Model preparation in *MOLREP* and examples of model improvement using X-ray data

Andrey A. Lebedev,* Alexei A. Vagin and Garib N. Murshudov

Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York YO10 5YW, England

Correspondence e-mail: lebedev@ysbl.york.ac.uk

The success of molecular replacement is critically dependent on the quality of the search model. Several model-preparation procedures are integrated in the molecular-replacement program *MOLREP*. These include model modification on the basis of amino-acid sequence alignment and model correction based on analysis of the solvent-accessibility of the atoms. The packing function used in *MOLREP* for the translational search is explained in the context of model preparation. In difficult cases, bioinformatics-based modifications are not sufficient for successful molecular replacement. An approach implemented in *MOLREP* for solving cases with translational noncrystallographic symmetry is an example of model preparation in which analysis of X-ray data plays an essential role. In addition, two examples are presented in which the X-ray data were used to refine partial models for subsequent use in molecular replacement.

1. Introduction

Molecular replacement (MR) is very sensitive to dissimilarity between the search model and the target protein. Therefore, despite the limited number of possible protein folds and the large number of already solved structures, the generation of a suitable search model remains a nontrivial task. This typically includes searching for homologous structures in the PDB (Berman *et al.*, 2002), their analysis and modification. The information important for MR includes the presence and conservation of oligomeric state(s) and domain structure in the family of homologous proteins. Data on oligomers can be obtained from PISA (Krissinel & Henrick, 2005) and domain descriptions are given, for example, by SCOP (Murzin *et al.*, 1995), which is linked to the PDB web resources. This information, combined with analysis of the self-rotation function (SRF), Patterson map, unit-cell parameters and symmetry of the crystal, allows the generation of a search model or a series of search models, including oligomers, monomers or domains. There is a wide range of tools available for modification of the selected model(s). A frequently used modification is the removal of those residues and atoms in the homologous model that have no match in the target sequence (Schwarzenbacher *et al.*, 2004). A three-dimensional superposition of homologous structures using, for example, *SSM* superposition (Krissinel & Henrick, 2004) integrated in *Coot* (Emsley & Cowtan, 2004) allows the identification of polypeptide segments that are variable within the given family of proteins. The removal of such segments from the search model can often prove critical for the MR search. More extensive

modifications sometimes help to solve difficult MR cases. These include homology modelling (Schwede *et al.*, 2003; Fiser & Sali, 2003) and scanning the possible conformations of the unknown protein using normal-mode analysis (Suhre & Sanejouand, 2004).

MOLREP is an automated program for MR (Vagin & Teplyakov, 1997) where, along with the default protocol, there are various search strategies as options. The program is part of the *CCP*4 (Collaborative Computational Project, Number 4, 1994) package. *AMoRe* (Navaza, 1994), *Phaser* (Read, 2001), the MR implementation in *CNS* (Brünger *et al.*, 1998) and *MOLREP* together cover more than 95% of structures solved by MR. A special feature of *MOLREP* is that it offers several built-in model-preparation functionalities. The integration of model-preparation and Patterson function techniques in one program has several advantages. Apart from convenience, such integration allows specific adjustment of the model-modification parameters for an efficient Patterson search. Moreover, the weighting parameters for the rotation (RF) and translation (TF) functions are more reliable if they are derived from the original sequence and atomic coordinates of the homologous protein. Such an integrated approach has proven to be efficient and has recently been implemented in several MR pipelines including *BALBES* (Long *et al.*, 2008), *MrBUMP* (Keegan & Winn, 2008) and *JSCG* (Schwarzenbacher *et al.*, 2008), in which *MOLREP* itself is used as a component. Currently, *MOLREP* is being updated to fulfill the requirements raised by *BALBES* development and benefits from its training.

In this paper, the default scheme of model preparation in *MOLREP* is presented. In addition, two examples of model modification using X-ray data are discussed that might be useful in solving complicated problems and might be relevant in designing automatic structure-solution pipelines.

2. Model preparation in *MOLREP*

This section describes the modifications of the search model that were shown to be quite efficient and were therefore included in the default model-preparation scheme of *MOLREP*.

2.1. Correction of the model using sequence alignment

In this subsection, it is assumed that the user has provided an input file containing the sequence of the target protein. In this case, the first stage of model preparation in *MOLREP* includes alignment of the sequence derived from the search model to the target sequence and, provisionally, deletion of residues and atoms of the search model that do not map onto the target sequence. This is a conservative approach to model correction, as no new atoms are added and the coordinates of preserved atoms are not changed. Besides, the sequence identity derived from the alignment of the two sequences is further used for weighting of the X-ray data (see §2.4).

