Indian J. Phys. **82** (1) 27-37 (2008)

High throughput structure determination for single-wavelength laboratory X-ray source anomalous diffraction data sets using iodinated tyrosines

S Selvanayagam[1], D Velmurugan[1,2*], T Yamane[3] and A Suzuki[3]

[1]Centre of Advanced Study in Crystallography and Biophysics, University of Madras, Guindy Campus, Chennai 600 025, India

[2]Venture Business Laboratory, Department of Biotechnology and Biomaterial Science, Graduate School of Engineering, Furo-Cho, Chikusa-Ku, Nagoya University, Nagoya 464-8603, Japan

[3]Department of Biotechnology and Biomaterial Science, Graduate School of Enqineering, Nagoya University, Furo-Cho, Chikusa-Ku, Nagoya 464-8603, Japan

E-mail d_velu@yahoo com

Received 12 September 2007, accepted 5 October 2007

Abstract : The availability of high-intensity synchrotron facilities, technological advances in data-collection techniques in synchrotron as well as laboratory source and improved user friendly crystallographic software have ushered in a new era in high-throughput macromolecular crystallography. Single-wavelength anomalous diffraction (SAD) phasing has become a useful tool for high-throughput structure determination. Attempts have been made to use SAD method using laboratory X-ray source in determining the three dimensional structure of an enzyme glucose isomerase (nearly 44 kDa molecular weight) using Cu $K\alpha$ and Cr $K\alpha$ anomalous scattering data sets corresponding to 1.8 and 2.4 Å resolutions, respectively. The tyrosine residues in this enzyme were iodinated with N-iodo-succinimide in crystallization PHENIX program was used to locate the iodine atom positions and also for phasing and model building. The model-building program ARP/wARP with REFMAC5 using SAD likelihood function can be used of to proceed further with the incomplete model built by PHENIX. In both the data sets, nine iodine positions initially located by PHENIX are sufficient enough to build the entire structure.

Keywords : SAD, PHENIX, glucose isomerase, ARP/wARP.

PACS Nos. : 61.10.Nz, 87.14.Ee

1. Introduction

Macromolecular crystallography has undergone tremendous progress in the last decade. The rapid advance of genomic sequencing projects has already produced remarkable results and the quantity of sequence information is rapidly expanding. The sequencing effort is now being followed by a concerted structural genomics effort hosted by at

*Corresponding Author

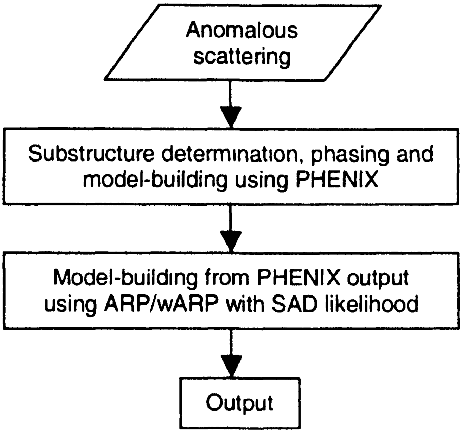
least 13 different consortia worldwide. The current bottleneck in the process of structural genomics appears to be in the production of X-ray quality crystals. In the era of structural genomics and high-throughput structural biology, the crystallographic community feels the need to solve structures in a relatively fast, accurate and automated fashion. Due to this reason, there has been an increasing need for software that can somehow bypass a great deal of human intervention and decide strategies automatically. The integrated choice of different softwares has become popular among crystallographers, especially when facing crystallographic cases that are not straightforward [1,2].

Anomalous scattering methods are currently used for phasing macromolecular structures. In the last decade, the multi-wavelength anomalous diffraction (MAD) method in conjunction with selenomethionine derivatization had become a powerful and commonly used tool to solve novel protein structures [3]. In this approach, two or three data sets are usually collected at various wavelengths around the absorption edge of the anomalous scatterers present in the crystal. In the last few years, due to the progress in data collection techniques and the current trend towards high-throughput structure determination, the single-wavelength anomalous diffraction (SAD) approach has acquired increasing popularity owing to its less demanding experimental requirements and favorably competes with MAD. Currently, the number of SAD structures exceeds the number of MAD based depositions in the PDB [4]. SAD phasing allows more structures to be solved with a less amount of time and this increased efficiency could prove useful for high-throughput structural genomic projects.

