Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Phased translation function revisited: structure solution of the cofilin-homology domain from yeast actin-binding protein 1 using six-dimensional searches

Boris V. Strokopytov,[a]* Alexander Fedorov,[a] Nicole M. Mahoney,[a]‡ Michael Kessels,[b] David G. Drubin[b] and Steven C. Almo[a]*

[a]Department of Biochemistry, Albert Einstein College of Medicine, Bronx, NY 10461, USA, and [b]Department of Molecular and Cell Biology, UC, Berkeley, USA

‡ Present address: Department of Pharmacology, UCSF, USA.

Correspondence e-mail: strokop@aecom.yu.edu, almo@aecom.yu.edu

Received 30 July 2004
Accepted 13 December 2004

PDB Reference: cofilin-homology domain of ABP1, 1hqz, r1hqzsf.

A modified molecular-replacement method is described that makes use of six-dimensional searches and the phased translation function, providing a systematic examination of all possible search-model orientations in an experimental electron-density map. As an example, the structure solution of the cofilin-homology domain of the *Saccharomyces cerevisiae* actin-binding protein 1 (ABP1) is presented in detail. Additional examples are presented in which these tools have significantly aided structure solutions in a variety of contexts. These results suggest that this approach might be of widespread utility for challenging structures involving weak phase information, complex asymmetric units and search models with weak structural homology. Furthermore, this approach supports an exhaustive molecular-replacement strategy in cases where an appropriate search model cannot readily be identified on the basis of sequence homology. The fully automated web-based implementation of this phased translation function is described.

1. Introduction

As the Protein Data Bank grows, molecular replacement (MR) is gaining increased importance for macromolecular structure solution (Rossmann & Blow, 1962; Crowther & Blow, 1967; Lattman & Love, 1970; Crowther, 1972). In general, MR seeks to relate a search model, with three rotational and three translational parameters, to each molecule in the asymmetric unit (ASU) (a total of $6N$ parameters, where N is the number of independent particles in the ASU). Hence, by its very nature MR takes the form of a multidimensional search; however, even the simplest six-dimensional searches can be computationally expensive. In practice, MR is typically parsed into two separate three-dimensional searches, the first determining rotational parameters, usually expressed as a set of three Eulerian angles, followed by a translational search that positions the correctly oriented model with respect to the unit-cell axes.

The assumptions and approximations implicit in MR are associated with practical difficulties in its successful implementation. Ideally, the rotation function would compare only self-Patterson vectors (*i.e.* vectors corresponding to pairs of atoms belonging to the same molecule); however, in practice this can only be realised for the Patterson function generated from the search structure. The observed Patterson calculated from experimental data will always contain a significant contribution from cross-vectors between symmetry-related molecules, even if limited to volumes close to the origin. Successful solution of the rotation function is further complicated if the ASU contains multiple copies of a molecule

or if additional molecular fragments with unknown structure are present (reviewed by Navaza, 1994). In unfavorable cases, an insufficient signal-to-noise ratio in the translation function prevents structure solution, even if the search model is correctly oriented (Read & Schierbeek, 1988). Furthermore, rotation and translation searches are highly sensitive to the structural/sequence similarity between the search model and the unknown structure, as the r.m.s. deviation typically must not exceed 1.0–1.5 Å.

To overcome the difficulties associated with unambiguously identifying the correct rotation-function solution, Navaza (1994) suggested the examination of all rotation function peaks greater than, for example, 50% of the maximum, while Tong (1996) proposed relaxing this limit to 10% of the maximum signal. These ideas led Sheriff *et al.* (1999) to consider the implementation of limited and exhaustive six-dimensional searches, entirely removing the dependence on finding a solution to the rotation function. However, even this approach is limited in cases where the ASU contains multiple molecules or the homology with the search model is low. In such cases, additional information, typically in the form of experimental phases, is required.

Well over a decade ago Read & Schierbeek (1988) described the phased translation function, which provides a powerful approach to overcome the limitations associated with conventional MR. This method seeks to maximize the overlap, as measured by some defined target function, of an experimental electron-density map with the electron density calculated from a single copy of a search model that has been correctly oriented in the unit cell *via* the rotation function. Given the correct orientation of the search model, even relatively poor phase information can dramatically aid in the placement of the search model in the ASU. A number of related criteria have been used to evaluate the phased translation function, including a standard linear correlation coefficient, CC(t), which is cast in terms of experimental and calculated electron density,

$$\mathrm{CC}(t) = \frac{\int_V [\rho_\mathrm{p}(\mathbf{x}) - \overline{\rho_\mathrm{p}}] \times [\rho_\mathrm{m}(\mathbf{x}-\mathbf{t}) - \overline{\rho_\mathrm{m}}]\,\mathrm{d}\mathbf{x}}{\left\{\int_V [\rho_\mathrm{p}(\mathbf{x}) - \overline{\rho_\mathrm{p}}]^2\,\mathrm{d}\mathbf{x} \times \int_V [\rho_\mathrm{m}(\mathbf{x}-\mathbf{t}) - \overline{\rho_\mathrm{m}}]^2\,\mathrm{d}\mathbf{x}\right\}^{1/2}} \quad (1)$$

where $\rho_\mathrm{p}(\mathbf{x})$ is the observed density and $\rho_\mathrm{m}(\mathbf{x})$ the model density; V is the unit-cell volume. Importantly, (1) can be recast in terms of reciprocal-space variables (*i.e.* structure factors) that allow the use of highly efficient Fast Fourier Transform (FFT) algorithms,

$$\mathrm{CC}(t) = \frac{(1/V)\sum_n m_\mathrm{p}|F_\mathrm{o}(\mathbf{h})|\exp(i\alpha_\mathrm{p}) \times F_\mathrm{m}^*(\mathbf{h})\exp(-2\pi i\mathbf{h}\mathbf{t})}{(1/V)\left\{\sum_n [m_\mathrm{p}|F_\mathrm{o}(\mathbf{h})|]^2 \times \sum_n [|F_\mathrm{m}(\mathbf{h})|]^2\right\}^{1/2}} \quad (2)$$

where $|F_\mathrm{o}(\mathbf{h})|$ is the observed structure-factor amplitude, $F_\mathrm{m}(\mathbf{h})$ is the model structure factor, α_p is the experimental phase, m_p is the figure-of-merit of the observed structure factor and n is the number of phased reflections.

