Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

COMO: a program for combined molecular replacement

Gerwald Jogl, Xiao Tao, Yingwu Xu and Liang Tong*

Department of Biological Sciences, Columbia University, New York, NY 10027, USA

Correspondence e-mail: tong@como.bio.columbia.edu

Received 13 February 2001
Accepted 23 April 2001

The combined molecular-replacement protocol uses a limited six-dimensional search to solve a structure by the molecular-replacement method, with the sampling of the rotational degrees of freedom guided by the rotation function. This protocol therefore automatically combines the information on the rotational and translational parameters of the search model. The combined molecular-replacement protocol has been implemented in a new computer program, *COMO*. The calculation of the Patterson correlation translation function has been optimized to improve its speed performance. A packing check is implemented that automatically removes impossible solutions and thereby increases the signal in the calculation. A family of atomic models can be used as the search model; the program will automatically select the model that gives the best result. The command interface is well organized and requires the definition of only a few critical parameters by the user. In addition, a graphical user interface has been constructed for the program. The program has been used to solve several difficult molecular-replacement problems. A case is presented where the program automatically determined the orientation and position of five copies of a search model in a high-symmetry space group.

1. Introduction

Structure solution by the molecular-replacement (MR) method requires the determination of the rotational and translational parameters of a search atomic model (Rossmann, 1990). This six-dimensional problem is traditionally solved in two steps: determination of the three rotational parameters by the rotation function (RF), followed by the determination of the three translational parameters of the search model by the translation function (TF; Rossmann, 1972). Although this protocol leads to tremendous savings in computing time, it suffers from the serious drawback that generally only a few rotation angles are examined by the TF. In other words, the rotational degrees of freedom are sampled extremely poorly by this protocol. This traditionally has placed tremendous pressure on the RF to produce the correct solution among its top few peaks. A wider sampling of the rotational degrees of freedom was implemented in the program package *AMoRe* (Navaza, 1994), where the top peaks in the RF are automatically examined by the TF.

We have recently described the combined molecular-replacement protocol (Tong, 1996), which provides a more general approach for the sampling of the rotational degrees of freedom. The previous MR protocols assume that the correct rotation solution is among or near the top peaks in the RF. In

contrast, the combined MR protocol makes the more general assumption that the correct rotation solution will produce a relatively high RF value irrespective of whether it corresponds to or lies near a peak in the RF. In other words, the combined MR protocol will examine all rotation angles by the TF that have RF values above a specified threshold. The combined MR is therefore a limited six-dimensional search protocol, with the sampling of the rotational degrees of freedom guided by the RF values. Such a protocol will be more expensive computationally than traditional protocols, but this is no longer a major problem owing to the power of modern computers.

Wider sampling of the rotational degrees of freedom has also been proposed by other researchers over the past few years (Sheriff *et al.*, 1999; Urzhumtsev & Podjarny, 1995). More recently, stochastic approaches have been implemented to solve structures by the MR method (Chang & Lewis, 1997; Glykos & Kokkinidis, 2000; Kissinger *et al.*, 1999), which represent an independent way of tackling this six-dimensional problem.

The combined MR protocol was originally implemented as a special feature of the *Replace* package of MR programs (Tong, 1993, 1996). This protocol has been used to solve several difficult MR problems (Maenaka *et al.*, 1999; Tong, 1996; Wu *et al.*, 1997). While this earlier implementation is functional, the speed performance and the user-friendliness were not optimal. We have now implemented the combined MR protocol in a new program called *COMO*. It incorporates new algorithms that improves significantly the speed performance of the protocol, supports many new features of MR calculation and improves the user-friendliness of the MR structure-determination process.

2. Description of the program

The program is written in standard Fortran and supports keyword-based free-formatted user inputs. It has been tested under SGI Irix6.3 and 6.5 and Linux Redhat7.0. It is available on request from the corresponding author. The program documentation and example input files are available on the web at http://como.bio.columbia.edu/tong. The program and the documentation are under continuous development, both for implementation of new features and for bug fixes.

