

Rapid Communication

SAD phasing by OASIS at different resolutions down to 0.30 nm and below*

Yao De-Qiang(姚德强)^{a)b)}, Li He(李鹤)^{c)}, Chen Qiang(陈强)^{d)},
Gu Yuan-Xin(古元新)^{a)}, Zheng Chao-De(郑朝德)^{a)}, Lin Zheng-Jiong(林政炯)^{a)c)},
Fan Hai-Fu(范海福)^{a)†}, Nobuhisa Watanabe(渡边信久)^{e)}, and Sha Bing-Dong(沙炳东)^{f)}

^{a)}Beijing National Laboratory for Condensed Matter Physics, Institute of Physics,
Chinese Academy of Sciences, Beijing 100080, China

^{b)}National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei 230026, China

^{c)}Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

^{d)}National Laboratory of Protein Engineering and Plant Genetic Engineering, Peking University, Beijing 100871, China

^{e)}Department of Biotechnology and Biomaterial Chemistry, Nagoya University, Nagoya 4648603, Japan

^{f)}Department of Cell Biology, University of Alabama at Birmingham, USA

(Received 20 July 2007)

Single-wavelength anomalous diffraction (SAD) phasing is increasingly important in solving *de novo* protein structures. Direct methods have been proved very efficient in SAD phasing. This paper aims at probing the low-resolution limit of direct-method SAD phasing. Two known proteins TT0570 and Tom70p were used as test samples. Sulfur-SAD data of the protein TT0570 were collected with conventional Cu-K α source at 0.18 nm resolution. Its truncated subsets respectively at 0.21, 0.30, 0.35 and 0.40 nm resolutions were used in the test. TT0570 Cu-K α sulfur-SAD data have an expected Bijvoet ratio $\langle |\Delta F| \rangle / \langle F \rangle \sim 0.55\%$. In the 0.21 nm case, a single run of OASIS-DM-ARP/wARP led automatically to a model containing 1178 of the total 1206 residues all docked into the sequence. In 0.30 and 0.35 nm cases, SAD phasing by OASIS-DM led to traceable electron density maps. In the 0.40 nm case, SAD phasing by OASIS-DM resulted in a degraded electron density map, which may be difficult to trace but still contains useful secondary-structure information. Test on real 0.33 nm selenium-SAD data of the protein Tom70p showed that even automatic model building was not successful, the combination of manual tracing and direct-method fragment extension was capable of significantly improving the electron-density map. This provides the possibility of effectively improving the manually built model before structure refinement is performed.

Keywords: OASIS, SAD phasing, dual-space fragment extension, proteins

PACC: 6110M, 8715

1. Introduction

Single-wavelength anomalous diffraction (SAD) method is increasingly important in solving *de novo* protein structures. In the Protein Data Bank (PDB),^[1] among the year 2006 entries, the number of protein structures solved by the SAD method was roughly the same as that solved by the multi-wavelength anomalous diffraction (MAD) method. The latter had been the first choice for solving *de novo* protein structures. The dual-space fragment extension procedure^[2] (see Fig.1) has been proved

very efficient for dealing with SAD data at about 0.2 nm or higher resolutions. It is important to know whether the procedure can be applied to SAD data at much lower resolutions and what is the limit for direct-method SAD phasing and for dual-space iteration. The present study aims at an insight into the problem. Tests were carried out with subsets truncated at 0.21, 0.30, 0.35 and 0.40 nm resolutions of the SAD data collected at 0.18 nm resolution with the protein TT0570 (TTHA1634 from *Thermus thermophilus* HB8).^[3] Tests were also carried out on real 0.33 nm resolution selenium-SAD data of the Protein Tom70p.^[4] Data of TT0570 are very good in qual-

*Project supported by the Innovation Project of the Chinese Academy of Sciences and the 973 Project (Grant No 2002CB713801) of the Ministry of Science and Technology of China.