Table 1
CHD data collection and reduction.

Values in parentheses correspond to the shell of data at highest resolution.

	Nat I	EMP	Nat II	Nat III
Unit-cell parameters				
a (Å)	159.4	160.3	156.9	156.5
b (Å)	68.0	68.3	66.6	66.7
c (Å)	127.6	127.6	125.8	125.1
β (°)	106.9	107.1	108.3	108.3
Data collection				
Wavelength (Å)	1.5418	1.5418	0.9794	1.0093
d_min (Å)	3.0	3.0	2.4	2.1
Completeness (%)	87.0 (59.0)	80.4 (47.0)	98.4 (82.4)	96.6 (92.6)
Completeness† (%)	98.6	98.3	—	—
Multiplicity	2.2 (1.6)	2.1 (1.8)	3.9 (3.1)	2.3 (2.2)
Mean $\langle I/\sigma(I)\rangle$	10.1 (2.0)	9.2 (1.7)	25.7 (9.8)	9.9 (1.9)
R_merge‡ (%)	7.2 (36.8)	8.4 (43.5)	3.3 (15.4)	6.9 (49.6)
No. of unique reflections	23937	21309	47054	68796

† Completeness in the 15–6 Å resolution range used for phased translation-function searches. ‡ The merging R factor, $R_\mathrm{merge} = \sum\sum_i |\langle I(h)\rangle - I(h)_i| / \sum\sum_i I(h)_i$, where $\langle I(h)\rangle$ is the mean intensity.

The application of the phased translation function requires both phase information and an appropriate search model and thus fills a unique niche where there is inadequate prior phase information for *de novo* structure solution, the lack of an appropriate search model for conventional MR or the ASU is too complex to allow conventional MR. Accordingly, its application has been rather limited and most investigators have focused on its application to fitting individual residues (Read & Moult, 1992) or elements of secondary structure (Cowtan, 1998). However, we believe that the phased translation function should be revisited as a major tool, as its application is likely to have significant utility in the era of structural genomics, which stresses high-throughput structure determination, as well as for the structure solution of multi-component macromolecular assemblies. We present a detailed description of the structure solution of the cofilin-homology domain (CHD; 141 amino acids) from the 65 kDa *Saccharomyces cerevisiae* actin-binding protein 1 (ABP1p; Drubin *et al.*, 1988), which has 22% sequence identity to yeast cofilin, a Ca^{2+}-insensitive pH-dependent actin-filament severing protein. This structure solution required the implementation of a six-dimensional search employing the phased translation function. In addition, we demonstrate the widespread utility of this approach to a number of structure solutions.

2. Methods

2.1. Crystallization and data collection

CHD crystals were grown by hanging-drop vapor diffusion using 22–28% PEG 4000, 0.3 *M* sodium citrate pH 5.6 as precipitant and have typical dimensions of 0.3 × 0.4 × 1.0 mm. Diffraction from these crystals is consistent with the monoclinic space group *C*2 (unit-cell parameters a = 159.4, b = 66.8, c = 127.6 Å, β = 106.9°) and extends to 2.8 Å using a rotating-anode X-ray source and to 2.0 Å using synchrotron radiation.

Native (Nat I) and ethylmercury phosphate (EMP) derivative data were collected at room temperature to 3.0 Å resolution on a Siemens X-1000 area detector using Cu $K\alpha$ radiation from a Rigaku RU-200 generator operating in fine-focus mode at 50 kV and 80 mA (Table 1). These data were processed using *XDS* (Kabsch, 1988). Two additional native data sets, Nat II and Nat III, were collected on beamline X9B at the National Synchrotron Light Source, Brookhaven National Laboratories (Upton, NY, USA) to 2.4 and 2.1 Å, respectively. Prior to synchrotron data collection, crystals were soaked in mother liquor containing 5% glycerol for 5–10 min and flash-cooled to 100 K using an Oxford low-temperature system. Synchrotron data were recorded on a MAR 345 image plate with 1.0° oscillations and typical exposure times of 120–180 s using a wavelength of 0.98 Å. Synchrotron data were processed using the *HKL* suite (Otwinowski & Minor, 1997). Further scaling and reduction of all data sets were performed with the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994). *TRUNCATE* (French & Wilson, 1978) was used (Nat II and Nat III) to replace negative intensities with small positive values according to Bayesian statistics.

