electronic reprint

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449
Editors: **E. N. Baker and Z. Dauter**

OASIS and molecular-replacement model completion

Yao He, De-Qiang Yao, Yuan-Xin Gu, Zheng-Jiong Lin, Chao-De Zheng and Hai-Fu Fan

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

OASIS and molecular-replacement model completion

Yao He,[a] De-Qiang Yao,[a,b] Yuan-Xin Gu,[a] Zheng-Jiong Lin,[a,c] Chao-De Zheng[a] and Hai-Fu Fan[a]*

[a]Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100080, People's Republic of China, [b]National Synchrotron Radiation Laboratory, University of Science and Technology of China, Hefei 230026, People's Republic of China, and [c]Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China

Correspondence e-mail: fanhf@cryst.iphy.ac.cn

Received 21 February 2007
Accepted 13 May 2007

A method of dual-space molecular-replacement model completion has been proposed which involves the programs *ARP*/*wARP*, *REFMAC*, *OASIS* and *DM*. *OASIS* is used in reciprocal space for phase refinement based on models built by *ARP*/*wARP*. For this purpose, the direct-method probability formula of breaking SAD/SIR phase ambiguities has been redefined. During the phase refinement, $\varphi''_{\mathbf{h}}$ in the expression $\varphi_{\mathbf{h}} = \varphi''_{\mathbf{h}} \pm |\Delta\varphi_{\mathbf{h}}|$ is redefined as a reference phase calculated from a randomly selected 5% of the atoms in the current structure model, while $|\Delta\varphi_{\mathbf{h}}|$ is defined as the absolute difference between the phase of the current model and $\varphi''_{\mathbf{h}}$. The probability formula $P_+(\Delta\varphi_{\mathbf{h}}) = \frac{1}{2} + \frac{1}{2}\tanh\{\sin|\Delta\varphi_{\mathbf{h}}| \times [\sum_{\mathbf{h}'} m_{\mathbf{h}} m_{\mathbf{h}-\mathbf{h}'} \kappa_{\mathbf{h}\,\mathbf{h}'} \sin(\Phi'_3 + \Delta\varphi_{\mathbf{h}'\mathrm{best}} + \Delta\varphi_{\mathbf{h}-\mathbf{h}'\mathrm{best}}) + \chi \sin\delta_{\mathbf{h}}]\}$ is then used to derive the sign of $\Delta\varphi_{\mathbf{h}}$. In this way the '0–$2\pi$' phase problem is reduced to a 'plus or minus' sign problem. The redefinition implies that during the refinement phases close to the true values will probably be kept unchanged, while those distant from the true values will probably undergo a large shift. This is the desired property of phase refinement. The procedure has been tested using protein diffraction data without SAD/SIR signals. The results show that dual-space MR-model completion making use of *OASIS* is much more efficient than that without.

1. Introduction

Dual-space fragment extension with SAD/SIR information (Wang, Chen, Gu, Zheng & Fan, 2004; Yao *et al.*, 2006) has proved to be very efficient for model completion in solving protein crystal structures using SAD/SIR data. Can this technique also be applied in the absence of SAD/SIR information? This has still to be explored. In a different context, the hybrid direct method proposed by Fan & Gu (1985) has been applied to improve SAD/SIR phasing. Is this method also useful for improving the molecular-replacement procedure? This is worth trying, as about 70% of protein structures submitted to the Protein Data Bank (Berman *et al.*, 2000) in recent years have been solved using the molecular-replacement method. The present work is aimed at answering these questions.

2. Method

2.1. Dual-space model completion

Dual-space model completion is an iterative procedure consisting of three parts: (i) real-space model building and refinement, (ii) reciprocal-space phase refinement based on a partial model and (iii) phase improvement by density modification. The first part is implemented by the programs *ARP*/

wARP (Perrakis *et al.*, 1999) and *REFMAC* (Murshudov *et al.*, 1997). These two programs are integrated in the *ARP/wARP* program suite. The second part is implemented by the program *OASIS*06 (Zhang *et al.*, 2007). The third part is implemented by the program *DM* (Collaborative Computational Project, Number 4, 1994; Cowtan, 1994). Fig. 1 shows a flowchart of the procedure. Further descriptions of the second part are given below. The program *OASIS*06 is freely available for academic purposes at the website http://cryst.iphy.ac.cn.

