Vol 16 No 10, October 2007
1009-1963/2007/16(10)/3022-07
Chinese Physics

SIR phasing by combination of SOLVE/RESOLVE and dual-space fragment extension involving OASIS*

He Yao(何 尧)[a], Gu Yuan-Xin(古元新)[a]†, Lin Zheng-Jiong(林政炯)[a,b], Zheng Chao-De(郑朝德)[a], and Fan Hai-Fu(范海福)[a]

[a] *Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing* 100080, *China*

[b] *Institute of Biophysics, Chinese Academy of Sciences, Beijing* 100101, *China*

(Received 16 July 2007; revised manuscript received 17 July 2007)

A new phasing procedure has been proposed for dealing with single isomorphous replacement (SIR) x-ray diffraction data. The procedure combines SOLVE/RESOLVE with the dual-space fragment extension involving OASIS. Two sets of SIR data at 0.28 nm resolution taken from the protein (R)-phycoerythrin (PDB code: 1LIA) were used in the test. For one of the two SIR data sets, a default run of SOLVE/RESOLVE based on the heavy-atom substructure found by SHLEXD led automatically to an interpretable electron density map. OASIS could not effectively improve the result. For the other set of SIR data, SOLVE/RESOLVE resulted in a fragmented model consisting of 454 of the total 668 residues, in which only 29 residues were docked into the sequence. Based on this model, 7 iteration cycles of OASIS-DM-RESOLVE (build only) yielded automatically a model of 547 residues with 133 residues docked into the sequence. The overall-averaged phase error decreased considerably and the quality of electron density map was improved significantly. Two more cycles of iterative OASIS-DM-RESOLVE were carried out, in which the output phases and figures of merit from DM were merged with that from the original run of SOLVE/RESOLVE before they were passed onto RESOLVE (build only). This led automatically to a model containing 452 residues with 173 docked into the sequence. The resultant electron density map is manually traceable. It is concluded that when results of SOLVE/RESOLVE are not sufficiently satisfactory, the combination of SOLVE/RESOLVE and OASIS-DM-RESOLVE (build only) may significantly improve them.

Keywords: SIR phasing, SOLVE/RESOLVE, OASIS, dual-space fragment extension for proteins
PACC: 6110M, 8715

1. Introduction

Multiple isomorphous replacement (MIR) method is one of the essential techniques of solving *de novo* protein structures. Single isomorphous replacement (SIR) method is an important supplement to MIR method. It needs fewer (only one) isomorphous heavy-atom derivative and hence less experimental work in both sample preparation and data collection. On the other hand, SIR method has the problem of intrinsic phase ambiguity and needs special treatment in phase derivation. This paper presents a new phasing procedure for SIR data, which combines two existing techniques. Test calculations showed that in a difficult case the new procedure yielded a result better than that obtainable with either of the two existing techniques alone. In a different context, the program SOLVE/RESOLVE[1−4] has been proved very efficient in solving protein structures with single/multi-wavelength anomalous diffraction (SAD/MAD) or SIR/MIR data. The dual-space fragment extension procedure combining OASIS,[5−7] DM[8,9] and ARP/wARP[10] or combining OASIS, DM and RESOLVE (build only)[3,4] has also been proved very efficient in dealing with protein SAD data.[11] In principle, such procedure can be applied to SIR data as well. The questions are, how well does the procedure perform when applied to the SIR case and, whether a combination of SOLVE/RESOLVE and dual-space fragment extension can do things better than either of them alone. In this paper, positive answers are given to both questions by a series of test calculations using two sets of SIR data at 0.28 nm resolution.

2. Data

Crystal structure of the protein R-phycoerythrin (PDB code: 1LIA) was originally determined at

*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: gu@cryst.iphy.ac.cn

0.28 nm resolution using MIR data of four heavy-atom derivatives.[12] The native and two heavy-atom derivatives were taken for the present study. One of the derivatives is the p-chloromercuriphenyl sulphonic acid derivative (hereafter referred to as Hg-derivative) and the other the K_2AuCl_4 derivative (hereafter referred to as Au-derivative). Crystallographic data of both derivatives and the native protein are summarized in Table 1.

