Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 20 July 2010
Accepted 2 November 2010

PVMR: assembling small helix fragments as structural solutions for molecular replacement

Fan Jiang* and Wei Ding

Institute of Physics, Beijing National Laboratory for Condensed Matter Physics, People's Republic of China. Correspondence e-mail: fjiang@aphy.iphy.ac.cn

A new real-space implementation of the molecular-replacement method is described. The method locates the search model in the target crystal by maximizing the matching between the search-model vectors and the Patterson self and cross vectors. In previous work, a new rotation function was introduced for the molecular-replacement method [Jiang (2008). *Acta Cryst.* D**64**, 561–566]. This rotation function is calculated by matching the search model directly with both the Patterson self and cross vectors in real space. All the matches are summed and averaged to enhance the overall signal-to-noise ratio for a given orientation of the search model. Recently, to avoid the dependence of the weights derived from the linear regression on the properties of the search model and the target crystal structure, such as secondary structures, space groups and cell parameters, a dynamic correlation coefficient has been designed and used as the total rotation function score [Jiang & Ding (2010). *Chin. Phys. B*, **19**, 106101]. This work further extends this idea to the implementation of translation search. A new real- or direct-space translation function has been implemented by matching the cross vectors between the symmetry mates of the search model to the Patterson cross vectors. This method enables effective searching for small helix fragments in the target crystal. Although the solution model assembled by using multiple fragments of helix is insufficient to start *ab initio* phasing of the target crystal, it can be used to identify the known protein folds in the Protein Data Bank that are homologous to the target structure. It can also be combined with other experimental and theoretical models to screen and select for better search models for molecular replacement.

1. Introduction

In our previous work, we have introduced a new rotation function for the molecular-replacement method (Jiang, 2008). This rotation function is calculated by matching the search model directly with both the Patterson self and cross vectors in real space. All the matches are summed and averaged to enhance the overall signal-to-noise ratio for a given orientation of the search model. The new rotation function uses a number of 'image-seeking' functions to estimate the goodness of the matches. The matching scores from all the 'image-seeking' functions are combined into a single score using the multiple linear regression method, which fits the rotation-angle distance from the correct solution. Our previous tests show that the correct solution usually appears in the top 100 solutions. Recently, to avoid the dependence of the weights derived from the linear regression on the properties of the search model and the target crystal structure, such as secondary structures, space groups and cell parameters, a dynamic correlation coefficient has been designed and used as the total rotation function score. This dynamic correlation coefficient is calculated on the fly, for each rotation sampled, from the differences of the true angle distances and the fitted angle distances. The data points used for fitting are obtained by using the neighbour rotations of the sampled rotation and fitting the angle distances of the neighbour rotations with their 'image-seeking' function scores (Jiang & Ding, 2010) as the free variables. Tests show that the dynamic correlation coefficient is a better rotation function and less dependent on particular search models or target crystal structures. Furthermore, by using the dynamic correlation coefficient, the best search parameters can be selected by sampling random rotations to find the parameter set with the least background noise, namely, the smallest dynamic correlation coefficient.

In this work, we will further extend the idea of the dynamic correlation coefficient to the implementation of translation search of the molecular-replacement method. We have implemented a real- or direct-space translation function by matching the cross vectors between the symmetry mates of the search model to the Patterson cross vectors. The same 'image-seeking' scoring functions are used as defined previously (Jiang, 2008; Jiang & Ding, 2010). These matching scores are combined by the linear regression method and then clustered locally to calculate a dynamic correlation coefficient. The

clustering here not only includes the neighbour rotations but also the nearby translations, namely a six-dimensional clustering. The result of the clustering gives both a rotation and a translation, which are obtained by averaging all the solutions included in the local cluster for calculating the local dynamic correlation coefficient. The test results will be described.

Then we implemented a rigid-body refinement method using one of the quasi-Newton methods (Press *et al.*, 1992). The gradient is calculated by using functional difference taken at small variable intervals. Since this is a real-space refinement method, we calculate the goodness-of-fit of the solution model to the electronic map phased by this model using a real-space map correlation coefficient. It is shown that a few cycles of such refinement will improve the accuracy of the orientation and position of the solution model, as well as its ranking in the solution list.