The choice of X-ray wavelength is one of the major decisions required for SAD data collection and this is made following evaluation of several wavelength-dependent factors and data-processing methods. Use of a wavelength of 2.1 Å was suggested as the best choice to obtain the highest anomalous signal-to-noise ratio using standard means of data collection and processing [5] and this can be performed in home laboratories with Cu or Cr X-ray sources. Several phasing techniques exploiting the anomalous signal contained in data collected with in-house generators have emerged during recent years. These include scatterers such as halides, sulfur, calcium and manganese. Iodine gives a significant amount of anomalous signal at wavelengths easily reachable at most synchrotron beam lines and at home sources. The iodination of tyrosine is an electrophilic aromatic substitution on the hydroxyl activated ortho carbon positions and results in either mono or di-substituted iodotyrosine [6]. The present work describes the high throughput structure determination of glucose isomerase using SAD method with iodinated tyrosine positions.

2. Description of programs used

The flowchart describes the present work for high throughput structure determination using various programs.


Anomalous scattering
Substructure determination, phasing and model-building using PHENIX
Model-building from PHENIX output using ARP/wARP with SAD likelihood
Output

PHENIX (Python-based Hierarchical Environment for Integrated Xtallography) is a software package developed for the automatic crystal structure determination. This provides the necessary algorithms to proceed from reduced intensity data to a refined molecular model [7,8]. For a given data, Hybrid Substructure Search (HYSS) makes use of Patterson and direct methods to locate anomalous scatterers or heavy atoms for experimental phasing. Once the substructure has been determined, SOLVE program enables rapid configuration of jobs for experimental phasing through interfaces. Maximum likelihood density modification algorithms implemented in RESOLVE produce minimally biased electron density maps using either the phases obtained from experimental phasing or by molecular replacement. These maps are then automatically interpreted using template matching implemented in RESOLVE and pattern recognition methods implemented in TEXTAL.

The procedures involved in automated model building program ARP/wARP [9] and RESOLVE [10] only consider the diffraction data obtained from the native crystal and neglect any available experimental phase information. The functional form of the likelihood refinement target encodes the prior phase information statically in the form of Hendrickson-Lattman coefficients and it assumes that the prior phase information is independent of the calculated model structure factor. But the incorporation of prior phase information into a maximum-likelihood formalism has been shown to strengthen model refinement [11]. Skubak *et al* [12] describe a multivariate SAD likelihood function that directly incorporates the measured Friedel pairs and the associated calculated model structure factors into structure retinement. This function does the simultaneous refinement of the heavy-atom and model parameters and thus directly considers the experimental phase information from a SAD experiment. The SAD likelihood function has been implemented in the refinement program REFMAC5 [13]. The SAD likelihood function in conjunction with the automated model-building procedures implemented in ARP/wARP leads to a correctly built model when current likelihood function fails for both data sets used here. All the computations mentioned here were carried out using Pentium IV PC.

3. Materials and methods

As detailed papers on the success of SAD phasing using Cu Kα radiation or Cr Kα radiation have frequently appeared in the literature [14–18], this paper mainly focuses on the SAD application using PHENIX and ARP/wARP with SAD likelihood function to an enzyme glucose isomerase (approximately 44 kDa molecular weight) using iodinated tyrosine atoms (Tyr OH replaced by I) with laboratory source (both Cu Kα and Cr Kα) anomalous scattering data sets corresponding to 1.8 and 2.4 Å resolutions. This enzyme contains 388 amino acids (9 Tyr) and two metal sites, one occupied by Mn^{2+} ion and the other by Mg^{2+}

