Applications of ACORN to data at 1.45 Å resolution

V. Rajakannan[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, [e]Japan Marine Science and Technology Center, 2-15 Natushima, Yokosuka 237-0061, Japan. E-mail: d_velu@yahoo.com

One of the main interests in the molecular biosciences is in understanding structure-function relations and x-ray crystallography plays a major role in this. ACORN can be used as a comprehensive and efficient phasing procedure for the determination of protein structures when atomic resolution data are available. Initial model can automatically be built by ARP/wARP followed by REFMAC for refinement. The α helices and β sheets occurring in many protein structures can be taken as starting fragments for structure solution in ACORN. ACORN along with ARP/wARP followed by REFMAC can be an *ab initio* method for solving protein structure for which data are better than 1.2 Å (atomic resolution). Attempts are here made in extending its applications to real data at 1.45 Å resolution and also to truncated data at 1.6 Å resolution. Two previously known structures congerin II and alkaline cellulase N257 were resolved using the above approach. Automatic structure solution, phasing and refinement for real data at still lower resolutions for proteins of various complexities are being carried out. Data mining of the secondary structural features using PDB is being carried out for this new approach for 'seed-phasing' to ACORN.

Keywords: ACORN; ARP/wARP; REFMAC; data-mining; phasing at lower resolution.

1. Introduction

Multiple Isomorphous Replacement (MIR) or Multiwavelength Anomalous Diffraction (MAD) or Molecular Replacement (MR) using a homologous model are the conventional techniques used in macromolecular structure solution when data is not at atomic resolution. Using the phases output by these techniques, an atomic model could be computed and refined. Direct methods are highly successful in solving small molecular crystal structures for which data are always available at atomic resolution (AR). These methods fail when applied as such to macromolecules due to the poor validity of the probabilistic estimates of phase relationships in this situation where the total number of atoms in the unit cell becomes very large. Although it was shown that integrating the conventional direct methods with Anomalous Scattering methods (AS) or Isomorphous Replacement method (IR) would yield favourable results in macromolecular applications (Hauptman 1982*a*; Hauptman 1982*b*; Hauptman *et al.*, 1982), no new structures were attempted using these techniques. The minimal principle approach was later proposed by Hauptman for finding probable phases (Hauptman *et al.*, 1991) which was later incorporated into *SnB* (Weeks *et al.*, 1994; Xu *et al.*, 2000). The Half-Baked approach in SHELXD (Sheldrick & Scheinder, 1997) is also used in macromolecular crystallography. For metalloproteins, applications were made using more conventional direct methods schemes like SAYTAN (Mukherjee, 1999; Mukherjee *et al.*, 1999). The use of the direct methods procedures incorporated in ACORN was described by Foadi *et al.*, (2000) as a flexible and efficient procedure for the solution and phase refinement of protein structures. As programmed at present, the method requires AR data. The starting fragment can be of various types like (i) set of experimentally determined anomalous scattering centres, (ii) a small idealised piece of secondary structure and (iii) fragments or ideal whole structures of homologous proteins. For types (ii) and (iii), MR technique has to be used for initial positioning. Recently it is shown that sufficiently accurate data can be measured using synchrotron beam lines to determine the positions of much lighter atoms such as S and Cl atoms from their small but significant anomalous signal (Dauter *et al*., 1999; Dauter *et al*., 2000). The position of sulphur can be located even at a wavelength of 1.75 Å for a data of 1.75 Å resolution. These sulphurs can be got either from *SnB* or SHELXD program. For the extension of direct methods to lower resolution, these excellent starting 'fragments' would be of great use. Faodi *et al.*, (2000) have detailed the ACORN program and the applications made for many structures. McAuley *et al.*, (2000) detailed how *ab initio* structure determination of a 19kDa metalloprotein is probable using ACORN when the data was collected at 1 Å resolution. Yao (2002) has detailed the ACORN in CCP4 along with applications. It was pointed out that the dynamic density modification (DDM) is a fast and powerful density modification approach that can be extended to work at lower resolutions (from 1.5 to 2.0 Å) with more sophisticated algorithms.

2. Applications

2.1. Description of the methods

ACORN uses the strong reflections with E>1.2 in the phase refinement by the DDM and Patterson superposition procedure. Both strong and weak reflections (E<0.1) are used in Sayre-equation refinement (SER). The medium reflections (0.1<E<1.2) are used to calculate the correlation coefficient (CC) for each potential solution of DDM. The CC describes the extent to which the magnitudes of the calculated normalised structure-factor (E_c) resemble the observed normalised structure factor amplitudes. CC for the medium reflections are strongly correlated to the mean phase error (MPE) and indicate a correct solution clearly. The first part of ACORN, namely ACORN-MR, deals with finding the position of a fragment of the structure, even a single atom that provides an initial set of estimated phases. This set is passed into ACORN-PHASE, where phase refinement by a number of real-phase processes is performed.

