Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Low-resolution molecular replacement using a six-dimensional search

Qun Liu, Arthur J. Weaver, Tao Xiang, Daniel J. Thiel and Quan Hao*

MacCHESS at Cornell High Energy Synchrotron Source, Cornell University, Ithaca, NY 14853-8001, USA

Correspondence e-mail: qh22@cornell.edu

Received 2 January 2003
Accepted 1 April 2003

A parallel-aware program *MPI_FSEARCH* has been implemented to perform molecular replacement using an exhaustive six-dimensional search. In particular, the program can be used to deal with diffraction data at very low resolution ($d > 10$ Å) that would normally not be appropriate for other molecular-replacement programs such as *AMoRe* and *CNS*. Although an envelope constructed from a PDB (Protein Data Bank) file was tested in the present study, the program can be used to perform low-resolution molecular replacement with an envelope derived from various sources such as electron microscopy or small-angle solution X-ray scattering.

1. Introduction

Molecular replacement is a commonly used method for initial phase determination by placing a model molecule of known structure into the unit cell of a crystal of an unknown homologous molecule. The heart of the method is a six-dimensional search task that aims to find the correct rotational and translational parameters for the designated search model. Since its proposal and first successful use (Rossmann & Blow, 1962), molecular replacement has evolved into a very powerful tool in protein crystal structure determination. At present, molecular-replacement methods may be grouped into two types based on the algorithm applied to the six search parameters.

The first type is the conventional method, in which the six-dimensional search problem is split into two sequential three-dimensional search steps using the rotational and translational Patterson functions. That is, the Patterson function of the model is rotated to match the Patterson function of the diffraction data before the translation function is evaluated. *AMoRe* (Navaza, 1994; Navaza & Saludjian, 1997) and *MOLREP* (Vagin & Teplyakov, 1997, 2000) are typical representatives of such an approach. Brünger (1992, 1997) modified this method by rotating a selected number of vectors from the Patterson function of the search model and then performing Patterson correlation refinements to improve the accuracy of the rotational parameters before the translational search. These two-step strategies can greatly improve the efficiency for most cases of molecular replacement, but they are likely to fail in difficult situations when the rotational peaks and translational peaks interfere with each other: this is common when only a low-resolution data set is available. Recently, owing to the advance of computing techniques, it became possible to test an exhaustive six-dimensional molecular replacement on a 12-processor SGI Onyx R8000 computer using the *CCP*4 (Collaborative Computational Project, Number 4, 1994) format translational function driven by several Perl scripts (Sheriff *et al.*, 1999). All these methods are based on the Patterson function and require discrimination between intermolecular vectors and intramolecular vectors. They often work well for molecular replacement with high-resolution diffraction data, but may be less successful when only low-resolution diffraction data or low-resolution search models are available. In cases where the search model is from electron microscopy or small-angle solution X-ray scattering, all the above methods have difficulties; as the density inside the envelope is uniform, there are no discrete intra-envelope Patterson vectors from the model to be matched with observed peaks. Alternative methods need to be explored.

The second type of molecular-replacement method is not based on a Patterson function search, but rather on R factors or correlation coefficients computed from standard or normalized structure factors. In addition, these methods all use stochastic methods such as the reverse Monte Carlo method (Glykos & Kokkinidis, 2000, 2001), the evolutionary approach (Kissinger *et al.*, 1999, 2001) and the genetic algorithm method (Chang & Lewis, 1997). In principle, they may have the ability to perform low-resolution molecular replacement, but global convergence cannot be achieved every time owing to the nature of the stochastic methods. In order to validate the result, multiple repetitions are required; this compromises the advantage of the methods.