2.2. Redefining the P_+ formula

The P_+ formula

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

was originally proposed by Fan & Gu (1985) to break the phase ambiguity intrinsic in SAD/SIR data, *i.e.* to estimate the probability of $\Delta\varphi_{\mathbf{h}}$ being positive. In the SAD case, the symbols in (1) are defined as follows.

$$\Delta\varphi_{\mathbf{h}} = \varphi_{\mathbf{h}} - \varphi''_{\mathbf{h}} \text{ or } \varphi_{\mathbf{h}} = \varphi''_{\mathbf{h}} \pm |\Delta\varphi_{\mathbf{h}}|, \quad (2)$$

where $\varphi_{\mathbf{h}}$ is the phase of a reflection associated with the reciprocal vector $\mathbf{h}$. $\varphi''_{\mathbf{h}}$ is the phase contributed from imaginary-part scattering of anomalous scatterers, *i.e.* the phase of

$$F''_{\mathbf{h}} = i\sum_{j=1}^{N} \Delta f''_i \exp(i2\pi\mathbf{h}\cdot\mathbf{r}_j), \quad (3)$$

where $\Delta f''_j$ is the imaginary-part correction to the atomic scattering factor of the jth atom. Both $\varphi''_{\mathbf{h}}$ and $|\Delta\varphi_{\mathbf{h}}|$ are known quantities provided that the anomalous-scattering substructure is known.

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

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 and $\sigma_{\Delta F_{\mathbf{h}}}$ is the standard deviation of Bijvoet difference $\Delta F_{\mathbf{h}}$ (see Wang, Chen, Gu, Zheng, Jiang *et al.*, 2004).

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

where $E_{\mathbf{h}}$ is the normalized structure-factor magnitude associated with the reciprocal vector $\mathbf{h}$ and Z_j is the atomic number of the jth atom in the unit cell.

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

is the three-phase structure invariant contributed from the imaginary-part scattering of anomalous scatterers.

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

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

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

where 'known' refers to the known substructure, 'unknown' refers to the unknown part of the unit cell and 'total' refers to 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.

The main points implied by the P_+ formula are as follows.

(i) Bimodal SAD/SIR phase distributions lead the '0–2π' phase problem to a much simpler 'plus or minus' sign problem. This is of great benefit to the reliability of phase derivation.

(ii) The concept of best phases and figures of merit in protein crystallography (Blow & Crick, 1959) is introduced into direct-methods phasing.

(iii) Partial structure information is incorporated into the phasing process.

In the absence of SAD/SIR information, (1) and (2) have to be redefined in order to retain the above advantages. The phase $\varphi''_{\mathbf{h}}$ in (2) is now defined as a reference phase, which is calculated from a randomly selected 5% of the total atoms in the current structure model. $|\Delta\varphi_{\mathbf{h}}|$ is then redefined as the absolute difference between the phase of the current model and $\varphi''_{\mathbf{h}}$. In order to use (1) to derive the sign of $\Delta\varphi_{\mathbf{h}}$, we can either simulate the SAD case or the SIR case. The SAD case is chosen in the present work. This ensures that the algorithm is equally efficient for noncentric reflections and centric reflections and there is no need to consider whether or not the 'heavy-atom substructure' is centrosymmetric. The randomly selected 5% of the atoms in the current structure model are now taken as 'anomalous scatterers'. An artificial imaginary-

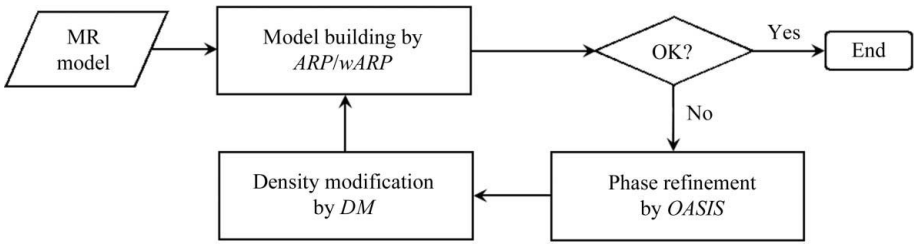

Figure 1
Flowchart of dual-space MR model completion.