Table 1. Summary of test data.

	Native	Au-derivative	Hg-derivative
Space group		R_3	
Unit cell parameters/nm, (°)	$a = b = 18.99$, $c = 6.01$; $\gamma = 120$	$a = b = 18.99$, $c = 5.98$; $\gamma = 120$	$a = b = 18.98$, $c = 6.00$; $\gamma = 120$
Resolution limit/nm	0.28	0.273	0.28
$R_{\rm merge}$(F)/%	4.2	4.41	4.84
Phasing power		1.51	1.36
Phasing power = $\langle F^2_{\mathbf{h},\,\text{heavy atom}}\rangle^{1/2}/\sigma_{\Delta F}$			

3. Direct-methods SIR phasing

In the SIR case, phases can be defined either as that associated with the native protein or that associated with the isomorphous derivative. In the following we define the phases as associated with the native protein. They can be expressed as

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

where $\boldsymbol{h}$ is the reciprocal vector; $\varphi'_{\mathbf{h}}$ is the phase of the heavy-atom (replacing-atom) substructure in the isomorphous derivative; $|\Delta\varphi_{\mathbf{h}}|$ is the absolute phase difference between the native and the heavy-atom substructure. Both $\varphi'_{\mathbf{h}}$ and $|\Delta\varphi_{\mathbf{h}}|$ are known quantities provided the heavy-atom substructure is known. The "plus or minus" sign preceding $|\Delta\varphi_{\mathbf{h}}|$ implies the SIR phase ambiguity, which can be resolved using the P_+ formula,[13] which gives the probability of $\Delta\varphi_{\mathbf{h}}$ being positive as follows:

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

Definitions of variables in formula (2) are as follows:

$$m_{\mathbf{h}} = \exp(-\sigma^2_{\mathbf{h}}/2)\Big\{\Big[2\Big(P_+ - \frac{1}{2}\Big)^2 + \frac{1}{2}\Big] \times (1 - \cos 2\Delta\varphi_{\mathbf{h}}) + \cos 2\Delta\varphi_{\mathbf{h}}\Big\}^{1/2} \tag{3}$$

with

$$\sigma^2_{\mathbf{h}} = \frac{2(n\sigma_{\Delta F_{\mathbf{h}}})^2|F_{\mathbf{h},D}|^2}{|F_{\mathbf{h},N}|^2|F_{\mathbf{h},H}|^2}, \tag{4}$$

where n is a scaling factor,[5] $\sigma_{\Delta F_{\mathbf{h}}}$ is the standard deviation of $\Delta F_{\mathbf{h}} = |F_{\mathbf{h},\mathrm{N}}| - |F_{\mathbf{h},\mathrm{D}}|$. $|F_{\mathbf{h},\mathrm{N}}|$, $|F_{\mathbf{h},\mathrm{D}}|$ and $|F_{\mathbf{h},\mathrm{H}}|$ are respectively the structure factor magnitudes of the native, derivative and heavy-atom substructure.

$$\kappa_{\mathbf{h},\mathbf{h}'} = 2\sigma_3\sigma_2^{-3/2}E_{\mathbf{h}}E_{\mathbf{h}'}E_{\mathbf{h}-\mathbf{h}'}, \quad \sigma_n = \sum_j Z_j^n, \tag{5}$$

where $E_{\mathbf{h}}$ is the normalized structure-factor magnitude derived from $|F_{\mathbf{h},\mathrm{N}}|$, Z_j is the atomic number of the jth atom in the unit cell.

$$\Phi'_3 = -\varphi'_{\mathbf{h}} + \varphi'_{\mathbf{h}'} + \varphi'_{\mathbf{h}-\mathbf{h}'} \tag{6}$$

is the three-phase structure invariant of the heavy-atom substructure.

$$\tan(\Delta\varphi_{\mathbf{h},\text{best}}) = 2\Big(P_+ - \frac{1}{2}\Big)\sin|\Delta\varphi_{\mathbf{h}}|/\cos\Delta\varphi_{\mathbf{h}}, \tag{7}$$

$$\varphi_{\mathbf{h},\text{best}} = \varphi'_{\mathbf{h}} + \Delta\varphi_{\mathbf{h},\text{best}}, \tag{8}$$

$$\chi = 2E_{\mathbf{h}}E_{\mathbf{h},\text{known}} \Big/ \Big(\sum_i^{\text{unknown}} Z_i^2 \Big/ \sum_j^{\text{total}} Z_j^2\Big), \tag{9}$$

