×1 vectors that, when applied to a point in real space, generate a point in vector space. Thus, in a single-site search (single vector sets) a transformation from a trial position $\mathbf{x}$ in real space to a vector $\mathbf{u}$ in vector space is defined as

$$\begin{aligned}\mathbf{u}_{ijkl}(\mathbf{x}) &= R_i^{\rm SGS}(R_k^{\rm NCS}\mathbf{x} + \mathbf{t}_k^{\rm NCS}) + \mathbf{t}_i^{\rm SGS} \\ &\quad - [R_j^{\rm SGS}(R_l^{\rm NCS}\mathbf{x} + \mathbf{t}_l^{\rm NCS}) + \mathbf{t}_j^{\rm SGS}] \\ &= (R_i^{\rm SGS}R_k^{\rm NCS} - R_j^{\rm SGS}R_l^{\rm NCS})\mathbf{x} + (R_i^{\rm SGS}\mathbf{t}_k^{\rm NCS} - R_j^{\rm SGS}\mathbf{t}_l^{\rm NCS}) \\ &\quad + (\mathbf{t}_i^{\rm SGS} + \mathbf{t}_j^{\rm SGS}) \\ &= T_{ijkl}\mathbf{x} + \mathbf{s}_{ijkl}, \end{aligned} \tag{1}$$

whereas in a cross vector search (more vector sets) aimed at finding additional sites by looking at the cross vectors between one or more known sites and the trial positions,

$$\begin{aligned}\mathbf{u}_{ijklm}(\mathbf{x}) &= R_i^{\rm SGS}(R_k^{\rm NCS}\mathbf{x} + \mathbf{t}_k^{\rm NCS}) + \mathbf{t}_i^{\rm SGS} \\ &\quad - [R_j^{\rm SGS}(R_l^{\rm NCS}\mathbf{x}_m^{\rm FIX} + \mathbf{t}_l^{\rm NCS}) + \mathbf{t}_j^{\rm SGS}] \\ &= (R_i^{\rm SGS}R_k^{\rm NCS})\mathbf{x} - (R_j^{\rm SGS}R_l^{\rm NCS})\mathbf{x}_m^{\rm FIX} \\ &\quad + (R_i^{\rm SGS}\mathbf{t}_k^{\rm NCS} - R_j^{\rm SGS}\mathbf{t}_l^{\rm NCS}) + (\mathbf{t}_i^{\rm SGS} + \mathbf{t}_j^{\rm SGS}) \\ &= T_{ik}\mathbf{x} + \mathbf{s}_{ijklm}. \end{aligned} \tag{2}$$

In these expressions, R and $\mathbf{t}$ are the rotation matrices and translation vectors needed to describe the SGS and/or NCS, $\mathbf{u}$ specifies a position in vector space, $\mathbf{x}$ is the search parameter (heavy-atom position), $\mathbf{x}^{\rm FIX}$ are the known heavy-atom positions and T and $\mathbf{s}$ are the generated matrices and translation vectors that map a trial position $\mathbf{x}$ in the crystal to vector positions $\mathbf{u}$ in the difference Patterson. Note that although the transformation matrix T is a 3×3 matrix, it will not necessarily be a rotation matrix.

To obtain the translate vector sets we start with a set of positions $\mathbf{x}$ with fixed interatomic vectors at an arbitrary location in the unit cell. We now need to find the translation $\mathbf{t}^x$ that will move the set of atoms as a rigid body to the correct location in the unit cell. Applying SGS and NCS to the translated positions $(\mathbf{x} + \mathbf{t}^x)$ and taking the difference to obtain vectors, we obtain

$$\begin{aligned}\mathbf{u}_{ijklmn}(\mathbf{t}^x) &= R_i^{\rm SGS}[R_k^{\rm NCS}(\mathbf{x}_m + \mathbf{t}^x) + \mathbf{t}_k^{\rm NCS}] + \mathbf{t}_i^{\rm SGS} \\ &\quad - \{R_j^{\rm SGS}[R_l^{\rm NCS}(\mathbf{x}_n + \mathbf{t}^x) + \mathbf{t}_l^{\rm NCS}] + \mathbf{t}_j^{\rm SGS}\} \\ &= (R_i^{\rm SGS}R_k^{\rm NCS} - R_j^{\rm SGS}R_l^{\rm NCS})\mathbf{t}^x \\ &\quad + (R_i^{\rm SGS}R_k^{\rm NCS}\mathbf{x}_m - R_j^{\rm SGS}R_l^{\rm NCS}\mathbf{x}_n) \\ &\quad + (R_i^{\rm SGS}\mathbf{t}_k^{\rm NCS} - R_j^{\rm SGS}\mathbf{t}_l^{\rm NCS}) + (\mathbf{t}_i^{\rm SGS} - \mathbf{t}_j^{\rm SGS}) \\ &= T_{ijkl}\mathbf{t}^x + \mathbf{s}_{ijklmn}. \end{aligned} \tag{3}$$