2.2. Overview of the approach

For locating a single atom, the above approach randomly generates thousands of positions in the asymmetric unit. For each random position, E_c values and corresponding CCs are calculated for all reflections. 1000 sets with highest CCs are saved as starting points. The solution is normally found in the top 100 sets. This approach can be used to determine a native protein structure from AR data if the structure contains at least one heavy atom (sulphur or heavier). For a heavy atom contribution of at least 5% of the total scattering power of the structure, ACORN provides good enough initial phases for ACORN-PHASE to successfully refine the single-atom fragment phases. ACORN approach can also be used to determine substructures with AS or IR data at much lower resolutions. DDM is the mean phase-refinement approach in ACORN-PHASE and can

Table 1

PROGRAM		SET - I		SET - II		SET - III	
ACORN	STARTING	R-FACTOR	CC	R-FACTOR	CC	R-FACTOR	CC
No. of reflections having large E = 5910	LARGE E (L)	0.442	0.1136	0.453	0.0803	0.459	0.0569
No. of reflections having medium E = 20166	MEDIUM E (M)	0.511	0.0724	0.517	0.0573	0.525	0.0424
INPUT*		S1 to S6 (48 a.a)		S2 to S5 (34 a.a)		S3 to S5 (27 a.a)	
	After 42 cycles of DDM			After 47 cycles of DDM		After 59 cycles of DDM	
	FINAL L	0.272	0.6344	0.272	0.6295	0.280	0.6107
	M	0.500	0.1236	0.498	0.1226	0.511	0.0797
ARP/wARP		R-FACTOR	Rfree	R-FACTOR	Rfree	R-FACTOR	Rfree
AUTOBUILDING : 10 Cycles REFMAC: 5 cycles for each auto building	INITIAL	0.416	0.391	0.418	0.402	0.439	0.428
Side dock after 7 cycles of auto building	FINAL	0.166	0.196	0.171	0.196	0.287	0.459
REMARKS	133 residues, 1 Chain, Connectivity index: 0.99			132 residues, 1 chain, Connectivity Index: 0.98		15 residues poly GLY Connectivity index: 0.67	
II nd ROUND OF ARP/warp							
[Same parameters are used as in I st Round]					INITIAL	0.377	0.422
					FINAL	0.169	0.199
						131 residues, Single Chain, Connectivity index: 0.98	

Congerin-II: 11 sheets [135 residues], Synchrotron Data: 48.795 - 1.452 Å, a=61.5, c=80.7Å, $P4_22_12$, Total no. of reflections : 26162, Wavelength 1Å, For Large E: E >1.2, Medium E, 0.1 < E < 1.2. *S1: 16-23 (8 a.a), S2: 29-35 (7 a.a), S3: 41-50 (10 a.a), S4: 55-63 (9 a.a), S5: 86-93 (8 a.a), S6: 121-126 (6 a.a).

Table 2

PROGRAM		RESULTS	SET 1	RESULTS	SET 2	RESULTS	SET 3	RESULTS	SET 4
ACORN	STARTING	R-FACTOR	CC	R-FACTOR	CC	R-FACTOR	CC	R-FACTOR	CC
No. of reflections having large E = 15810	Large E (L)	0.378	0.2632	0.389	0.2351	0.393	0.2263	0.403	0.2134
No. of reflections having medium E = 51591	Medium E (M)	0.501	0.1487	0.505	0.1305	0.513	0.1206	0.514	0.1129
INPUT*		S1 to S7 (132 a.a)		S1, S3-S8 (120 a.a)		S1-S4, S6-S8 (112 a.a)		S1,S2,S4-S6,S8(102 a)	
	After 32 cycles of DDM			After 31 cycles of DDM		After 35 cycles of DDM		After 33 cycles of DDM	
	FINAL L	0.260	0.6464	0.261	0.6461	0.262	0.6440	0.261	0.6445
	M	0.514	0.1109	0.514	0.1102	0.516	0.1066	0.517	0.1043
ARP/wARP		R-FACTOR	Rfree	R-FACTOR	Rfree	R-FACTOR	Rfree	R-FACTOR	Rfree
AUTOBUILDING : 10 Cycles REFMAC: 5 cycles for each auto building	INITIAL	0.450	0.438	0.430	0.418	0.432	0.415	0.431	0.428
Side dock after 7 cycles of auto building	FINAL	0.202	0.236	0.141	0.174	0.144	0.176	0.148	0.179
REMARKS	342 residues, 2 Chains, Connectivity Index: 0.99			370 residues, 1 Chain, Connectivity Index: 0.99		370 residues, 1 Chain, Connectivity Index: 0.99		347 residues, 2 chains, Connectivity Index: 0.99	