In summary, all the commonly used molecular-replacement methods based on the Patterson function have difficulty handling models derived from solution scattering or electron microscopy because of the low resolution of the models, which is usually worse than 10 Å. Even for a search model derived from the PDB (Berman *et al.*, 2000), conventional molecular replacement may not be able to yield correct solutions when the diffraction data resolution is lower than 10 Å. Here, we report an exhaustive six-dimensional search method for low-resolution molecular replacement, using the molecular envelope derived from a PDB file and diffraction data from 20 to 12.0 Å resolution. Based on previously published work (Hao *et al.*, 1999; Ockwell *et al.*, 2000; Hao, 2001), the program *MPI_FSEARCH* has been implemented to locate an envelope in a crystallographic unit cell; it performs a simultaneous six-dimensional search on orientation and translation to find the best match between experimental structure factors, $F_{\rm obs}$, and calculated structure factors, $F_{\rm calc}$. The program can handle any envelope and no longer requires non-crystallographic symmetry information. Recent advances in parallel computer clusters (Diederichs, 2000) and parallel libraries have enabled us to greatly expand the usability of *MPI_FSEARCH* in the field of protein crystallography.

2. Parallel computing environment

All computations in the present work were performed on a 64-processor Linux-based cluster dedicated to crystallographic and synchrotron-related applications. A full description of this machine (SIRIUS) can be found at http://www.macchess.cornell.edu/~weaver/sirius.html. There are 31 diskless 'client' nodes connected to a physically separate 'server' node with a SCSI disk array. Each machine has dual 1.2 GHz AMD processors, 1 Gb RAM, dual Ethernet ports and a Myrinet network interface. The SCSI array on the server node provides 137 Gb of RAID 0 disk storage. The 31 client nodes were purchased from APPRO (http://www.appro.com) as pre-configured model 1124 'servers' and assembled into a single heavy-duty moveable rack. Two of 16-port switches provide 100 Mb Ethernet connections for client boot-up and background NFS communications. Gigabit optical networking is provided by a 32-port Myrinet switch (http://www.myricom.com) for message-passing during execution of parallel-aware code.

Table 1
A list of MPI subroutines used in *MPI_FSEARCH*.

Name	Meaning
MPI_Init	Setting up an MPI execution environment
MPI_Comm_Rank	Determining the rank of the calling process in a communicator
MPI_Comm_Size	Determining the size of the group associated with a communicator
MPI_Bcast	Broadcasting a message from the process with rank 'root' to all other processes of the same group
MPI_Send	Sending data to a specific process
MPI_Recv	Receiving data from a specific process
MPI_Finalize	Terminating an MPI execution environment
MPI_Wait	Waiting for an MPI send or receive to complete
MPI_Wtime	Returning elapsed time on the calling processor

The most distinctive feature of SIRIUS is its diskless architecture. On boot-up, each client node relies on the server node to supply it with its network identity *via* DHCP and its Linux kernel *via* trivial ftp. Further, each client mounts its own root filesystem over ClusterNFS (http://clusternfs.sourceforge.net/). In the diskless client design, system administration tasks are reduced to management of a single server. If one or more clients 'fail' for any reason (other than a server problem), the trivial and most practical solution is to simply execute a server script to rebuild the entire client filesystem and reboot the client. This process requires less than 15 s per client.

SIRIUS is currently running optimized versions of the Linux 2.4.14 kernel under the RedHat 7.1 distribution environment. To take best advantage of the optical network, parallel programs coded using the MPI (Message Passing Interface) library specification (http://www-unix.mcs.anl.gov/mpi/) are executed under Myrinet MPICH-GM (http://www.myri.com/scs/), a port of MPICH to GM, Myricom's low-level message-passing system. Parallel programs written using PVM library calls can also be run over the high-speed optical network under PVM-GM (http://www.markus-fischer.de/myrinetproject.htm).

