The use of *ACORN* in solving a 39.5 kDa macromolecule with 1.9 Å resolution laboratory source data

V. Rajakannan,[a] S. Selvanayagam,[a] T. Yamane,[b] T. Shirai,[c] T. Kobayashi,[d] S. Ito[e] and D. Velmurugan[a]*

[a]Department of Crystallography and Biophysics, University of Madras, Guindy Campus, Chennai 600 025, India, [b]Department of Biotechnology and Biomaterial Science, Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan, [c]Department of Computational Biology, Biomolecular Engineering Research Institute, Furuedai 6-2-3, Suita, Osaka 565-0874, Japan, [d]Tochigi Research Laboratories of Kao Corporation, 2606 Alkabane, Ichikai, Haga, Tochigi 321-3497, Japan, and [e]Japan Marine Science and Technology Center, 2-15 Natushima, Yokosuka 237-0061, Japan. E-mail: d_velu@yahoo.com

Data from the alkaline cellulase apo form were collected at a resolution of 1.9 Å using an in-house X-ray source (Cu $K\alpha$). By using different fragments of helices from the model solved by macromolecular crystallographic means, the direct-methods program *ACORN* was used to arrive at the complete model. Attempts have been made to use various percentages of input phasing information from these helices. The minimum input phasing required in feeding the fragments was about 14% of the whole structure. The phases obtained from *ACORN* were of superb quality, allowing automated model building to be carried out using *ARP/wARP*. Minimal manual model building was required and the structure determination was completed using the maximum-likelihood refinement program *REFMAC*. The whole process, starting from the running of *ACORN* and ending with the refined model, took nearly 15 h of CPU time using a Pentium III PC.

Keywords: alkaline cellulase; *ACORN*; 1.9 Å resolution; macromolecules; *ARP/wARP*; in-house X-ray sources.

1. Introduction

Ab initio solutions of the crystal structures of small molecules are possible by using atomic-resolution diffraction data, usually at ~0.8 Å. Most of these small molecular crystal structures are usually solved using direct-methods programs. The estimation of the linear relationships among phases forms a major part of direct methods. However, the probabilistic estimates of these relationships become invalid as the number of atoms in the unit cell becomes very large (*e.g.* thousands of atoms). Macromolecules come under this category. Also, the diffraction data available for macromolecular crystals are not usually at an atomic resolution. For these reasons, direct methods cannot be used to solve macromolecules.

During the last decade, admirable advances have taken place in the data-collection facilities and techniques available to the macromolecular crystallographer. With advances such as more intense X-ray sources, in particular dedicated synchrotron beamlines, highly efficient two-dimensional detectors in the form of imaging plates and, more recently, charged-couple devices, and cryogenic nitrification to alleviate the effects of radiation damage and extend the resolution of data accessible, the average number of structures deposited in the Protein Data Bank (Berman *et al.*, 2002) per week is around 75 at present, and around 25000 structures have so far been deposited. With the above advances, more data sets appear to be coming from atomic-resolution data. The above possibility of gaining atomic-resolution data even for macromolecules prompted the direct-methods practitioners to make attempts to extend the direct methods using other macromolecular techniques (using anomalous scattering/density modification approaches *etc.*) to enable them to tackle the structure solution of macromolecules. *ACORN* is a comprehensive and efficient phasing procedure involving direct methods for the determination of protein structures when atomic-resolution data are available (better than 1.2 Å).

2. Description of the program

Foadi *et al.* (2000), Foadi (2003) and Yao (2002) have described how *ACORN* can generate phases that are good enough to reveal the whole molecule from a slightly 'better-than-random' starting set. This initial phasing can be generated: (i) from a correctly positioned structural fragment (obtained from molecular replacement methods), (ii) from a few heavier atoms such as a metal or the sulphur atoms of a molecule, (iii) in favourable cases by testing many randomly positioned atoms or selecting the most promising for phase extension, (iv) from a small fragment of a known structure with some sequence homology or (v) from an idealized secondary structural element such as an alpha helix or small section of a beta sheet. The use of *ACORN* to determine substructures has recently been detailed (Dodson & Yao, 2003). It has even been successful in locating 155 Se sites. Foadi (2003) has detailed the general underlying concepts of *ACORN* for the solution of protein structures. Using anomalous scattering data, a substructure can be determined (*e.g.* one containing S or Se atoms) in order to be used for the determination of whole protein structures. *ACORN* needs an atomic resolution higher than 1.3 Å, but for determination of substructures the resolution can be as low as 3 Å (see Mukherjee *et al.*, 1989). Dauter & Adamiak (2001) and Dauter *et al.* (2002) have described the question of accuracy and the attainment of this in anomalous scattering measurements in macromolecules at higher resolutions where the errors in measurements are often of the same order as the free signal.