†E-mail: fanhf@cryst.iphy.ac.cn

ity. Test calculations using them are intended for probing the low-resolution limit of our procedure in favourable cases. On the other hand, data of Tom70p are marginally useable owing to the sensitivity of sample crystals to x-ray radiation. Test calculations using

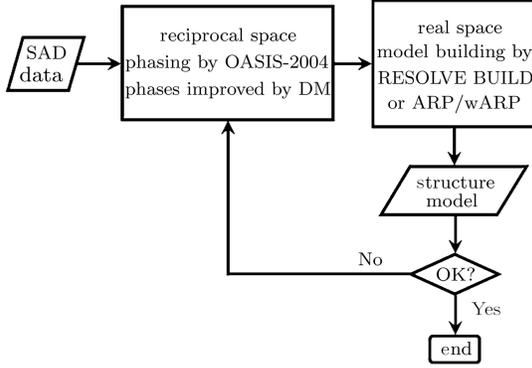


Fig.1. Flowchart of the dual-space fragment extension.

them are intended to show the performance of our procedure in a difficult case. In the latter case, the combination of SAD phasing by OASIS^[5]-DM^[6,7] with manual tracing can significantly improve the electron density map produced by SOLVE/RESOLVE. This provides the possibility of replacing automatic model building by manual tracing in dual-space fragment extension in low resolution cases.

2. Direct-methods SAD phasing

The theory of our phasing procedure is summarized in this section. For further details the reader is referred to the relevant papers cited below. In the SAD case, phases can be expressed as

$$\varphi_{\mathbf{h}} = \varphi''_{\mathbf{h}} \pm |\Delta\varphi_{\mathbf{h}}|, \quad (1)$$

where \mathbf{h} is the reciprocal vector, $\varphi''_{\mathbf{h}}$ is the phase contributed by the imaginary-part scattering of anomalous scatterers, i.e. the phase of

$$F''_{\mathbf{h}} = i \sum_{j=1}^N \Delta f''_j \exp(i2\pi\mathbf{h}_j \cdot \mathbf{r}_j), \quad (2)$$

where $\Delta f''_j$ is the imaginary-part correction to the atomic-scattering factor of the j th atom. $|\Delta\varphi_{\mathbf{h}}|$ is the absolute phase difference between the protein and $\varphi''_{\mathbf{h}}$. Both $\varphi''_{\mathbf{h}}$ and $|\Delta\varphi_{\mathbf{h}}|$ are known quantities provided the anomalous-scattering substructure is known. The ‘plus or minus’ sign preceding $|\Delta\varphi_{\mathbf{h}}|$ implies the SAD phase ambiguity. This can be resolved using the P_+ formula,^[8] which gives the probability of $\Delta\varphi_{\mathbf{h}}$ being positive:

$$P_+ = \frac{1}{2} + \frac{1}{2} \tanh \left\{ \sin |\Delta\varphi_{\mathbf{h}}| \left[\sum_{\mathbf{h}'} m_{\mathbf{h}'} m_{\mathbf{h}-\mathbf{h}'} \kappa_{\mathbf{h},\mathbf{h}'} \sin (\Phi'_3 + \Delta\varphi_{\mathbf{h}',\text{best}} + \Delta\varphi_{\mathbf{h}-\mathbf{h}',\text{best}}) + \chi \sin \delta_{\mathbf{h}} \right] \right\}. \quad (3)$$

Definitions of variables in formula (3) are as follows:

$$m_{\mathbf{h}} = \exp(-\sigma_{\mathbf{h}}^2/2) \left\{ \left[2 \left(P_+ - \frac{1}{2} \right)^2 + \frac{1}{2} \right] (1 - \cos(2\Delta\varphi_{\mathbf{h}})) + \cos(2\Delta\varphi_{\mathbf{h}}) \right\}^{1/2} \quad (4)$$

with

$$\sigma_{\mathbf{h}}^2 = \frac{(n\sigma_{\Delta F_{\mathbf{h}}})^2}{2|F''_{\mathbf{h}}|^2}, \quad (5)$$

where n is a scaling factor,^[9] $\sigma_{\Delta F_{\mathbf{h}}}$ is the standard deviation of Bijvoet difference $\Delta F_{\mathbf{h}}$.

$$\kappa_{\mathbf{h},\mathbf{h}'} = 2\sigma_3\sigma_2^{-3/2} E_{\mathbf{h}} E_{\mathbf{h}'} E_{\mathbf{h}-\mathbf{h}'}, \quad \sigma_n = \sum_j Z_j^n, \quad (6)$$

where $E_{\mathbf{h}}$ is the normalized structure-factor magnitude derived from $|F_{\mathbf{h}}|$, Z_j is the atomic number of

the j th atom in the unit cell.

$$\Phi'_3 = -\varphi''_{\mathbf{h}} + \varphi''_{\mathbf{h}'} + \varphi''_{\mathbf{h}-\mathbf{h}'} \quad (7)$$

is the three-phase structure invariant contributed by the imaginary-part scattering of the anomalous-scattering substructure.