The glucose isomerase crystals were grown by the hanging-drop vapor diffusion method using protein concentration of 1 2% w/v with a reservoir solution of 20% MPD, 0 2 M $MgCl_2$ and 10 mM Tris/HCL buffer pH 8 0 These native crystals were subsequently soaked for approximately 20 min in mother liquor with N-iodosuccimide (2 2 mM) The diffraction data sets were collected using in-house copper and chromium X-ray sources at High Intensity X-ray laboratory, Nagoya University, Nagoya, Japan using Rigaku R-axis VII IP detector system Table 1 shows the crystallographic details of these two data sets

Table 1 Crystallographic details of glucose isomerase for Cu Kα and Cr Kα data

For Cu Kα data	
a (Å)	93 044
b (Å)	98 564
c (Å)	102 585
Space group	I222
Resolution range (Å)	10–1 8 (1 859–1 8)
Completeness (%)	97 52 (94 33)
$I/\sigma(I)$	69 35 (31 85)
Anomalous signal-to-noise ratio	1 45
For Cr Kα data	
a (Å)	92 962
b (Å)	98 109
c (Å)	102 554
Space group	I222
Resolution range (Å)	10–2 4 (2 479-2 4)
Completeness (%)	95 99 (95 47)
$I/\sigma(I)$	48 76 (15 58)
Anomalous signal-to-noise ratio	5 68

3 1 Glucose isomerase, Cu Kα data

The average anomalous signal-to-noise ratio for the experimental data is 1 45. In order to locate the anomalous scatterers and to carry out the phasing and model building, the program PHENIX was used PHENIX was run with the inputs of scalepack format for this data set, single letter sequence file and substructure present in this data as iodine (I). The imaginary component of the anomalous scattering (f'') of iodine at this wavelength is 6.9 electron units. Initially, HYSS algorithm in PHENIX located nine iodine positions and phasing was done using SOLVE algorithm with these iodine atom positions. After density modification and map interpretation using RESOLVE algorithm in PHENIX, the program finally built 147 residues without side chains in 22 chains out of a total of 388 residues. At this stage, PHENIX gave the best score of 31.1; Figure of Merit (FOM) as 0.18; R and R_f as 50% and 54%, respectively. The map correlation coefficient for the overall model at this stage was 0.32. A map was calculated using the SOLVE output phases and 390 peaks were above 3σ cut-off.

This model was then fed to regular ARP/wARP [9] for automatic model building. After 50 cycles of auto-building, it was able to build only 89 residues in 11 chains. ARP/wARP with SAD likelihood function script for automatic model building using the option 'model building using the existing model' was then attempted. This was run with the data set inputs of mtz format (with F, F^+ and F^-), phenix model (147 residues), single letter sequence file and nine iodine positions located by HYSS. After 100 cycles of auto-building, it was able to build 381 residues with side chains (out of a total of 388 residues) and located 350 water atoms. At this stage, the R_w and R_f values were 19.2% and 22.2%, respectively. The map indicated the difference densities of the missing regions and the remaining residues were modeled into this. After the manual model building, the water atoms were checked and included if the density of water atoms was above 3σ level and its distance with protein atoms was 2.4–3.6 Å. The occupancies of the iodine sites in the final model were refined with MLPHARE [19] according to procedure that the calculated structure factor amplitudes of the protein model without the iodine atoms are treated as native data and the measured amplitudes are treated as derivative data. The resulting occupancies lie between 0.2 and 0.4. Then 20 cycles of normal maximum-likelihood refinement were performed using REFMAC5 [13]. The final R_w and R_f values were 17.1 and 19.8%, respectively. Table 2 details the results of PHENIX and ARP/wARP.

3.2. Glucose isomerase Cr $K\alpha$ data :

The average anomalous signal-to-noise ratio for the experimental data is 5.68. PHENIX was run with the inputs of scalepack format for this data set, single letter sequence file and substructure present in this data as iodine (I). The imaginary component of the anomalous scattering (f'') of iodine for this wavelength is 12.9 electron units. Initially, HYSS algorithm in PHENIX located nine iodine positions. Then the phasing and model building procedure built 274 residues out of 388 residues in 31 chains (14 residues with side chains). The map correlation coefficient for the overall model at this stage was 0.51. A map was calculated using the SOLVE output phases and 216 peaks were above 3σ cut-off.