Since we have developed the method to use small search models such as a helix fragment, we need to assemble several fragments positioned by the translation search to obtain a bigger solution model so that its phasing power is increased. Currently, the solution model obtained is neither big nor accurate enough to phase the target crystal so that automatic model building can be applied. The initial rough phases of the solution model cannot be used to bootstrap the *ab initio* phasing process by the current methods we know of. Instead we use the solution model consisting only of helices to search the PDB (Protein Data Bank) to find similar known structures that are structurally homologous to this solution model. The known structures found can then be tested by further molecular-replacement searches. We will show that the solution model obtained by our method can pick out the correct native crystal structure with a different space group from the PDB in top rankings. For this database search, we use our own search program package *SDEPS*. Combining with other structural determination methods, our method is almost ready for practical application.

2. Materials and methods

2.1. Real-space rotation function using Patterson vectors (*PVMR_ROT*)

The method of *PVMR_ROT* rotation function has been described previously (Jiang, 2008). Briefly, the search model vectors are directly matched to the Patterson vectors in real space using five 'image-seeking' functions, allowing partial matching between the search model and the Patterson vectors. The rotation function for a given orientation of the search model is the average values of the five 'image-seeking' functions. These average values are further combined to calculate a single rotation function value using a dynamic correlation coefficient (Jiang & Ding, 2010). The five 'image-seeking' functions for matching the model and Patterson vectors are as follows.

The first is the product of the weights, namely the peak heights of the Patterson vectors, p_i, and the heights of the model atomic vectors, q_i,

$$R_{\rm prod} = \frac{1}{N}\sum_{i}^{\{\rm MP\}}(p_i \cdot q_i), \tag{1}$$

where {MP} denotes the set of matched points or vectors, and N is the total number of matched points.

The second is the correlation coefficient between the weights, with the same p's and q's as in equation (1),

$$R_{\rm corr} = \frac{\langle pq\rangle - \langle p\rangle\langle q\rangle}{[(\langle p^2\rangle - \langle p\rangle^2)(\langle q^2\rangle - \langle q\rangle^2)]^{1/2}}, \tag{2}$$

where $\langle p\rangle$ and $\langle q\rangle$ denote the average over the matched points or vectors.

The third is the relative R factor of the weights, with the same p's, q's, {MP} and N as in equation (1),

$$R_{\rm rfac} = \frac{1}{N}\sum_{i}^{\{\rm MP\}}\left[\frac{|p_i - q_i|}{(\langle p\rangle + \langle q\rangle)/2}\right]. \tag{3}$$

The fourth is the relative entropy between the model and Patterson vectors, with the same p's, q's, {MP} and N as in equation (1),

$$R_{\rm ent} = \frac{1}{N}\sum_{i}^{\{\rm MP\}} p_i \ln\left(\frac{p_i}{q_i}\right) - \langle p\rangle \ln\left(\frac{\langle p\rangle}{\langle q\rangle}\right). \tag{4}$$

The fifth is the mean-square residues between the model and Patterson vectors, with the same p's, q's, {MP} and N as in equation (1),

$$R_{\rm res} = \frac{1}{N}\sum_{i}^{\{\rm MP\}}(p_i - q_i)^2. \tag{5}$$

Since each 'image-seeking' function can be used for searching the matching peaks, there are five peak modes. For each peak mode, there are five 'image-seeking' function scores; so, for a given orientation, there are 25 scores. Two more scores are added to each peak mode, and they are the number of matched vectors averaged over all matching peaks and the average number of matching peaks. These two numbers are found to be proportional to the signal of the rotation function. Therefore, there are $5 \times 7 = 35$ scores for each orientation of the search model.

For each orientation sampled, its neighbour orientations within an angle distance of say 15° are pre-calculated. From these neighbours, a given fixed number of them are randomly selected, say 260. The angle distances from these orientations to the 'centre' orientation sampled are then used as the 'experimental' value for the multivariate linear regression method. The variables for the regression are the 35 function scores. The correlation coefficient between the fitted value and the 'experimental' value is then defined as the dynamic correlation coefficient for the orientation sampled and assigned to be the rotation-function value. This method of calculating the rotation function has been tested (Jiang & Ding, 2010) and has several advantages. Its most important advantage, compared to the previously defined 'static' rotation function (Jiang, 2008), is that it is independent of the crystal

properties such as cell parameters and space groups, as well as the specific types of search models and target structures.