$$\tan(\Delta\varphi_{\mathbf{h},\text{best}}) = 2 \left(P_+ - \frac{1}{2} \right) \sin |\Delta\varphi_{\mathbf{h}}| / \cos \Delta\varphi_{\mathbf{h}}, \quad (8)$$

$$\varphi_{\mathbf{h},\text{best}} = \varphi''_{\mathbf{h}} + \Delta\varphi_{\mathbf{h},\text{best}}, \quad (9)$$

$$\chi = 2E_{\mathbf{h}}E_{\mathbf{h},\text{known}} \left/ \left(\sum_i^{\text{unknown}} Z_i^2 \left/ \sum_j^{\text{total}} Z_j^2 \right. \right) \right., \quad (10)$$

where ‘known’ means the known partial structure of the protein, ‘unknown’ means the unknown part of the unit cell and ‘total’ means the whole unit cell.

$$\delta_{\mathbf{h}} = \varphi'_{\mathbf{h}} - \varphi''_{\mathbf{h}}, \quad (11)$$

where $\varphi'_{\mathbf{h}}$ is the phase contributed from real-part scattering of the known substructure. In practice, values of $\Delta\varphi_{\mathbf{h},\text{best}}$ and $m_{\mathbf{h}}$ to be substituted into formula (3) are first calculated respectively by formulae (8) and (4) with the initial P_+ set to $\frac{1}{2}$. The values of P_+ are updated by formula (3) in each cycle of the iterations.^[2,5] In the initial cycle, the “known” part of the protein is the anomalous-scattering substructure. Then from the next cycle onward, the ‘known’ part of the protein should be updated with the partial model found in the preceding cycle.

3. Data

SAD data from two proteins were used in the present test. They are summarized in Table 1. The crystal structure of TT0570^[3] was originally solved with Cr-K α sulfur-SAD data at 0.22 nm resolution using OASIS. In the present test diffraction data of

TT0570 were collected with conventional Cu-K α radiation (Rigaku FR-E SuperBright, 45 kV, 45 mA; Osmic Confocal Red optics; Rigaku R-Axis VII imaging plate detector modified for longer wavelength^[10]) at a temperature of 93 K. The post-sample helium path of the system was not used. A total of 1440 images of 0.5° oscillation were collected with one minute exposure time per image. A crystal of dimensions about 0.2 mm \times 0.5 mm \times 0.1 mm was mounted using the standard cryo-loop in cryoprotectant consisting of 20% (w/v) PEG 3350, 0.15 M lithium nitrate, pH 7.5 and 20% (v/v) glycerol. All data were indexed, integrated, and scaled with HKL2000.^[11] The data were truncated to four subsets with cut-off resolutions of 0.21, 0.30, 0.35 and 0.40 nm, respectively. Each subset was used in the present test separately. The crystal structure of Tom70p^[4] was originally solved by a combination of SOLVE/RESOLVE^[12–15] and OASIS in Ref.[5]. SeMet SAD data of Tom70p were collected in beam line SER-CAT at APS. The crystals were flash-frozen at 100 K in a nitrogen gas stream in cryoprotectant consisting of 100 mM MES buffer (pH 6.0), 30% (w/v) PEG 4000, 0.2 M ammonium acetate and 20% (v/v) ethylene glycol. SeMet Tom70p crystals are very sensitive to x-ray radiation and only single wavelength data with rather low redundancy can be collected using the same crystal. 0.33 nm resolution selenium-SAD data of Tom70p were used for the present test. Calculations here are independent of that of the original work in Ref.[4].

Table 1. Summary of test data.

protein	TT0570 (native)	Tom70p (SeMet)
x-ray wavelength/nm	0.15418 (Cu-K α)	0.09789 (synchrotron radiation)
space group	P 2 ₁ 2 ₁ 2	P 2 ₁
unit-cell parameters/nm; °	$a = 10.0572, b = 10.9096, c = 11.4862$	$a = 4.4894, b = 16.8774, c = 8.3407;$ $\beta = 102.74$
number of residues in ASU	1206	1234
number of monomers in ASU	2	2
resolution cut-off/nm	0.18 truncated respectively to 0.21, 0.30, 0.35 and 0.40	0.33
completeness/%	97.6 (95.4)	90.0 (62.7)
redundancy	29.2 (28.0)	3.3 (1.5)
I/sigma	43.9 (12.3)	16.3 (2.0)
Rmerge	0.044 (0.304)	0.080 (0.366)
anomalous scatterers in ASU	S (22)	Se (24)
estimated $\langle \Delta F \rangle / \langle F \rangle$ (%)	0.55	4.3
reference	[3]	[4]