Ramagopal *et al.* (2003) have detailed how even a small Bijvoet ratio of 0.6% arising from sulfur can be used favourably in solving a macromolecule. By using any of the above models, *ACORN* then uses a combination of approaches, most importantly dynamic density modification (DDM), to set up a refined set of phases. Foadi (2003) has given a detailed explanation of the reasons for the failure of *ACORN* when the resolution is below 1.2 Å. In *ACORN* the DDM modifies the density by sharpening or broadening. The whole purpose of doing this is to build up individual atomic peaks. Such a process is clearly suitable when the map shows details at atomic resolution; it is less clear how the density can be properly modified when it does not show atomic details. At atomic resolution, two neighboring atomic peaks will be two separate entities and the DDM will enhance both of them. At lower resolutions, these two peaks will merge into a single peak and the DDM, when acting on this peak, will just enhance it and hence no positive phase refinement can be expected in this situation. As the height of the various initial density peaks will be different at different resolutions, the shape of the density modification curve will have to be properly modified and, as such, this part in *ACORN* is likely to be difficult at resolutions which do not show 'peakiness'. Our present work, reported in this paper, overcomes this problem at low resolution (1.9 Å) by using fragments as 'seed phasing' information. As detailed papers on the *ACORN* program have already appeared in the literature (McAuley *et al.*, 2001; Banumathi *et al.*, 2002; Raja-

  DOI: 10.1107/S0909049504011185

kannan *et al.*, 2002, 2003; Velmurugan *et al.*, 2002), including its various options and its use in determining substructures, and the *ab initio* structure determination of macromolecules (Rajakannan *et al.*, 2003), this paper mainly focuses on the discussion of the applications of *ACORN* to the structure elucidation of a 357 residue containing the alkaline cellulase apo form diffracting at 1.9 Å resolution with laboratory source Cu $K\alpha$ data (Shirai *et al.*, 2001).

3. Materials and methods

3.1. Crystallization

A crystal was grown using the hanging-drop vapor-diffusion method at 291 K. The initial conditions were 1 ml of 40 m*M* cadmium sulfate hydrate and 0.5 *M* sodium acetate in 0.1 *M* HEPES buffer (pH 7.5) for the reservoir, and a mixture of 4 ml of the reservoir solution and 4 ml of 2% (*w*/*v*) protein solution for the drop. The crystal grew to an approximate size of 0.3 mm × 0.3 mm × 0.4 mm in one week.

3.2. Intensity data collection

The X-ray diffraction experiment was carried out by using an R-AXIS IV imaging-plate detector with a mirror–monochromator mounted on a Rigaku RU-300 copper rotating-anode X-ray generator (wavelength = 1.54 Å). The crystal was flash frozen in liquid nitrogen and kept in a 100 K dry-N_2 stream (Oxford Cryosystems) during the experiment. The crystal belonged to trigonal space group $P3_121$ and diffraction images were processed using the *DENZO* and *SCALEPACK* programs (Otwinowski & Minor, 1997). The statistics for the data collection are summarized in Table 1.

4. Overview of the method

Table 1 lists the crystallographic data and the various helices used as 'seeds' in *ACORN*, along with the total number of residues in each of these helices. This table also defines the various sets, which contain different combinations of these helices, along with the CPU time for running *ACORN* in each case. *ACORN* output which yielded successful and unsuccessful models are denoted by Y and N, respectively, in the 'Result' column.