Many of the input commands have parallels in the earlier *Replace* package (Tong, 1993), which should facilitate the use of this new program. Reasonable default values have been provided for the input variables of the program. These default values are generally suitable for most of the MR calculations. Moreover, some of the variables have 'smart' defaults that depend on the particulars of the search model. For example, the radius of integration in the rotation-function calculation, as well as the size of the triclinic unit cell for generating the model structure factors, can be automatically assigned by the program based on the size of the search model. Therefore, the user only needs to define the crystal space group and unit-cell parameters and the names of the files that contain the reflection data and atomic coordinates of the search model. The program can automatically carry out the RF and TF calculations to determine the MR solution, including cases where there is more than one copy of the search model in the crystallographic asymmetric unit.

The program calculations generally proceed in the following steps.

(i) Define the crystal information.

(ii) Define the model information.

(iii) Calculate RF, if necessary.

(iv) Select rotation angles (grid points in the RF map) based on the RF information and perform combined MR search.

(v) Select the solution to the MR problem.

(vi) Repeat to search for another copy of the same molecule in the asymmetric unit (go back to step iv) or to search for a different molecule (back to step ii), if necessary.

To reflect these steps in the calculations, the input commands to *COMO* have been grouped into blocks. Each input block defines the parameters that are important for one of the steps of the calculation. Therefore, the program currently supports the following input blocks: crystal, model, rotation, translation, solution and structure-factor calculation. Each of these input blocks is described in some detail below.

3. The crystal block

This block of input commands defines the space-group symmetry, the unit-cell parameters and the reflection data for the crystal. The program can directly read the reflection file from common data-reduction programs such as *HKL* (Otwinowski & Minor, 1997). It can also read the reflection data files for structure-refinement programs such as *SHELXL* (Sheldrick, 1990) and *CNS* (Brunger *et al.*, 1998). It currently does not support the MTZ format.

This block also defines the resolution range within which the reflection data will be saved in the program. Subsequent RF and TF calculations can define their individual resolution ranges, but these should be subsets of the range specified here. By default, the resolution range defined here will be used in both the RF and TF calculations.

An example input to this block is given in Fig. 1. In this and future examples, the required portion of each command is given in upper case. Additional letters for the command are included for easy readability of the input file.

```
CRystal
   CEll 112.1 112.1 361.5 90 90 120  ! cell parameters
   SYmmetry p6522                    ! space group symmetry
   FOBs-file mydata.sca              ! reflection file name
   FORmat hkl                        ! reflection file format
   REsolution 10 3.5                 ! resolution range
   ENd                               ! end of input to block
```

Figure 1
An example input to the crystal block.

4. The model block

This block of input commands defines the atomic models to the program. The input coordinate files must be in the Protein Data Bank (PDB) format. The program can modify the

temperature factors, residue numbers, chain name, orientation and position of the model as it is read. The occupancy parameters of the atoms are not modified by the program. Segments with zero occupancy are ignored during the MR calculations, but their coordinates are output together with the solution. This feature can be used, for example, to handle regions where large conformational differences are expected between the model and the actual structure (for example flexible loops and regions of low sequence homology).

Similar to the *Replace* package, the atomic models in the program are assigned different flags to define their different status in the MR calculation. Five flags are supported in the *COMO* program and they are T, R, C, U and X, given in decreasing level of information that is available for the model (Table 1). Only the T, R and X flags are available in the *Replace* package. Models with the T flag are solutions to the MR problem with known rotational and translational parameters and they will contribute to the search for additional molecules. Most search models will carry the U flag, whereby the program will determine both the rotational and translational parameters. In cases where the orientation of the model has been determined beforehand, for example from a different program, it can be assigned the flag R. The program will only search for the translational parameters for models with the R flag.

The *COMO* program also supports new features for the handling of the atomic models. In cases where the crystal contains more than one copy of the search model in the asymmetric unit, the program can attempt to automatically locate all the copies, one after another. Another important feature of the *COMO* program is that it can automatically carry out MR calculations with a family of atomic models. The program will select the best result among all the models as the solution. For example, this family of models can be Fab molecules with different elbow angles, or protein kinases with the small lobe showing different degrees of opening, or a collection of homologous structures. With the current implementation, the program does not automatically support the independent sampling of different conformers of an NMR model in a single PDB file. Experiences show that such independent sampling is generally inferior to using the entire assembly (Chen *et al.*, 2000).