2.2. Real-space translation function using Patterson vectors (*PVMR_TRAN*)

The translation function uses the Patterson vectors in real space, whose general principles have been described (DeLano & Brünger, 1995; Grosse-Kunstleve & Adams, 2001; Jiang & Rao, 2001; Kissinger *et al.*, 2001; Rossmann, 1990). Briefly, the matching of the model cross vectors of any symmetry mate to the Patterson vectors can only determine two components of a translation vector. Therefore, at least two symmetry mates must be used and the overlapping components must have the same translation to ascertain that the same translation vector has been found in matching the cross vectors. This method can be implemented in a computer algorithm so that as many sets of cross vectors can be checked as there are different symmetry mates. The consistent peaks from this matching of cross vectors are output.

A fast translation algorithm has been used in the matching of cross vectors [see Fig. 1 in Jiang (2008)]. The peak is not very sharp, as shown in the figure. But, in practice, when the matching scores are weighted by the 'image-seeking' functions the peaks are more sharpened than without the weighting.

The matching between the model cross vectors and the Patterson vectors is scored by the same 'image-seeking' functions described above in equations (1)–(5). For each translation solution, including a rotation and a translation, a dynamic correlation coefficient is also calculated and used as the translation-function value. The neighbours of the 'centre' solution are searched in the six-dimensional space instead of only in the three-dimensional rotation space as in calculating the rotation function. Again, a fixed number of neighbours are randomly selected and a dynamic correlation coefficient is calculated by using the linear regression method, with the 'experimental' values being the centre-of-mass distance between the neighbour solutions and the centre solution.

2.3. Refinement of the orientation and position of translation solutions

After obtaining the translation solutions of the molecular replacement, the orientation and position of the solution models need to be refined. The relatively low accuracy of the solutions is partially due to the fact that the search model is of very small size. It is also due to the fact that the accuracy of the real-space grid-based translation search depends on the grid size used, but using smaller grids will increase the computation time considerably. Therefore, after locating a potential solution, it needs to be refined locally using better orientation and position sampling and a more sensitive scoring function.

For a scoring function, the real-space map correlation coefficient is used. For a given solution model, the corresponding phases are calculated. Then these phases are used to calculate two electron-density maps: one using the experimental diffraction amplitudes and the other the calculated amplitudes based on the solution model. The correlation coefficient between these two maps is calculated. The definition of the correlation coefficient used is

$$R = \frac{\langle \rho_1 \rho_2 \rangle - \langle \rho_1 \rangle \langle \rho_2 \rangle}{\left(\langle {\rho_1}^2 \rangle - \langle \rho_1 \rangle^2\right)^{1/2} \left(\langle {\rho_2}^2 \rangle - \langle \rho_2 \rangle^2\right)^{1/2}} \qquad (6)$$

where ρ_1 and ρ_2 are two electron densities, and $\langle\ \rangle$ denotes averaging. The range of R is usually between 0 and 1. A value of 1 means complete matching with each other. It has been shown that the R value is directly correlated with the average phase error between the model and the target crystal structure (Lunin & Woolfson, 1993). However, when the target structure is unknown, the R value is only a relatively inaccurate index of how the solution model matches with the target structure.

There are two ways to improve the sampling of the orientations and positions of a translation solution. One is random sampling within a radius of the angle distance and the centre-of-mass distance. This is called the perturbation method. The other is to use a multivariate optimization method. For this strategy the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method (Press *et al.*, 1992) is used.

For the perturbation method, 60 orientations are sampled within an angle distance of 15°. For these 60 orientations, the grid-based translation search is performed again and the translation solutions clustered and sorted again. The top ten solutions are saved as the candidate solutions for further refinement.