The sequence alignment implemented in *MOLREP* is a modified version of the dynamic alignment algorithm (Needleman & Wunsch, 1970), which takes into account the known three-dimensional structure of the search model. Thus, buried residues contribute to the total alignment score more than residues at the surface. In addition, it is assumed that gaps and insertions are impossible within sequence segments corresponding to helices and strands. Based on this sequence alignment and provided that sequence identity is higher than 20%, the search model is modified as follows. Firstly, residues that align with gaps in the target sequence are deleted. Secondly, the atoms in the aligned pairs of residues are analysed and atoms in the search model that have no counterpart in the target are deleted. For example, if Val in the target sequence corresponds to Leu in the search model, then N, CA, C, O, CB and CG of Leu are kept in the model and CD1 and CD2 are removed. Finally, the residues and atoms of the modified search model are renamed and renumbered. In the above example, CG is renamed CG1, Leu renamed Val and the residue number of Leu in the model is changed to the number of Val in the target sequence.
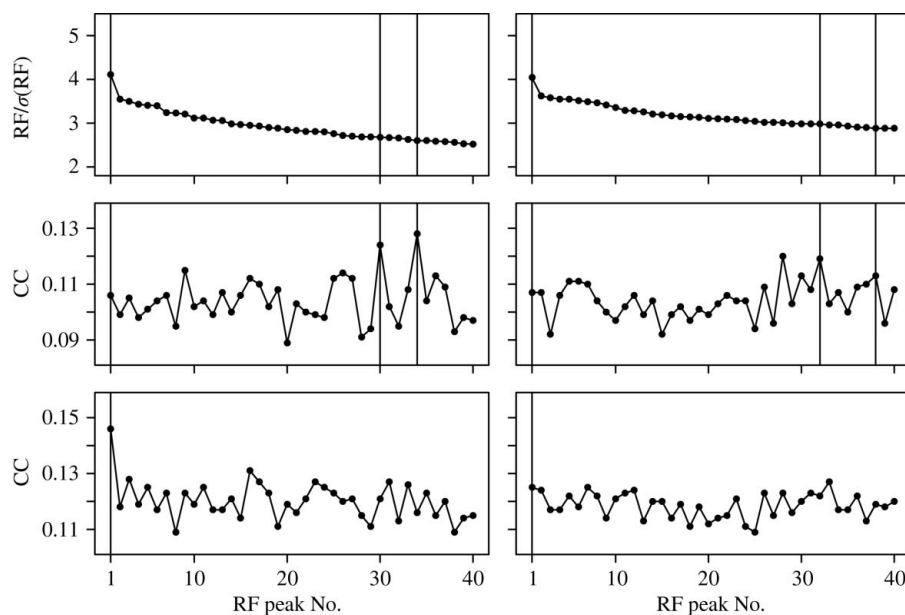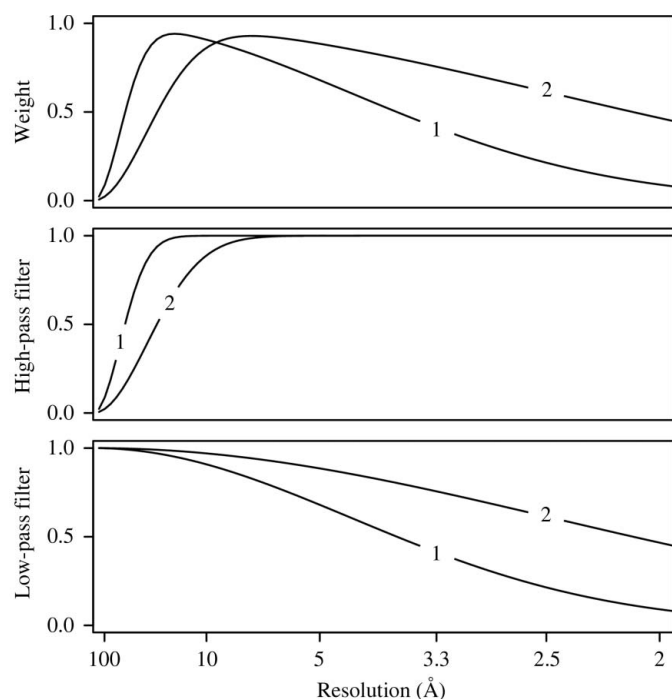

Figure 1
The effect of model correction according to the target sequence on MR demonstrated using the crystal structure of hydrolase from *Ophiostoma novo-ulmi*. Top, RF; middle, TF without fixed partial model; bottom, TF with fixed partial model containing one of two monomers. Two search models are compared: left, corrected according to the target sequence; right, taken from the homologous structure without modifications. RF and TF steps are represented by plots of RF/σ(RF) and by the correlation coefficient (CC), respectively, *versus* the RF peak number. The vertical thin lines indicate RF peaks corresponding to the correct orientations for a given search. There are two close RF peaks consistent with the orientation of the first found monomer.