An example input to this block is given in Fig. 2.

```
MOdel
   BOverall 35                ! overall B value
   INput model.pdb            ! input model file name
   CEnter                     ! center model at (0, 0, 0)
   ROtate -30 -60 -30 euler   ! rotate model
   COpies 6                   ! number of copies in ASU
   FLag u                     ! model flag
   ENd
```

Figure 2
An example input to the model block.

5. The RF block

This block of input commands defines the parameters that are needed for the RF calculation with the fast RF (Crowther, 1972). RF calculations are performed only when a search model with flag U is used in the MR. The unique region of the rotation space can be automatically assigned by the program (Rao *et al.*, 1980). The grid intervals for the fast RF calculation are 3° along θ_2 and 2.8125° (360/128) along θ_1 and θ_3 for Euler angles. The program can automatically resample the RF map and change to coarser grids of 3.6 or 5° intervals along the three angles. Using a coarser grid can significantly reduce the number of rotation angles that need to be searched by the TF. This accelerates the execution of the program, although at the risk of missing the correct solution owing to potentially larger angular errors. In the structure determination of an SH2 domain, a full six-dimensional search using 5° grid intervals for rotation angles successfully found the correct MR solution after rigid-body refinement of the TF results (Hoedemaeker *et al.*, 1999).

To improve the quality of the RF results, RF maps calculated using different resolution ranges and radii of integration are often compared. Peaks that consistently appear in the different RF maps are more likely to be correct. This was normally performed by manual comparisons of the different RF results. In the *COMO* program, the sampling of two resolution ranges and three different radii can be performed automatically. With this new algorithm, the sampling of the different radii at a given resolution range is carried out by combining the appropriate a_{lmn} coefficients in the fast RF (Crowther, 1972). For sampling of the different resolution ranges, the two RF maps are calculated first and brought to the same scale. The average of the RF values is subtracted from all the grid points and the RF map values are then scaled such that the standard deviation is 100. The two RF maps are then combined by taking either the minimum or the average values.

An example input to this block is given in Fig. 3.

```
RF
   CUtoff 0.25 0.5            ! large term cut-off
   NOrmalize true 10          ! normalize structure factors
   ORigin true                ! remove Patterson origin
   SAngle euler               ! search in Euler angles
   REsolution 10 3.5 10 4     ! sample 2 resolution ranges
   RAdius 30 5                ! sample 3 radii (25, 30, 35)
   SAve 10                    ! identify top 10 peaks in the RF
   ENd
```

Figure 3
An example input to the RF block.

6. The TF block

This input block defines the parameters for TF calculations, including resolution range, large-term cutoff, packing check criterion, rotation-angle selection criterion and solution selection instruction. Most of these input parameters have reasonable default values. The criterion for the selection of rotation angles, however, will generally depend on the individual search calculation, especially the quality of the search model.

The combined MR-search protocol that is used by the program is as follows (Tong, 1996). A set of rotation angles is selected and then sorted based on their RF values. For each rotation angle, the program will first calculate the TF based on

Table 1
Supported atomic model flags.

Flag	Description
T	A model with known rotational and translational parameters. This model will be kept stationary during the MR calculation.
R	A model with known orientation but unknown position. The translational parameters will be determined during the MR calculation.
C	A model with a pre-calculated RF map. The rotational and translational parameters will be determined during the MR calculation.
U	A model with unknown orientation and position. Both the rotational and translational parameters will be determined during the MR calculation.
X	A model that will be ignored by the program.

Patterson correlation (Harada *et al.*, 1981; Tong, 1993) by fast Fourier transform (FFT). If phase information is available for the reflections, the phased translation function can also be used (Read & Schierbeek, 1988). In cases where there is more than one molecule in the asymmetric unit, the phase information can be based on those molecules that have already been correctly placed into the unit cell. The unique region of the TF search (the Cheshire group) is assigned automatically by the program (Hirshfeld, 1968). The top peaks (up to 200) in the TF are identified and a packing check is performed for each of these peaks (see below for more details on packing check). A correlation coefficient (CC) and *R* factor (both based on structure-factor amplitudes) are calculated for those peaks that do not have packing problems in the unit cell. The results from the searches for all the rotation angles are sorted based on the CC or the *R* factor.