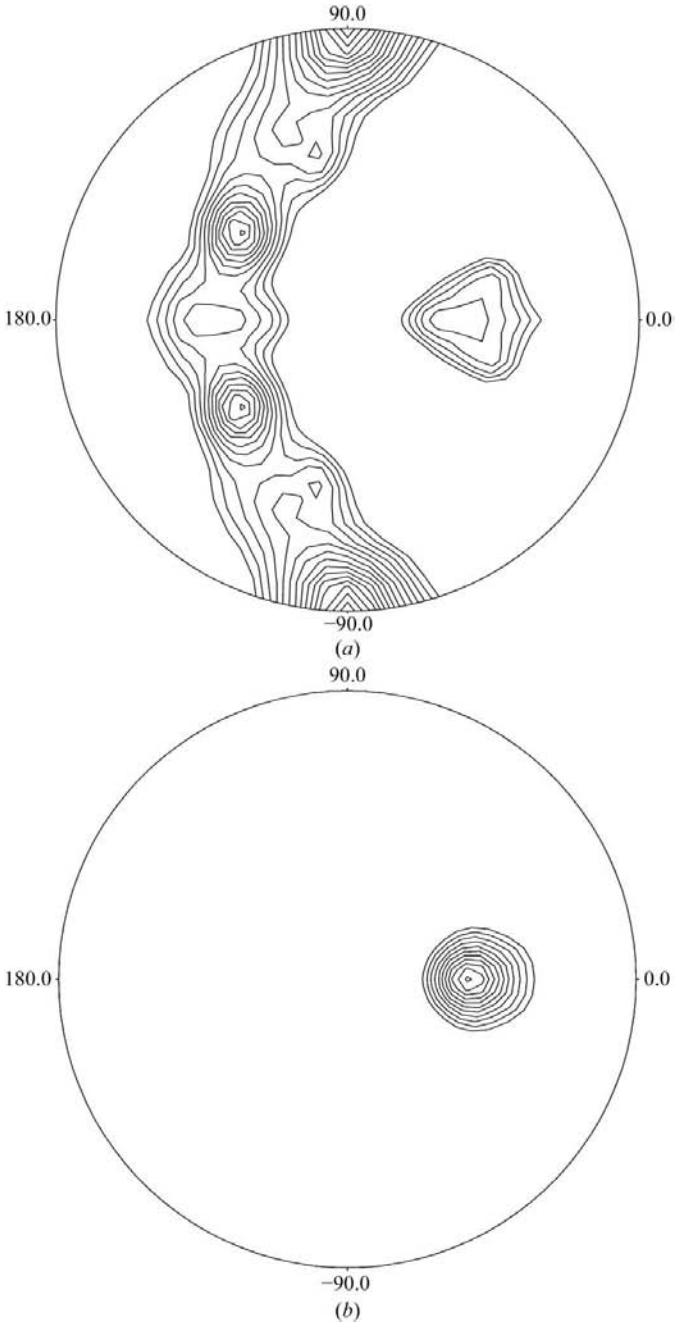

Figure 1
Self-rotation function in polar coordinates based on 20–4.0 Å data. The integration radius is 16 Å. (*a*) The $\kappa = 180^\circ$ section clearly shows the presence of a twofold non-crystallographic axis. (*b*) $\kappa = 120^\circ$ section. Threefold axis and orthogonal twofold axes simulate possible 32 point-group symmetry.

3. Results

3.1. Preliminary molecular-replacement analysis

The large unit-cell parameters and relatively small size of the protein suggested that the ASU contained 6–10 copies of the CHD. The self-rotation function (*POLARRFN*; Collaborative Computational Project, Number 4, 1994) calculated with data in the 10–4.0 Å resolution range revealed both twofold and threefold non-crystallographic symmetry (NCS) axes (Fig. 1). A peak on the native Patterson (*u*0*w* section) with height equal to 10% of the origin peak was consistent with a local twofold axis parallel to the crystallographic twofold axis (Fig. 2).

Initial attempts to solve the structure by conventional MR with *AMoRe* (Navaza, 1994) using the structure of yeast cofilin (PDB code 1cof; Fedorov *et al.*, 1997) as a search model failed, probably as a consequence of both the low homology of the search model (22% sequence identity) and the complex nature of the ASU. Examination of various parameters including

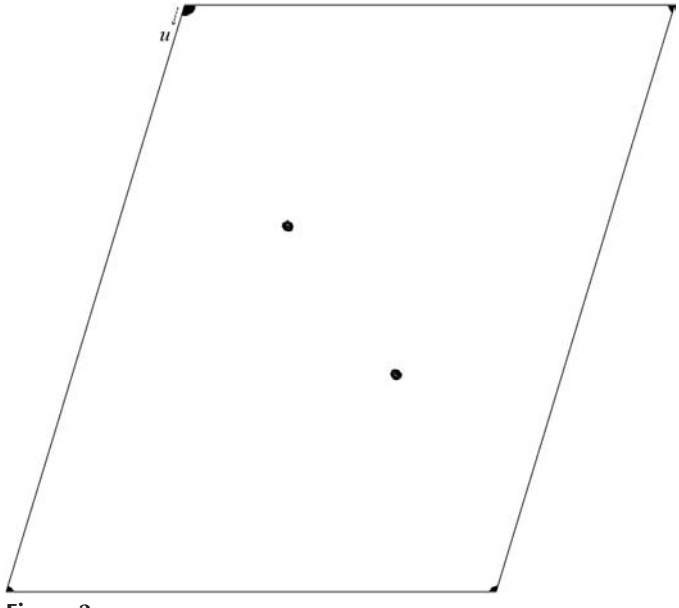

Figure 2
Native Patterson (15–6.0 Å data). Section $v = 0$. The second largest peak has height 10% of the origin peak. This indicates the presence of translational NCS symmetry or a twofold NCS axis that is almost parallel to the *y* axis.

Table 2
Phasing statistics for the EMP derivative at two different resolutions.

Resolution (Å)	15–3.0	15–6.0
$R_{\rm deriv}$†	0.154	0.140
$R_{\rm Cullis}$‡	0.83/0.61	0.53/0.39
$R_{\rm Cullis\ ano}$§	1.02	0.95
Mean figure of merit¶	0.263	0.510
Phasing power††	1.24/1.31	2.59/2.41

† The derivative-to-native scaling R factor, $R_{\rm deriv} = \sum \big||\mathbf{F}_{\rm PH}| - |\mathbf{F}_{\rm P}|\big| / |\mathbf{F}_{\rm P}|$, where $|\mathbf{F}_{\rm PH}|$ and $|\mathbf{F}_{\rm P}|$ are the derivative and native structure-factor amplitudes, respectively. ‡ The Cullis R factor, $R_{\rm Cullis} = \sum \big||\mathbf{F}_{\rm PH} - \mathbf{F}_{\rm P}| - |\mathbf{F}_{\rm H}|\big| / |\mathbf{F}_{\rm PH} - \mathbf{F}_{\rm P}|$, where $\mathbf{F}_{\rm H}$ is the calculated heavy-atom structure factor. Pairs of values are for centric and centric reflections, repectively. § The Cullis anomalous R factor, $R_{\rm Cullis\ ano} = \langle \text{lack-of-closure}_{\rm ano} \rangle / \langle \| \mathbf{F}_{\rm PH(+)} - \mathbf{F}_{\rm PH(-)} \| \rangle$, where $|\mathbf{F}_{\rm PH(+)}|$ and $|\mathbf{F}_{\rm PH(-)}|$ are the amplitudes of the positive and negative counterparts of the Bijvoet pair. ¶ The mean figure of merit is equal to the estimated mean cosine of the phase error. †† The phasing power is defined by $\langle |\mathbf{F}_{\rm H}| \rangle / \langle (\text{lack-of-closure}) \rangle$ summed over the reflections used in heavy-atom refinement. Pairs of values are for acentric and centric reflections, respectively.