4. Sulfur-SAD data of TT0570 truncated at 0.21, 0.30, 0.35 and 0.40 nm resolutions

Sulfur sites were located by SHELXD^[16] for each of the four truncated data sets separately. Root-mean-square deviations of positional parameters of sulfur atoms calculated against the final parameters are listed in Table 2. As is expected, the lower is the high-resolution cut-off, the larger is the root-mean-square deviation. This further enlarges the phasing errors and thus the map degrades as the high-resolution cut-off went down. Table 3 lists the cumulative phase errors resulting from a single run of OASIS-DM with TT0570 data sets at different resolutions. As is seen in the table, phasing errors increase as resolution decreases, even at the same F_{obs} shell. Corresponding electron-density maps are shown in Fig.2 and Fig.3. The map at 0.21 nm resolution

is readily traceable. With this map ARP/wARP^[17] yielded automatically a model containing 1178 of the total 1206 residues, all docked into the sequence. This means that, even Cu-K α sulfur-SAD data at 0.21 nm resolution with an extremely low Bijvoet ratio ($\langle |\Delta F| \rangle / \langle F \rangle \sim 0.55\%$) is sufficient for automatic solution of a big protein by a single run of OASIS-DM-ARP/wARP. For maps at 0.30, 0.35 and 0.40 nm resolutions, while automatic model building was not successful, it is clearly seen that the electron-density map at 0.30 nm resolution would not be difficult for manual tracing, because apart from good connectivity, most carbonyl groups are recognizable. The map at 0.35 nm resolution possesses also good connectivity, and many carbonyl groups are accompanied with reasonable electron densities. Hence it is still traceable by an experienced worker. As for the map at 0.40 nm resolution, there is lack of electron densities adjacent to most carbonyl groups, but the connectivity is still reasonable and some side chains are revealed clearly. Thus it has still chance for successful manual tracing.

Table 2. Root-mean-square deviations of positional parameters of sulfur atoms in TT0570 obtained with SAD data at different cut-off resolutions calculated against the final parameters.

resolution cut-off/nm	0.21*	0.30	0.35	0.40
root-mean-square deviation of heavy-atom sites/nm	0.041	0.041	0.048	0.055

*In the 0.21 nm case, heavy-atom sites were located using reflections with the high-resolution cut-off at 0.30 nm.

Table 3. Cumulative phasing errors/($^{\circ}$) of OASIS-DM for TT0570 data sets at different resolutions.

number of reflections	resolution cut-off/nm			
	0.21	0.30	0.35	0.40
500	16.9	16.9	17.5	20.8
1000	20.4	18.9	20.1	24.0
5000	28.4	26.1	28.1	33.0
10000	33.5	31.4	35.0	44.8
15000	37.7	36.4	43.8	
20000	41.3	41.5		
25000	44.0	47.6		
30000	46.5			
40000	51.0			
50000	55.1			
60000	59.1			

Reflections are sorted in descending order of F_{obs} and cumulated in groups as listed in the first column.