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

$$\delta_{\mathbf{h}} = \varphi_{\mathbf{h},\text{known}} - \varphi'_{\mathbf{h}}. \tag{10}$$

In practice, values of $\Delta\varphi_{\mathbf{h},\text{best}}$ and m_{h} are first calculated respectively via formulae (7) and (3) with the initial P_+ set to 1/2. These values are then substituted into formula (2) to calculate new values of P_+. The process can be made iterative. In the initial cycle, the 'known' part of the protein consists of nothing, while during fragment extension the 'known' part of the protein should be updated in each cycle with the partial model found in the preceding cycle.

4. Phasing and model building protocols

Four protocols of phasing and model building were used in the present test.

1) SOLVE/RESOLVE: Intensity data and heavy-atom parameters were input to the program SOLVE/RESOLVE for SIR phasing, density modification and automatic model building.

2) OASIS-DM-RESOLVE (build only): Intensity data and heavy-atom parameters were input to the program OASIS for SIR phasing, DM for density modification and RESOLVE (build only) for automatic model building. Calculations were done iteratively until no further improvement on the output model could be made. For details of iterative OASIS-DM-RESOLVE (build only) the reader is referred to the original paper.[11]

3) SOLVE/RESOLVE + OASIS-DM-RESOLVE (build only): Intensity data and heavy-atom parameters were input to the program SOLVE/RESOLVE for SIR phasing, density modification and automatic model building. Then OASIS-DM-RESOLVE (build only) were used for fragment extension based on the model given by SOLVE/RESOLVE. The fragment extension was done iteratively until no further improvement on the output model could be made.

4) Merging iteration of OASIS-DM-RESOLVE (build only): Output phases and figures of merit from DM were merged with that from the original run of SOLVE/RESOLVE before they were passed onto RESOLVE (build only). The merged phases and figures of merit are defined as

$$[m_{\mathbf{h}}\exp(\mathrm{i}\varphi_{\mathbf{h},\mathrm{best}})]_{\mathrm{merged}} = \{[m_{\mathbf{h}}\exp(\mathrm{i}\varphi_{\mathbf{h},\mathrm{best}})]_{\mathrm{DM}} + [m_{\mathbf{h}}\exp(\mathrm{i}\varphi_{\mathbf{h},\mathrm{best}})]_{\mathrm{SOLVE/RESOLVE}}\}/2.$$

The fragment extension was done iteratively until no further improvement on the output model could be made. This protocol was used in combination with and after protocol 3. The purpose is to introduce some disturbance to the result of protocol 3 in case it converges to a result which is not quite satisfactory.

5. Heavy-atom substructure and NCS

Since the protein R-phycoerythrin was solved before the release of SHELXD[14,15] and SOLVE/RESOLVE, in the present test the heavy-atom substructures of Hg-derivative and Au-derivative were re-determined by SHELXD, then refined and searched for NCS by SOLVE/RESOLVE. Refined heavy-atom parameters and the twofold NCS generators obtained from SOLVE/RESOLVE (which are listed in Table 2) were used in subsequent test calculations.

Table 2. Summary of heavy-atom substructures.

	Au-derivative					Hg-derivative				
Heavy atom	x	y	z	q	B	x	y	z	q	B
1	0.8558	0.4437	0.0590	0.56	40.95	0.5917	0.1478	0.0530	0.32	39.60
2	0.0855	0.1909	0.0730	0.61	37.76	0.0871	0.1891	0.0574	0.35	38.61
3	0.1612	0.0130	0.0960	0.38	60.00	0.1870	0.0919	0.1000	0.14	2.70
4	0.5181	0.1707	0.0298	0.27	54.65	0.4832	0.2145	0.0141	0.13	11.55
5						0.2819	0.0662	0.2583	0.14	22.04
6						0.0705	0.1241	0.1306	0.08	1.00
7						0.2964	0.5479	0.3069	0.09	13.18
NCS operator										
R_{11} R_{12} R_{13}		−0.6428	0.7659	−0.0104			−0.6096	0.7926	0.0141	
R_{21} R_{22} R_{23}		0.7659	0.6425	−0.0223			0.7926	0.6091	0.0287	
R_{31} R_{32} R_{33}		−0.0104	−0.0223	−0.9997			0.0141	0.0287	−0.9995	
t_1 t_2 t_3		0.1101	−0.2155	−12.1095			−0.1051	0.2884	−13.2586	