Using this output, ARP/wARP job failed with normal likelihood functions. Hence ARP/wARP with SAD likelihood function script for automatic model building using the option 'model building using the existing model' was attempted. After 150 cycles of

Table 2. Details of PHENIX and ARP/wARP : glucose isomerase; Cu Kα data; 9 iodine atom positions as input; R_w and R_f are in %.

PROGRAM	Resolution limit	10–1.8 Å			
	Nine peaks	Score = 31.1	FOM = 0.18	SOLVE MCC = 0.26; 390 peaks > 3σ	
PHENIX	RESOLVE Built	147 (without side chains)	R_w = 50.0	R_f = 54.0	Overall model MCC = 0.32
ARP/wARP with SAD likelihood function using REFMAC5	Initial	R_w = 50.3	R_f = 49.6		
	No of auto building cycles	20			
	No. of Refmac cycles in each auto building cycle	5			
	Final	R_w = 19.2	R_f = 22.2		
	Connectivity index	0.99			
	No chains	2			
	No Res Built	381 (with side chains)			
	Water atoms	350			
	Final model with solvent atoms	R_w = 17 1 r.m.s. deviation of backbone atoms (1OAD) : 0.16 Å (1MNZ) : 0.25 Å	R_f = 19.8		

auto-building, it built 382 residues out of 388 residues with side chains and located 224 water atoms. Manual model building was carried out for the missing residues and the water atoms were checked and included if necessary. The occupancies of the iodine sites in the final model were refined with MLPHARE (lying between 0.2 and 0.4). Then 20 cycles of normal maximum-likelihood refinement were performed using REFMAC5. The final R_w and R_f values were 16.3 and 22.7%, respectively. All these details are presented in Table 3.

4. Results and discussion

4.1. *Glucose isomerase, Cu Kα data :*

Figures 1a and 1b show the cartoon diagrams of the PHENIX output and the final model. A main chain superposition of the present model with glucose isomerase structures deposited in the protein data bank (PDB-ID : 1OAD; space group $P2_12_12$; PDB-ID : 1MNZ; space group I 222 – atomic resolution structure) indicates that the overall tertiary fold is similar. The root-mean-square deviation is 0.16 Å with PDB (1OAD) and 0.25 Å with PDB (1MNZ). Figure 1c shows a region [16–19] of the final model superposed with SOLVE map and also the final $2|F_o| - |F_c|$ map. The map

Table 3 Details of PHENIX and ARP/wARP glucose isomerase Cr Kα data, 9 iodine atom positions as input, R_w and R_f are in %

PROGRAM	Resolution limit	10–2 4 Å			
	Nine peaks	Score = 52 7	FOM = 0 3	SOLVE MCC = 0 32, 216 peaks > 3σ	
PHENIX	RESOLVE Built	274 (14 with side chains)	R_w = 46 0	R_f = 49 0	Overall model MCC = 0 51
ARP/wARP	Initial	R_w = 45 3	R_f = 45 3		
with SAD	No of auto building cycles	30			
likelihood	No of Refmac cycles in each	5			
function	auto building cycle				
using	Final	R_w = 19 1	R_f = 24 3		
	Connectivity index	0 99			
REFMAC5	No chains	3			
	No Res Built	382 (with side chains)			
	Water atoms	224			
	Final model with solvent atoms	R_w = 16 3	R_f = 22 7		
		r m s deviation of backbone atoms (1OAD) 0 20 Å (1MNZ) 0 34 Å			

correlation coefficient between the SOLVE map and the final map is 0 26 The average thermal factor of the current model is 13 7 Å^2 and the estimated overall coordinates error is 0 11 Å A Ramachandran (ϕ, ψ) map assessed using PROCHECK in CCP4 for the final model shows that 92 1% of all non-glycine residues are in the core region and 7 3% residues are in the additionally allowed regions

4 2 Glucose isomerase, Cr Kα data .