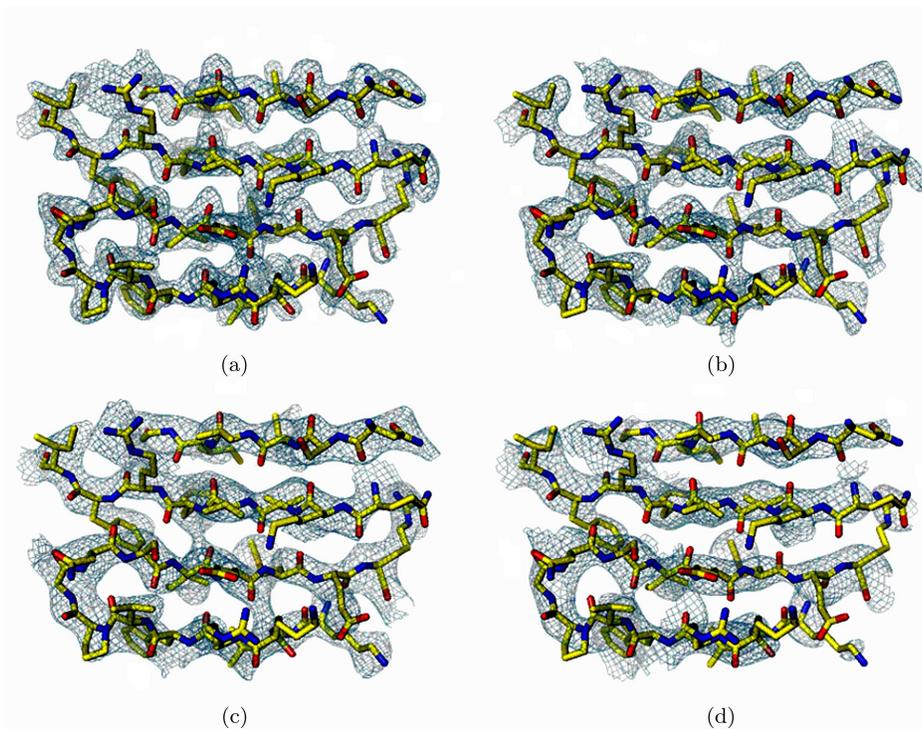


Fig.2. Partial electron-density maps (1σ) of TT0570 at different resolutions showing some β -sheets. (a) 0.21 nm resolution; (b) 0.30 nm resolution; (c) 0.35 nm resolution; (d) 0.40 nm resolution.

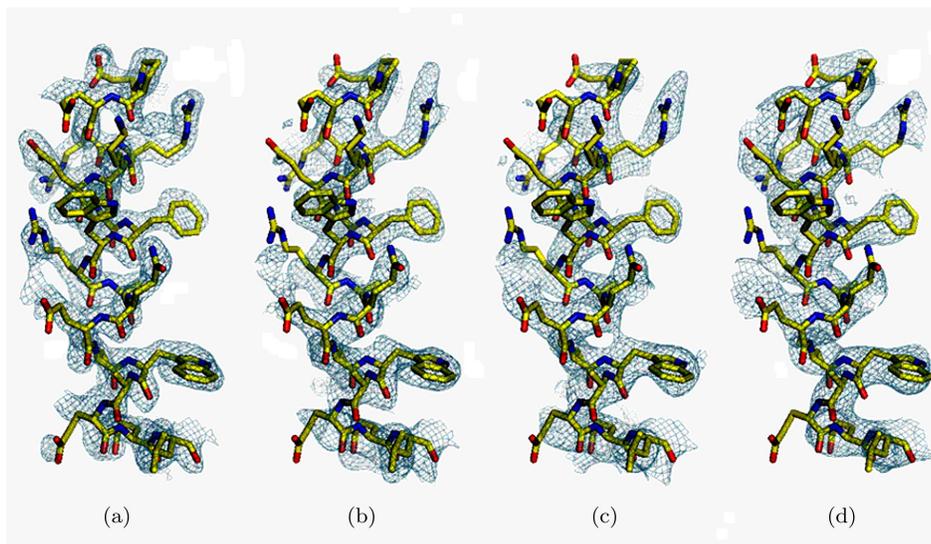


Fig.3. Partial electron-density maps (1σ) of TT0570 at different resolutions showing a segment of α -helix. (a) 0.21 nm resolution; (b) 0.30 nm resolution; (c) 0.35 nm resolution; (d) 0.40 nm resolution.

5. Real 0.33 nm resolution selenium-SAD data of Tom70p

0.33 nm resolution selenium-SAD data of Tom70p treated by SOLVE/RESOLVE based on the heavy-atom substructure found by SHELXD^[16] yielded an electron-density map showing clearly α -helices but with poor evidence on side chains. Automatic dual-

space fragment extension has been proved unsuccessful. Three models were constructed manually. Two of them are independent results of initial manual tracing for the two molecules in the asymmetric unit based on the SOLVE/RESOLVE map. They have some broken parts on main chains. The third model is a poly-alanine model mutated from the final model with artificial broken parts on main chains. All the models are far from complete as can be seen from the corre-

sponding R factors and phase errors listed in Table 4. SigmaA^[18] weighted Fourier calculation and OASIS fragment extension were performed based on the

three models separately. DM was then used for density modification in all cases. The resultant overall-averaged phase errors are listed in Table 5.

Table 4. Summary of manually constructed models for Tom70p.

model	I	II	III
broken locations on main chains	A263–A272	B317–B323	A263–A272 and B317–B323
R -factor	0.463	0.426	0.402
overall-averaged phase error/($^{\circ}$)	73.5	69.7	47.8

Table 5. Overall-averaged phase errors in degrees of different phasing methods based on different models.

method	model		
	I	II	III
SigmaA map + DM	66.9	60.9	46.2
OASIS + DM	61.2	59.0	53.1

*The overall-averaged phase error resulted from the initial SAD phasing by SOLVE/RESOLVE is 62.5° .