Vectors $\mathbf{u}_{ijklmn}(\mathbf{t}^x)$ are computed for all combinations of symmetry for each pair of atoms $\mathbf{x}_m$ and $\mathbf{x}_n$ in the search fragment. The search parameter now is the translation vector $\mathbf{t}^x$ that has to be added to the starting positions to move them to the correct location in the unit cell. If $m = n$, vectors between symmetry-related positions result (Harker vectors when $k = l = 1$ so that $R^{\rm NCS} = I$, pseudo-Harker vectors when $i = j = 1$ so that $R^{\rm SGS} = I$), whereas when $m \neq n$, cross vectors are generated. [With $m = n = 1$ and $\mathbf{x}_m = 0$, (3) reduces to (1) with $\mathbf{t}^x$ in place of $\mathbf{x}$ as the search parameter, showing that the single-site vector set is a special case of the translate vector set.] When $i = j$ and $k = l$, T_{ijkl} becomes a null matrix and intra-fragment vectors are generated (dashed arrows in Fig. 1). These are simply copies of the original $\mathbf{x}_m - \mathbf{x}_n$ vectors and are independent of the position of the fragment and thus need not be considered. Note that in all cases where a cross vector between two atoms can be identified, a two-atom search fragment can simply be constructed by placing one atom at the origin and another at the end of the cross vector. If NCS is to be used in the search, it is of course important that all positions in the search fragment belong to the same NCS copy of the molecule. For this reason, it is sensible to select a short vector when generating a two-atom fragment from a cross vector, since a short vector is more likely to link atoms from the same copy of the molecule.

It is often the case that NCS axes are more or less parallel to crystallographic symmetry axes (Wang & Janin, 1993). When the aligned axes share the same minimal point symmetry, then two or more copies of the molecule in the unit cell will be in the same orientation. For example, an NCS dimer with the twofold axis parallel to a two-, four- or sixfold crystal symmetry axis will generate two copies of the dimer in identical orientation. These molecules will then be related by a simple translation vector giving rise to a strong translation peak in the Patterson function. In such cases, for some non-trivial combination of *i, j, k, l* ($i \neq j$, $k \neq l$), $R_i^{\rm SGS}R_k^{\rm NCS} = R_j^{\rm SGS}R_l^{\rm NCS}$. If this condition is applied to the vector expressions above, it will be seen that a subset of the vectors are independent of the atomic coordinates and only depend on the position of the NCS axis in the unit cell. Such structure-invariant vectors may thus be used to find the position of molecules with NCS. For example, in the case of a single-site search, (1) above becomes

$$\mathbf{u}_{ijkl} = (R_i^{\rm SGS}\mathbf{t}_k^{NCS} - R_j^{\rm SGS}\mathbf{t}_l^{\rm NCS}) + (\mathbf{t}_i^{\rm SGS} - \mathbf{t}_j^{\rm SGS}). \tag{4}$$

Now, $\mathbf{t}_k^{\rm NCS} = (I - R_k^{\rm NCS})\mathbf{t}^x$, where $\mathbf{t}^x$ is the position of the NCS axis and I is the identity matrix, so

Table 1
Summary of available search options in *RSPS* version 4.0.

Vector set	SGS	NCS	Action
Single atoms (equation 1)	On	On/Off	Perform single-site search using vectors between symmetry-related positions. When only SGS is used, this amounts to a search using Harker vectors only. NCS symmetry will in addition generate pseudo-Harker cross vectors between NCS-related positions and cross vectors between positions on different NCS copies of the protein molecule.
	Off	On using only rotational part	Locate positions related by NCS from the translation-independent pseudo-Harker cross vectors between NCS-related positions. The positions will be displaced from their true position by a vector **t** which may be found in a TRANSLATE ATOMS scan.
More atoms (equation 2)	On/Off	On/Off	Given one or more fixed positions, find additional sites by looking at cross vectors to the fixed sites. Harker vectors for potential solutions may then be examined by using the VLIST command.
Translate atoms (equation 3)	On/Off	On/Off	Translate two or more positions as a rigid body. These positions may come from (i) single-site search with SGSYMM OFF, NCSYMM ON, (ii) cross vector ⇒ two-site search or (iii) known (oriented) fragment.
Single molecules (equation 5)	On	On using only rotational part	Find location of symmetric molecule using the structure-invariant subset of Harker vectors.
More molecules (equation 6)	On	On using only rotational part	Given the position of one or more molecules with NCS, find the position of additional symmetric molecules using the structure-invariant subset of cross vectors.
Translate molecules (equation 7)	On	On using only rotational part	Translate two or more NCS molecules with a fixed separation as a rigid body.