resolution and integration radii did not improve the low contrast of this Patterson-based translation function. [A retrospective analysis with *PHASER*-II (Read, 2001; Storoni *et al.*, 2004) and *EPMR* (Kissinger *et al.*, 1999, 2001) failed to yield a solution.] At this stage, it was apparent that methods involving the use of heavy-atom derivatives were required in order to progress with the structure determination.

3.2. SIRAS analysis

After extensive screening, a single EMP derivative was obtained by soaking crystals in mother liquor containing 0.1 m*M* ethylmercury phosphate for 24 h. These data were collected at room temperature. Heavy-atom positions were identified by isomorphous difference Patterson syntheses (Fig. 3) and heavy-atom parameters were refined with *MLPHARE* (Otwinowski, 1991). Phases were calculated with the inclusion of the anomalous scattering contribution, which was particularly important as the heavy-atom constellation was found to be centrosymmetric. Owing to non-isomorphism between this derivative and the native data set (Nat1), the SIRAS map at 3.0 Å resolution was of poor quality (figure of merit of 0.26) and precluded direct model building. Attempts to locate NCS symmetry by visual inspection failed and hence precluded further attempts to improve the experimental density. However, the higher quality of the SIRAS phases at 6 Å resolution (figure of merit of 0.51; Table 2) and the availability of a model (albeit with only ∼20% identity) suggested that the phased translation function might have utility.

3.3. Phased translation function

Typically, the implementation of the phased translation function requires that the search model be correctly oriented. In the present case, however, the complexity of the ASU and the weak homology (∼20% identity) between the CHD and the cofilin search model prevented successful interpretation of the cross-rotation function. To circumvent this difficulty, a standard *X-PLOR* protocol (Brünger, 1992) for the phased translation function was modified to allow for the systematic examination of all possible orientations of a search model consisting of the polyalanine trace of yeast cofilin. The three search angles were incremented in 10° steps, with a translational increment of $d_{\rm min}/3$ (*e.g.* 2 Å grid at 6.0 Å resolution). Utilizing data (room-temperature Nat1 and EMP derivative) in the 15–6 Å resolution range in space group *P*1 required the calculation of 23 328 (36 × 18 × 36) individual phased translation functions. Owing to the formulation of the phased translation function as an FFT, the implementation of this exhaustive six-dimensional search required 22 min of CPU time on a PowerEdge 8450 server equipped with eight 700 Mhz Pentium III processors; each processor had sufficient physical memory (1 Gb) to keep the entire map in memory, which enhances the overall performance of the calculation.

The phased translation maps were normalized according to their σ values and the five highest peaks from each phased translation-function map were stored for further analysis. This six-dimensional search was performed twice as the correct handedness of the map was not known; the top peaks for the second run with swapped phases ($\alpha_{hkl} \rightarrow -\alpha_{hkl}$) were systematically higher than those obtained with the original phase set (8.6–8.2σ *versus* 6.3–6.2σ). After rejection of the weakest solutions (<5.5σ) and filtering of symmetry-related solutions, the top eight solutions (molecules 1–8) were consistent with the self-rotation function and the native Patterson. These eight molecules packed to yield four independent dimers in the ASU, all of which shared a similar dimer interface. Additional solutions were analyzed and rejected as unacceptable owing to their overlap with higher ranked solutions.

A number of algorithms designed to filter out the numerous false solutions were essential for the realisation of this new approach. Potential solutions are assigned a weight that is proportional to their real-space correlation coefficient and a

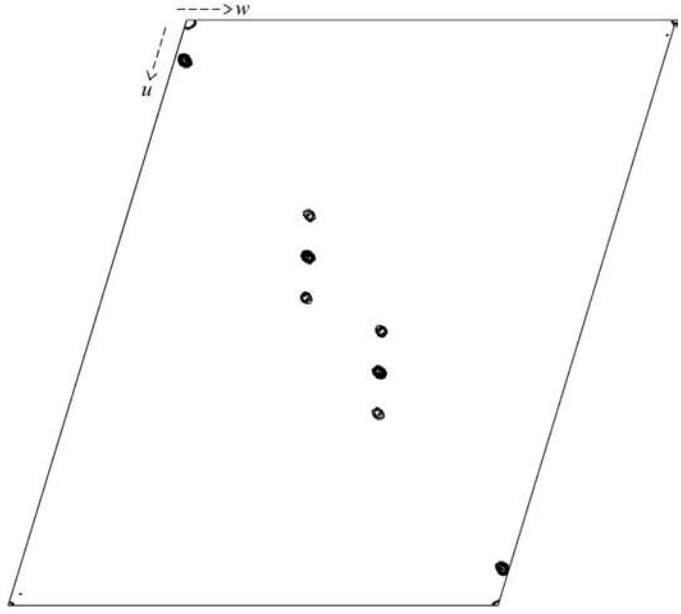

Figure 3
Heavy-atom difference Patterson. Harker section v = 0, showing both heavy atoms lying in the same plane. Stronger peaks correspond to cross-vectors between heavy-atom sites. Weaker sets of peaks correspond to self-vectors.

real-space packing function utilizing the true geometric features (*i.e.* shape) of the search model allows the accurate calculation of the volume overlap between potential solutions. Subsequently, a procedure similar to the greedy maximum independent set search (Feo *et al.*, 1994) is employed. This process involves the examination of all pairs of solutions having a volume overlap above a certain threshold value (typically around 6%) and results in the rejection of the solution having the lower weight (correlation coefficient). An important assumption in this protocol is that the correct solution contributing to an overlapping pair will always have a higher correlation coefficient than the false solution. This process rapidly identifies a set of non-overlapping solutions that maximizes the sum of weights (real-space coefficients). Like all greedy algorithms, this approach does not guarantee an exact solution; however, it has proven sufficiently robust to solve numerous structures in a nearly fully automated fashion. A final check is performed by evaluating the Patterson correlation coefficient as each individual solution (molecule) is added to the solution set; if the addition of a particular molecule results in a significant decrease in the Patterson correlation coefficient then it is excluded from the final set.