As a result, if MR is successful then the residue numbering in the solution corresponds to the target sequence.

Tests carried out with difficult MR problems show that the correction of a search model at a low level of sequence identity is not very reliable and is unlikely to increase the probability of finding a solution. In addition, such a correction will result in the removal of too many buried atoms and the resultant sparse model will not be treated properly at the surface-modification stage (see §2.3). Therefore, if the sequence identity between the model and the target is less than 20%, then, by default, the model is kept intact. Nevertheless, even if the model has not been modified, the value of the sequence identity is used in defining the weighting scheme for the RF and TF.

In the following example, the correction of the search model according to the sequence alignment was critical for structure determination by MR. The crystal structure of hydrolase from *Ophiostoma novo-ulmi* (Isupov *et al.*, 2004) contains two molecules per asymmetric unit (AU). It was solved using a search model derived from the PDB entry 1ci9, which had a sequence identity to the target of 24.5%. We analysed the behaviour of the TF with two search models: a monomer from the above homologous structure without modification and this monomer modified according to the sequence of the target

protein (Fig. 1). In both TF runs with and without the fixed model containing one monomer in the correct position the unmodified search model gave no contrast between correct and incorrect peaks, whereas the TF with modified search model produced the complete structure in two steps by selecting the highest TF peak at each step.

2.2. Packing function

By default, the search model corrected according to the target sequence undergoes further modifications. Two slightly different models are created at this stage. One of these is used in the RF and TF and the other is used in the packing function (PF; Vagin, 1983; Stubbs & Huber, 1991). The details of the TF and PF used in *MOLREP* are described by Vagin & Teplyakov (1997). Here, we give a brief overview of how they work together.

The overlap between two molecules can be defined as a product of their electron densities. The total overlap is a function of the position of the search model, r. It is computed using a single FFT run, but it accounts for all the symmetry-generated contacts. The packing function PF(r) is derived from the total overlap and is an estimate of the likelihood of the molecular packing corresponding to a given r. It takes values from 0 to 1, with PF(r) = 1 if there are no clashes. Therefore, the peaks of the combined translation function, CTF(r) = TF(r) PF(r), corresponding to unlikely crystal packing are down-weighted and excluded from the list of top hits and from further analysis. For the highest CTF(r) peaks, the correlation coefficient based on intensities, CC(r), is computed. This function is considered to be a better indicator than TF(r), albeit computationally more expensive. The CC(r) is also weighted using PF(r) to give Score(r) = PF(r) CC(r). The peak with the highest score is selected as a potential solution.

2.3. Modification by surface accessibility

Usually, residues on the surface of biomolecules are less conserved than those buried inside it. Moreover, it can be expected that atoms that are accessible to the solvent have a higher mobility than the rest. These factors can be taken into account by smearing the electron density of the atoms exposed to solvent, *e.g.* by assigning higher atomic displacement parameters (ADPs) to them. A new ADP is computed for each atom using the equation ADP = $U + VS$, where $U = 15$ Å^2 and $V = 20$ are empirical parameters and S is the accessible surface area of the atom (Lee & Richards, 1971) computed using a fast algorithm. This modification is applied to the model used in the RF and TF. Note that at this stage only the search model is modified and the additional ADPs assigned to the surface atoms do not depend on the experimental data. The Wilson B factor of the data is accounted for later as a parameter of the weighting function.