The BFGS method is a very powerful and stable gradient-based algorithm for finding the local minimum of a multivariate function. It usually requires the calculation of the functional gradient at a given variable point. For the map correlation coefficient, the analytical gradient calculation is not possible. The common way to substitute for the direct calculation of the functional gradient is to calculate the gradient numerically. For our purpose, the function difference method is used with the formula

$$g_{(k)} = \frac{f[x^{(k)}] - f[x^{(k)} + \Delta x^{(k)}]}{\Delta x^{(k)}}, \qquad (7)$$

where $g_{(k)}$ is the functional gradient at step k of the optimization. $f[x^{(k)}]$ is the functional value at the variable point $x^{(k)}$. Δx is the difference step size.

2.4. Assemble the solution models and search for the homologous folds in the PDB

The solution model is generated by assembling the fragments positioned by the translation solutions and the resulting model is checked for steric clash in crystal packing. Usually many solutions are possible and these models are further screened. They can either be further refined using the BFGS method or other rigid-body-based refinement methods, or they can be used to search in the PDB to find structurally homologous folds so that better search models for molecular replacement can be found.

For the BFGS method, both the position and the ranking of the correct solutions can be improved. However, in this work

the phases generated with the assembled solution models are not good enough to lead to traceable electron-density maps using 2 Å data. As a second strategy, the solution models are used to search in the PDB to find homologous folds that are structurally similar to the correct native structure, namely myoglobin. In our tests, another myoglobin structure in a different space group has been found and ranked high.

For searching the PDB, a program package *SDEPS* was used. It is a six-dimensional structural alignment method and has not been published. Owing to limitations of space, it will be described elsewhere. It can align two structures using Cα atoms without constraints on the sequential order of the fragments aligned. This is a required feature in our work.

Table 1
Optimization of parameters for the rotation search.

No. of residues†	Grid sampling of parameters			Best parameters (grid size, Nsum, Nsum_unit)‡	Noise level§
	Grid size (Å)	Nsum	Nsum_unit		
8	(4.0 4.2 4.4 4.6 4.8 5.0)	(2 4 8)	(2 4 8 12 16)	(4.4, 8, 8)	0.5728
12	(4.0 4.2 4.4 4.6 4.8 5.0)	(2 4 8)	(2 4 8 12 16)	(4.2, 8, 12)	0.5272
16	(4.0 4.2 4.4 4.6 4.8 5.0)	(2 4 8)	(2 4 8 12 16)	(4.2, 2, 16)	0.4947

† Number of residues in the helix fragment used as the search model. ‡ The best parameter set was found by finding the lowest noise level in the grid sampling. § The noise level is the average dynamic correlation coefficient for the 30 random rotations.

2.5. Data used for tests

The target crystal used for the positive tests was the myoglobin crystal structure (PDB code 1J3F) (Ueno *et al.*, 2005). The space group is $P2_12_12_1$, with unit-cell parameters a = 49.163, b = 40.002, c = 80.011 Å and with one molecule in an asymmetric unit. The experimental diffraction data for this entry 1J3F were available to 1.45 Å resolution. Another crystal structure (PDB code 1BXW) (Pautsch & Schulz, 1998) was used for the negative control. Its space group is $P2_12_12_1$, with unit-cell parameters a = 32.972, b = 58.787, c = 76.237 Å and with one molecule in an asymmetric unit. The diffraction data for this entry were available to 2.50 Å resolution. This crystal structure was used to construct an artificial crystal with the same crystal parameters as the 1J3F crystal.

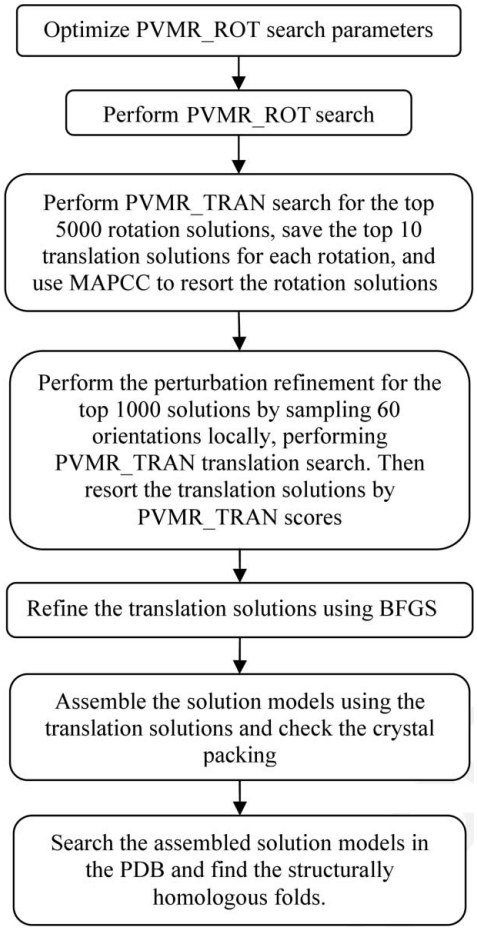

Figure 1
The workflow of the *PVMR* programs.