The rotation angles for the combined MR calculation are selected from the grid points in the RF map that have values within specified selection criteria. In the *COMO* program, the criteria are defined as fractions of the highest value in the RF map. The optimal values for the selection criteria depend on the specific parameters of individual MR searches and are generally difficult to predict beforehand. A reasonable practice is to err on the side of caution, selecting many more rotation angles than necessary and terminating the search (manually) when the correct solution has been found (see below). In cases where there are several copies of the search model in the asymmetric unit, the selection criteria for searching for the different copies can be different, for example with those for the later copies having lower RF values.

A new algorithm is implemented in the *COMO* program for the handling of the rotation angles during the combined MR calculation. At the beginning of the calculation, the rotation angles are sorted based on their RF values. Once a rotation is found to produce the best solution (highest CC or lowest *R*), the program will immediately examine all the neighboring grid points of this rotation that have been selected for the calculation. This provides a way of automatically following a correct solution to its optimal rotational and translational parameters and should lead to faster identification of the correct MR solution (see the test case reported below).

In the original implementation of the combined MR protocol in the *Replace* package (Tong, 1996), the algorithm used for the TF calculation minimized the memory usage at the expense of CPU performance. This was the appropriate protocol as the *Replace* package was originally intended to examine only one rotation angle by the TF in each calculation (Tong, 1993). In the current program, a new algorithm is implemented that optimizes the CPU performance, with the use of additional memory, in the TF calculation. Moreover, the program supports two separate large-term cutoff values (Tollin, 1966), which enables the use of a higher cutoff value in the Patterson correlation TF calculations. The large terms are selected based on a cutoff value (Δ). A reflection is selected as a large term if its squared amplitude (I_h) satisfies the condition $I_h \geq \Delta \times \langle I_h \rangle$, where the average squared amplitude $\langle I_h \rangle$ is calculated in spherical shells of equal reciprocal-space volume. As the Patterson correlation TF is based on the squared amplitudes, a higher cutoff value can generally be used to select only the strongest reflections in each resolution shell. Extensive experiences suggest that accurate TF results can be obtained with a cutoff value of 2. This selects only about 10% of the observed reflections, leading to significant acceleration of the computation. On the other hand, the CC and *R* factors are based on amplitudes and the program uses a second, lower cutoff value for their calculation. The default for this value is 0.5, which usually leads to the selection of about 50% of the observed reflections. Therefore, the combined MR protocol generally ignores the contribution of the weaker reflections, which may also help improve the signal-to-noise ratio in the MR calculation.

The combined MR protocol often needs to examine thousands of rotation angles by the TF. Even with modern computers, it still can take a significant amount of time to complete this calculation. However, once the correct MR solution is found, the TF searches with additional rotation angles become unnecessary. For this purpose, the program periodically (after searching every 10% or 100 rotation angles) outputs the current best five MR solutions from the calculation. If this list suggests that the correct solution has been found, the program can be instructed to terminate the search for the current molecule and continue with the next step of the calculation.

The correct MR solution can be recognized by its high CC and/or low *R* value, as well as reasonable packing in the unit cell. In addition, the combined MR protocol often gives rise to clusters of MR solutions with similar rotational and translational parameters, since neighboring grid points in the RF map are often selected by the given criteria. Our experiences have shown that the correct MR solution is often part of a large cluster of solutions. This represents a detailed sampling of the rotational and translational parameters near the correct MR solution, which should further increase the signal in the search.

Once the correct solution is identified from the MR search, rigid-body optimization of the solution can be carried out by the program. This will improve the rotational and translational

parameters of the search model, which will help the search for additional molecules.

An example input to this block is given in Fig. 4.

```
TF
   CUtoff 2 0.5        ! large term cut-off
   REsolution 8 4      ! resolution ranges
   DCutoff 2.5 2       ! packing check criterion
   RIgid true          ! rigid-body optimization
   SElect 0.5 0.6      ! RF selection for copy 1
   SElect 0.3 0.5      ! RF selection for copy 2
   SElect 0.2 0.4      ! RF selection for copy 3
   SElect 0.1 0.3      ! RF selection for copy 4
   SElect 0.1 0.3      ! RF selection for copy 5
   SElect 0.0 0.2      ! RF selection for copy 6
   SAve 200            ! examine top 200 peaks in TF
   ENd
```

Figure 4
An example input to the TF block.