Conformations are especially variable in the contact areas, thus it is necessary to allow some overlap between neighbouring molecules. Technically, this is achieved by removing exposed atoms from the model used in the PF. This criterion is

Figure 2
Weighting of X-ray data in *MOLREP*. Top, the resolution-dependence of the weights which would be used (1) for a model with 20% identity and radius 20 Å and (2) for a model with 40% identity and radius 10 Å. Each weighting function is a band-pass filter defined as the product of the corresponding (middle) high-pass and (bottom) low-pass filters (Gonzales & Woods, 2002). The high-pass filter is defined by the size of the search model. The low-pass filter is defined by overall B factor (20 Å^2 in the examples) and the sequence identity. The low-pass filter is a constant function in the case of 100% identity. Weights are applied to both the observed and calculated structure amplitudes, which are initially scaled to zero overall B factor.

also based on the estimate of the accessible surface area S. Atoms with $S > 0$ are removed, provided that the modified model would contain more than five residues and more than 10% of the initial amount of atoms.

2.4. Weighting of the X-ray data

Producing a search model is only one aspect of model preparation. Another aspect is to define the weighting scheme that is most suitable for a given model. During model preparation, *MOLREP* estimates a number of parameters for the search model. Two of them, the radius of gyration of the model and its sequence identity with the target protein, are translated into the parameters of Gaussian low-pass and high-pass filters (see, for example, Gonzales & Woods, 2002), which define the resolution-dependence of reflection weights in the RF and TF (Fig. 2). It is assumed that the reflections of low resolution, the intensities of which mostly depend on large-scale details of the crystal and are almost independent of the internal features of the molecules forming it, produce only noise in the RF and TF. High noise and low signal are also produced by high-resolution reflections, defining structural details of finer scale than the r.m.s.d. between the search and target molecules. The exact value of this r.m.s.d. is unknown until the structure is solved, but it correlates with the known sequence identity (Chothia, 1992).

The dependence of the filter parameters on the search-model parameters were adjusted using a test set of difficult MR problems. However, explicit definition of filter parameters by a user may be needed in some cases where, for example, low three-dimensional similarity between the search and the target proteins is in disagreement with high sequence identity between them.

2.5. Model modification in the presence of translational NCS

The treatment of translational NCS, otherwise known as pseudo-translation, is a special case of model modification because it is applied to the TF but not to the RF. In *MOLREP*, this modification is applied implicitly in reciprocal space and therefore can be considered as either model modification or as weighting of the X-ray data. Translational NCS is detected and the NCS translation vector is derived using the experimental Patterson function. *MOLREP* assumes that translational NCS is present if there is a non-origin peak in the Patterson function with a height of 1/8 or more of the origin peak height. An additional requirement is that this peak is sufficiently distant from the origin (three-quarters of the diameter of the search model) to ensure that it is not caused, for example, by regular structural patterns in the target molecule. Such an approach is simple, works in most cases and is therefore used by default. However, neither false positives nor false negatives can be excluded. For example, translational NCS will not be detected
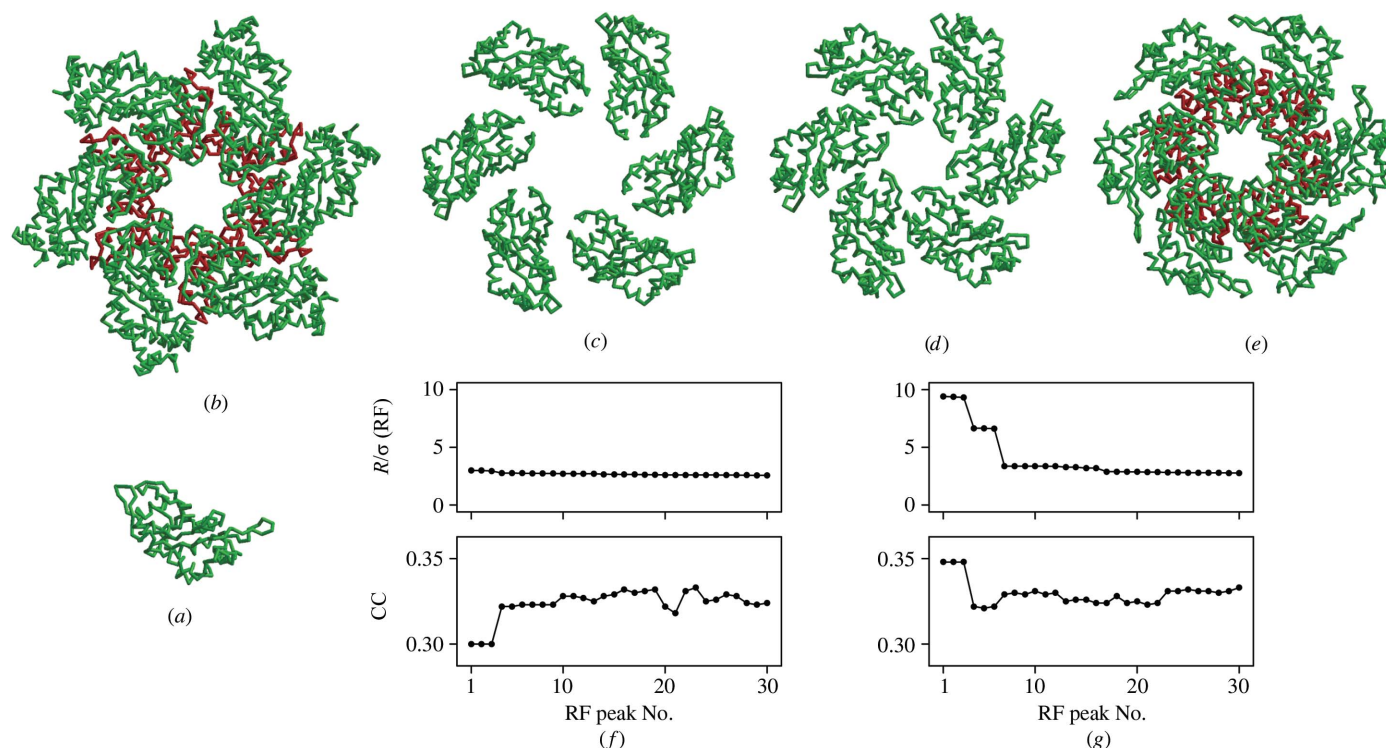
Figure 3
Determination of the crystal structure of BPV-1 E1 helicase. (*a*) The AAA+ domain of HPV-18 helicase and (*b*) the hexamer of SV40 helicase, which were used to generate a synthetic model (*c*). (*d*) Refined hexamer. (*e*) Final structure of BPV-1 E1 helicase. Colours indicate (red) oligomerization and (green) AAA+ domains. The r.m.s.d.s for C$^{\alpha}$ atoms between the last three models were 5.6 Å (synthetic and final models), 4.5 Å (synthetic and refined models) and 2.5 Å (refined and final models). In the final hexamers the sixfold symmetry is significantly perturbed. Therefore, the r.m.s.d. between the refined and symmetrized final hexamers is only 1.2 Å. (*f*) The behaviour of MR for synthetic and (*g*) refined hexamers. RF and TF steps are represented by plots of RF/σ(RF) and by correlation coefficient (CC), respectively, *versus* RF peak number.