PROGRAM		RESULTS	SET 5	RESULTS	SET 6	RESULTS	SET 7
ACORN	STARTING	R-FACTOR	CC	R-FACTOR	CC	R-FACTOR	CC
No. of reflections having large E = 15810	Large E (L)	0.412	0.1881	0.423	0.1543	0.430	0.1453
No. of reflections having medium E = 51591	Medium E (M)	0.520	0.0909	0.525	0.0738	0.531	0.0720
INPUT*		S3-S5,S7,S8 (87 a.a)		S4,S5,S7,S8 (73 a.a)		S3,S5,S7, S8 (66 a.a)	
	After 37 cycles of DDM			After 41 cycles of DDM		After 41 cycles of DDM	
	FINAL L	0.262	0.6418	0.262	0.6382	0.264	0.6336
	M	0.518	0.1039	0.520	0.0961	0.521	0.0937
ARP/wARP - I		R-FACTOR	Rfree	R-FACTOR	Rfree	R-FACTOR	Rfree
AUTOBUILDING : 10 Cycles REFMAC: 5 cycles for each auto building	INITIAL	0.432	0.415	0.435	0.424	0.435	0.435
Side dock after 7 cycles of auto building	FINAL	0.134	0.165	0.146	0.177	0.311	0.398
REMARKS	370 residues, 1 Chain. Connectivity Index: 0.99			344 residues, 3 Chains, Connectivity Index: 0.98		poly GLY Connectivity Index: 0.78	
ARP/wARP – II						R-FACTOR	Rfree
[Same parameters are used as in I st Round]					INITIAL	0.374	0.443
					FINAL	0.193	0.229
REMARKS						289 residues, 3 Chains, Connectivity Index : 0.98	

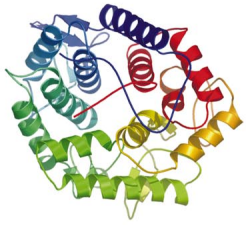
Alkaline cellulase N257**:** 377 residues, $P2_12_12_1$, a=62.54, b=71.76, c=88.86 Å, Resolution: 20.0 - 1.45 Å, For Large E: E >1.2, Medium E, 0.1 < E < 1.2. *S1: 24-39 (16 a.a), S2: 64 – 75 (12 a.a), S3: 124-137 (14 a.a), S4: 144-164 (21 a.a), S5: 208-227 (20 a.a), S6: 262-278 (17 a.a), S7: 284-299 (16 a.a), S8: 357-372 (16 a.a).

CONGERIN II
Figure 1.

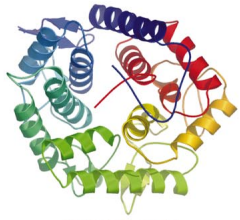
SET 1
Input: 48 residues
Auto built: 133 residues, Single chain
Breaks: residue no. 1and 135
Fig. 1*a*

SET 2
Input: 34 residues
Auto built: 132 residues, Single chain
Breaks: residue no. 1,2 and 135
Fig. 1*b*

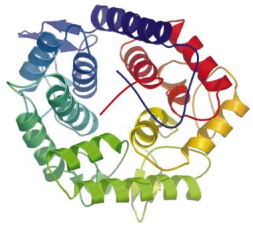
SET 3
Input: 27 residues
Auto built: 131 residues, Single chain
Breaks: residue no. 1-3 and 135
Fig. 1*c*



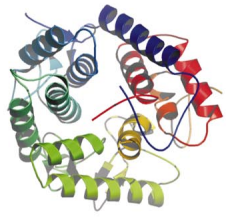
ALKALINE CELLULASE
Figure 2.