3. Parallel-aware *FSEARCH* through MPI

The single-processor version of *FSEARCH* (Hao, 2001) was parallelized by introducing MPI library calls into the original Fortran code. Message Passing Interface (MPI) is one of the most favored packages for parallel computation. It was designed for high performance on both parallel machines and on network clusters. By calling a series of MPI library routines (see Table 1 for details), it was straightforward to split the computing task into approximately equal pieces and distribute them to all CPUs. In the case of a six-dimensional search involving three Eulerian angles, each ranging from 0 to 180°, and three translational parameters, each covering one half of the unit-cell dimension, it is natural to split one of these Eulerian angles into n pieces with an interval of $180°/n$ each. Here, the division of the Eulerian angles, n, is based on the resolution of the data and the dimensions of the search model. We adopted Master–Slave mode instead of Shared mode for all the data distribution and collection, *i.e.* one CPU as the server node for the control of the whole process. In order to make the program *CCP*4 compatible, we did not use the new features of the MPI-2 standard for parallel I/O libraries. Hence, *MPI_FSEARCH* can use unaltered standard *CCP*4 libraries. All the processes were controlled by a shell script, in which one can define runtime parameters such as the resolution, σ cutoff value, unit-cell parameters, search range, grid size and so on. After obtaining these control parameters, the server reads the diffraction data and the search model for initial processing including the envelope construction and the diffraction data analysis. The server then sends all the necessary data to each individual slave CPU and waits for the return of the results. At the completion of the required calculation, each slave CPU sends a series of rotational and translational parameters, sorted on R factor, to the server and then stops. Finally, the server resorts all these results and gives a final tabled list. Any intermolecular clashes can be checked during or after the computation by turning on or off an option in the control file. Our experience with coding *MPI_FSEARCH* may prove to be very useful in any future efforts to make computationally intensive *CCP*4 programs run faster by utilizing distributed computer clusters.

4. Results and discussion

Table 2
Molecular-replacement solution for phytase.

The three Eulerian angles and three translational parameters were exhaustively searched using the *MPI_FSEARCH* program. The results are given in ascending order of the R factor and the top solution was chosen to place the envelope in the unit cell.

α (°)	β (°)	γ (°)	x (Å)	y (Å)	z (Å)	R factor
63	20	145	28	10	32	0.412
60	20	145	28	12	32	0.444
60	15	150	28	12	32	0.461
51	15	160	26	12	34	0.462
57	20	150	28	12	32	0.463
60	15	150	28	12	34	0.463
48	15	160	28	14	34	0.467
63	15	145	28	12	34	0.468
60	20	150	26	10	32	0.469
69	15	140	30	10	32	0.472

The *MPI_FSEARCH* program was tested with X-ray diffraction data from the protein *Aspergillus fumigatus* phytase. Phytase is an enzyme that catalyzes the release of phosphate from phytic acid in microorganisms and plants. This protein crystallizes in space group $P4_32_12$, with unit-cell parameters $a = b = 70.30$, $c = 186.68$ Å. There is one molecule (48 kDa) per asymmetric unit. The diffraction data set to 1.6 Å resolution was collected at the Advanced Photon Source (APS). The structure had been solved using *AMoRe* (Navaza, 1994) with a homologous search model (PDB code 1ihp; Kostrewa *et al.*, 1997). The data processing and structural details will be reported elsewhere (PDB code 1n4z; Xiang *et al.*, 2003). The r.m.s. deviation between the search model (1ihp) and the final refined model (1n4z) is 0.73 Å for all backbone atoms calculated using the program *Swiss-PdbViewer* (Guex & Peitsch, 1997). In the present work, only 149 reflections, representing a data completeness of 98%, between 20 and 12 Å were used for low-resolution phasing and the results were then compared with the original results (Xiang *et al.*, 2003) obtained from molecular replacement at a much higher resolution (3 Å). It should be noted that *AMoRe* and *CNS* (Brünger *et al.*, 1998) failed to find the correct solution when the data resolution was lower than 10 Å. A coarse envelope was constructed from the search model (Kostrewa *et al.*, 1997) to simulate an envelope that would come from other sources. Firstly, the unit cell was covered with a cubic grid (initial grid-point values were set to zero). The grid size used for computation was 64 × 64 × 128, which is comparable with the unit-cell parameters in ångströms.