Phe A162

Pro A159

(a)   (b)   (c)   (d)
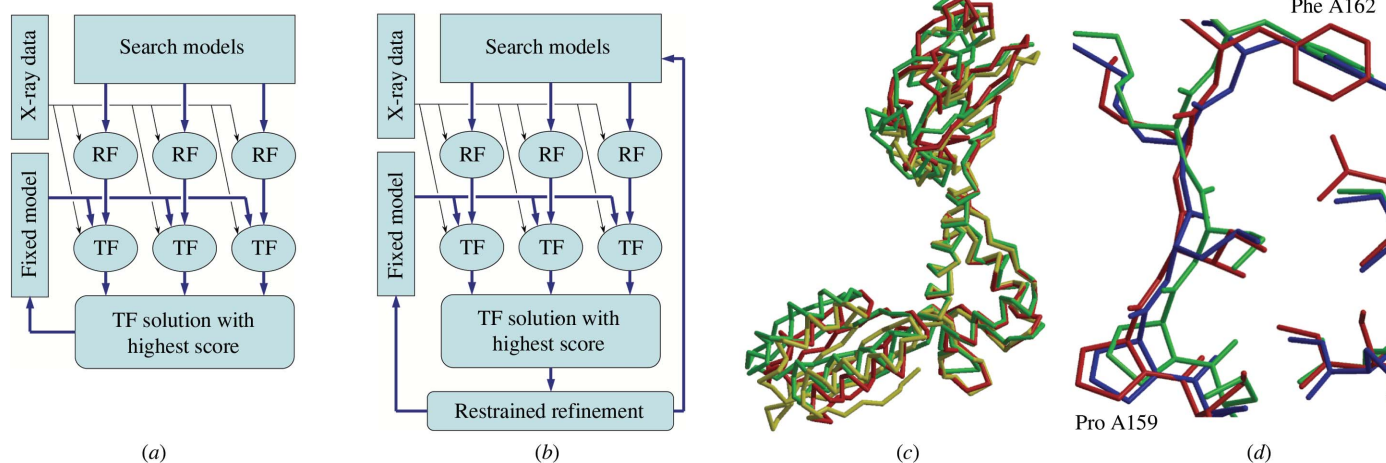
Figure 4
MR solution of the MTH685 protein crystal structure. (*a*) The one-by-one MR protocol in which the RF is computed only once for each single-domain search model. (*b*) The MR protocol with two feedbacks, in which each new partial model is refined and therefore search models and lists of their possible orientations are updated along with the partial structure. (*c*) Superposition of (red, yellow) two molecules of MTH685 protein forming the AU and (green) the homologous protein Af0491 (PDB code 1p9q) by second domains, showing that MR solution using the whole molecule as a search model is impossible. (*d*) Enlarged superposition of second domains (red) from chain *A* of the final structure, (blue) from a refined partial structure containing two of six domains and (green) from the homologue corrected according to the target sequence.