3.4. Structure refinement

Manual inspection of the eight phased translation-function solutions with the program *O* (Jones *et al.*, 1991) verified that the overall packing was good, with the only bad contacts involving an extended loop corresponding to residues 73–76 of cofilin in some of the molecules. These residues were removed from the model and subsequently rebuilt during refinement. Given the placement of these eight molecules, molecular envelopes were readily determined (*NCSMASK* from the *CCP*4 package) and eightfold NCS averaging with *DM* (Cowtan & Main, 1998) was used to extend the phases to 3.0 Å resolution. The averaging procedure decreased $R_{\rm free}$ from 0.54 to 0.29, with an increase of the average correlation coefficient of the masked regions from 0.29 to 0.86, and resulted in a high-quality map based on the Nat I room-temperature data set. Rigid-body refinement and additional cycles of averaging/phase extension against the 2.4 Å synchrotron data (Nat II) yielded a map (Fig. 4) which readily allowed the entire primary sequence to be built into molecule 1 using the yeast cofilin as a guide during tracing. The NCS transformations were applied to the chain trace for molecule 1 to provide starting models for molecules 2–8.

At this point, data set Nat III, which extended to 2.1 Å, became available and was used for further refinement with *X-PLOR* (Brünger, 1992) and *CNS* (Brünger *et al.*, 1998). Rigid-body refinement and crystallographic refinement including alternating cycles of simulated annealing (Brünger *et al.*, 1990) and conventional least-squares refinement resulted in an $R_{\rm cryst}$ of 0.301 and an $R_{\rm free}$ of 0.401 and somewhat surprisingly the refinement appeared to have stalled. However, analysis of *SIGMAA*-weighted ($mF_{\rm o} - DF_{\rm c}$) maps (Read, 1986) revealed a large feature of density corresponding to an entire additional CHD. Re-examination of the list of six-dimensional search solutions identified a solution corresponding to the missing molecule with a peak height of 5.6σ, which was below the arbitrary threshold level of 6.2σ set for further consideration. Inclusion of the ninth molecule (molecule 9) in the refinement, correction of loop regions, isotropic bulk-solvent correction and addition of solvent molecules resulted in an $R_{\rm cryst}$ of 0.213 and an $R_{\rm free}$ of 0.259 for all data between 20 and 2.1 Å resolution (Table 3). The phase-angle differences between the final model and various stages in the structure solution are given in Table 4. The overall structure of the CHD is similar to that of cofilin (Fig. 5), with an r.m.s. deviation of 2.0–2.2 Å (see Table 5). Pairwise comparison of all independent dimers shows that dimer *A*, formed by molecules 1 and 2, has the least compact interface (Tables 7*a* and 7*b*), suggesting why heavy-atom binding sites are only present in the *A* dimer (Tables 6 and 7). Dimer *E*, formed by molecule 9 and its symmetry mate, is roughly parallel to both dimer *A* and the crystallographic twofold axis, consistent with the significant peak on the native Patterson. Proper threefold rotation and threefold screw axes are not present in the ASU; however, certain pairs of dimers (*e.g.* dimers *A* and *B*) in the ASU are related by approximately 120° rotation, consistent with the apparent threefold local symmetry suggested by the self-rotation function. The final packing in the ASU is shown in Fig. 6. Following this initial application, the six-dimensional phased translation function has recently been successfully applied to other challenging structure determinations and has significantly facilitated the determination of a number of more routine structures (see Table 8).

Table 3
Refinement statistics.

Resolution (Å)	20.0–2.1
Completeness (%)	96.7
Data cutoff	$F/\sigma(F) > 2$
No. of non-H atoms	9669
No. of solvent atoms	521
B value from Wilson plot (Å^2)	29.50
R.m.s. deviations from ideal values	
Bond lengths (Å)	0.006
Bond angles (°)	1.2
Dihedral angles (°)	23.70
Improper angles (°)	0.79
Completeness after cutoff (%)	94.9
No. of reflections used	67961
R value	0.213
$R_{\rm free}$	0.259
Statistics for the highest resolution shell (2.23–2.10 Å)	
Completeness (%)	87.8
R value	0.309
$R_{\rm free}$	0.358
Estimated mean error in coordinates from *SIGMAA* (Å)	0.320

3.5. Poly(A)-binding protein

The poly(A)-binding protein (PABP) consists of two homologous 81-residue RNA-binding domains, termed RRM1 and RRM2. Crystals of PABP exhibit diffraction consistent with the monoclinic space group $P2_1$ (unit-cell parameters a = 79.6, b = 176.8, c = 82.5 Å, β = 117.7°) with 16 RRM

Table 4
Phase-angle differences for structure determinations.

(*a*) CHD structure.

Resolution (Å)	MIR *versus* refined model (°)	8 molecules poly-Ala model *versus* refined† (°)	NCS-averaged MIR density for 9 molecules *versus* refined model‡ (°)
20–6	65	69 (68)	40
20–5	68	72 (71)	39
20–4	72	75 (75)	35
20–3	77	81 (81)	37

(*b*) Poly(A)-binding protein.

Resolution (Å)	MIR *versus* refined model (°)	16 RRM model§ *versus* refined (°)	NCS-averaged MIR density *versus* refined (°)
20–6	60	64	50
20–5	64	68	53
20–4	71	71	61
20–3.4	77	74	69

† Phase differences for nine molecules are given in parentheses. ‡ Calculated *post factum*. § Reconstructed using *MODELLER* 3.