x, y, z: fractional coordinates; q: occupancy; B: temperature factor; R_{ij}: components of the NCS rotation matrix; t_j: components of the NCS translation vector.

6. Comparison of results of SOLVE/RESOLVE using different SIR data sets

To see how the data quality affects the SIR phasing, SOLVE/RESOLVE results using Hg-SIR data and Au-SIR data are compared. As is shown in Table 1, Hg-derivative data have larger $R_{\rm merge}$ and lower phasing power. Besides, the heavy-atom substructure of Hg-derivative contains more sites with lower occupancies (see Table 2). Consequently, Hg-SIR data is less favourable than Au-SIR data for solving the protein structure. In the second and last columns of Table 3 there are listed cumulative phase errors resulting from SOLVE/RESOLVE using Hg-SIR data and Au-SIR data respectively. As can be seen, the accuracy of resultant phases from Au-SIR data is much higher than that from Hg-SIR data. This led to different results of automatic model building as shown in the second and the last column of Table 4. With Hg-SIR data SOLVE/RESOLVE yielded a model consisting of 454 of the total 668 residues, of which only 29 were docked into the sequence. On the other hand, much better result was obtained with the Au-SIR data, which led to a model consisting of 552 residues, of which 150 were docked into the sequence. Figures 1(a) and 1(b) show ribbon models obtained by SOLVE/RESOLVE with Hg-SIR and Au-SIR data respectively. In comparison with the final ribbon model (Fig.1(e)) it is seen that all α-helixes in the two models well match the final model (see Figs.2(a) and 2(b)), but the model from Au-SIR data provides much more structural information. Two portions of electron density maps with the final model superimposed are shown respectively in Figs.3 and 4, in which (a) is derived from Hg-SIR data, while (d) is from Au-SIR data, both phased by SOLVE/RESOLVE. As is seen, the electron density map derived from Au-SIR data is much easier to interpret than that from Hg-SIR data. Furthermore, OASIS could not effectively improve the result of SOLVE/RESOLVE with Au-SIR data, but was able to improve significantly the result of SOLVE/RESOLVE with Hg-SIR data, as will be seen in the next section.

Table 3. Cumulative phase errors for different SIR data and different phasing protocols.

Number of reflections	SOLVE/RESOLVE	Hg-SIR data						Au-SIR data
		OASIS-DM-RESOLVE (build only)		SOLVE/RESOLVE+OASIS-DM--RESOLVE (build only)				SOLVE/RESOLVE
		Cycle 0	Cycle 3	Cycle 3	Cycle 5	Cycle 7	Cycle 9*	
500	31.9	39.4	30.9	25.4	24.6	23.2	20.8	27.5
1000	35.2	42.0	32.6	27.4	26.8	25.4	24.2	30.1
5000	46.2	52.0	42.7	40.4	37.5	36.5	34.9	39.8
10000	53.0	57.2	49.5	47.2	44.7	44.4	42.2	46.5
15000	57.8	61.6	54.9	52.8	50.9	50.3	48.8	51.6
17500	60.0	63.8	57.7	56.0	54.1	53.7	52.0	54.5

Reflections were arranged in descending order of $F_{\rm obs}$ and cumulated into groups as listed in the first column. *In cycles 8 and 9, output phases and figures of merit from DM were merged with that from the original run of SOLVE/RESOLVE before they were passed onto RESOLVE (build only).

Table 4. Number of residues found automatically for different SIR data with different phasing protocols.