On the other hand, that from the original treatment by SOLVE/RESOLVE was 62.5° . It turns out that, starting from any of the three models OASIS-DM managed to return an overall phase error lower than that of SOLVE/RESOLVE, while SigmaA-DM failed to do the same in the case of model I. For model III SigmaA-DM gave much smaller overall phase error than that by OASIS-DM. This means that when

the model is sufficiently close to the true structure SigmaA-DM may work better than OASIS-DM. However, as will be seen later, even in this case, the electron density map of OASIS-DM is still significantly better than that of SigmaA-DM in regions where the model is broken. Two parts of electron-density maps corresponding to two broken regions of the models are shown for comparing results from different phasing methods. Figure 4 shows at 0.6σ the parts of electron-density maps covering residues A263–A272. There are no residues of model I placed within this region. SOLVE/RESOLVE map with the final model superimposed is shown in Fig.4(a). SigmaA map based on model I is shown in Fig.4(b), which is not much better than the SOLVE/RESOLVE map. However OASIS-DM map in Fig.4(c) based on the same model is better than either SigmaA map or SOLVE/RESOLVE map in giving clearer evidence on side chains.

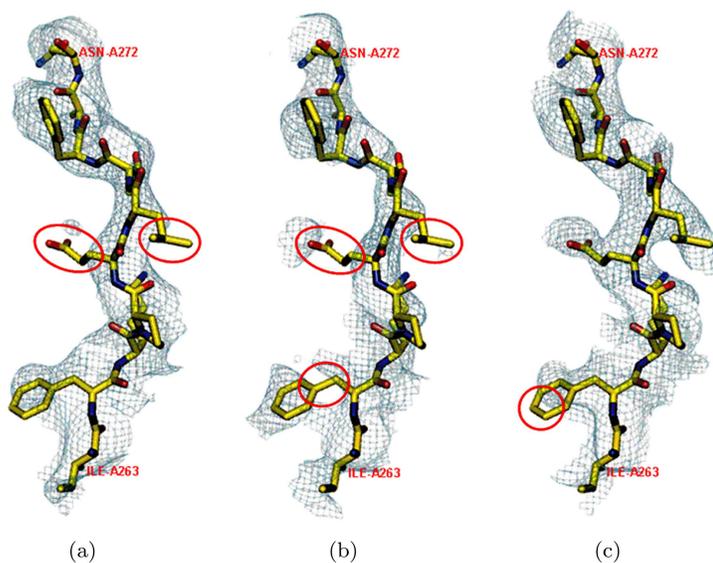


Fig.4. Partial electron density maps of Tom70p at 0.6σ covering residues A263–A272. (a) SOLVE/RESOLVE map; (b) SigmaA map based on model I; (c) OASIS-DM map based on the same model. The final model is superimposed on (a), (b) and (c) respectively. Regions where electron densities not matching well with the final model are circled in red.

Figure 5 shows electron-density maps at 1σ covering residues B313–B325. SOLVE/RESOLVE map with model II superimposed is shown in Fig.5(a), while the same map with the final model superimposed is shown in Fig.5(b). It is seen the SOLVE/RESOLVE map is not good in side-chain regions. SigmaA map Fig.5(c) based on model II failed to build up good enough electron densities in the broken region of the model. On the other hand, OASIS-DM map in Fig.5(d) based on the same model gives reasonable electron densities in that region. In addition, the map shows considerable improvement on the missing side-chain electron densi-

ties in comparison with the SOLVE/RESOLVE map. Figures 6 and 7 show the same part of electron density maps as that in Figs.4 and 5, respectively. Both SigmaA map and OASIS-DM map in Figs.6 and 7 were calculated based on model III. We see here again that OASIS-DM works better than SigmaA in retrieving electron densities corresponding to broken parts of the model and in improving electron densities corresponding to side chain regions. All figures in this paper except Fig.1 were plotted using the program PyMOL.^[19]

