Journal of Advanced Bioinformatics Applications and Research
ISSN 0976-2604.Online ISSN 2278–6007
Vol 6, Issue1, 2015, pp1-7
http://www.bipublication.com

High Throughput Technique in Structural Bioinformatics

Sivashanmugam Selvanayagam[1] and Devadasan Velmurugan[2]

[1]Department of Physics and International Research Centre,
Kalasalingam University, Krishnankoil – 626 126, India
[2]Centre of Advanced Study in Crystallography and Biophysics,
University of Madras, Guindy Campus, Chennai – 600 025, India
[*]Corresponding author: Email: s_selvanayagam@rediffmail.com,
Tel: +91-4563289042 ; Fax: +91-4563289322

[Received 23/12/2014, Accepted-09/01/2015]

ABSTRACT:

High Throughput techniques are necessary for solving the three dimensional structures of macromolecules because the macromolecular sequences available exceed far in number than the available three-dimensional structure. Experimental phasing using anomalous data is essential to solve non-homologous protein structures. Sulfur atom is commonly present in protein structures in the form of cysteine and methionine residues. With recent progress in data collection, Sulfur – SAD phasing methods are extensively used for phasing macromolecular structures. S-SAD phasing has been tested on experimental Cr Kα data set, 2.4 Å, of an enzyme glucose isomerase of molecular weight 44 kDa. The combination of PHENIX and ARP/wARP built 379 residues out of 388 residues using 11 sulfurs located initially from the SAD data. Truncate the data to 3 Å resolution for testing the power of S-SAD. The above combination built 283 residues out of 388 residues using 11 sulfurs. Successive model could be built using OASIS which built 375 residues out of 388 residues. Minimal manual model building was required at this stage and the structure determination was completed using REFMAC5.

Keywords: Sulfur – SAD, Phasing, model-building, r.m.s.d., truncated data

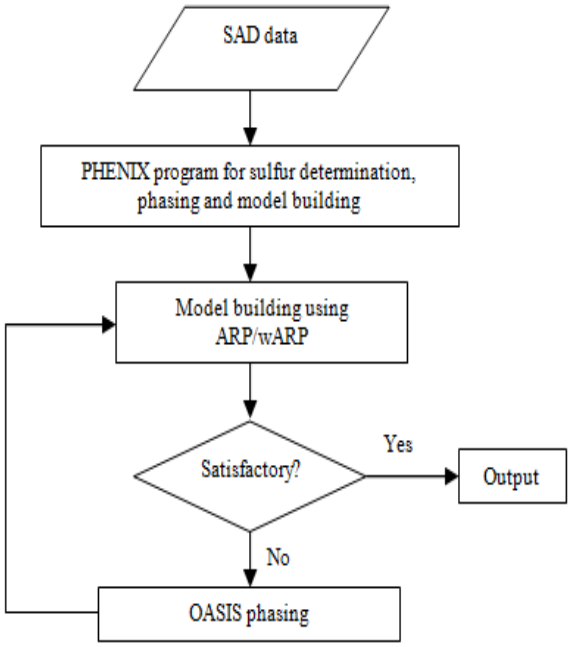
[I] NTRODUCTION

The sequence information available for macromolecules exceeds far in number than the three dimensional structures in the protein data bank [1, 2]. Thus 'High Throughput' techniques are needed in all aspects of macromolecular crystallography so that structural genomics project aiming at rapidly solving a large number of new structures in a short time would be successful. Experimental phasing is essential to solve protein structures which cannot be determined by molecular replacement. The use of tunable synchrotron radiation has led to the