	Hg-SIR data				Au-SIR data
Protocol	I	II	III	IV	I
Number of residues found	454 (29)	456 (49)	547 (133)	452 (173)	552 (150)

Numbers of residues that have been docked into the sequence are shown in parentheses. Protocols: I — SOLVE/RESOLVE; II — 3 cycles of OASIS-DM-RESOLVE (build only); III — SOLVE/RESOLVE + 7 cycles of OASIS-DM-RESOLVE; IV — SOLVE/RESOLVE + 7 cycles of OASIS-DM-RESOLVE + 2 merging cycles of OASIS-DM-RESOLVE.

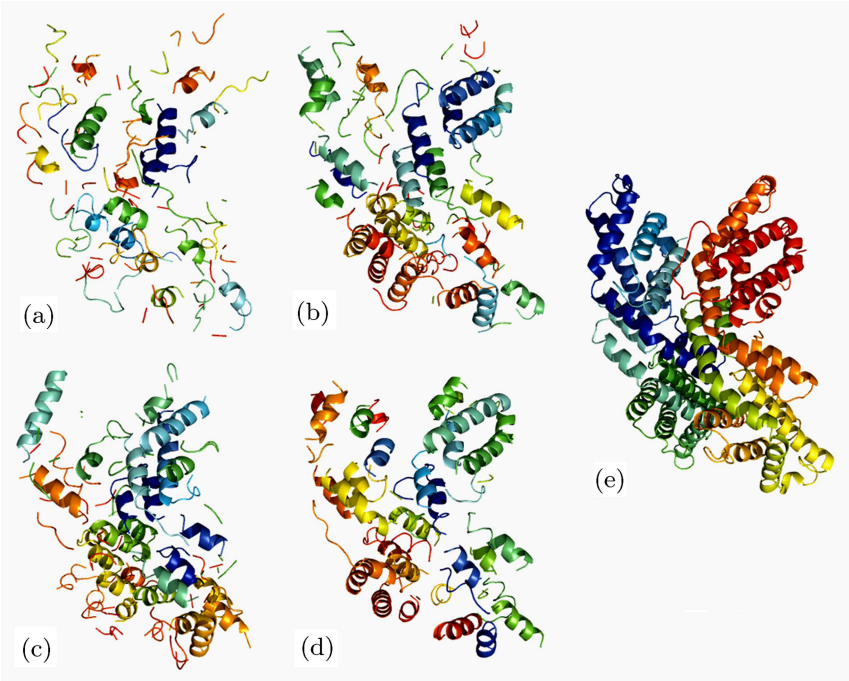

Fig.1. Ribbon models of the protein R-phycoerythrin. (a) Hg-SIR data phased by SOLVE/RESOLVE; (b) Au-SIR data phased by SOLVE/RESOLVE; (c) Hg-SIR data phased by SOLVE/RESOLVE followed by 7 cycles of OASIS-DM-RESOLVE; (d) Hg-SIR data phased by SOLVE/RESOLVE followed by 7 cycles of OASIS-DM-RESOLVE plus 2 merging cycles of OASIS-DM-RESOLVE; (e) final model.

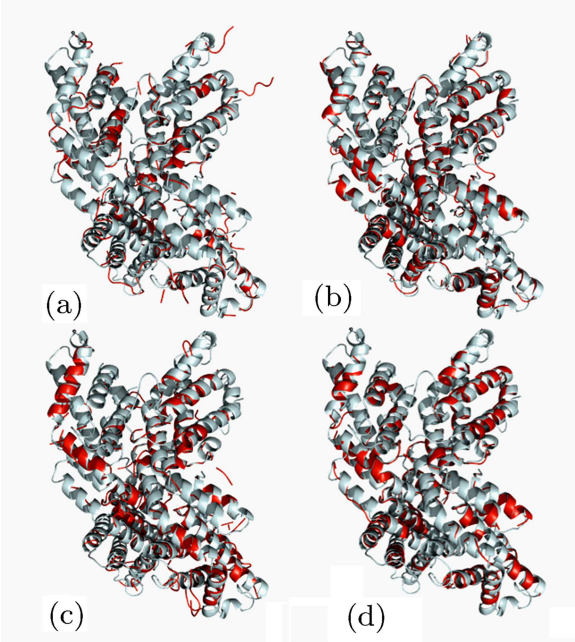

Fig.2. Ribbon models (red) from different phasing protocol and different SIR data matching with the final ribbon model (grey). (a) Hg-SIR data phased by SOLVE/RESOLVE; (b) Au-SIR data phased by SOLVE/RESOLVE; (c) Hg-SIR data phased by SOLVE/RESOLVE followed by 7 cycles of OASIS-DM-RESOLVE; (d) Hg-SIR data phased by SOLVE/RESOLVE followed by 7 cycles of OASIS-DM-RESOLVE plus 2 merging cycles of OASIS-DM-RESOLVE.