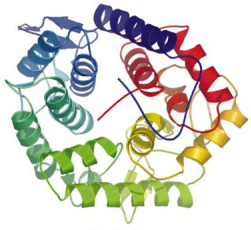
SET 2 (1.45A)
Input: 120 residues
Auto built: 370 residues, Single chain
Breaks: residue no. 1-6 and 377
Fig. 2*a*



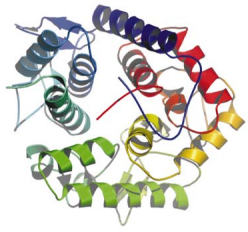
SET 5 (1.45A)
Input: 87 residues
Auto built: 370 residues, Single chain
Breaks: residue no. 1-6 and 377
Fig. 2*b*



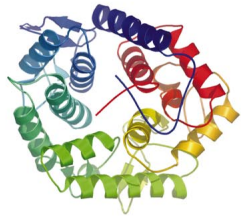
SET 6 (1.45A)
Input: 73 residues
Auto built: 344 residues, Three chain
Breaks: residue no. 1-6, 43-45, 278-300 and 377
Fig. 2*c*



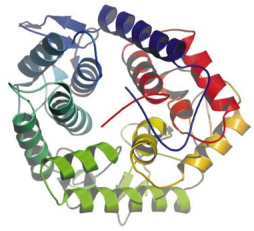
SET 4 (1.5A)
Input: 102 residues
Auto built: 370 residues, Single chain
Breaks: residue no. 1-6 and 377



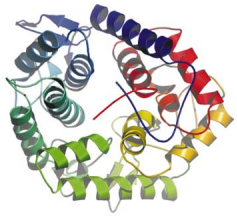
SET 5 (1.5A)
Input: 87 residues
Auto built: 357 residues, Two chains
Breaks: residue no. 1-6,142-156 and 377
Fig. 2*e*



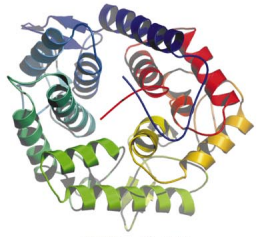
SET 2 (1.55A)
Input: 120 residues
Auto built: 370 residues, Single chain
Breaks: residue no. 1-6 and 377
Fig. 2*f*



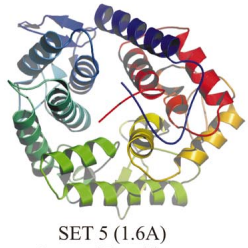
SET 6 (1.55A)
Input: 73 residues
Auto built: 370 residues, Single chain
Breaks: residue no. 1-6 and 377
Fig. 2*g*



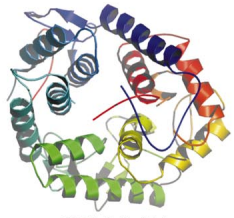
SET 7 (1.55A)
Input: 66 residues
Auto built: 347 residues, Single chain
Breaks: residue no. 1-6, 278-300 and 377
Fig. 2*h*



SET 1 (1.6A)
Input: 132 residues
Auto built: 370 residues, Single chain
Breaks: residue no. 1-6 and 377
Fig. 2*i*



SET 5 (1.6A)
Input: 87 residues
Auto built: 370 residues, Single chain
Breaks: residue no. 1-6 and 377
Fig. 2*j*



SET 6 (1.6A)
Input: 73 residues
Auto built: 350 residues, Single chain
Breaks: residue no. 1-6,100-127,
328-332 and 377
Fig. 2*k*

develop an initial phase set with the MPE approaching 80° to a final MPE around 15°. Each cycle of DDM starts from a set of phases with weights and calculates a weighted E map with strong E values. The map is modified and back-transformed by FFT to obtain a new set of structure factors. In DDM, all negative densities are set to zero; positive densities are modified according to the ratio $\rho/\sigma(\rho)$; modified densities are truncated to a value of $kn\sigma(\rho)$, where k is the constant (default value is 3); n is the number of DDM but after five cycles n always equals to 5. A couple of cycles of SER can help DDM to reach a global rather than local minimum.

Use of ACORN along with ARP/wARP (Perrakis *et al.*, 1999) followed by REFMAC (Murshudov *et al.*, 1999) in the structure elucidation of the triple mutant K53,56,121M of bovine pancreatic phospholipase A_2 under various options has been detailed elsewhere

(Banumathi *et al.*, 2002; Rajakannan *et al*., 2002; Velmurugan *et al*., 2002) and a manuscript is under preparation. This paper extends the application of the above procedures to real data of two macromolecules at 1.45 Å resolution. Results are also presented for the truncated data sets upto 1.6 Å resolution. Majority of the two structures (congerin II and alkaline cellulase N257) can be seen in the map and in the missing region of the automatically fitted map, densities exist and manual model building and iterative refinement would bring the rest of the structure also.

2.3. Congerin II

Shirai *et al*., (2002) have presented the structural details of *conger eel* galectin (congerin II) at 1.45 Å resolution. The PDB ID 1C1L was used. The synchrotron data corresponds to 1.45 Å resolution. Table 1 presents the consolidated results in the running of ACORN, ARP/wARP and REFMAC for different types of seed inputs to ACORN from the different beta sheets. Each model building cycle in ARP/wARP was always followed by 5 cycles of REFMAC refinement. Figs. 1 to 1*c* show the structures from the 1IS3 PDB and from the final outcomes of the three sets.