Table 1
Summary of the test data.

	Acidic phospholipase A_2	E7_C–Im7_C complex
PDB code	1m8r	1ujz
Space group	$P6_1$	$I222$
Unit-cell parameters (Å, °)	$a = b = 83.27$, $c = 32.80$, $\gamma = 120$	$a = 62.88$, $b = 74.55$, $c = 120.44$
Resolution limit (Å)	1.9	2.1
X-rays	Cu $K\alpha$	Synchrotron
No. of reflections	8967	17107
No. of residues in the ASU	124	215 (molecule *A*, 87; molecule *B*, 128)
No. of residues in the starting model	Model 1, 63; model 2, 51; model 3, 40	46; 13 with side chains, all in molecule *A*
Reference	Xu *et al.* (2003)	Kortemme *et al.* (2004)

Table 2
Summary of model completion of acidic phospholipase A_2.

The number of residues found is given, with the number of residues docked in sequence given in parentheses. A, single run of *ARP/wARP*; D-A, *DM–ARP/wARP* iteration; O-D-A, *OASIS–DM–ARP/wARP* iteration.

Start	63-residue model	51-residue model			40-residue model		
Protocol	A	A	D-A	O-D-A	A	D-A	O-D-A
Cycle 0	119 (119)	37 (37)			32 (0)		
Cycle 1			33 (0)	41 (0)		14 (0)	42 (0)
Cycle 2			38 (27)	121 (121)		5 (0)	43 (0)
Cycle 3			72 (59)			0 (0)	57 (0)
Cycle 4			121 (121)				66 (0)
Cycle 5							121 (121)

part correction $\Delta f'' = 2.0$ to the atomic scattering factor is assumed for these atoms. The reference phase $\varphi''_{\mathbf{h}}$ can then be calculated using (3). Note that the value of $\Delta f''$ does not affect the resultant $\varphi''_{\mathbf{h}}$. For MR model completion in the absence of SAD/SIR signals there is no lack-of-closure error. Hence, in principle, (4) should reduce to

$$m_{\mathbf{h}} = \left\{\left[2\left(P_+ - \tfrac{1}{2}\right)^2 + \tfrac{1}{2}\right](1 - \cos 2\Delta\varphi_{\mathbf{h}}) + \cos 2\Delta\varphi_{\mathbf{h}}\right\}^{1/2}. \quad (12)$$

However, an alternative treatment is to keep (4) unchanged by replacing $\sigma_{\Delta F_{\mathbf{h}}}$ with $\sigma_{F_{\mathbf{h}}}$ in (5) so that

$$\sigma^2_{\mathbf{h}} = \frac{(n\sigma_{F_{\mathbf{h}}})^2}{2|F''_{\mathbf{h}}|^2}. \quad (13)$$

Originally, $\sigma^2_{\mathbf{h}}$ was the variance of two Gaussian functions used to approximate the bimodal SAD/SIR distribution (see Fan *et al.*, 1984); the use of (13) thus implies the introduction of an artificial bimodal SAD distribution with the 'sharpness' depending on the errors of experimentally measured structure-factor magnitudes. (4) and (13) are actually used in the program *OASIS*06 to calculate figures of merit. No further modification is necessary for formulae (6)–(11), keeping in mind that the term 'known' in (10) now means the current structure model and $\varphi'_{\mathbf{h}}$ in (11) is now the phase from the current structure model. Finally, (4), (8) and (9) are used to update figures of merit and 'best phases' during iteration.