For set 1, 11 helices containing 102 residues were given as input to *ACORN*. Here, the *ACORN PHASE* option was selected for the structure solution. The *R*-factor and correlation coefficient for the medium reflections with the normalized structure factors *E* from 0.1 to 1.2 of the initial model are 49.5% and 0.1034, respectively. Within ten cycles of the DDM the *R*-factor and correlation coefficient attained 47.9% and 0.1437 indicating a good solution. A map was calculated for the *ACORN* output phases and we were able to find 546 peaks which were above the 5σ cut-off. The phases were then fed to *ARP*/*wARP* (Perrakis *et al.*, 1999) and *REFMAC* (Murshudov *et al.*, 1999). After the initial model building by *ARP*/*wARP*, the *Rw* and *Rf* values were 41.8 and 42.3%. This initial model was refined and ten cycles of auto-building along with five cycles of *REFMAC* in each auto-building cycle were performed. Finally, *ARP*/*wARP* was able to build 346 out of 357 residues in five chains. At this stage the *Rw* and *Rf* values were 19 and 24.5%, respectively. The map also showed the densities in the missing region, so the manual model building was carried out for the remaining 11 residues. After the manual model building, 20 cycles of maximum-likelihood refinement were performed using *REFMAC* and solvent atoms were updated after the refinement using the *ARP*/*wARP* 'build solvent atoms' script. The final *Rw* and *Rf* values were 15.7 and 18.4%, respectively. The backbone of this final model was superimposed with structure conventionally solved by the molecular replacement method. The root-mean-square deviation was 0.29 Å and the details are shown in Table 2. The results for sets 2 to 8 are also shown.

Table 1
Details of the crystallographic data, helices and sets.

Crystallographic parameters	
a (Å)	97.835
b (Å)	97.835
c (Å)	121.529
α (°)	90
β (°)	90
γ (°)	120
Space group	$P3_221$
Data collection statistics	
Resolution limits (Å)	40.00–1.90
No. of unique reflections ($F > 0$)	49176
Mean $I/\sigma I$	20.1
R_{merge} (%)	7.7
Completeness (%)	90.9
Model contents	
Amino acid residues	357
Cd^{2+}	10
CH_3COO^-	5

Input helix details (as per the Protein Data Bank entry)		
Helix 1,	232 to 235	4 residues
Helix 2,	276 to 283	8 residues
Helix 3,	308 to 321	14 residues
Helix 4,	346 to 357	12 residues
Helix 5,	388 to 398	11 residues
Helix 6,	400 to 408	9 residues
Helix 7,	418 to 421	4 residues
Helix 8,	424 to 429	6 residues
Helix 9,	467 to 476	10 residues
Helix 10,	499 to 511	13 residues
Helix 11,	559 to 569	11 residues

Input set details				
Set	Input helix	Residues	*ACORN* CPU time (s)	Result
1	1–11	102	272.6	Y
2	1–6 and 8-11	98	288.4	Y
3	2–6 and 8–11	94	293.0	Y
4	2–6 and 9–11	88	269.8	Y
5	3–6 and 9–11	80	281.0	Y
6	3–5 and 9–11	71	283.7	Y
7	3–5 and 10–11	61	241.2	Y
8	3–5 and 10	50	271.8	Y
9	3, 4 and 10	39	–	N
10	3 and 10	27	–	N
11	3	14	–	N

5. Results and discussion

Figs. 1, 2 and 3 show sections of the final electron density map for sets 1, 4 and 8, respectively. Table 2 lists the *ACORN* statistics and the *ARP*/*wARP* running details for these cases. The final results obtained in each case are also given. The results show that a minimum of 321 residues can be built automatically when the 'seed phasing' information comes from the helices corresponding to sets 1 to 7. Although only 16 residues in three chains were initially built in the case of set 8 when given as 'seed', an iterative cycle carried out with these output phases (16 residues plus 4012 dummy atoms) revealed around 346 residues with a high value of 0.98 for the connectivity index. Table 2 also lists the root-mean-square deviations of the backbone atoms with the reported structure for each final model. The correlation coefficients of the maps before and after *ACORN* for the final map for all the sets are also shown in Table 2.

Table 2
Details of *ACORN* phasing, *ARP*/*wARP* model building and *REFMAC* refinement.

CC: correlation coefficient. MCC: map correlation coefficient before and after *ACORN* for the final map. a.a.: amno acids.