when high symmetry of the crystal produces more than eight symmetry equivalent non-origin peaks, as their heights will be less than the threshold value of 1/8. On the other hand, some special structural features (*e.g.* helices in DNA structures) can generate strong Patterson peaks. One should bear in mind that such peculiarities as partial disorder or twinning by reticular merohedry can also generate non-origin peaks in the Patterson function (Rye *et al.*, 2007) that have nothing to do with translational NCS.

3. Examples of model correction using X-ray data

In some cases MR allows users to build only a partial model (this includes the case where the correct orientation and therefore a *P*1 substructure are found). Frequently, some kind of refinement of this partial model produces a better search model, which can be sufficiently good for subsequent MR to build the complete structure. Two such cases are presented below.

3.1. Rigid-body refinement *versus* RF

The crystal structure of BPV-1 E1 helicase (PDB code 2v9p; Sanders *et al.*, 2007)[1] belongs to space group $P2_12_12_1$, with unit-cell parameters $a = 181.3$, $b = 187.0$, $c = 135.3$ Å. The AU contains two hexamers related by translational NCS. Each monomer is composed of an AAA+ domain (∼200 amino acids) and oligomerization domain (∼75 amino acids). At the time of structure determination, the closest homologue in the PDB was the AAA+ domain of HPV-18 helicase, which had 51% sequence identity with the AAA+ domain of the target

[1] The structure determinations described here and in the cited paper differ as another structure of the same protein became available before the cited paper was published.

protein (PDB code 1tue). In this structure the oligomerization domain is absent and the AAA+ domain exists in a monomeric form. The closest homologue with a known hexameric structure was SV40 helicase (PDB code 1n25), which shared only 16% of its amino-acid sequence with the full length of the target protein. Attempts to find a solution with the monomer from 1tue or with the hexamer from 1n25 failed.

The structure was solved starting from a synthetic model containing six AAA+ domains from 1tue, corrected according to the target sequence and fitted to six subunits of the hexamer from 1n25 using *SSM* superposition (Figs. 3*a*, 3*b* and 3*c*). For this model, the first three peaks in the RF were consistent with the self-rotation function and had a small but appreciable contrast compared with the other peaks (Fig. 3*f*). However, the TF did not give a reasonable solution. We assumed that the hexamer in the unknown structure had a slightly different organization and undertook refinement of the synthetic hexamer model. During this procedure four parameters were refined: three angles defining the orientation of the monomer *A* and the distance between the centre of monomer *A* and the sixfold axis. The remaining five monomers were generated from monomer *A* by the sixfold symmetry. The target function was the value of RF/σ(RF) for the highest RF peak. Maximization of the target function was performed iteratively using a *tcsh* script. For a given current hexamer, eight new hexamers were generated where the distance was incremented by ±1 Å or one of the angular parameters was incremented by ±1°. *MOLREP* was used to compute the RF for each of new hexamers. The values of RF/σ(RF) for the first RF peaks were extracted from the log files. The new hexamer with the highest value of the target function became the current model in the next iteration. The procedure was terminated when none of the new models gave an increase in the target function compared with the current model. The refinement rotated the

Table 1
The sequence of MR searches that led to the solution of the MTH685 protein crystal structure.

The composition of the models is given in terms of domains comprising residues 1–89 (domain 1), 90–162 (domain 2) and 163–232 (domain 3).