7. Comparison of results from Hg-SIR data with different protocols

Here we shall see that with the combination of SOLVE/RESOLVE and dual-space fragment extension by OASIS-DM-RESOLVE (build only) much better electron density maps can be obtained from Hg-SIR data. Four phasing and model-building protocols described in section 3 were applied to Hg-SIR data. The resultant cumulative phase errors are listed in Table 3. It is seen that SIR phasing by OASIS followed by three iteration cycles of OASIS-DM-RESOLVE (build only) fragment extension yielded slightly better results (column 4 of Table 3) than that of SOLVE/RESOLVE (column 2). However the electron density map is still not easy to trace. On the other hand, the combination of SOLVE/RESOLVE and dual-space fragment extension of OASIS-DM-RESOLVE (build only) yielded much better results (columns 5–8 of Table 3). Results of automatic model building from different protocols are listed in Table 4.

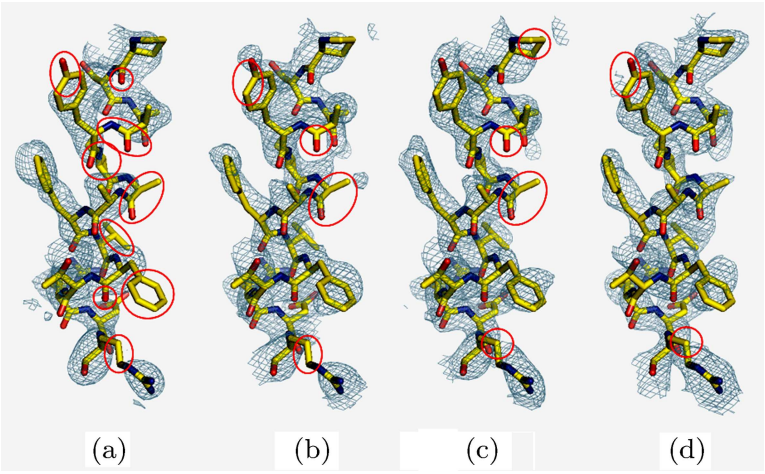

Fig.3. Partial electron density maps (1σ) covering residues A125-139 with the final model superimposed. (a) Hg-SIR data phased by SOLVE/RESOLVE; (b) Hg-SIR data phased by SOLVE/RESOLVE followed by 7 cycles of OASIS-DM-RESOLVE; (c) Hg-SIR data phased by SOLVE/RESOLVE followed by 7 cycles of OASIS-DM-RESOLVE plus 2 merging cycles of OASIS-DM-RESOLVE; (d) Au-SIR data phased by SOLVE/RESOLVE. Regions where electron densities not well matching the final model are circled in red.