Program		Set 1			Set 2			Set 3			Set 4		
		R-factor (%)	CC	MCC	*R*-factor (%)	CC	MCC	*R*-factor (%)	CC	MCC	*R*-factor (%)	CC	MCC
ACORN													
No. of reflections having large E = 11119	Starting: Large E	0.409	0.1843	0.3700	0.410	0.1778	0.3633	0.413	0.1703	0.3536	0.416	0.1628	0.3500
No. of reflections having medium E = 37700	Medium E	0.495	0.1034		0.496	0.0987		0.498	0.0993		0.498	0.0913	
Input		Helix 1–11 (102 a.a.)			Helix 1–6 and 8–11 (98 a.a.)			Helix 2–6 and 8–11 (94 a.a.)			Helix 2–6 and 9–11 (88 a.a.)		
		After 9 cycles of DDM			After 56 cycles of DDM			After 58 cycles of DDM			After 53 cycles of DDM		
	Final: Large E	0.263	0.6413	0.5331	0.264	0.6412	0.5245	0.263	0.6409	0.5166	0.264	0.6427	0.5311
	Medium E	0.479	0.1437		0.498	0.0873		0.499	0.0867		0.497	0.0930	
		R-factor (%)	R_{free}		*R*-factor (%)	R_{free}		*R*-factor (%)	R_{free}		*R*-factor (%)	R_{free}	
ARP/*wARP*													
Auto-building: 10 cycles; *REFMAC*: 5 cycles for each auto-building; side dock after 7 cycles of auto-building	Initial	0.418	0.423		0.418	0.443		0.418	0.423		0.418	0.421	
	Final	0.190	0.245		0.186	0.237		0.181	0.232		0.192	0.245	
Details of *ARP*/*wARP* result		346 residues, five chains, missing residues 1, 113, 270–274, 298, 314–316, dummy atoms 756, connectivity index 0.97			344 residues, five chains, missing residues 1, 113, 270–274, 297, 313–316, dummy atoms 796, connectivity index 0.97			344 residues, five chains, missing residues 111, 112, 270–274, 298, 299, 314–316, dummy atoms 812, connectivity index 0.97			343 residues, five chains, missing residues 1, 111–113, 270–274, 298, 299, dummy atoms 772, connectivity index 0.97		
		R-factor (%)	R_{free}		*R*-factor (%)	R_{free}		*R*-factor (%)	R_{free}		*R*-factor (%)	R_{free}	
Without dummy atoms made by *ARP*/*wARP*		29.3	31.7		28.6	31.0		29.4	31.7		29.4	31.9	
After manual model building for missing residues and solvent building		15.7	18.4		15.7	18.4		15.9	18.8		15.8	18.5	
r.m.s. deviations with the same structure which is conventionally solved by molecular replacement		0.29 Å			0.27 Å			0.30 Å			0.30 Å		

Program		Set 5			Set 6			Set 7			Set 8		
		R-factor (%)	CC	MCC	*R*-factor (%)	CC	MCC	*R*-factor (%)	CC	MCC	*R*-factor (%)	CC	MCC
ACORN													
	Starting: Large E	0.420	0.1547	0.3377	0.424	0.1411	0.3164	0.432	0.1248	0.2842	0.440	0.1037	0.2606
	Medium E	0.502	0.0842		0.506	0.0697		0.507	0.0646		0.511	0.0521	
Input		Helix 3–6 and 9–11 (80 a.a.)			Helix 3–5 and 9–11 (71 a.a.)			Helix 3–5 and 10–11 (61 a.a.)			Helix 3–5 and 10 (50 a.a.)		
		After 58 cycles of DDM			After 61 cycles of DDM			After 51 cycles of DDM			After 63 cycles of DDM		
	Final: Large E	0.261	0.6431	0.5209	0.265	0.6347	0.5110	0.264	0.6427	0.5199	0.266	0.6353	0.4929
	Medium E	0.498	0.0899		0.501	0.0840		0.497	0.0920		0.502	0.0782	
		R-factor (%)	R_{free}		*R*-factor (%)	R_{free}		*R*-factor (%)	R_{free}		*R*-factor (%)	R_{free}	
ARP/*wARP*													
Auto-building: 10 cycles; *REFMAC*: 5 cycles for each auto-building; side dock after 7 cycles of auto-building	Initial	0.369	0.446		0.419	0.425		0.421	0.423		0.423	0.431	
	Final	0.198	0.251		0.190	0.240		0.229	0.294		0.288	0.465	
Details of *ARP*/*wARP* result		343 residues, 5 chains, missing residues 1, 2, 113, 267–274, 298, 357, dummy atoms 798, connectivity index 0.97			343 residues, 5 chains, missing residues 1, 2, 113, 269–274, 299, dummy atoms 780, connectivity index 0.97			321 residues, 7 chains, missing residues 1–3, 111, 112, 266, 267, dummy atoms 922, connectivity index 0.96			16 residues, 3 chains, dummy atoms 4012, connectivity index 0.68		