Step	1	2	3	4	5	6
Composition of fixed model						
Chain A	—	1	1	1, 2	1, 2	1, 2
Chain B	—	—	2		1, 2	1, 2, 3
Search model that gave the best TF score	1	2	2†	1†	3	3†
Refinement of partial structure						
R_{cryst}	0.526	0.495	0.459	0.447	0.404	0.358
R_{free}	0.548	0.534	0.503	0.486	0.445	0.425

† The search model was taken from the refined partial structure from the previous step.

subunits by 10° and translated them by 6 Å (Fig. 3d). Using the refined hexamer, the behaviour of conventional MR improved dramatically (Fig. 3g). The refined hexamer (Fig. 3d) and the hexamer from the final structure (Fig. 3e) were very similar to each other and differed significantly from the initial synthetic hexamer (Fig. 3c).

If the initial hexamer is fitted to the final crystal structure, severe clashes between symmetry-related molecules occur and conventional rigid-body refinement does not converge to the correct structure. Overlaps between symmetry-related molecules are a likely reason why MR did not work with the nonrefined hexamer. The difference between refinement against the RF and conventional rigid-body refinement is that in the first case the target function is based on intensities and in the second case it is based on structure amplitudes. Also, the weighting scheme in the second case is adjusted for complete structures with a smaller deviation of coordinates from the final values. This example shows that rigid-body refinement is potentially a very powerful tool for improvement of models at the stage where the complete structure has not yet been built. It is worth mentioning that a combination of locked rotation and translation functions (Tong, 2001) is likely to be an alternative method of building a hexamer model from single subunits.

3.2. MR with feedback from the refined partial model

Elements (monomers or domains) forming the AU do not always obey proper noncrystallographic symmetry. In this case, the only solution is a standard one-by-one search with very incomplete search models. Two problems are usually encountered in this approach. Firstly, a minor problem is the lack of contrast in the TF when positioning the last few elements. The major problem is that the RF is calculated only once for each original search model representing only a small fraction of the AU (Fig. 4a). Even if the search model is adequately modified, some of the correct RF peaks may remain weak owing to the specific configuration of the interatomic vectors in the actual crystal structure. Such peaks are therefore absent in the list of top RF peaks provided for

the further TF search. As a result, the corresponding elements of the AU are not positioned at all. If, however, a partial MR solution is found, restrained refinement of this partial structure allows the search model(s) and the list of RF peaks to be updated (Fig. 4b).

This technique was instrumental in the determination of the crystal structure of the hypothetical protein MTH685 from the archaeon *Methanothermobacteria thermautotrophicus* (C. L. Ng and A. A. Antson, private communication). This crystal, with unit-cell parameters $a = 68.3$, $b = 72.1$, $c = 146.8$ Å, belongs to space group $P222_1$. X-ray data were collected to a resolution of 1.8 Å. The AU contains two identical monomers in different conformations (Fig. 4c), which are not related by any proper rotation. Each monomer contains three domains. At the date of structure determination, the PDB contained a structure (PDB code 1p9q) of a homologous protein from *Archaeoglobus fulgidus* with a sequence identity of 50%. Because of the domain mobility (Fig. 4c), the target structure could not be solved using the complete monomer as a search model. The problem turned out not to be a simple MR problem despite the high sequence identity. None of the possible search models were perfect; the monomer model because the three-dimensional similarity was too low and the single-domain model because the completeness was too low.

The protocol presented in Fig. 4(a) allowed *MOLREP* to find correct MR solutions for domain 1 from chain A and domain 2 from chain B (steps 1 and 2 in Table 1). However, it was not obvious whether this partial model was correct, as the orientation of domain $B2$ corresponded to only the 24th highest peak in the RF and a search for the remaining domains was unsuccessful. Moreover, this model could not be validated on the basis of connectivity considerations, as the two domains found belonged to different polypeptide chains.

In contrast, a protocol which included refinement of the partial structure using *REFMAC* (Murshudov *et al.*, 1997) allowed the complete model to be built (steps 3–6 in Table 1). Step 3 was critical for structure determination and we analysed it in more detail. Despite the partial structure $A1 + B2$ being only about 30% complete, the restrained refinement performed quite efficiently: most of the atoms moved closer to their final positions (Fig. 4d) and the r.m.s.d. for C^α atoms between domain $A2$ in the final structure and the corresponding search model reduced from 1.42 to 0.98 Å. This improvement completely changed the behaviour of the RF. It turned out that the correct orientation of domain $A2$ was not in the list of 200 highest RF peaks until the corresponding search model was updated. The impact of the search-model improvement on the TF was not so significant. Additional tests showed that if the correct orientation was known, the improvement of the search model would only cause a 15% increase in contrast. Starting from step 3, the models were validated by the connectivity between neighbouring domains and by the decrease in R_{free} (Table 1). It is likely that after step 3, when 50% of the complete structure had been defined, it was already possible to switch to searching for the remaining domains in the electron density using, for example, *SAPTF* (Vagin & Isupov, 2001) implemented in *MOLREP*.