$$\begin{aligned}\mathbf{u}_{ijkl}(\mathbf{t}^x) &= R_i^{\rm SGS}(I - R_k^{\rm NCS})\mathbf{t}^x - R_j^{\rm SGS}(I - R_l^{\rm NCS})\mathbf{t}^x + (\mathbf{t}_i^{\rm SGS} - \mathbf{t}_j^{\rm SGS}) \\ &= [R_i^{\rm SGS} - R_j^{\rm SGS} - (R_i^{\rm SGS}R_k^{\rm NCS} - R_j^{\rm SGS}R_l^{\rm NCS})]\mathbf{t}^x \\ &\quad + (\mathbf{t}_i^{\rm SGS} - \mathbf{t}_j^{\rm SGS}) \\ &= (R_i^{\rm SGS} - R_j^{\rm SGS})\mathbf{t}^x + (\mathbf{t}_i^{\rm SGS} - \mathbf{t}_j^{\rm SGS}) \\ &= T_{ij}\mathbf{t}^x + \mathbf{s}_{ij}. \end{aligned} \tag{5}$$

Similarly, one may derive expressions for structure-invariant vectors between a molecule with NCS at a fixed (known) position and a second molecule, or for a pair of NCS molecules at an arbitrary position in the unit cell separated by a known translation vector. Given the condition stated above that $R_i^{\rm SGS}R_k^{\rm NCS} = R_j^{\rm SGS}R_l^{\rm NCS}$, the set of structure-invariant cross vectors between a molecule with NCS at a fixed position and a second molecule is given by

$$\begin{aligned}\mathbf{u}_{ijk}(\mathbf{t}^x) &= R_i^{\rm SGS}\mathbf{t}^x - R_j^{\rm SGS}\mathbf{t}_k^{\rm NCS,FIX} + (\mathbf{t}_i^{\rm SGS} - \mathbf{t}_j^{\rm SGS}) \\ &= R_i^{\rm SGS}\mathbf{t}^x + \mathbf{s}_{ijk} \\ &= T_i\mathbf{t}^x + \mathbf{s}_{ijk} \end{aligned} \tag{6}$$

and the set of structure-invariant vectors between a pair of NCS molecules with a fixed separation is given by

$$\begin{aligned}\mathbf{u}_{ijk}(\mathbf{t}^x) &= (R_i^{\rm SGS} - R_j^{\rm SGS})(\mathbf{t}^x + \mathbf{t}_k^{\rm NCS,FIX}) + R_j^{\rm SGS}\mathbf{t}^{\rm SEP} \\ &\quad + (\mathbf{t}_i^{\rm SGS} - \mathbf{t}_j^{\rm SGS}) \\ &= T_{ij}\mathbf{t}^x + \mathbf{s}_{ijk}. \end{aligned} \tag{7}$$

In (6) and (7), $\mathbf{t}^{\rm NCS,FIX}$ represents the initial position of the molecule(s) and $\mathbf{t}^{\rm SEP} = \mathbf{t}_2^{\rm NCS,FIX} - \mathbf{t}_1^{\rm NCS,FIX}$ is the fixed separation between two molecules in the unit cell. By choosing one of the molecule vector sets in *RSPS*, structure-invariant vectors are selected and may be used to find the position of NCS molecules in the unit cell.