Table 2 (continued)

		Set 5		Set 6		Set 7		Set 8	
		R-factor (%)	R_{free}	*R*-factor (%)	R_{free}	*R*-factor (%)	R_{free}	*R*-factor (%)	R_{free}
ARP/*wARP*									
Auto-building: 10 cycles; *REFMAC*: 5 cycles for each auto-building; side dock after seven cycles of auto-building	Initial							0.364	0.465
	Final							0.170	0.217
Details of *ARP*/*wARP* result								346 residues, 4 chains, missing residues 1–3, 111, 112, 271–274, 294–297, dummy atoms 824, connectivity index 0.98	
		R-factor (%)	R_{free}	*R*-factor (%)	R_{free}	*R*-factor (%)	R_{free}	*R*-factor (%)	R_{free}
Without dummy atoms made by *ARP*/*wARP*		29.5	31.9	29.5	31.8	29.7	31.7	30.1	31.9
After manual model building for missing residues and solvent building		16.5	19.4	16.4	19.3	15.6	18.4	15.8	18.7
r.m.s. deviations with the same structure which is conventionally solved by molecular replacement		0.28 Å		0.31 Å		0.25 Å		0.28 Å	

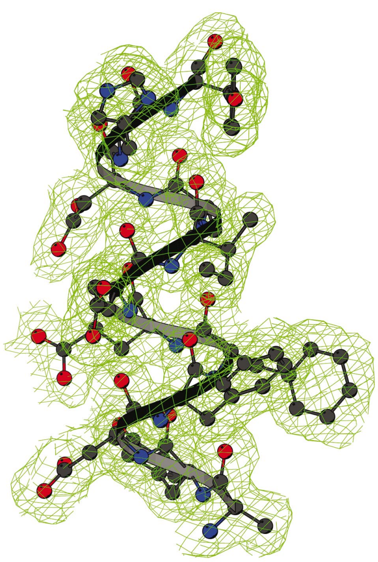

Figure 1
Section of the final electron density map with a α-helix region of residues 124–135 of the final model.

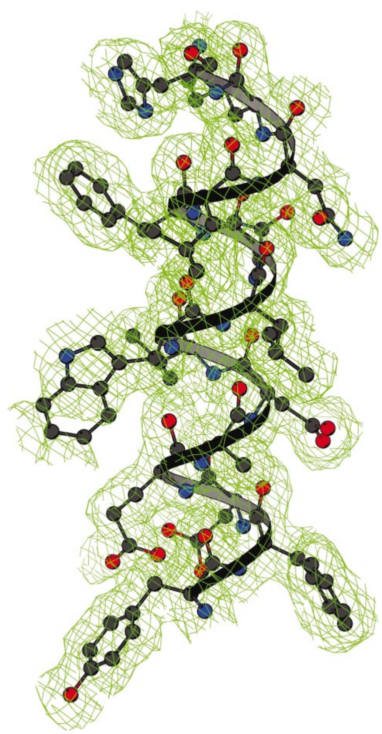

Figure 2
Section of the final electron density map with a α-helix region of residues 276–290 of the final model.

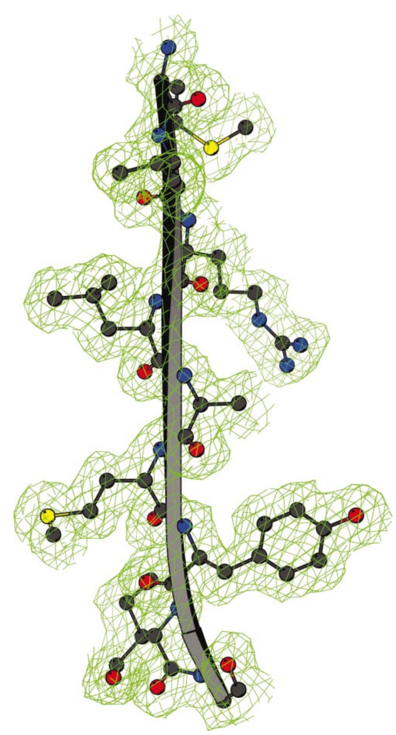

Figure 3
Section of the final electron density map level with a β-sheet region of residues 68–76 of the final model.