2.3. Implications of the redefinition

Phase derivation in *OASIS* is associated with both 'best phases' and 'figures of merit'. For reasons of simplicity, the discussion in this section will focus only on 'phases' and will disregard 'figures of merit'. The redefinition of (1) implies the following phasing properties. At the beginning of model completion, according to (2), the initial phases will be those calculated from the starting model. By applying (1) and (2), a new phase will be given to each reflection. There are only two choices for each new phase: (i) if the sign of $\Delta\varphi_{\mathbf{h}}$ is positive then $\varphi_{\mathbf{h}}$ will be kept unchanged and (ii) if the sign of $\Delta\varphi_{\mathbf{h}}$ is negative then $\varphi_{\mathbf{h}}$ will be changed by $-2|\Delta\varphi_{\mathbf{h}}|$. Since $|\Delta\varphi_{\mathbf{h}}|$ ranges from 0° to 180° with a mean of ∼90°, $2|\Delta\varphi_{\mathbf{h}}|$ will mostly be a large angle. In principle, (1) tends to move $\varphi_{\mathbf{h}}$ towards its true value. Hence, if the initial $\varphi_{\mathbf{h}}$ is close to the true value, (1) will probably give a plus sign to keep $\varphi_{\mathbf{h}}$ unchanged. On the other hand, if the initial $\varphi_{\mathbf{h}}$ is distant from the true value, (1) will probably give a minus sign and let $\varphi_{\mathbf{h}}$ have a large shift. This property is what is expected for a good phase-refinement procedure.

2.4. Comparison of the redefined P_+ formula with the tangent formula in Karle recycling

The direct method used in the present study is in fact a phase refinement incorporated with partial structure information. A similar type of method is Karle recycling (Karle, 1968), which has long been used successfully for model completion in small-molecule crystallography. However, Karle recycling is not as efficient for proteins as for small molecules. In comparison with Karle recycling, the present method has the following advantages.

(i) The '0–2π' phase problem reduces to a much simpler 'plus or minus' sign problem. Obviously, the latter will provide much more reliable results.

(ii) Experimental errors in measured structure-factor magnitudes have been taken into account.

Figure 2
Model completion of acidic phospholipase A_2 with the 63-residue starting model. Left, starting model; middle, model built after a default run of *ARP/wARP*; right, refined (final) model.

(iii) In Karle recycling, partial structure information is implicitly included in the tangent formula by associating partial structure phases with normalized structure-factor magnitudes. This may cause strong model bias in three-phase structure invariants. On the other hand, in the P_+ formula (1), partial structure information is explicitly included in the term $\chi \sin \delta_{\mathbf{h}}$, which is separated from the term involving three-phase structure invariants. Model bias in phase relationships is thus weakened.

Table 3
Correlation between input phase errors and output signs of $\Delta\varphi_{\mathbf{h}}$ from *OASIS* during model completion of acidic phospholipase A_2 using the 40-residue starting model.

Nr, number of reflections with input phase error in the range specified in the first column; %, percentage of reflections with a plus sign for $|\Delta\varphi_{\mathbf{h}}|$ resulting from *OASIS*.

	Cycle 1		Cycle 2		Cycle 3		Cycle 4		Cycle 5	
Range of phase errors (°)	Nr	%	Nr	%	Nr	%	Nr	%	Nr	%
0–30	1828	55.6	1954	59.0	1839	58.9	2044	58.4	2092	61.9
30–60	1614	48.5	1575	48.8	1612	50.2	1542	46.8	1556	51.2
60–90	1316	34.7	1313	35.7	1278	37.6	1319	34.6	1332	34.2
90–120	1186	22.8	1112	25.1	1230	25.0	1117	24.2	1080	24.0
120–150	1133	20.3	1118	20.2	1111	20.3	1058	18.5	1112	18.4
150–180	1112	18.0	1120	18.5	1120	18.5	1110	16.1	1012	14.8