Table 5
Comparison of ABP1p CHD (1hqz) with native cofilin (1cof).

Chain	1	2	3	4	5	6	7	8	9
No. of C^{α}—C^{α} pairs aligned	128	128	126	127	126	127	124	124	126
R.m.s.d.† (Å)	2.2	2.1	2.1	2.2	2.1	2.1	2.1	2.0	2.1

† As given by the DALI server (Holm & Sander, 1993).

Table 6
R.m.s.d. (Å $\times 10^3$) of superimposed monomers of ABP1p CHD.

Residues 3–135 were aligned in Tables 7 and 8.

(*a*) C^{α} atoms only.

Chain	1	2	3	4	5	6	7	8	9
1	—	412	382	438	330	360	599	326	371
2		—	637	648	462	469	611	451	416
3			—	355	322	349	602	271	392
4				—	358	374	624	380	384
5					—	313	581	238	209
6						—	491	259	320
7							—	478	615
8								—	301
9									—

(*b*) All atoms.

Chain	1	2	3	4	5	6	7	8	9
1	—	670	743	869	730	785	1010	697	737
2		—	822	957	697	808	931	719	667
3			—	745	616	737	848	514	693
4				—	731	709	824	618	768
5					—	687	863	568	593
6						—	755	627	700
7							—	735	869
8								—	599
9									—

domains in the asymmetric unit (eight copies of a two-domain protein; Deo *et al.*, 1999). MIR phase information extending to 4 Å resolution, with SIR phases to 3.4 Å resolution, yielded an overall figure-of-merit from *SHARP* (de La Fortelle & Bricogne, 1997) of 0.37 for acentric and 0.30 for centric reflections to 3.4 Å. Consistent with these statistics, the quality of the resulting map was poor and did not lend itself to facile interpretation.

To implement the six-dimensional phased translation function, a search model was constructed with the program *MODELLER*-3 (Sali & Blundell, 1993), using an alignment of the PABP RRM1 with the previously solved RRM1 of hnRNP A1 (22% sequence identity; PDB code 1ha1). The six-dimensional phased translation function search with data in the 15–5.0 Å resolution range initially identified eight strong solutions (between 8.0 and 14σ) that gave an elegant octameric structure which ultimately corresponded to the eight RRM1 domains in the ASU. Additional non-overlapping solutions could not be distinguished from noise. To simplify the search for the remaining solutions, a mask was constructed around the first eight RRM1s (*SFALL*; Collaborative Computational Project, Number 4, 1994) and these features were subtracted from the original electron density. The resultant density was then used to back-calculate structure factors and phases, which were employed for a second six-dimensional phased translation search with the same search model and resolution range. This search yielded eight additional solutions (peak heights 12–14σ) which exhibit the same NCS symmetry as the first eight solutions and corresponded to the RRM2 domains of PABP. Model building was readily completed using eightfold averaged $2F_o - F_c$ maps produced by *CNS* (Brünger *et al.*, 1998). In the final refined structure, the RRM2 domains exhibited weaker electron density and higher *B* factors, consistent with the inability to identify them in the initial search. The final R_{cryst} and R_{free} were 0.230 and 0.304, respectively (PDB code 1cvj). The phase-angle differences between phases for the

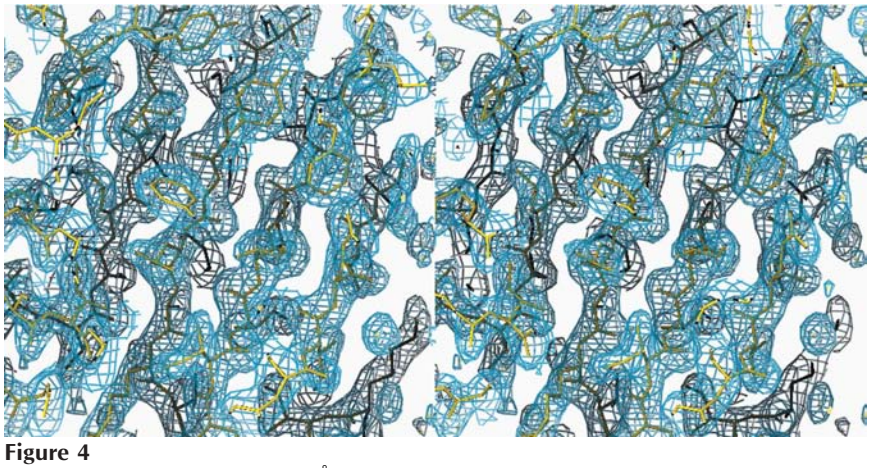

Figure 4
Averaged electron density at 2.1 Å resolution. A part of the CHD central β-sheet is shown formed by residues 32–37, 43–49, 65–71 and 80–86.

final model and various steps in the structure solution are given in Table 4. The r.m.s.d. between the final model and the search model is 1.5 Å for the RRM1 and 1.8 Å for the RRM2 domain (based upon 78 aligned C^{α} atom pairs for both cases). Additional details regarding this structure solution and the refined model are reported in Deo *et al.* (1999).

3.6. Internet development

The six-dimensional phased translation-function search described in this paper has been successfully applied to the solution of over 20 structures (see additional selected examples in Table 8). In addition to the challenging structures described above, this approach has significantly aided in a number of more routine structure solutions. In particular, it has allowed the rapid and unbiased placement of multiple copies of a model in an ASU for subsequent averaging or as a guide for model building.

Figure 5
Superimposed structures of cofilin (in yellow) and ABP1p CHD (in green) shown as C^{α} traces.