development of multiwavelength anomalous dispersion (MAD) and single-wavelength anomalous dispersion (SAD), powerful phasing methods that are based on anomalous scattering of certain atoms [3]. With recent progress in data collection techniques and the current trend towards high-throughput structure determination, the SAD approach has acquired increasing popularity and favorable competes with MAD. SAD may need more accurate intensity estimations than MAD, especially when weak anomalous scatterers such as phosphorous or sulfur are used [4]. The long wavelength of the sulfur K absorption edge (5.02 Å) hampers data collection at wavelengths close to the absorption edge of sulfur and the small anomalous effect in the usual wavelength range limits the use of S atoms for SAD phasing [5]. The introduction of chromium rotating anode and designs of new synchrotron beam lines especially for the long-wavelength are deliberately intended for the application of Sulfur-SAD phasing (S-SAD) because f" for sulfur is larger for longer wavelengths [6]. At longer wavelengths the absorption coefficients of materials are larger so that these materials prevent the detection of tiny anomalous signals with a high degree of accuracy. Therefore, use of the longer wavelength is still controversial. Successful applications of the SAD technique have been reported using a longer wavelength from a chromium target [7, 8, 9, 10]. Attempts are here made in extending the S-SAD applications to the high throughput structure determination of an enzyme glucose isomerase (approximately 44 kDa molecular weight) using lab source Cr Kα (λ=2.2909 Å) anomalous scattering data corresponding to 2.4 Å resolution and also 3 Å resolution truncated data from the original data. The structure determination of this enzyme is already discussed [11] using SAD approach with the help of iodine atoms present in the enzyme (form of iodinated tyrosine). The present work is focusing on the high throughput structure determination of this enzyme with the help of 11 sulfurs present in the form of cysteine and methionine residues.

[II] MATERIALS AND METHODS

The diffraction data of glucose isomerase presented in this work was collected using in house chromium X-ray sources at High intensity X-ray laboratory, Nagoya University, using a Rigaku R-axis VII IP detector system at 2.4 Å resolution. 3 Å resolution truncated data of this enzyme were prepared using 2.4 Å data. The crystallographic details of these two set is given in Table 1. The flowchart describes the use of present approach to high throughput structure determination for these two data sets. PHENIX [12] is a software package developed for automatic crystal structure determination. ARP/wARP software suite [13] is a package utilities aimed at delivering an essentially complete macromolecular atomic model from a given electron density map. OASIS [14] is a computer program used for phasing based on direct methods procedure and it break the phase ambiguity intrinsic to SAD or Single Isomorphous Replacement (SIR) data.



Flowchart

Parameter	2.4 Å data	3 Å truncated data
a (Å)	92.962	
b (Å)	98.109	
c (Å)	102.554	
Space group	I222	
Resolution range	10-2.4 (2.459-2.4)	10-3 (3.071-3)
Completeness	97.58 (95.33)	98.82 (97.82)

Table-1 Crystallographic Details

[III] OVERVIEW OF THE METHOD

3.1. Glucose isomerase, 2.4 Å data

HYSS in PHENIX program initially located 11 sulfur positions and using these 11 sulfur positions SOLVE/RESOLVE in PHENIX successfully built 342 residues at 2.4 Å resolution. Using these 342 residues, 25 auto-building cycles of ARP/wARP built 379 residues out of 388 residues in one chain and located 259 water atoms. At this stage, R_{work} and R_{free} values were 19.4 and 25.9%, respectively. The map indicated 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 25 cycles of refinement were then performed using REFMAC5 [15]. The final R_{work} and R_{free} values were 16.7 and 23.7%, respectively. The details are shown in Table 2.