Figures 2a and 2b show the cartoon diagrams of the PHENIX output and the final model The r m s deviation of this model with the PDB-ID 1OAD is 0 20 Å and that with PDB-ID 1MNZ is 0 34 Å Figure 2c shows a section [16–19] of the final model superposed with SOLVE map and also the final $2|F_o| - |F_c|$ map The map correlation coefficient between the SOLVE map and the final map is 0 32 The average thermal factor of the current model is 15 7 Å^2 and the estimated overall coordinates error is 0 35 Å. A Ramachandran (ϕ, ψ) map for the final model show that 90.6% of all non-glycine residues are in the core region and 8.5% residues are in the additionally allowed regions.

Figure 3a shows the superposition of the C$^\alpha$ atoms of the current model with the final model of Cu Kα data, final model of Cr Kα data with PDB-ID : 1 OAD. It

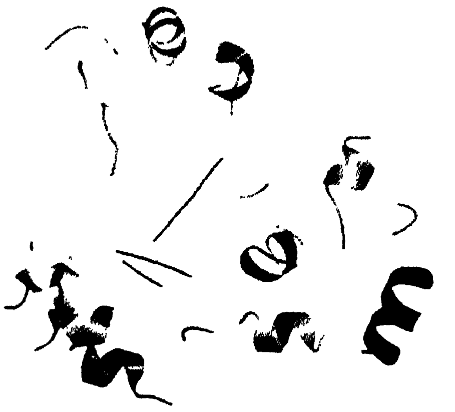


Figure 1a. PHENIX model 147 a.a (without side chains) from 9 iodine atoms for Cu Kα data

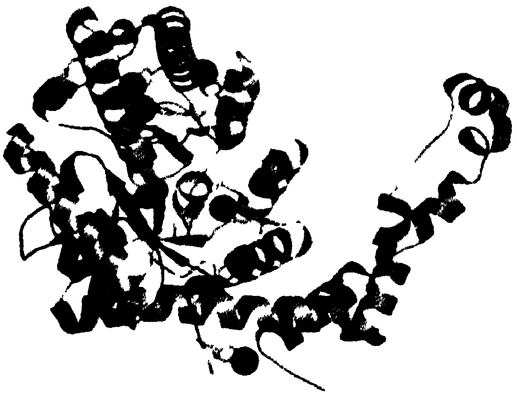


Figure 1b. Final model (Cu Kα data) using 9 iodine atoms Auto Built · 381 a a with side chains using ARP/wARP from PHENIX output (red colour atoms correspond to iodine).

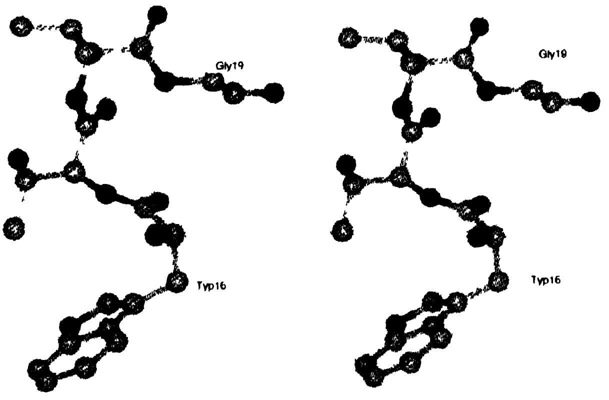


Figure 1c. Final model superposed with SOLVE map (0.6σ) and final $2|F_o| - |F_c|$ map (0.7σ)

shows that the overall fold of the models obtained using Cu Kα and Cr Kα data sets are similar to that of PDB i.d. 1OAD. Figure 3b shows the superposition of the current model with the final model using Cu Kα data, final model of Cr Kα data and PDB-ID : 1 MNZ.