Three poly-Ala models were used as separate search models for the positive tests. They consisted of a helix from residues A3 to A10, A3 to A14 and A3 to A18 in the myoglobin structure of 1A6M and had been rotated by a polar angle (45, 60, 180°) relative to the 1J3F crystal. The same 16-residue model was also used as the search model for the negative control.

Both the coordinates and diffraction data were extracted from the PDB (http://www.rcsb.org). The Patterson maps were calculated using CCP4 programs (Collaborative Computational Project, No. 4, 1994) and the resolution cutoffs for the diffraction data were 2.0 Å. The Patterson vectors selected for matching were map points with values greater than or equal to 1σ. One unit cell of the Patterson map was scanned for matching.

The PDB fold database was obtained from *ASTRAL-SCOP* version 1.63 (Brenner *et al.*, 2000; Casbon *et al.*, 2006). It consisted of 762 unique fold domains.

3. Results

The criterion for a correct solution is when the angle distance of the rotation is no greater than 25° and the centre-of-mass distance of the translation is within a radius of 15 Å away from the true native solution. The workflow of the *PVMR* programs is shown in Fig. 1, where each box corresponds to a major step. The results of each step for a single-fragment search model are described below.

3.1. Step 1: optimize the rotation-search parameters

This was done by sampling 30 random rotations globally to estimate the level of the background noise. The lower the background noise is, the better the corresponding search parameters (Jiang & Ding, 2010). Three parameters are screened by grid sampling, namely the grid size, the Nsum cutoff and the Nsum_unit cutoff. The 'Nsum' cutoff denotes the minimum number of matched vectors for peak searching;

Table 2
Rotation search using *PVMR_ROT*.

No. of residues	No. of top rotation solutions	No. of correct solutions	Best ranking of correct solutions	Ranking of the best solution	Error of the best solution (°)
8	5000	5	870	4967	13.535
12	5000	3	728	3366	4.769
16	5000	6	1200	1228	9.851

Table 3
Ranking of correct solutions after re-sorting by the map correlation coefficient.

No. of residues	No. of top rotation solutions	No. of correct solutions	Best ranking of correct solutions	Ranking of the best solution	Error of the best solution (°)
8	1000	4	297	679	13.535
12	1000	2	422	468	11.583
16	1000	3	634	905	12.502

Table 4
The results of the perturbation refinement by using *PVMR_TRAN*.

No. of residues	No. of sampled rotations	No. of top translation solutions	No. of correct solutions	Best ranking of correct solutions	Errors and ranking of the best solution (°, Å, ranking)
8	60000	6000	52	94	7.908, 10.244, 3569
12	60000	6000	33	81	10.980, 10.569, 3614
16	60000	6000	34	56	10.788, 8.314, 5882

Table 5
Improvement of translation solutions by BFGS refinement.

No. of residues	No. of translation solutions refined	No. of correct solutions	Best ranking of correct solutions	Errors and ranking of the best solution (°, Å, ranking)
8	6000	47	352	5.935, 9.091, 5144
12	6000	57	27	4.431, 3.534, 255
16	6000	64	4	14.257, 2.713, 2861

'unit' refers to the unit cell. For the tests presented here the best parameters for the rotation search used are listed in Table 1.

3.2. Step 2: rotation search

All the rotations sampled were sorted according to the descending order of the dynamic correlation coefficients and the top 5000 solutions were saved for the next step. The results are shown in Table 2.