6. Conclusions

Based on the published work and the work carried out by our group (Rajakannan *et al.*, 2004), it has now become clear that very little initial information is needed to determine the structure of a protein using *ACORN*. Among the multiple solutions, the correct solutions can be obtained in all trials with high reliability by the working of the correlation coefficient and hence high resolution and fairly complete diffraction data enable one to solve a protein *ab initio*, in a relatively short amount of time. Thus *ACORN* has the great potential to establish itself as a program for high-throughput structure determination.

The success of *ACORN* relies mainly on the DDM, and the work carried out by us so far clearly substantiates the fact that automatic model building of *ACORN* output phases was possible in most cases because the quality of the electron density map output by *ACORN* was superb. To substantiate our finding we have superposed the final model obtained with the electron density from *ACORN* with the background. The final model in most of the cases fits with the electron density output from *ACORN*. The essence of this paper is in the applicability of *ACORN* to real data from laboratory sources even at a resolution of 1.9 Å. Success was obtained when the initial percentage of input information was just 14%. Currently, in order to extend the applicability of *ACORN* to lower resolutions, the seed phasing has been obtained from the native structure itself (as the structure had already been solved by traditional macromolecular crystallographic methods). We are at present working on the seed-feeding aspect of *ACORN* using a data-mining approach of the secondary structural elements which can be obtained from the data deposited in the Protein Data Bank.

Given the success of *ab initio* structure determination using *ACORN* at medium high resolution (1.9 Å) with laboratory source Cu $K\alpha$ data, this result, if reproduced with other protein crystals needing to be solved *de novo*, would alter the nature of the balance between synchrotron- and laboratory-based protein crystalography research.

VR and DV thank the Department of Science and Technology, Government of India, for funding a major project, which enabled them to carry out this work. Part of this work was carried out at the Department of Biotechnology, School of Engineering, Nagoya University, Nagoya, Japan, under the VBL Visiting Professorship Scheme awarded to one of the authors (DV).

References

Banumathi, S., Rajakannan, V., Velmurugan, D., Dauter, Z., Dauter, M., Tsai, M. D. & Sekar, K. (2002). *Japanese Crystallographic Society Meeting*, Poster, P3-II-27, 123.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucl. Acids Res.* **28**, 235–242.
Dauter, Z. (2002). *Acta Cryst.* D**58**, 1958–1967.
Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* D**57**, 990–995.
Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* D**58**, 494–506.
Dodson, E. J. & Yao, J. X. (2003). *Crystallogr. Rev.* **9**, 67–72.
Foadi, J. (2003). *Crystallogr. Rev.* **9**, 43–65.
Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *Acta Cryst.* D**56**, 1137–1147.
McAuley, K. E., Yao, J. X., Dodson, E. J., Lehmbeck, J., Astergaard, P. R. & Wilson, K. S. (2001). *Acta Cryst.* D**57**, 1571–1578.
Mukherjee, A. K., Helliwell, J. R. & Main, P. (1989). *Acta Cryst.* A**45**, 715–718.
Murshudov, G. N., Lebedev, A., Vagin, A. A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. M. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Rajakannan, V., Velmurugan, D., Yamane, T., Dauter, Z., Dauter, M., Tsai, M. D. & Sekar, K. (2002). *Japanese Crystallograhic Society Meeting*, Poster P3-I-22, 84.
Rajakannan, V., Yamane, T., Shirai, T., Kobayashi, T., Ito, S. & Velmurugan, D. (2003). *International Symposium on Diffraction Structural Biology*, Tsukuba, Japan, 28–31 May 2003, Poster P-085.
Rajakannan, V., Yamane, T., Shirai, T., Kobayashi, T., Ito, S. & Velmurugan, D. (2004). *J. Synchrotron Rad.* **11**, 64–67.
Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 868–875.
Shirai, T., Ishida, J., Noda, H., Yamane, T., Ozaki, K., Hamada, M. & Ito, S. (2001). *J. Mol. Biol.* **310**, 1079–1087.
Velmurugan, D., Rajakannan, V., Yamane, T., Dauter, Z. & Sekar, K. (2002). *Japanese Crystallographic Society Meeting*, Poster P3-II-26, 122.
Yao, J. X. (2002). *Acta Cryst.* D**58**, 1941–1947.

 Received 9 February 2004 • Accepted 7 May 2004