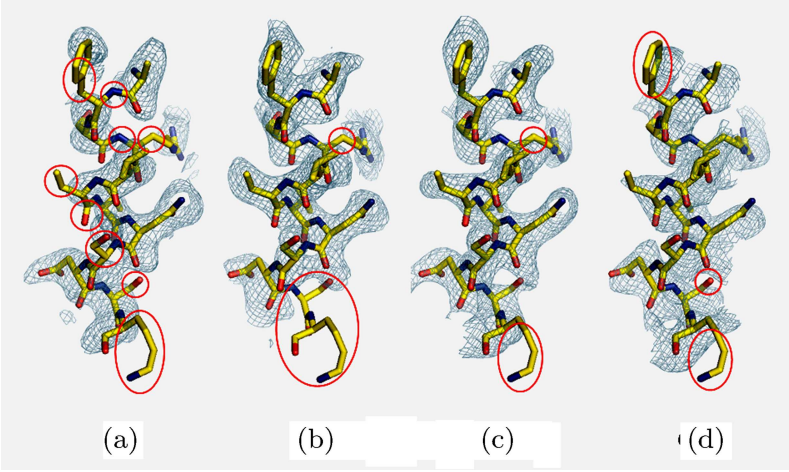

Fig.4. Partial electron density maps (1σ) covering residues B4-B15 with the final model superimposed. (a) Hg-SIR data phased by SOLVE/RESOLVE; (b) Hg-SIR data phased by SOLVE/RESOLVE followed by 7 cycles of OASIS-DM-RESOLVE; (c) Hg-SIR data phased by SOLVE/RESOLVE followed by 7 cycles of OASIS-DM-RESOLVE plus 2 merging cycles of OASIS-DM-RESOLVE; (d) Au-SIR data phased by SOLVE/RESOLVE. Regions where electron densities not well matching the final model are circled in red.

Protocol III yielded a model of 547 residues of which 133 were docked into the sequence, while Protocol IV yielded a model of 452 residues with 173 docked into the sequence. The latter result is comparable with that from Au-SIR data by protocol I, which consists of 552 residues with 150 docked into the sequence. Ribbon models from the Hg-SIR data phased by protocols III and IV are shown respectively in Figs.1(c) and 1(d). Two portions of the electron density maps deduced from different protocols are compared respectively in Figs.3 and 4. As is seen, the quality of electron density maps of Hg-derivative is continuously improving from (a) to (c). The quality of (c) is comparable with that from Au-SIR data phased by SOLVE/RESOLVE (d). All figures in this paper were plotted using the program PyMOL.[16]

8. Concluding remarks

Direct methods have been proved successful in SIR phasing and in fragment extension with SIR data at 0.28 nm resolution of a protein of considerable size. In case the quality of SIR data is not good enough and the SOLVE/RESOLVE result is not sufficiently satisfactory, the combination of SOLVE/RESOLVE with dual-space fragment extension by OASIS-DM-RESOLVE (build only) may lead to a better result.

Acknowledgement

The authors would like to thank Professor Chang Wen-rui and Dr Jiang Tao for making available the SIR data of R-phycoerythrin and for helpful discussions.

References

[1] Terwilliger T C and Berendzen J 1999 *Acta Cryst.* D **55** 849
[2] Terwilliger T C 2000 *Acta Cryst.* D **56** 965
[3] Terwilliger T C 2003 *Acta Cryst.* D **59** 38
[4] Terwilliger T C 2003 *Acta Cryst.* D **59** 45
[5] Wang J W Chen J R Gu Y X Zheng C D Jiang F and Fan H F 2004 *Acta Cryst.* D **60** 1987
[6] Wang J W Chen J R Gu Y X Zheng C D and Fan H F 2004 *Acta Cryst.* D **60** 1991
[7] Zhang T, He Y, Gu Y X, Zheng C D, Hao Q, Wang J W and Fan H F 2007 OASIS06 - a direct-method program for SAD/SIR phasing and reciprocal-space fragment extension (Institute of Physics, Chinese Academy of Sciences, Beijing 100080, China) (Program available at http://cryst.iphy.ac.cn and at http://www.ccp4.ac.uk/prerelease_page.php)
[8] Collaborative Computational Project No 4 1994 *Acta Cryst.* D **50** 760
[9] Cowtan K 1994 *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31** 34
[10] Perrakis A Morris R and Lamzin V S 1999 *Nature Struct. Biol.* **6** 458
[11] 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
[12] Chang W R Jiang T Wan Z L Zhang J P Yang Z X and Liang D C 1996 *J. Mol. Biol.* **262** 721
[13] Fan H F and Gu Y X 1985 *Acta Cryst.* A **41** 280
[14] Usón I and Sheldrick G M 1999 *Curr. Opin. Struct. Biol.* **9** 643
[15] Schneider T R and Sheldrick G M 2002 *Acta Cryst.* D **58** 1772
[16] DeLano W L 2002 *The PyMOL Molecular Graphics System* (San Carlos, CA: DeLano Scientific)