3.3. Step 3: re-sort the rotation solutions by calculating the map correlation coefficient

The 5000 rotations saved from the last step were used for the translation search using *PVMR_TRAN*. The translation solutions for each rotation were sorted by the dynamic correlation coefficients and the top ten solutions were saved. Then R was calculated for these ten translation solutions and the highest R was used as the score for this rotation. Next, the 5000 rotations were re-sorted according to the R values. After the re-sorting, only the top 1000 rotations were saved for the next step. Alternatively, the R-value calculation was not used, but the sorting order according to the dynamic correlation coefficients was used to save the top 1000 rotations. The results by using R values are listed in Table 3.

3.4. Step 4: perturbation refinement of the rotation and translation solutions

Sixty random rotations within an angle distance of 15° were selected to perturb the rotations saved from the last step. For each rotation sampled, including those sampled by the perturbation, the translation search was performed using *PVMR_TRAN* and the top ten translation solutions were saved. Then all the translation solutions were pooled together and sorted according to the dynamic correlation coefficients. At the end of this step, 6000 translation solutions were saved (see Table 4).

3.5. Step 5: refine the translation solutions using the BFGS minimization method

After this step, both the ranking and the accuracy of the solution were improved (see Table 5).

3.6. Step 6: assemble the solution models using the translation solutions and check the crystal packing

Because the search model was relatively small, the solution models were constructed by combining several translation solutions and assembling the transformed fragments. Then the crystal packing was checked for steric clash. Usually a few hundreds of assembled solution models were able to pass the crystal-packing check (see Table 6).

3.7. Step 7: search the PDB for structurally homologous folds

The solution models were then searched against a non-redundant fold database provided by *ASTRAL-SCOP* version 1.63 (Brenner *et al.*, 2000; Casbon *et al.*, 2006), consisting of 762 unique folds. The results are shown in Table 6.

3.8. Negative control using a β-barrel structure 1BXW from the PDB

The target crystal structure for the negative control was 1BXW, which is a β-barrel structure. From Table 6 it can be seen that the best ranking obtained using the 16-residue model for the alignment with fold 1A6M was 73, much worse than the positive tests, with rankings of 23, 1 and 3 for the 8-, 12- and 16-residue models, respectively. This proves the effectiveness of the *PVMR_TRAN* solutions.

The computing time for each step is shown in Table 7. Because of the low speed of *PVMR_TRAN* and BFGS

Table 6
The best ranking of the alignment of the solution models with the fold 1A6M.

No. of residues	No. of fragments	No. of solution models	Maximum RMS of aligned folds	Minimum No. of aligned residues per fragment	No. of alignments	Ranking of alignment with fold 1A6M	Statistics of best alignment (No. of residues, RMS, No. of fragments)
8	8	3280	8	4	178	23	46, 7.01, 8
12	8	199	5	4	11	1	49, 4.76, 7
16	4	649	5	8	803	3	50, 4.27, 5
16†	4	328	5	8	446	73	41, 3.60, 4

† The negative control is performed using the same 16-residue search model and an artificial target crystal using the same crystal parameters as 1J3F but using a β-barrel structure from the PDB entry 1BXW. The workflow performed on this artificial crystal is the same as the positive tests.

Table 7
Computing times.

Step	CPU hour (one processor at 2.0 GHz)
1	0.5
2	8
3	643
4	7000
5	2100
6	0.02
7	135

refinement, the execution had to be distributed on a Linux cluster.

4. Discussion

In this work, a complete procedure for molecular replacement (*PVMR*) has been described. The rotation function of the previous work (Jiang, 2008; Jiang & Ding, 2010) has been improved upon. The translation search has been implemented in a more automatic fashion than was done previously (Jiang & Rao, 2001). The resulting program package *PVMR* includes both the rotation search and the translation, and is almost automatic and ready for practical application. It has a number of unqiue features. First, for the rotation function, it matches the vectors of the search model with the Patterson vectors in real space using both the self and the cross vectors, not just the self vectors as is done for most other reciprocal-space implementations of the rotation function. This enhances the signal-to-noise ratio and enables the use of very small search models such as a helix fragment. Second, the translation function clusters the solutions in the six-dimensional space and uses the dynamic correlation coefficient to sort the solutions. Third, it can search the PDB database for structurally homologous folds to derive new and better search models for further structure determination by molecular replacement. Fourth, it is capable of searching very small models such as helix fragments consisting of only eight residues. These are just a few among other new features associated with the *PVMR* method. Therefore, this is a complete novel implementation of the molecular-replacement method in real space.