3.2. Glucose isomerase, 3 Å truncated data

3 Å resolution data (truncated) set was prepared using 2.4 Å resolution data. The HYSS - SOLVE/RESOLVE - ARP/wARP approach was failed to this data. The structure determination was successfully completed at 3 Å resolution using the approach mentioned in the flow chart with 11 sulfurs. HYSS initially located 11 sulfur positions for this 3 Å data. Using these sulfur positions, SOLVE/RESOLVE built 284 residues in 25 chains out of 388 residues. As shown in the flow chart, these 284 residues were fed to ARP/wARP and run in phased likelihood mode option. 100 auto-building cycles of ARP/wARP built 283 residues in 3 chains and located 126 water atoms. Now OASIS was run in fragment extension option with these 283 residues and 11 sulfur positions for phasing. After OASIS phasing, 100 auto-building cycles of ARP/wARP built 375 residues out of 388 residues in 1 chain and located 60 water atoms. At this stage, R_{work} and R_{free} values were 21.8 and 29.8%, respectively. The map indicated difference densities of the missing regions and the remaining residues were modeled into this. After the manual model building, water atoms were checked and 20 cycles of refinement were performed using REFMAC5. The final R_{work} and R_{free} values were 19.9 and 28.9%, respectively. The details are shown in Table 2.

[IV] RESULTS AND DISCUSSION

4.1. Glucose isomerase, 2.4 Å data

Figs: 1a, 1b and 1c show the cartoon diagram of the PHENIX output, ARP/wARP output and the final model. Fig: 2 shows a section of the final model superposed with the corresponding electron density map of final $2|F_O| - |F_C|$ map. The average B factor for the current model is 15.5 $Å^2$ and the overall coordinate error is 0.255 Å.

3.2. Glucose isomerase, 3 Å truncated data

Figs: 3a, 3b and 3c show the cartoon diagram of the PHENIX output, ARP/wARP output and the final model. Fig: 4 shows a section of the final model superposed with the corresponding electron density map of PHENIX output, OASIS output and also the final $2|F_O| - |F_C|$ map. It can clearly be seen that phases are developed slowly from PHENIX output to final output. The average B factor for the current model is 5.3 $Å^2$ and the overall coordinate error is 0.519 Å.

Fig. 5 shows the superposition of the final model of 2.4 Å data with 3 Å truncated data and it results that the r.m.s. deviation between these two data sets is 1.18 Å.

[V] CONCLUSION

As suggested by Wang [16], the longer wavelength from a chromium target is advantageous for S-SAD phasing because the f" value of the S atom is larger (1.14 e^-). This can be confirmed from the results obtained here. The above work emphasizes the applicability of the S-SAD technique to solve a macromolecular structure using lab source data using Cr Kα radiation when data extends to 3 Å resolution. Inspection of PDB statistics shows that the average quality of structures deposited from the structural genomics centres does not differ from the quality of structures elucidated in a more traditional way and hence this method of solving a macromolecular structure using lab source data is one of the high throughput techniques in Structural Bioinformatics. The combination of PHENIX/ARP/wARP/OASIS/ARP/wARP is one of the method to identify a new folds from X-ray diffraction data in the fastest way for non-homologous structures.

ACKNOWLEDGEMENT

SS acknowledges the Department of Science and Technology (DST), India for providing computing facilities under DST-Fast Track Scheme and also thanks the Vice Chancellor and Management of the Kalasalingam University, Anand Nagar, Krishnankoil for their support and encouragement. The authors acknowledge Prof. T. Yamane and Prof. A. Suzuki for providing the data set for the present work.