2.4. Alkaline cellulase N257

For the application of alkaline cellulase N257, the synchrotron data corresponds to 1.45 Å resolution (Shirai *et al.*, 2003). Various alpha helices containing different number of residues in this structure were used as initial fragments for ACORN. Table 2 shows the consolidated results for the various inputs to ACORN and figs. 2 to 2*c* present some of the final structures.

To look into the feasibility of using the above procedure to data at lower resolutions, artificial truncation to various resolutions were made using the data at 1.45 Å resolution. Similar calculations were then carried out for data at 1.5, 1.55 and 1.6 Å resolutions. Figs. 2*d* to *2k* present some of the final structures for the above resolutions. The breaks are mentioned in figure. In these missing regions, manual model building could be done as densities were present. In some cases the second iterative run has brought the rest of the structure also in an automated way.

3. Conclusion

From the above studies it is clearly shown how ACORN along with ARP/wARP followed by REFMAC can be used to solve structures for which data are at resolutions near to 1.45Å. Data mining using PDB is being carried out for the secondary structural features like alpha helices and beta sheets for seed information to ACORN instead of the present assumption of these features from the same structures. Attempts are also being made to tackle proteins of larger molecular weights using the above approach. Recent work carried out by us in alkaline cellulase K (Shirai *et.al.*, 2001) (357 amino acids) strongly suggests the success of this technique to a real data at 1.9 Å resolution also. All the computations were carried out using a Pentium III PC.

We would like to thank the Department of Science and Technology (DST), India for a major project and one of the authors (DV) thanks the Venture Business Laboratory, Nagoya University, Japan for the visiting professorship during which this work was carried out.

References

Banumathi, S., Rajakannan, V., Velmurugan, D., Dauter, Z., Dauter, M., Tsai, M. D. & Sekar, K. (2002). *Japanese Cryst. Society meeting*, Poster, P3-II-27, 123.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricongne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol*. 289, 83-92.
Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst*. D56, 232-237.
Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Yao, J. X. & Zheng , C, D. (2000).*Acta Cryst.* D56, 1137-1147
Hauptman, H. A. (1982*a*). *Acta Cryst.* A38, 289-294.
Hauptman, H. A. (1982*b*). *Acta Cryst.* A38, 632-641.
Hauptman, H. A., Potter, S., Weeks, C. M. (1982). *Acta Cryst.* A38, 294-300.
Hauptman, H. A., Velmurugan, D. & Han Fusen. (1991). *Direct Methods of solving crystal structures*, Edited by Schenk, H. Plenum press, New York, pp 403-406.
McAuley,K. E.,Yao,J. X., Dodson,E. J., Lehmbeck, J., Astergaard, P. R.& Wilson,K. S.(2001). *Acta Cryst.* D57,1571-1578
Mukherjee, M. (1999). *Acta Cryst.* D55, 820-825.
Mukherjee, M., Ghosh, S. & Woolfson, M. M. (1999). *Acta Cryst.* D55, 168-172.
Murshudov, G. N., Lebedev, A., Vagin, A. A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D55, 247-255.
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 Cryst. Society meeting*, Poster P3-I-22, 84.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* 277, 319-343.
Shirai, T., Ishida, H., Noda, J., Yamane, T., Ozaki, K., Hakamada, Y. & Ito, S. (2001). *J. Mol. Biol*. 310, 1079-1087.
Shirai, T., Matsui, Y., Shionyu-Mitsuyama, C., Yamane, T., Kamiya, H., Ishii, C., Ogawa, T. & Muramoto, K. (2002). *J. Mol. Biol.* 321, 879-889.
Shirai, T., Yamane, T., Kobayashi, T. & Ito, S. (2003). Personal Communication.
Velmurugan, D., Rajakannan, V., Yamane, T., Dauter, Z. & Sekar, K. (2002). *Japanese Cryst. Society meeting*, Poster P3-II-26, 122.
Weeks, C. M., De Titta, G. T., Hauptman, H. A., Thuman, P. & Miller, R. (1994). *Acta Cryst.* A50, 210-220.
Xu, H., Hauptman, H. A., Weeks, C. M. & Miller, R. (2000). *Acta Cryst.* D56, 238-240.
Yao, J. X. (2002). *Acta Cryst.*. D58, 1941-1947.

 Received 30 April 2003 • Accepted 15 October 2003