5. Conclusion

The results of this work clearly demonstrate that with minimal user intervention, SAD data collected using laboratory source can be used for *ab initio* structure determination through largely automated methods. The longer wavelength from a chromium target is advantageous for iodine SAD phasing because the f'' value of the iodine atom is larger (12.9 e^-) than that obtained with copper radiation (6.9 e^-). This can be confirmed from the results obtained for Cr Kα data; using nine iodine positions,

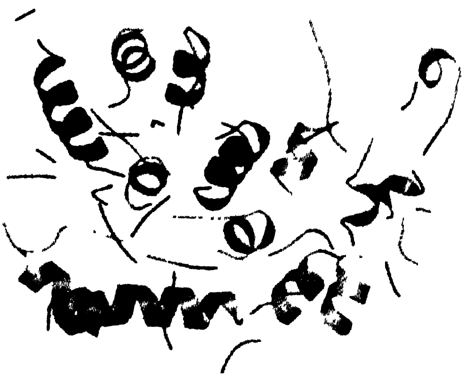


Figure 2a. PHENIX model : 274 a.a. (14 a.a with side chains) from 9 iodine atoms for Cr Kα data.

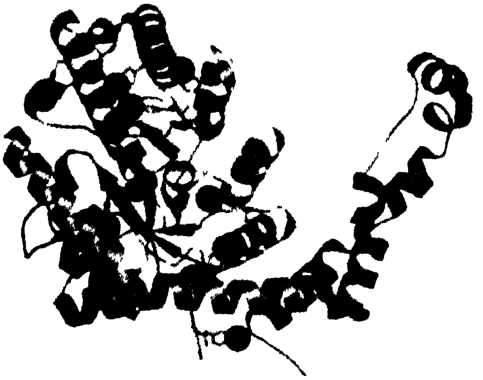


Figure 2b. Final model (Cr Kα data) using 9 iodine atoms Auto Built : 382 a.a. with side chains using ARP/wARP from PHENIX output (red colour atoms correspond to iodine).

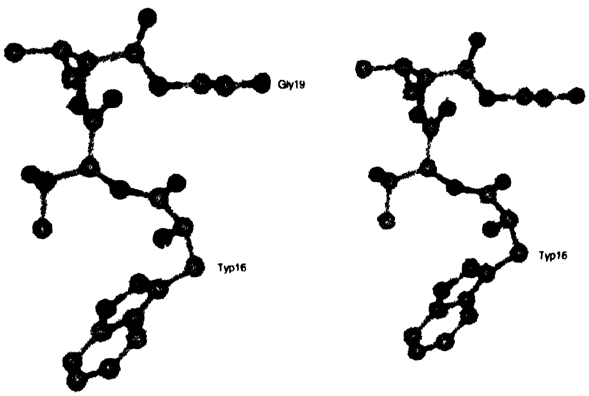


Figure 2c. Final model superposed with SOLVE map (0.5σ) and final $2|F_o| - |F_c|$ map (0.6σ).

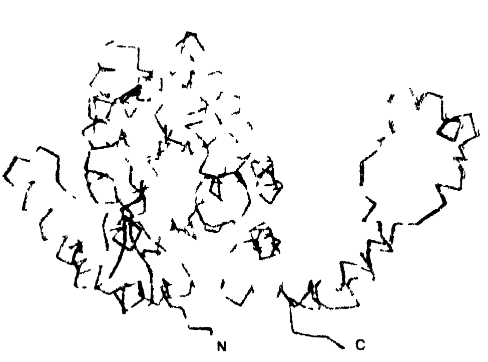


Figure 3a. Superposition of the C^α atoms of the final model of Cu Kα data (blue), Cr Kα data (cyan) along with PDB-ID : 1 OAD (red).

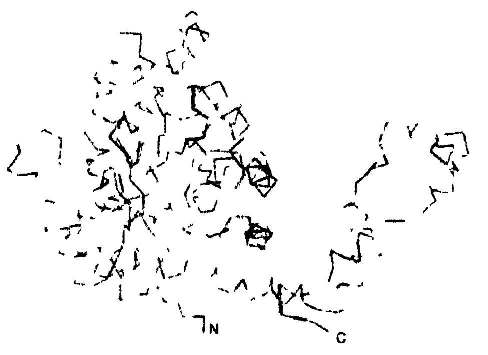


Figure 3b. Superposition of the C^α atoms of the final model of Cu Kα data (blue), Cr Kα data (cyan) along with PDB-ID : 1 MNZ (red)