REFERENCES

1. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T and Tasumi, M. (1977). J. Mol. Biol. 112, 535-542.
2. Selvanayagam, S., Velmurugan, D. and Yamane, T. (2005). Acta Cryst. A61, C151.
3. Cha, S-S., An, Y. J., Jeong, C-S., Kim, M-K., Lee, S-G., Lee, K-H. and Oh, B-H. (2012). Acta Cryst. D68, 1253-1258.
4. Dauter, Z. and Adamiak. D. A. (2001). Acta Cryst. D57, 990-995.
5. Kitago, Y., Watanabe, Y. and Tanaka, I. (2005). Acta Cryst. D61, 1013-1021.
6. Dauter, Z. (2006). Acta Cryst. D62, 1-11.
7. **C**hen, L., Chen, L., Zhou, X. E., Wang, Y., Kahsai, M. A., Clark, A. T., Edmondson, S. P., Liu, Z., Rose, J. P., Wang, B. –C., Meehan, E. J. and Shriver, J. W. (2004). J. Mol. Biol. 341**,** 73-91.
8. Rose, J. P., Liu, Z. –J., Tempel, W., Chen, D., Lee, D., Newton, M. G. and Wang, B. –C. (2004). Rigaku J. 21, 1-9.
9. Watanabe, N., Kitago, Y., Tanaka, I., Wang, J., Gu, Y., Zheng, C. and Fan, H. (2005). Acta Cryst. D61, 1533-1540.
10. Selvanayagam, S., Velmurugan, D., Yamane, T. and Suzuki, A. (2008). Indian. J. Phys. 82(1), 27-37.
11. Selvanayagam, S., Velmurugan, D., Yamane, T. and Suzuki, A. (2009). Indian. J. Phys. 83(12), 1671-1679.
12. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, D. N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010). Acta Cryst. D66, 213-221.
13. Lamzin, V. S., Perrakis, A. and Wilson, K. S. (2012). International Tables for Crystallography. Volume F: Crystallography of biological macromolecules (Amold E, Himmel DM and Rossmann MG eds), 2nd edition, ISBN: 978-0-470-66078-2, Dordrecht, Kluwer Academic Publishers, The Netherlands, 525-528.
14. Wang, J. W., Chen, J. R., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2004). Acta Cryst. D60, 1991-1996.

15. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. and Vagin, A. A. (2011). Acta Cryst. D67, 355-367.

16. Wang, B. –C. (1985). Methods Enzymol. 115, 90-112

Figures and Tables:

DATA SETS		**2.4 Å data**		**3 Å truncated data**	
PROGRAM	Resolution Limit	10-2.4		10-3	
PHENIX	Using 11 sulfur atoms located from HYSS in PHENIX	342 resdiues built R_{work} = 28.5 R_{free} = 32.3	MCC = 0.688	284 residues built R_{work} = 38.8 R_{free} = 45.0	MCC = 0.375
ARP/wARP - I	**Initial**	R_{work} = 31.8	R_{free} = 35.5	R_{work} = 38.4	R_{free} = 48.9
	No. of auto building cycles	5		20	
	No. of REFMAC cycles in each auto building cycle	5		5	
	Final	R_{work} = 19.4	R_{free} = 25.9	R_{work} = 31.4	R_{free} = 38.6
	No. of Chains	1		3	
	No. of Res. Built	379 residues / 388 residues		283 residues / 388 residues	
	Water atoms	259		126	
OASIS	Fragment extension option using ARP/wARP output	---		283 residues + 11Sulfur atoms	
ARP/wARP - II	**Initial**			R_{work} = 31.3	R_{free} = 38.6
	No. of auto building cycles			20	
	No. of REFMAC cycles in each auto building cycle			5	
	Final			R_{work} = 21.8	R_{free} = 29.8
	No. of Chains			1	
	No. of Res. Built			375 residues / 388 residues	
	Water atoms/Dummy atoms			60	
Final model with solvent atoms		**R_{work} = 16.7**	**R_{free} = 23.7**	**R_{work} = 19.9**	**R_{free} = 28.9**
r.m.s. deviation in bond lengths		0.021		0.013	
r.m.s. deviation in bond angles		1.951		1.548	