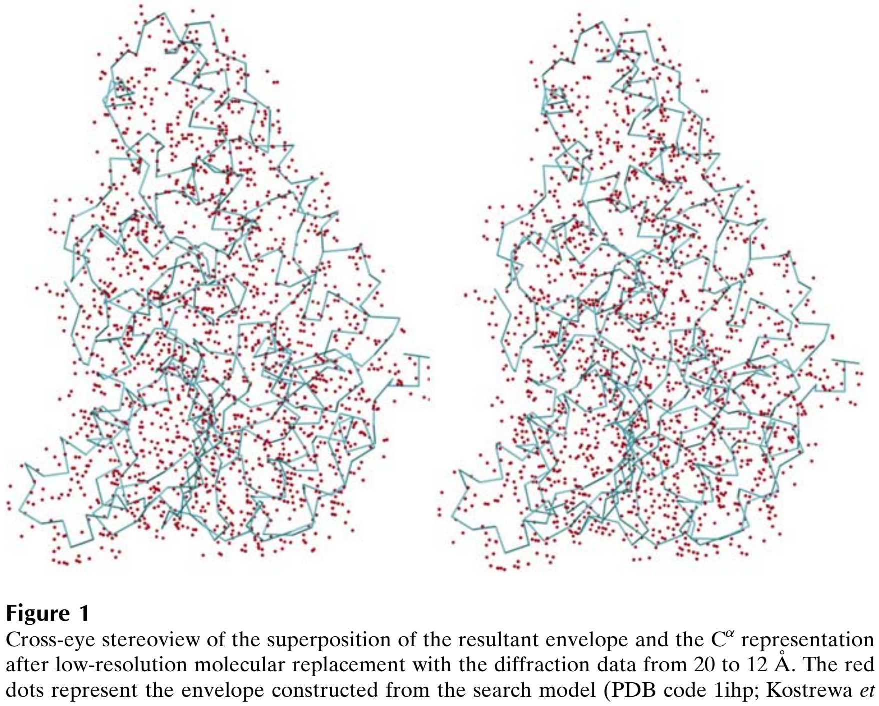

Figure 1
Cross-eye stereoview of the superposition of the resultant envelope and the C^α representation after low-resolution molecular replacement with the diffraction data from 20 to 12 Å. The red dots represent the envelope constructed from the search model (PDB code 1ihp; Kostrewa *et al.*, 1997), rotated and translated according to the top solution in Table 2. The blue C^α trace is drawn from the refined crystal structure of phytase (PDB code 1n4z; Xiang *et al.*, 2003).

Secondly, values for the grid points at the positions of each atom of the search model and its six nearest neighbors (~1 Å from the atom) were set to one. As mentioned in the previous section, we split the total task into 60 pieces according to the α angle in the Eulerian space with an increment of 3°. Each of these sub-tasks was then sent to one of the 60 slave processors by the server. At the same time, the server took charge of collecting and sorting all the results from the 60 slave processors. This was equivalent to performing a five-dimensional search on each slave processor. In each five-dimensional search, the initial search step was set to 5° for the Eulerian angles β and γ and 2 Å for all translational parameters. An *R*-factor filter was used to shorten the output lists in the temporary file before sorting. In our computation, we used 16 801 grid points to construct the envelope and all 149 measured low-resolution (20–12 Å) reflections. The entire computation took 322.5 min, less than 6 h. We believe this time frame is acceptable and *MPI_FSEARCH* can be used as a general method to perform low-resolution initial phasing with envelopes derived from PDB files, solution scattering or electron microscopy.

A number of peaks with low *R* factors were obtained after the exhaustive six-dimensional search. After checking the possible molecular-packing clashes, we kept those solutions with overlap less than 0.1% of the total unit-cell volume. Not surprisingly, in our test case the molecular overlap was zero for the correct solution. The search results are listed in Table 2. The correct solution stands out at the top of the list with an *R* factor of 0.412. There is a clear gap in *R* factors between the first two solutions, which is a good indication that the search has been successful. The top solution was further refined to $\alpha = 62$, $\beta = 19$, $\gamma = 146°$, $x = 28$, $y = 11$, $z = 32$ Å. The quality of the top solution can also be viewed (Fig. 1) directly by superimposing the search envelope transformed by *MPI_FSEARCH* and the final crystal structure of phytase. Although we only used low-resolution data, the envelope and the crystal structure match very well. This result suggests that the exhaustive six-dimentional search can be a successful molecular-replacement method at low resolution. In fact, the six-dimensional search is not sensitive to the data resolution. When the search was performed with data in a conventional molecular-replacement resolution range, say 10–4.0 Å, the correct solution could also be obtained at the cost of a longer computation time. A bulk-solvent correction (Fokine & Urzhumtsev, 2002) was attempted. However, neither an exponential scaling model nor a mask bulk-solvent model could improve our solution.