Searches are carried out in *RSPS* by letting the search parameter (atom position **x** or translation vector **t** in 1–7) vary over a grid covering the unit cell and computing a score for each trial position by combining the value of the Patterson function at the points predicted using the vector sets defined above. At present, three basic scoring functions are available in *RSPS*: the sum, the product and the harmonic mean functions. All of these may be modified to use only the N smallest values at the predicted vector positions [min(N) function; High & Kraut, 1966]. When $N = 1$, this amounts to the classical minimum function. The selection of $N > 1$ allows the problem of missing Patterson peaks to be circumvented. By setting N to less than the number of vectors per site, the dominating effect of very strong peaks on the score may be limited. The scores may also be modified by a simple weighting function based on the multiplicity of the peaks. *RSPS* automatically assigns a multiplicity to each vector by counting the number of times each transformation occurs. In the case of the sum function, this amounts to a simple correlation function between the predicted and observed Patterson peaks. A cut-off level setting the minimum acceptable Patterson density at a predicted vector position may be specified, together with a limit for the maximum number of predicted vectors allowed to be less than the cut-off. When this latter limit is passed, the trial position is rejected and given a score of zero. The combined use of this rejection scheme and judicious use of the minimum function leads to a flexible set of scoring options that can gracefully accommodate the varying degree of 'sharpness' of a protein difference Patterson function. The value of the chosen score function at each trial position is stored in a *CCP*4 format map file as an average over all predicted vectors in units of the standard deviation of the Patterson map.

To illustrate how *RSPS* may be used, consider the example in Fig. 1. A single-site search will place a trial heavy atom **x** at each point of a grid covering the asymmetric unit of the *P*2 cell. For each trial position, the value of the difference Patterson function at predicted vector positions will be combined using one of the available scoring functions. If only SGS is used, one Harker vector will be generated for each trial position. Assuming the search is over the 'top' half of the cell in Fig. 1, predicted vectors should coincide with high values in the Patterson when the trial position is at $\mathbf{x}_1$, $\mathbf{x}_2$ or $\mathbf{x}_3$ (generating the vectors $\mathbf{x}_1 - \mathbf{x}_1'$, $\mathbf{x}_2 - \mathbf{x}_2'$ and $\mathbf{x}_3 - \mathbf{x}_3'$; solid arrows in Fig. 1) or at positions related to these by inversion or a shift of

origin. However, since there is only one vector per position, many trial positions will give relatively high scores because the corresponding vector happens to fall in a region of high noise in the Patterson. NCS may then be used to advantage to increase the number of vectors per position and thus decrease the probability of false 'hits'. If only the rotational part of the NCS is known, a single-site search with the SGS turned off may be carried out. In the current example, this would generate three unique vectors between NCS-related heavy atoms (vectors $\mathbf{x}_1 - \mathbf{x}_2$, $\mathbf{x}_2 - \mathbf{x}_3$ and $\mathbf{x}_3 - \mathbf{x}_1$; dashed arrows in Fig. 1). However, by leaving out the translational part of the NCS, we have shifted the NCS rotation axis to the origin. Predicted vectors will, therefore, coincide with observed peaks between heavy atoms in the trimer when the trial position is shifted from the real positions by the same amount. A search using the translate single atoms vector set may then be used to find the correct location of the three heavy atoms by shifting them simultaneously through the asymmetric unit and looking at Harker vectors (solid arrows in Fig. 1) and cross vectors between heavy atoms on different molecules (dotted arrows in Fig. 1). If the location of the NCS axis is already known, we may instead carry out a single-site search using both the SGS and the NCS from the start. In this case, nine unique vectors would be generated for each trial position: three Harker vectors (one for each NCS-related position), three pseudo-Harker cross vectors between each of the NCS-related positions and three unique cross vectors between heavy atoms on different trimers. The probability of finding false hits where all nine predicted vectors fall in high-noise regions of the Patterson map is obviously very much smaller than in the case of using only the SGS.

The Patterson map is read from an external file that may be in either *CCP*4 (Collaborative Computational Project, Number 4, 1994) or *PROTEIN* (Steigemann, 1974) format. The whole Patterson map is stored and held in core as a three-dimensional array of real values. For low-symmetry space groups one asymmetric unit will suffice, whereas for space groups with trigonal or higher symmetry a larger volume is normally required. Origin peaks or any other undesired peaks may be removed or truncated by a simple resetting function. The Patterson map may be picked using the *RSPS* picking options to produce a list of Patterson peaks. These peaks are not used to determine the heavy-atom positions, but may be used in the checking options to compare the predicted and the observed difference Patterson function.

The score map produced by a search may be picked using built-in commands to generate a list of potential heavy-atom sites that can also be written to a PDB (Bernstein *et al.*, 1977) format file. Coordinates of potential heavy-atom sites may also be read from an external PDB format file. Each entry is given a 'position' number, which is just a sequential numbering of the stored coordinates, and a 'site' number that groups together homomorphous positions. These are defined here as positions that generate the same set of Harker vectors and, in space groups with a polar direction, are related by an inversion halfway along the polar axis or have identical coordinates in the polar direction. The site number thus groups positions related by inversion and/or origin shifts (or SGS).