Table-2 Details of phasing and model building for two data sets

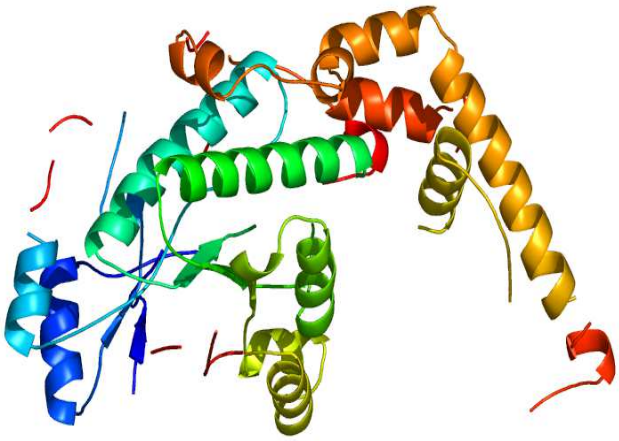


Fig. 1a: PHENIX output
Input: 11 sulfurs
Output: 342 a.a

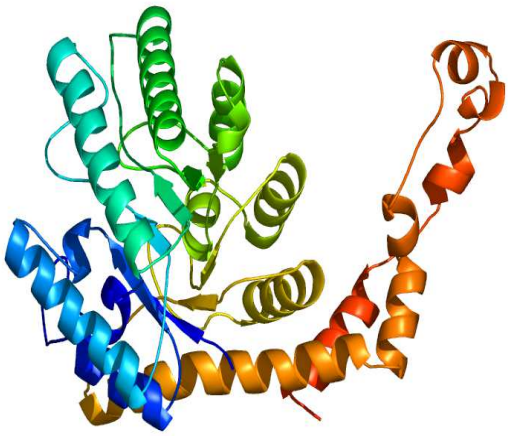


Fig. 1b: ARP/wARP output
Input: 342 a.a (PHENIX o/p)
Output: 379 a.a

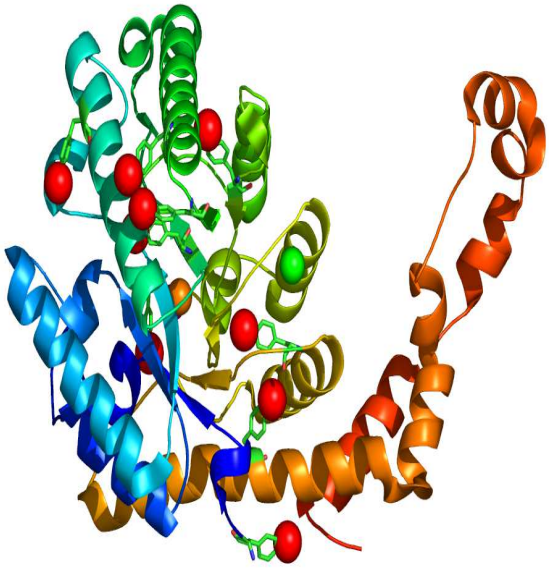


Fig. 1c: Final Model

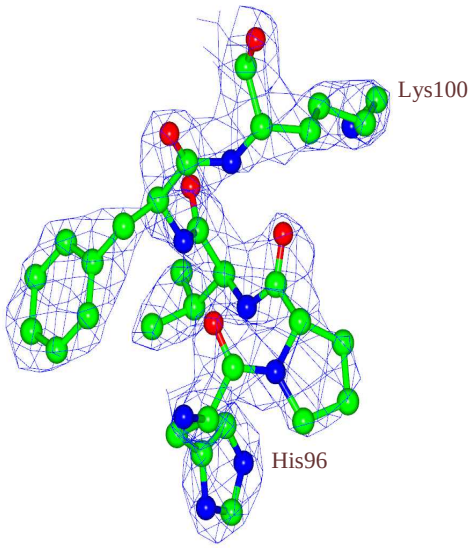


Fig. 2: Superposition of final model with final $2|F_O|-|F_C|$ map

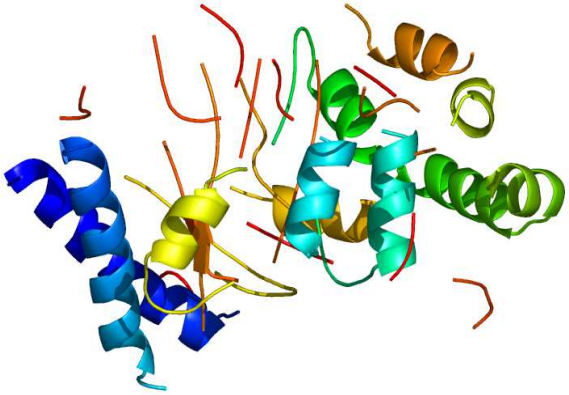