7. The solution block

This input block instructs the program to output the current MR solution, corresponding to all molecules with the T flag. In addition, a final packing check will be performed for this solution.

To calculate structure factors based on this solution (or an input atomic model), the SF input block can be used to set up parameters for this calculation.

8. Packing check

Checks on the packing of the atomic model in the crystal are an important part of a structure determination by the MR method. A set of rotational and translational parameters that cause serious steric clashes of the search model in the crystal cannot be a correct MR solution, but such packing arrangements sometimes produce high CC values. Such spurious solutions must be ignored in selecting the correct MR solution. This is often performed by manual examination of the packing with a graphics program. An automated packing-check procedure is crucial for the combined MR protocol, as thousands of rotation angles and tens of thousands of possible MR solutions can be examined in each calculation.

Packing functions have been proposed that calculate the overlap among crystallographically related molecules (Harada *et al.*, 1981; Hendrickson & Ward, 1976). Such functions could be used as weighting factors for the Patterson correlation TF (Harada *et al.*, 1981) to reduce the TF values of those positions that have serious overlap of the search molecules. However, these functions only offer a global examination of the packing in the crystal, while a more detailed analysis is needed before potential MR solutions can be automatically discarded for packing problems.

For this detailed packing analysis, the procedure used in the current program is based on counting the number of close C^{α}–C^{α} contacts among the protein molecules in the crystal (Tong, 1993). The distance cutoff for this packing check can be set at 2–3 Å. Only those rotational and translational parameters that have a small number of crystal packing clashes are saved as potential MR solutions. For nucleic acid structures, the P, C4′, N1 and C4 (for A, U and T bases) or N9 (for G and C bases) atoms of each nucleotide can be selected for the packing check.

9. The graphical user interface

To further improve the user-friendliness of the program, we have developed a graphical user interface (GUI) to the program. The GUI is implemented in standard Tcl/TK version 8.0 and has been tested under SGI Irix6.3 and Linux Redhat7.0.

The interface can be configured in three modes. The 'novice' mode requires only minimal user inputs, including space-group symmetry, unit-cell parameters, reflection data and search-model file names, resolution range and the number of copies of the search model in the asymmetric unit (Fig. 5). The GUI can then generate the entire input file to the program, taking advantage of the default values that have been set up in the program. The 'standard' and the 'expert' modes are set up for finer control of the program execution and allow configuration of all available commands in the program. Alternatively, a text editor can be used to modify the input file generated from the 'novice' mode.

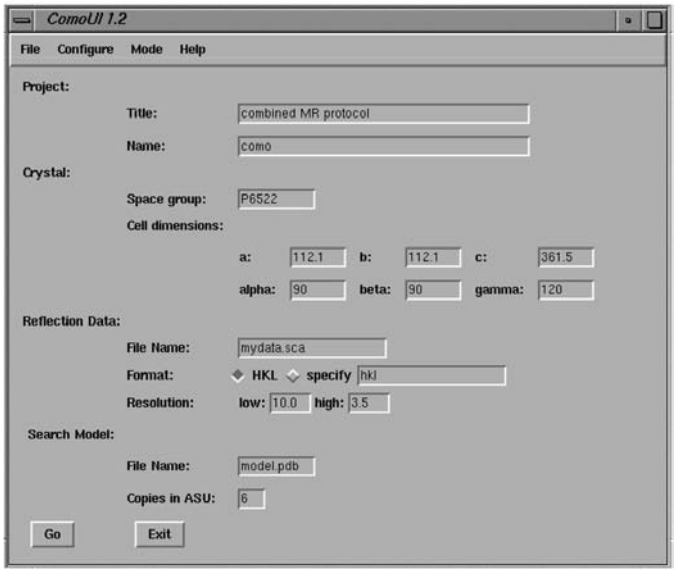

Figure 5
The 'novice' mode of the graphical user interface (GUI) to the *COMO* program. Once the 'Go' button is clicked, the GUI will produce the complete input file and launch the program with this input.