PHENIX built only 147 residues for Cu Kα data whereas 274 residues were built for Cr Kα data. The above work emphasizes the applicability of the SAD technique to solve a macromolecular structure using laboratory source data using Cr Kα radiation when data extends to 2.4 Å resolution. For both the cases described here, ARP/wARP with SAD likelihood function can refine and improve PHENIX model and heavy-atom parameters and it successfully built 99% of the total residues. High-throughput crystallography in the post-genomic era requires a method by which the protein structure could be solved both quickly and easily. The present results confirm that the combination of PHENIX and ARP/wARP with SAD likelihood function is a powerful and routine tool to solve novel structures in the fastest way at lower resolutions and hence this method of solving a macromolecular structure using lab source data is beneficial to those who do not have access to synchrotron data collection.

Acknowledgement

SS thanks Council of Scientific and Industrial Research (CSIR) for providing Senior Research Fellowship. DV acknowledges Department of Science and Technology (DST), Government of India for major project supporting this work and thanks Venture Business Laboratory authorities, Nagoya Univeristy, Nagoya, Japan for the visiting Professorship assignments in short terms and also acknowledges financial support to the Department under UGC-CAS program.

References

[1] Joseph S Brunzelle, Padram Shafaee, Xiaojing Yang, Steve Weigand, Zhong Ren and Wayne F Anderson *Acta Cryst.* **D59** 1138 (2003)

[2] Vito Calderone *Acta Cryst.* **D60** 2150 (2004)

[3] W A Hendrickson *J. Synchrotron Rad.* **6** 845 (1999)

[4] Jiawei Wang, Miroslawa Dauter and Zbigniew Dauter *Acta Cryst.* **D62** 1475 (2006)

[5] C Mueller-Dieckmann, S Panjikar, P A Tucker and M Weiss *Acta Cryst.* **D61** 1263 (2005)

[6] Petrus H Zwart, Sankaran Banumathi, Miroslawa Dauter and Zbigniew Dauter *Acta Cryst.* **D60** 1958 (2004)

[7] P D Adams, R W Grosse-Kunstleve, L Hung, T R Ioerger, A J McCoy, N W Moriarty, R J Read, J C Sacchettini, N K Sauter and T C Terwilliger *Acta Cryst.* **D58** 1948 (2002)

[8] P D Adams, K Gopal, R W Grosse-Kunstleve, L Hung, T R Ioerger, A J McCoy, N W Moriarty, R K Pai, R J Read, T D Romo, J C Sacchettini, N K Sauter, L C Storoni and T C Terwilliger *J. Synchrotron Rad.* **11** 53 (2004)

[9] A Perrakis, R Morris and V S Lamzin *Nature Struct. Biol.* **6** 458 (1999)

[10] T C Terwilliger *Acta Cryst.* **D59** 1174 (2003)

[11] N S Pannu, G N Murshudov, E J Dodson and R J Read *Acta Cryst.* **D54** 1285 (1998)

[12] Pavol Skubak, Garib N Murshudov and Navraj S Pannu *Acta Cryst.* **D60** 2196 (2004)

[13] G N Murshudov, A Lebedev, A A Vagin, K S Wilson and E J Dodson *Acta Cryst.* **D55** 247 (1999)

[14] W A Hendrickson and M M Teeter *Nature* (London) **290** 107 (1981)

[15] Christopher T Lemke, G David Smith and P Lynne Howell *Acta Cryst.* **D58** 2096 (2002)

[16] L Chen, L R Chen, X E Zhou, Y Wang, M A Kahsai, A T Clark, S P Edmondson, Z -J Liu, J P Rose, B -C Wang, E J Meehan and J W Shriver *J. Mol. Biol.* **341** 73 (2004)

[17] Nobuhisa Watanabe, Yu Kitago, Isao Tanaka, Jia-Wei Wang, Yuan-xin Gu, Chao-de Zheng and Hai-fu Fan *Acta Cryst.* **D61** 1533 (2005)

[18] S Selvanayagam, D Velmurugan, T Yamane and A Suzuki *Indian J Phys.* **80** 969 (2006)

[19] Z Otwinowski *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering,* (eds) W Wolf, P R Evans and A G W Leslie 80 (Warrington : Daresbury Laboratory) (1991)