Fig. 3a: PHENIX output
Input: 11 sulfurs
Output: 284 a.a

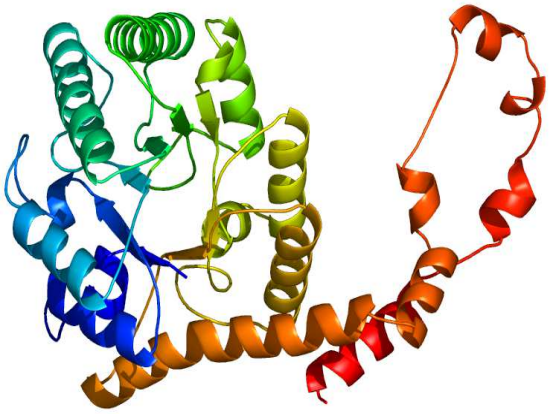


Fig. 3b: ARP/wARP output
Input: 284 a.a (PHENIX o/p)
Output: 283 a.a

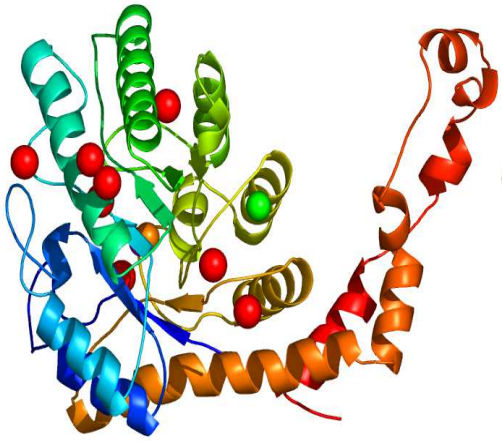


Fig. 3c: Final Model

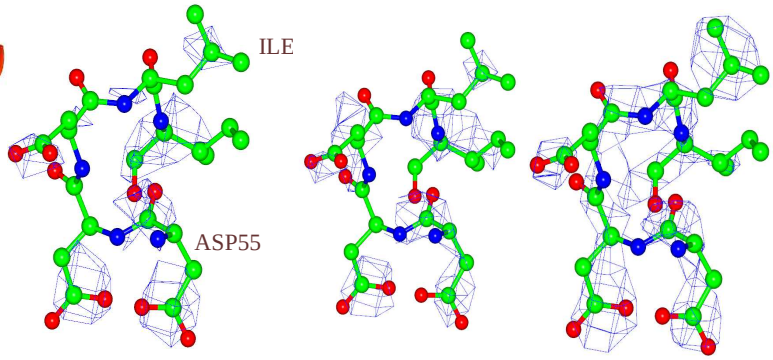


Fig. 4: Final Model superposed with PHENIX map, OASIS map and final $2|F_O|-|F_C|$ map

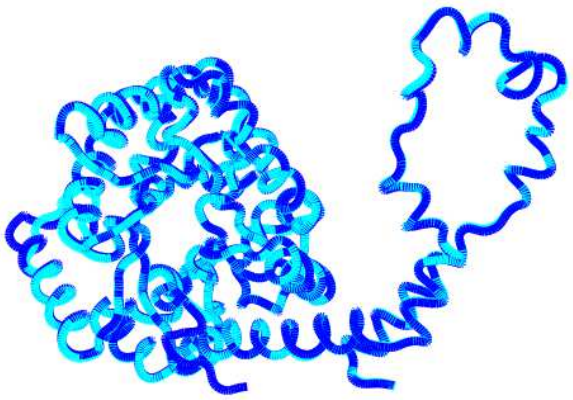


Fig. 5: Superposition of the C^{α} atoms of final model of 2.4 Å data (cyan) with 3 Å data (blue)