The GUI can also display the status of the combined MR calculation. If the correct solution has been found, the GUI can be used to instruct the program to terminate the search for the current molecule.

Table 2
Summary of structure solution of TIR domain.

Mol.	Rotation angles	CPU (min)	% solutions rejected	Rotation† (°) θ_1	θ_2	θ_3	Translation x	y	z	CC	R	Cont‡	TF§	RF¶	CC†† (refined)
1	Selection: 0.5–0.6	77	93.6	28.1	60.0	30.9	0.533	0.589	0.120	20.5	46.9	0	1	69	21.4
	No. selected: 1146			30.9	60.0	28.1	0.528	0.583	0.120	20.5	47.4	0	1	77	
	No. examined: 200			30.9	57.0	30.9	0.533	0.589	0.120	19.9	47.2	0	1	67	
				33.8	60.0	28.1	0.528	0.583	0.120	19.8	48.0	0	1	71	
2	Selection: 0.3–0.5	86	94.7	11.2	54.0	278.4	0.639	0.800	0.224	26.1	45.5	0	1	89	27.2
	No. selected: 7056			11.2	51.0	281.2	0.639	0.800	0.222	26.1	45.6	0	1	91	
	No. examined: 200			14.1	48.0	278.4	0.639	0.800	0.224	25.7	45.6	0	1	136	
				8.4	51.0	284.1	0.639	0.794	0.224	25.5	45.5	0	1	122	
3	Selection: 0.2–0.4	340	96.6	47.8	45.0	118.1	0.556	0.311	0.157	34.9	43.1	0	1	593	35.6
	No. selected: 12260			50.6	48.0	118.1	0.556	0.311	0.157	34.6	43.0	0	1	588	
	No. examined: 700			50.6	42.0	115.3	0.556	0.311	0.157	34.3	34.1	0	1	583	
				47.8	48.0	118.1	0.556	0.311	0.157	43.2	43.2	0	1	594	
4	Selection: 0.1–0.3	1000	98.8	11.2	54.0	22.5	0.567	0.800	0.033	36.8	42.1	0	1	1905	39.9
	No. selected: 19337			8.4	51.0	25.3	0.567	0.800	0.033	36.3	42.2	0	1	1911	
	No. examined: 2100			8.4	54.0	22.5	0.567	0.800	0.033	36.3	42.1	0	1	1912	
				14.1	54.0	22.5	0.567	0.800	0.033	36.0	42.4	0	1	1916	
5	Selection: 0.1–0.3	107	99.2	39.4	48.0	160.3	0.389	0.494	0.270	42.2	40.0	0	1	104	42.3
	No. selected: 19337			42.2	48.0	160.3	0.389	0.494	0.270	41.8	40.0	0	1	115	
	No. examined: 200			42.2	51.0	160.3	0.389	0.494	0.270	41.6	40.1	0	1	116	
				45.0	48.0	157.5	0.389	0.494	0.270	41.6	40.2	0	1	122	

† The top four solutions for each calculation are shown. ‡ Number of close contacts (<2.5 Å) of among C^{α} atoms of different molecules in the unit cell. § The peak number in the Patterson correlation TF map. ¶ The rotation-angle number. †† The CC after rigid-body optimization.

10. Test cases

The program has already been used to solve several difficult MR problems. This includes the structure of human ornithine transcarbamylase (OTCase) in a cubic crystal form, where previous attempts at solving this structure by other programs failed to give the correct solution (Shi *et al.*, 2001). Similarly, the structure of the coenzyme B_{12} binding subunit of glutamate mutase was solved with this new program, which has resisted many earlier attempts with other programs (G. Jogl *et al.*, unpublished results). This crystal, in space group $P2_13$, contains three copies of the molecule in the asymmetric unit; the orientation and position of all three molecules were determined automatically by the program overnight.