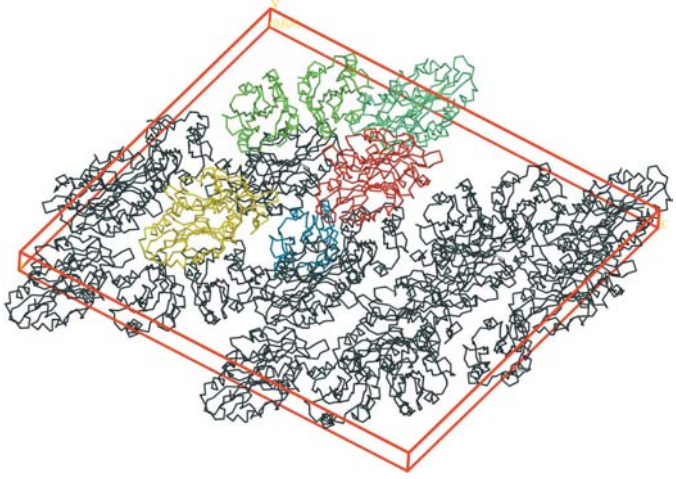

Figure 6
Packing of protein molecules in the unit cell of CHD ABP1 crystals. ASU contains nine molecules arranged as four and a half dimers (shown in colour). The ninth CHD molecule is positioned along the crystallographic twofold axis, resulting in a 'perfect' dimer.

Table 7
R.m.s.d. (Å $\times 10^3$) of superimposed dimers of ABP1p CHD.

Dimer definitions: dimer *A* includes monomers 1 and 2, *B* monomers 3 and 4, *C* monomers 5 and 6, *D* monomers 7 and 8 and *E* monomers 9 and 10. Monomer 10 was generated by applying operator $(-x, y, -z)$ to monomer 9 and has been shifted by (1, 0, 1). Cyclical permutation was applied to the second dimer, *i.e.* instead of comparing monomers 1 *versus* 3 and 2 *versus* 4 we superimpose monomers 1 *versus* 4 and 2 *versus* 3 *etc*. This 'permuted' r.m.s.d. is presented in the lower left table triangle. There is no unique way to compare dimers consisting of (even slightly) different monomers. The last column and the last row contain the same results owing to strict crystallographic symmetry imposed on dimer *E*. In this case permutation does not produce different results since the two monomers in the dimer are identical.

(*a*) C^{α} atoms only.

Dimer	*A*	*B*	*C*	*D*	*E*
A	—	2220	2215	2132	2154
B	2182	—	489	834	507
C	2221	486	—	820	289
D	2131	806	779	—	836
E	2154	507	289	836	—

(*b*) All atoms.

Dimer	*A*	*B*	*C*	*D*	*E*
A	—	2344	2310	2225	2222
B	2288	—	757	997	808
C	2309	804	—	1032	660
D	2212	937	962	—	1023
E	2222	808	660	1023	—

To make this powerful method readily accessible to the crystallographic community, we have developed a new package called *BRUTEPTF* (brute-force phased translation function). *BRUTEPTF* has been implemented as an internet-based menu-driven program and parallelized to run on an eight-processor PowerEdge 8450 server operating under Red Hat Linux. *BRUTEPTF* is accessible at http://russel.bioc.aecom.yu.edu/server/NYSGRC.html and requires only the downloading of the appropriate reflection file and search model in MTZ and PDB format, respectively. The robust filtering procedures described above minimize the intervention by the experimentalist and result in the almost complete automation of structure determination by this implementation of the phased translation function. Some features of *BRUTEPTF* are similar to *FFFEAR* developed by K. Cowtan (1998). To date, the server has logged more than 6000 hits (since July, 2000) and run more than 1000 jobs, reflecting a high level of interest in this promising method of structure solution.

3.7. Other potential areas for application of the phased translation function

The present implementation of the phased translation function is expected to have a direct impact on ongoing

Table 8
Selected structure determination using six-dimensional phased translation searches.

Protein name	ASU MW (kDa)	Space group	Unit-cell parameters (Å, °)	Sequence identity of search model† (%), model PDB code	Current status	Comments
Human macrophage-capping protein	38.5, three related ~13 kDa domains	$P6_322$	a = 206.7, c = 56.2	55.5 (average), 1eqy	PDB code 1j72, R = 0.243, R_{free} = 0.299 at 2.5 Å	Search map: d = 4.5 Å‡, FOM = 0.65; three 15 kDa domains correctly positioned
Receptor-binding domain of human B7-2	Two copies of 12.8 kDa	$P2_12_12$	a = 56.7, b = 63.0, c = 58.6	11.0, 1qa9	PDB code 1ncn, R = 0.214, R_{free} = 0.278 at 2.7 Å	Search map: d = 3.0 Å, FOM = 0.63
Human CTLA-4 dimer	Two copies of 13.1 kDa	$P2_12_12_1$	a = 42.8, b = 53.32, c = 115.83	67.0, 1qdt	PDB code 1lx7, R = 0.182, R_{free} = 0.215 at 2.0 Å	Manually built one monomer (27 kDa) into SAD density, *BRUTEPTF* was used to place second monomer into 4 Å density.
S. cerevisiae myoinositol phosphate synthase	Four copies of 60 kDa	$P2_12_12_1$	a = 155.7, b = 187.3, c = 98.9	100.0	PDB code 1la2, R = 0.224, R_{free} = 0.280 at 2.65 Å	Manually built one monomer (60 kDa) into SAD density, *BRUTEPTF* was used to place the remaining three molecules into density at 4.5 Å resolution
Guanyltransferase	Two copies of 44 kDa	$C2$	a = 47.1, b = 125.0, c = 92.7, β = 100.5	27.0	In refinement, R = 0.222, R_{free} = 0.287	Used polyserine model to dock two monomers into interpretable 4 Å SAD density
A. thaliana fimbrin	Two copies of ~60 kDa, each chain composed of four tandem CH domains	$P2_1$	a = 63.4, b = 104.1, c = 106.3, β = 104.3	~30–50% for the various CH domains, 1aoa	PDB code 1pxy	The first two human CH domains were fitted to 3.0 Å MAD density
Chicken vinculin	Complete chicken vinculin molecule, ~120 kDa	$C222_1$	a = 56, b = 127, c = 352	1qkr, 1dov, 1h6g	PDB code 1st6	Fitted three independent domains into 4.5 Å MAD density

† With the nearest homologue available. ‡ High-resolution limit for experimental electron-density map.

structural genomics efforts, as the primary focus of these high-throughout structure-discovery initiatives is the structural characterization of new sequence families. Based on our results, it is likely that models arising from a particular sequence family (*i.e.* members of which all share >30% sequence identity) will be useful in the structure solution of other sequence families (*i.e.* <30% sequence identity with model sequence family) that reside within a common fold family. The phased translation function is also predicted to find increasing use in the structure solution of large multi-component complexes, especially as the number of unique individual protein structures continues to grow. Furthermore, these approaches are likely to facilitate the fitting of substructures into moderate- and low-resolution EM reconstructions and Chiu *et al.* (1999) have implemented a real-space six-dimensional search method for this purpose.