Given a list of potential heavy-atom positions, the GETSETS option may be used to search for sets of positions. This is performed by looking at the cross vectors between all pairs of atoms (and their symmetry-related equivalents) in the list. If a pair of positions passes the specified rejection criteria it is flagged as 'connected'; otherwise it is flagged as 'unconnected'. In this way, an $N \times N$ matrix is built up with the value 'true' at indices (i, j) if positions i and j are connected and 'false' otherwise. An iterative filtering algorithm is then applied to pick out all sets where all pairs of positions are flagged as 'connected'. The output from GETSETS lists, for each set, the coordinates of the positions in the set and a score table giving the score for the Harker and cross vectors generated by these positions. A summary of the 'quality' of the set is also provided, showing what fraction of the predicted vectors are close to peaks in the Patterson map.

A single-site search in a polar space group will result in a list of potential positions where one coordinate has not been determined (this will have been arbitrarily set to zero by the program). The POLARSCAN option may then be used to try and relate different solutions from the search to the same origin by fixing one position and translating the rest, one at a time, along the polar axis. Scores based on all the cross vectors between the fixed and the translated position (and their symmetry-related equivalents) are computed as a function of the displacement along the polar axis and are stored. This type of search does not produce a score map; instead, the coordinates are put directly in the main positions storage area where they will overwrite any previously stored coordinates.

The VLIST option is used to examine potential solutions. Given a list of one or more positions (if cross vector analysis is selected, of course, at least two are expected), the Harker or cross vectors for these positions are listed. If the Patterson map has been picked, the list of stored peaks is searched to find peaks close to the predicted vectors. The largest peak within 2.5 grid divisions of a predicted vector is listed, together with the distance from the predicted vector.

RSPS is written in Fortran77 with a few commonly accepted extensions that are detailed in the program manual. The program is designed so that it can easily be run interactively, although, depending on the symmetry, the size of the cell and the computer, the response may be far from interactive. Thus, a vector search for heavy-atom positions would normally be run as a batch job, whereas checking of results can in most cases be performed interactively.

The structure of the *RSPS* program is highly modular in order to allow for flexibility in debugging and future development. At the lowest level are a number of library routines that handle matrix and vector algebraic operations as well as elementary operations on positions and peaks in direct and vector space, respectively. At the heart of the program is the command interpreter which is based on the *CCP*4 parser and

terminal input/output routines from the library package *FORLIB* written by Kraulis (1989). Higher level routines carry out the various search and checking options available in *RSPS*. Definitions of default values for parameters, dimensioning statements and common block statements have been collected in a number of include files and may thus easily be modified.

The program, with a manual and test examples, is distributed with the *CCP*4 program system and is also available directly from the author on request (e-mail: stefan@xray.bmc.uu.se).

The author wishes to thank Carl-Ivar Brändén for providing the initial incentive to write this program and Alwyn Jones, Lars Liljas and others at the Department of Molecular Biology, Uppsala, for useful ideas and comments. Sherry Mowbray is thanked for critical reading of the manuscript. This work was supported by grants from the Swedish Natural Science Research Council NFR.

References

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Buerger, M. J. (1959). *Vector Space.* New York: John Wiley.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

High, D. F. & Kraut, J. (1966). *Acta Cryst.* **21**, 88–96.

Knight, S. (1989). Thesis, Swedish University of Agricultural Sciences, Uppsala, Sweden.

Knight, S., Andersson, I. & Brändén, C.-I. (1990). *J. Mol. Biol.* **215**, 113–160.

Kraulis, P. J. (1989). *J. Magn. Reson.* **84**, 627–633.

Sheldrick, G. M. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 23–38. Warrington: Daresbury Laboratory.

Steigemann, W. (1974). PhD thesis, Technische Universität München, Germany.

Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. (1987). *Acta Cryst.* A**43**, 1–5.

Tickle, I. J. (1983). *Proceedings of the CCP4 Study Weekend. Location and Refinement of Heavy-Atom Sites*, edited by P. Machin, A. Wonacott & D. S. Moss, pp. 5–10. Warrington: Daresbury Laboratory.

Wang, X. & Janin, J. (1993). *Acta Cryst.* D**49**, 505–512.