The structure solution of a new crystal form of the Toll/interleukin-1 receptor (TIR) domain is presented here in some detail (Xu *et al.*, 2000). The crystal belongs to space group $P6_122$ or $P6_522$, with unit-cell parameters $a = b = 112.1$, $c = 361.5$ Å. The X-ray diffraction data were collected at the X4A beamline of Brookhaven National Laboratory. The atomic coordinates of this domain in a different crystal form was used as the search model (Xu *et al.*, 2000). The earlier crystal has very high solvent content, with a V_M of 5.4 Å^3 Da^{-1}. The unit-cell volume of the new crystal is 3.4 times that of the earlier crystal, which was in space group $P6_222$, suggesting that the new crystal may have three to seven molecules in the asymmetric unit, depending on the actual V_M value. Several attempts at solving this structure using *AMoRe* (Navaza, 1994) or *X-PLOR* (Brünger, 1992) were not successful.

All the calculations were performed on a PC with an AMD Athlon 1 GHz CPU, 512 MB memory and running Linux 7.0. The first step in the structure determination is to define which enantiomorph is the correct space group. A combined MR search is carried out to look for the first molecule in both space groups. The search atomic model contained 130 amino-acid residues. A clear solution (best CC = 20.5) was found for space group $P6_522$, after examining about 1100 rotation angles (266 min CPU). The rotational parameters of the solution have an RF value that is only about 60% of the highest value in the RF map. The same calculation in space group $P6_122$ showed no clear solution (best CC = 16.6). This suggests that the correct space group is $P6_522$.

With the knowledge of the correct space group, a new combined MR search was set up to look for six copies of the search model in the asymmetric unit. The program inputs for this search are shown above as the examples for each of the program blocks. The correct solutions produced significant changes in the CC and the R factor (Table 2). Moreover, the correct MR solution is invariably the top peak in the TF map. In addition, many neighboring grid points of the correct rotation angle in the RF map can also produce the correct TF solution, leading to clustering of the solutions in the program output. This is another indication that the solution obtained is likely to be correct. It took a total of about 27 h of CPU time to locate the first five molecules (Table 2). About 3500 rotation angles were examined for this calculation, or roughly about 0.5 min for examining each rotation angle by TF. For all five molecules, once the correct solution was identified the program was instructed to move on to the search for the next molecule. (The program periodically checks for the presence of a scratch file and will terminate the search for the current molecule if this file exists.) The search for the sixth molecule did not produce any clear solution, even after screening more

than 10 000 rotation angles. It is likely that the asymmetric unit contains only five copies of this TIR domain. This was later confirmed from a selenomethionyl anomalous difference electron-density map.

The calculations showed that the packing check removed most of the possible MR solutions, especially after several of the molecules have already been placed in the asymmetric unit. More than 99% of the possible solutions were rejected when the fifth molecule was located (Table 2). Surprisingly, more than 93% of the possible solutions were rejected for locating the first molecule, even though the solvent content at this time is very high. This indicates the power and the necessity of the packing check. Rigid-body optimization after the MR calculation can introduce significant improvements in the CC. For example, a 3% increase in the CC was obtained by rigid-body optimization of molecule 4 (Table 2), which caused shifts of 6° in θ_2 and θ_3.

The fourth and fifth molecules were found with the same collection of rotation angles (selection criterion 0.1–0.3). At the beginning of the search for each molecule, the rotation angles are sorted based on their RF values. In this regard, it is interesting to note that about 2100 rotation angles were examined to find the fourth molecule, whereas only 200 were examined to find the fifth molecule. This illustrates another important feature of the program. If a rotation angle produces the best solution in the search so far, the program will immediately examine the neighbors of this rotation angle in the collection. This provides a way for the program to automatically follow a possible solution to determine the rotational and translational parameters at its optimum. For the search for molecule 5, the second rotation in the collection has about 26° error in θ_1 and 3° error in θ_2. However, this was enough to produce a high CC value. By following this rotation angle, the program came upon the correct rotation angle, even though it was the 11 325th entry in the rotation-angle collection at the beginning of the search. When searching for the fourth molecule, this second angle was also examined. However, as there was not as much information known about the structure, this rotation did not produce a high CC value owing to its large angular errors. Therefore, this protocol of automatically following the neighbors of a rotation angle that produces the best CC value appears to be quite powerful in revealing the correct MR solution.