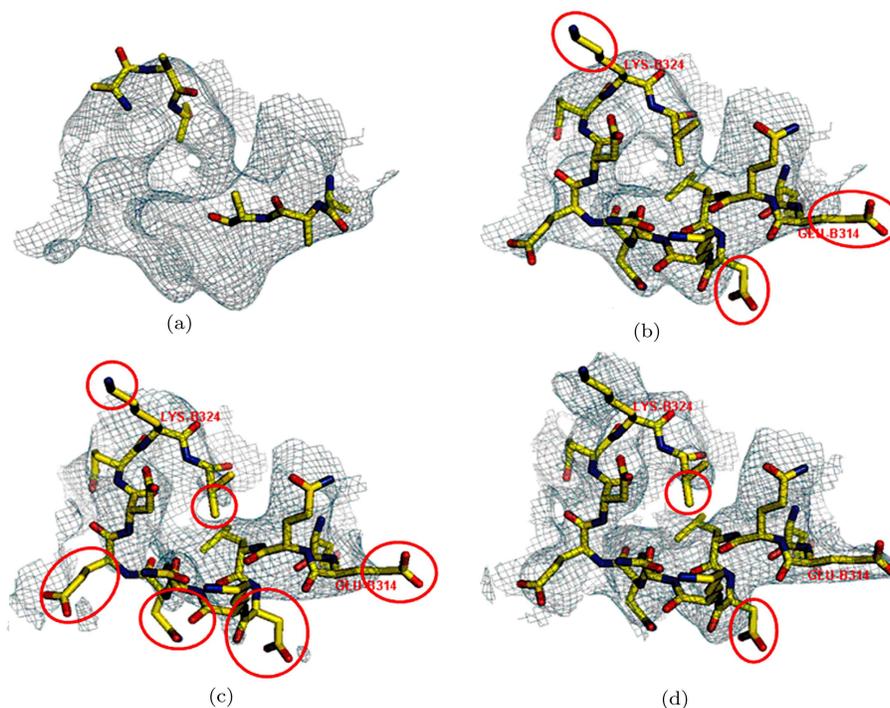


Fig.5. Partial electron density maps of Tom70p at 1σ covering residues B313–B325. (a) SOLVE/RESOLVE map with model II superimposed; (b) SOLVE/RESOLVE map with the final model superimposed; (c) SigmaA map based on model II; (d) OASIS-DM map based on the same model. The final model is superimposed on (c) and (d) respectively. Regions where electron densities not matching well with the final model are circled in red.

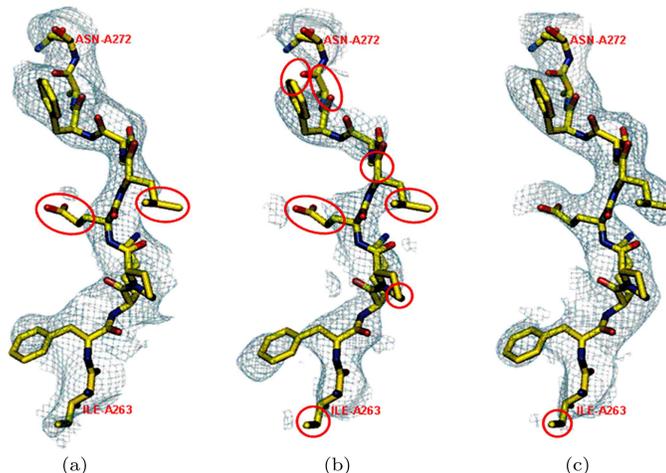


Fig.6. Partial electron density maps of Tom70p at 0.6σ covering residues A263–272. (a) SOLVE/RESOLVE map (b) SigmaA map based on model III; (c) OASIS-DM map based on the same model. The final model is superimposed on (a), (b) and (c) respectively. Regions where electron densities not matching well with the final model are circled in red.