To further examine the capabilities of the program, the search model was broken into two equal pieces according to the amino-acid sequence and only one piece was fed into *MPI_FSEARCH*. Although the correct solution was not obvious, it was still among the first few peaks. This result suggests the potential of the program for locating more than one search model in the asymmetric unit or for use when only a partial model is available at low resolution.

The *FSEARCH* program was originally written for phasing from SAXS envelopes (Hao, 2001) instead of envelopes calculated from PDB files as in the present case, where the effective search-model resolution is much higher than the data resolution. We intend to develop a molecular-replacement technique that is capable of handling a variety of challenging situations when the data resolution is low and/or the search-model resolution is low. We expect that the *MPI_FSEARCH* program may be used to place an envelope from any source into the crystallographic unit cell and therefore to provide initial phase information. It is worth noting that if a non-crystallographic symmetry axis (normally determined by a self-rotation function) exists, the six-dimensional search is reduced to four: *MPI_FSEARCH* would also be applicable in this case and the computation would only take a few seconds on our SIRIUS cluster.

We are very grateful to Frank Labonte for his efforts in installation and maintenance of the SIRIUS cluster. We thank Marian Szebenyi for improving the manuscript. This work is supported by award RR-01646 from the National Institutes of Health, through its National Center for Research Resources.

References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Brünger, A. T. (1992). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR.* New Haven, Connecticut, USA: Yale University Press.

Brünger, A. T. (1997). *Methods Enzymol.* **176**, 558–580.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J., 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.

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

Diederichs, K. (2000). *J. Appl. Cryst.* **33**, 1154–1161.

Fokine, A. & Urzhumtsev, A. (2002). *Acta Cryst.* A**58**, 72–74.

Glykos, N. M. & Kokkinidis, M. (2000). *Acta Cryst.* D**56**, 169–174.

Glykos, N. M. & Kokkinidis, M. (2001). *Acta Cryst.* D**57**, 1462–1473.

Guex, N. & Peitsch, M. C. (1997). *Electrophoresis*, **18**, 2714–2733.

Hao, Q. (2001). *Acta Cryst.* D**57**, 1410–1414.

Hao, Q., Dodd, F. E., Grossmann, J. G. & Hasnain, S. S. (1999). *Acta Cryst.* D**55**, 243–246.

Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* D**55**, 484–491.

Kissinger, C. R., Gehlhaar, D. K., Smith, B. A. & Bouzida, D. (2001). *Acta Cryst.* D**57**, 1474–1479.

Kostrewa, D., Gruninger-Leitch, F., D'Arcy, A., Broger, C., Mitchell, D. & VanLoon, A. P. (1997). *Nature Struct. Biol.* **4**, 185–190.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Navaza, J. & Saludjian, P. (1997). *Methods Enzymol.* **277**, 581–594.

Ockwell, D. M., Hough, M. A., Grossmann, J. G., Hasnain, S. S. & Hao, Q. (2000). *Acta Cryst.* D**56**, 1002–1006.

Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.

Sheriff, S., Klei, H. E. & Davis, M. E. (1999). *J. Appl. Cryst.* **32**, 98–101.

Vagin, A. & Teplyakov, A. (1997). *J. Appl.Cryst.* **30**, 1022–1025.

Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* D**56**, 1622–1624.

Xiang, T., Deacon, A. M., Koshy, M., Kriksunov, I., Lei, X. & Thiel, D. J. (2003). In preparation.