It should be noted that this is a rather large test case. The TF calculations used reflections between 8 and 4 Å resolution. Even at this resolution, the Patterson correlation TF required a Fourier transform of 180 × 180 × 540 grids, owing to the large unit-cell parameters. Therefore, a significant amount of the CPU time was used to calculate this Patterson correlation TF transform and the CPU time required to examine each rotation angle will be shorter for more typical MR cases. Moreover, the top 200 peaks in the Patterson correlation TF were examined for each rotation angle. At least for this test case, this appears to be unnecessary as the correct solution is invariably the top peak in the TF (Table 2). Examining a smaller number of TF peaks will also reduce the CPU time requirement.

The grid interval in the RF map is 2.8125° along θ_1 and θ_3, and 3° along θ_2 in the calculation described above. To test the effect of using a coarser grid, a grid interval of 3.6° is used for all three angles. As expected, this reduced the number of rotation angles by a factor of two for the search of each copy of the molecule. Somewhat surprisingly, the amount of CPU time needed to locate all five copies is not reduced significantly, as similar number of rotation angles were examined for locating each copy as with the finer grid. Moreover, the clustering of the solutions is not as obvious with this coarser grid and the correct solution often is no longer the top peak in the TF. Therefore, it appears that at least for this structure determination using a 3.6° rotation grid is detrimental to the MR calculation.

The combined MR protocol represents a more general approach for sampling the rotational and translational degrees of freedom in a structure determination by the MR method. This should be a more powerful way of resolving this six-dimensional problem, as illustrated by the number of difficult MR structures that have been solved over the past few years. The *COMO* program is an efficient and user-friendly implementation of the combined MR protocol and should facilitate structure determination by the MR method.

We thank H. Wu for many useful discussions, D. Shi for the OTCase test case and a grant from the National Science Foundation (DBI-98-76668) for financial support.

References

Brünger, A. T. (1992). *The X-PLOR Manual*. Yale University, New Haven, Connecticut, USA.
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.
Chang, G. & Lewis, M. (1997). *Acta Cryst.* D**53**, 279–289.
Chen, Y. W., Dodson, E. J. & Kleywegt, G. J. (2000). *Structure Fold. Des.* **8**, R213–R220.
Crowther, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 173–178. New York: Gordon & Breach.
Glykos, N. M. & Kokkinidis, M. (2000). *Acta Cryst.* D**56**, 169–174.
Harada, Y., Lifchitz, A. & Berthou, J. (1981). *Acta Cryst.* A**37**, 398–406.
Hendrickson, W. A. & Ward, K. B. (1976). *Acta Cryst.* A**32**, 778–780.
Hirshfeld, F. L. (1968). *Acta Cryst.* A**24**, 301–311.
Hoedemaeker, F. J., Siegal, G., Roe, S. M., Driscoll, P. C. & Abrahams, J. P. (1999). *J. Mol. Biol.* **292**, 763–770.
Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.
Maenaka, K., Juji, T., Stuart, D. I. & Jones, E. Y. (1999). *Structure*, **7**, 391–398.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Rao, S. N., Jih, J. H. & Hartsuck, J. A. (1980). *Acta Cryst.* A**36**, 878–884.
Read, R. J. & Schierbeek, A. J. (1988). *J. Appl. Cryst.* **21**, 490–495.
Rossmann, M. G. (1972). Editor. *The Molecular Replacement Method.* New York: Gordon & Breach.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Sheldrick, G. M. (1990). *Acta Cryst.* A**46**, 467–473.

Sheriff, S., Klei, H. E. & Davis, M. E. (1999). *J. Appl. Cryst.* **32**, 98–101.
Shi, D., Morizono, H., Yu, X., Tong, L., Allewell, N. M. & Tuchman, M. (2001). *Acta Cryst.* D**57**, 719–721.
Tollin, P. (1966). *Acta Cryst.* **21**, 613–614.
Tong, L. (1993). *J. Appl. Cryst.* **26**, 748–751.
Tong, L. (1996). *Acta Cryst.* A**52**, 782–784.
Urzhumtsev, A. & Podjarny, A. (1995). *Acta Cryst.* D**51**, 888–895.
Wu, H., Kwong, P. D. & Hendrickson, W. A. (1997). *Nature (London)*, **387**, 527–530.
Xu, Y., Tao, X., Shen, B., Horng, T., Medzhitov, R., Manley, J. L. & Tong, L. (2000). *Nature (London)*, **408**, 111–115.