However, the current *PVMR* method cannot generate solution models that are accurate enough to bootstrap the phasing of an unknown target crystal. This is because the size of the search model is too small to be able to refine the orientation and position of a solution model effectively using either *PVMR_TRAN* or BFGS. However, there are a number of ways in which the solution models produced by *PVMR* can be used in facilitating structure determination. First, the solution models can help select the best models derived from small-angle X-ray scattering data (Petoukhov & Svergun, 2007). Second, the fold domains found from the PDB search using the solution models can be scored for fold threading using the target protein sequence. This way the best fold found, which should lead to new and bigger search models, can then be used as the new search model for further molecular replacement. Finally, the solution models can help narrow the range of possibilities of the predicted structures obtained with prediction methods such as ROSETTA (Bystroff & Shao, 2002; Raimondo *et al.*, 2007; Schonbrun *et al.*, 2002). It has been shown that the predicted structure can also be used as a search model for molecular replacement (Rigden *et al.*, 2008).

PVMR has not been tested using large search models because this class of problems has been solved successfully by the current methods (*MOLREP*, *PHASER*, *AMORE*, *EPMR etc.*). Nevertheless, the ability of the *PVMR* method to tolerate errors in the search models needs to be tested in the future. The problem with using bigger search models is that the computing time will increase with the size of the search models.

In future work, the method of rigid-body refinement will be further improved using new scoring functions. They can include information from both reciprocal space (phase refinement) and real space (electron-density-map pattern recognition), so that more accurate solution models can be obtained to facilitate the *ab initio* phasing process. A better optimization algorithm can also be implemented using hill-climbing algorithms such as simulated annealing.

This work was supported by the National Natural Science Foundation of China (grant Nos. 10674172 and 10874229).

References

Brenner, S. E., Koehl, P. & Levitt, M. (2000). *Nucleic Acids Res.* **28**, 254–256.
Bystroff, C. & Shao, Y. (2002). *Bioinformatics*, **18**, S54–S61.
Casbon, J. A., Crooks, G. E. & Saqi, M. A. (2006). *BMC Bioinformatics*, **7**, 10.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
DeLano, W. L. & Brünger, A. T. (1995). *Acta Cryst.* D**51**, 740–748.
Grosse-Kunstleve, R. W. & Adams, P. D. (2001). *Acta Cryst.* D**57**, 1390–1396.
Jiang, F. (2008). *Acta Cryst.* D**64**, 561–566.
Jiang, F. & Ding, W. (2010). *Chin. Phys. B*, **19**, 106101.

Jiang, F. & Rao, Z. (2001). *J. Synchrotron Rad.* **8**, 1051–1053.
Kissinger, C. R., Gehlhaar, D. K., Smith, B. A. & Bouzida, D. (2001). *Acta Cryst.* D**57**, 1474–1479.
Lunin, V. Yu. & Woolfson, M. M. (1993). *Acta Cryst.* D**49**, 530–533.
Pautsch, A. & Schulz, G. (1998). *Nat. Struct. Biol.* **5**, 1013–1017.
Petoukhov, M. V. & Svergun, D. I. (2007). *Curr. Opin. Struct. Biol.* **17**, 562–571.
Press, W., Teukolsky, S., Vetterling, W. & Flannery, B. (1992). *Numerical Recipes in C.* Cambridge University Press.
Raimondo, D., Giorgetti, A., Bosi, S. & Tramontano, A. (2007). *Proteins Struct. Funct. Bioinf.* **66**, 689–696.
Rigden, D. J., Keegan, R. M. & Winn, M. D. (2008). *Acta Cryst.* D**64**, 1288–1291.
Rossmann, M. G. (1990). *Acta Cryst.* A**46**, 73–82.
Schonbrun, J., Wedemeyer, W. J. & Baker, D. (2002). *Curr. Opin. Struct. Biol.* **12**, 348–354.
Ueno, T., Koshiyama, T., Ohashi, M., Kondo, K., Kono, M., Suzuki, A., Yamane, T. & Watanabe, Y. (2005). *J. Am. Chem. Soc.* **127**, 6556–6562.