The extreme economy of the phased translation function also provides the opportunity for the generalization of 'exhaustive' molecular replacement when no obvious model exists, as a large number of representative folds in the PDB can be used as search models in cases where some phase information is available. In preliminary work at the New York Structural Genomics Research Consortium (NYSGRC), we examined the feasibility of this 'exhaustive' approach on the newly solved structure (PDB code 1b54) of an open reading frame from *S. cerevisiae* that displayed the classical TIM barrel, but which had no detectable sequence similarity with any structure in the PDB. Ten different TIM-barrel search models that displayed widely varying sequence and structural similarity to the unknown structure (2–13% sequence identity, r.m.s. deviations of 2.7–4.5 Å and DALI Z scores that ranged from 18.7 to 3.5) were used as search models against a map calculated from a single derivative (FOM = 0.31). All calculations were performed with data in the 15–6 Å range and poly-Ala traces derived from the search models. These preliminary efforts emphasize the remarkably robust nature of the translation function, as clear solutions were generally obtained for models that displayed as much as a 3.5 Å r.m.s. deviation from the unknown structure. This approach is strengthened by virtue of the fact that multiple independent models with no sequence homology yield similar solutions, thus providing an assessment of internal consistency. While it remains a major challenge to produce a refined structure from starting models with r.m.s. deviations in excess of 3.0 Å, these observations highlight the utility of the phased translation function as a possible tool for fold recognition. Further tests to examine the limits of this approach are ongoing.

4. Conclusion

We have shown that a systematic examination of search-model orientation allows successful utilization of the phased translation function in several challenging structure determinations, as well as in more routine applications. Although this

approach requires phase information for the calculation of an experimental map, the quality of the observed phases may be relatively low, with even a single poor derivative proving to be sufficient for initial low-resolution work. It is quite satisfying that the method is able to locate molecules at low resolution as this reduces the need for high-quality isomorphous derivatives and allows the use of a coarse search grid (*e.g.* 10° rotational, 2.0 Å translational increments), which significantly reduces CPU requirements. Of particular significance is the effective utilization of weakly homologous search models (*i.e.* 20–30% identity) with r.m.s. deviations in the range >2–3 Å. This is in contrast to conventional MR, which requires a much greater degree of similarity between the search model and the unknown (*i.e.* typically <1.0–1.5 Å r.m.s. deviation).

In summary, we have successfully applied a computationally efficient exhaustive six-dimensional search to challenging structure determinations with complex ASUs, weakly homologous search models and relatively poor phase information. This approach is likely to positively impact on the structural analysis of multi-component assemblies, as well as some classes of structural genomics targets. The availability of a web-based tool allows easy utilization with a minimum of user intervention.

We thank anonymous referees for their useful comments. Special thanks go to Dr Z. Dauter for help with data collection and processing at NSLS station X9B.

References

Brünger, A. T. (1992). *X-PLOR Version 3.1. A System for X-ray Crystallography and NMR*. New Haven: Yale University Press.
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.
Brünger, A. T., Krukowski, A. & Erickson, J. W. (1990). *Acta Cryst.* A**46**, 585–593.
Chiu, W., McCough, A., Sherman, M. B. & Schmidt, M. F. (1999). *Trends Cell Biol.* **9**, 154–159.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1998). *Acta Cryst.* D**54**, 750–758.
Cowtan, K. & Main, P. (1998). *Acta Cryst.* D**54**, 487–493.
Crowther, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 183–187. New York: Gordon & Breach.
Crowther, R. A. & Blow, D. M. (1967). *Acta Cryst.* **23**, 544–548.
Deo, R. C., Bonanno, J. B., Sohenberg, N. & Burley, S. K. (1999). *Cell*, **98**, 835–845.
Drubin, D. G., Miller, K. G. & Botstein, D. (1988). *J. Cell Biol.* **107**, 2551–2561.
Fedorov, A. A., Lappalainen, P., Fedorov, E. V., Drubin, D. G. & Almo, S. C. (1997). *Nature Struct. Biol.* **4**, 366–369.
Feo, T. A, Resende, M. G. C. & Smith, S. H. (1994). *Oper. Res.* **42**, 860–878.
French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.
Holm, L. & Sander, C. (1993). *J. Mol. Biol.* **233**, 123–138.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
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.
La Fortelle, E. de & Bricogne,G. (1997). *Methods Enzymol.* **276**, 472–494.
Lattman, E. E. & Love, W. E. (1970). *Acta Cryst.* B**26**, 1854–1857.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphus Replacement and Anomalous Scattering*, edited by W. Wolf & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–325.
Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.
Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.
Read, R. J. & Moult, J. (1992). *Acta Cryst.* A**48**, 104–109.
Read, R. J. & Schierbeek, A. J. J. (1988). *J. Appl. Cryst.* **21**, 490–495.
Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
Sali, A. & Blundell, T. L. (1993). *J. Mol. Biol.* **234**, 779–816.
Sheriff, S., Herbert, E. K. & Davis, M. E. (1999). *J. Appl. Cryst.* **32**, 98–101.
Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* D**60**, 432–438.
Tong, L. (1996). *Acta Cryst.* A**52**, 782–784.