4. Conclusions

MOLREP is an automated stand-alone molecular-replacement software package that incorporates various options for model preparation and correction. Such an integrated approach to molecular replacement has proved to be very efficient. Model correction using sequence alignment, weighting of X-ray data based on the sequence identity, modelling of surface flexibility and modelling of the translational NCS increase the chances of solving difficult molecular-replacement problems. However, there is significant room for improvement. For example, the conversion of sequence identity into three-dimensional similarity, which is required for weighting, is not a straightforward task. It may be necessary to analyse the class of proteins that the target molecule belongs to and to predict the difference between the target and search molecules according to the variability of the proteins within the class.

In this paper, we have demonstrated that the methods used for model preparation should not be restricted to bioinformatics techniques. A combination of restrained refinement and molecular replacement seems to be promising in cases where several copies of a multidomain molecule are present in the AU in different conformations, especially when high-resolution experimental data are available. Our example shows that the rotation search performs much better and the translation search is improved if the partial structure is refined and the search model is derived from it. Another example showed that rigid-body refinement with a specialized target function could be successfully applied to a substructure with an as yet unknown position in the crystal, but with an approximately known organization and orientation. If accurate algorithms are designed and implemented, the refinement of multidomain or oligomeric models before the translation search may become a powerful technique.

MOLREP can be downloaded from either the *CCP*4 website (http://www.ccp4.ac.uk) or from that of AAV (http://www.ysbl.york.ac.uk/~alexei/molrep.html). The command-line interface user's guide can also be found at AAV's website.

References

Berman, H. M. *et al.* (2002). *Acta Cryst.* D**58**, 899–907.
Brünger, 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.
Chothia, C. (1992). *Nature (London)*, **357**, 543–544.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fiser, A. & Sali, A. (2003). *Methods Enzymol.* **374**, 461–491.
Gonzales, R. C. & Woods, R. E. (2002). *Digital Image Processing.* Upper Saddle River, New Jersey: Prentice Hall.
Isupov, M. N., Brindley, A. A., Hollingsworth, E. J., Murshudov, G. N., Vagin, A. A. & Littlechild, J. A. (2004). *Acta Cryst.* D**60**, 1879–1882.
Keegan, R. M. & Winn, M. D. (2008). *Acta Cryst.* D**64**, 119–124.
Krissinel, E. & Henrick, K. (2004). *Acta Cryst.* D**60**, 2256–2268.
Krissinel, E. & Henrick, K. (2005). *CompLife 2005*, edited by M. R. Berthold, R. Glen, K. Diederichs, O. Kohlbacher & I. Fischer, pp. 163–174. Berlin, Heidelberg: Springer-Verlag.
Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.
Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. (2008). *Acta Cryst.* D**64**, 125–132.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Needleman, S. B. & Wunsch, C. D. (1970). *J. Mol. Biol.* **48**, 443–453.
Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.
Rye, C. A., Isupov, M. N., Lebedev, A. A. & Littlechild, J. A. (2007). *Acta Cryst.* D**63**, 926–930.
Sanders, C., Kovalevskiy, O., Sizov, D., Lebedev, A., Isupov, M. N. & Antson, A. A. (2007). *Nucleic Acids Res.* **35**, 6451–6457.
Schwarzenbacher, R., Godzik, A., Grzechnik, S. K. & Jaroszewski, L. (2004). *Acta Cryst.* D**60**, 1229–1236.
Schwarzenbacher, R., Godzik, A. & Jaroszewski, L. (2008). *Acta Cryst.* D**64**, 133–140.
Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. (2003). *Nucleic Acids Res.* **31**, 3381–3385.
Stubbs, M. T. & Huber, R. (1991). *Acta Cryst.* A**47**, 521–526.
Suhre, K. & Sanejouand, Y. H. (2004). *Nucleic Acids Res.* **32**, W610–W614.
Tong, L. (2001). *Acta Cryst.* D**57**, 1383–1389.
Vagin, A. A. (1983). PhD thesis. Institute of Crystallography, Moscow, USSR.
Vagin, A. A. & Isupov, M. N. (2001). *Acta Cryst.* D**57**, 1451–1456.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.