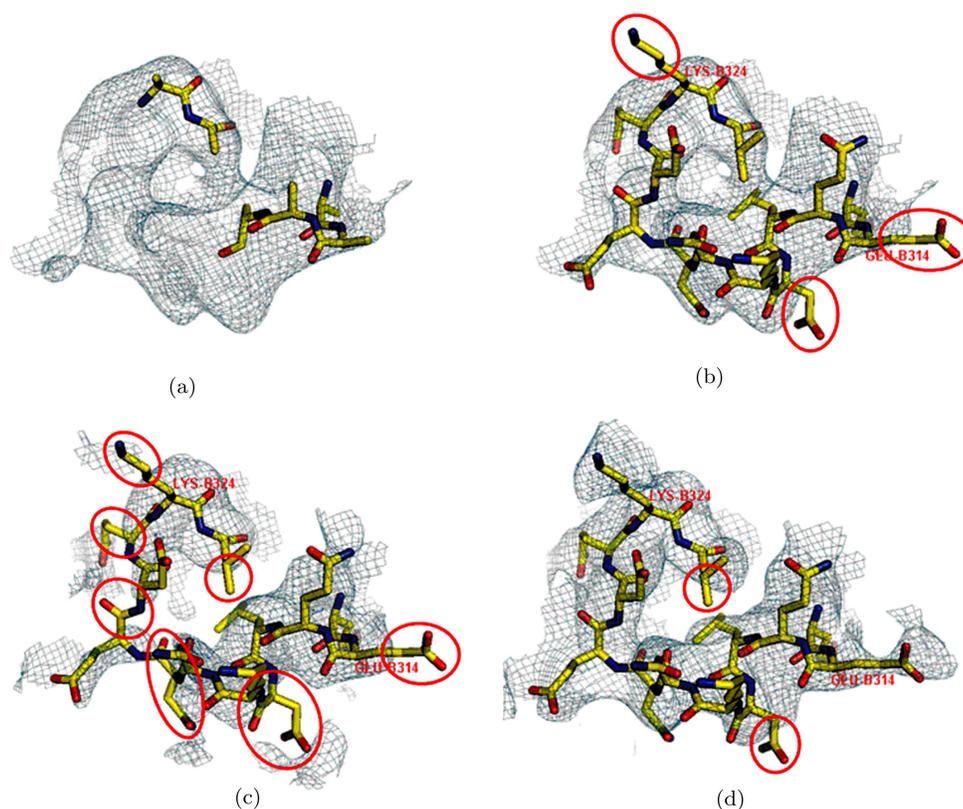


Fig. 7. Partial electron density maps of Tom70p at 1σ covering residues B313–B325. (a) SOLVE/RESOLVE map with model III superimposed; (b) SOLVE/RESOLVE map with the final model superimposed; (c) SigmaA map based on model III; (d) OASIS-DM map based on the same model. The final model is superimposed on (c) and (d) respectively. Regions where electron densities not matching well with the final model are circled in red.

6. Concluding remarks

Test with TT0570 shows that in favourable cases, OASIS-DM is capable of dealing SAD data down to 0.35 nm resolution, yielding interpretable electron density maps. Even at 0.40 nm resolution OASIS-DM can still provide useful structural information. Test with Tom70p shows that even in a rather difficult case, fragment extension by OASIS-DM based on a manually built model is capable of dealing with SAD data at lower than 0.30 nm resolution, yielding significantly improved electron density maps. This means

that the dual-space fragment extension technique can work efficiently even at resolutions lower than 0.30 nm provided the automatic model building is replaced by manual tracing or the automatic model building programs are improved to work at lower resolution. Dual-space fragment extension with either automatic model building or manual tracing provides an efficient tool for improving a rough model before an ordinary structure refinement is carried out.

The direct-methods program OASIS used in this study is available on the Web at <http://cryst.iphy.ac.cn> and at http://www.ccp4.ac.uk/prerelease_page.php.

References

- [1] Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N and Bourne P E 2000 *Nucleic Acids Research* **28** 235
- [2] Yao D Q, Huang S, Wang J W, Gu Y X, Zheng C D, Fan H F, Watanabe N and Tanaka I 2006 *Acta Cryst. D* **62** 883

- [3] Watanabe N, Kitago Y, Tanaka I, Wang J W, Gu Y X, Zheng C D and Fan H F 2005 *Acta Cryst. D* **61** 1533
- [4] Wu Y and Sha B 2006 *Nat. Struct. Mol. Biol.* **13** 589
- [5] Wang J W, Chen J R, Gu Y X, Zheng C D and Fan H F 2004 *Acta Cryst. D* **60** 1991
- [6] Collaborative Computational Project No 4 1994 *Acta Cryst. D* **50** 760
- [7] Cowtan K 1994 *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31** 34
- [8] Fan H F and Gu Y X 1985 *Acta Cryst. A* **41** 280
- [9] Wang J W, Chen J R, Gu Y X, Zheng C D, Jiang F and Fan H F 2004 *Acta Cryst. D* **60** 1987
- [10] Kitago Y, Watanabe N and Tanaka I 2005 *Acta Cryst. D* **61** 1013
- [11] Otwinowski Z and Minor W 1997 *Methods Enzymol.* **276** 307
- [12] Terwilliger T C and Berendzen J 1999 *Acta Cryst. D* **55** 849
- [13] Terwilliger T C 2000 *Acta Cryst. D* **56** 965
- [14] Terwilliger T C 2003 *Acta Cryst. D* **59** 38
- [15] Terwilliger T C 2003 *Acta Cryst. D* **59** 45
- [16] Schneider T R and Sheldrick G M 2002 *Acta Cryst. D* **58** 1772
- [17] Perrakis A, Morris R and Lamzin V S 1999 *Nature Struct. Biol.* **6** 458
- [18] Read R J 1986 *Acta Cryst. A* **42** 140
- [19] DeLano W L 2002 *The PyMOL Molecular Graphics System* (San Carlos, CA: DeLano Scientific)