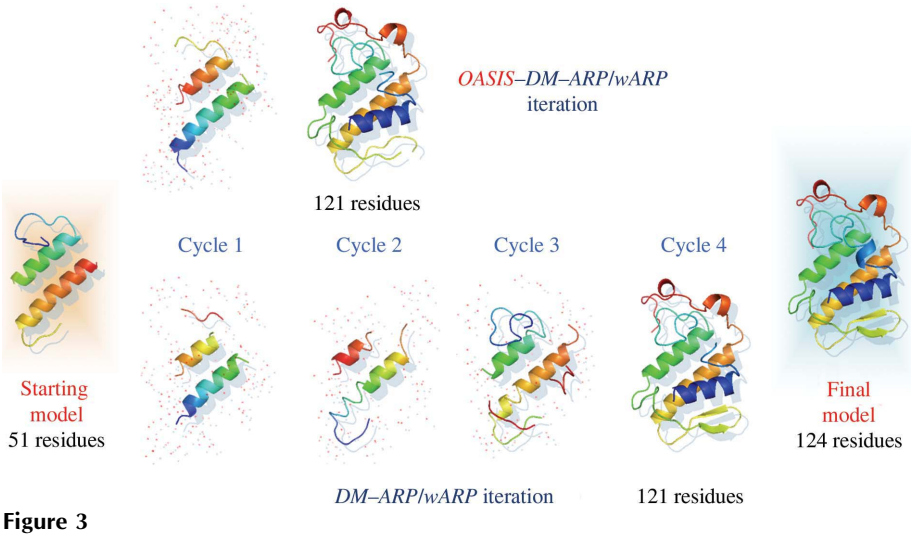

Figure 3
Model completion of acidic phospholipase A_2 with the 51-residue starting model. Upper row, models from two cycles of *OASIS–DM–ARP/wARP* iteration; lower row, models from four cycles of *DM–ARP/wARP* iteration (bypassing *OASIS* in the flowchart in Fig. 1). The starting and refined (final) models are shown on the left and right, respectively.

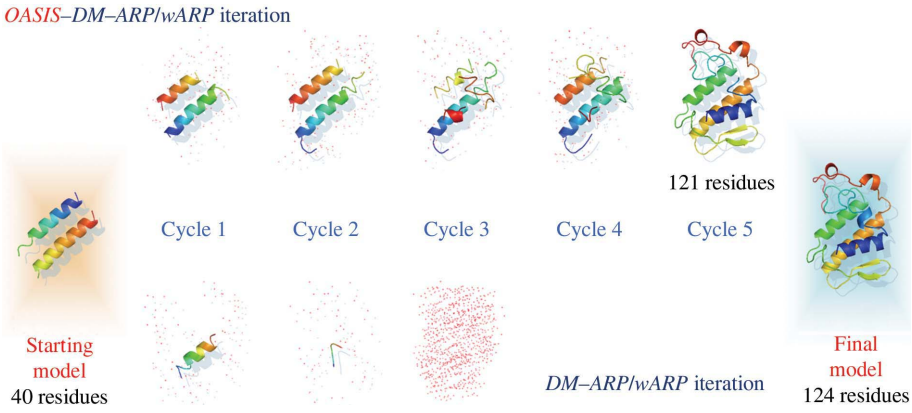

Figure 4
Model completion of acidic phospholipase A_2 with the 40-residue starting model. Upper row, models from five cycles of *OASIS–DM–ARP/wARP* iteration; lower row, models from three cycles of *DM–ARP/wARP* iteration (bypassing *OASIS* in the flowchart in Fig. 1). The starting and refined (final) models are shown on the left and right, respectively.

3. Test data

Two known proteins were selected for testing. They are summarized in Table 1. Structure-factor amplitudes without SAD/SIR signals were taken from the Protein Data Bank (Berman *et al.*, 2000).

4. Results and discussion

4.1. Model completion of the protein acidic phospholipase A_2

Acidic phospholipase A_2 from the venom of *Agkistrodon halys* Pallas is a 124-residue protein; its structure was originally solved by the molecular-replacement method (Wang *et al.*, 1996; Xu *et al.*, 2003). Three starting models were used in this test. They were obtained from a partial model truncated to 63, 51 and 40 residues, which correspond to 51, 41 and 32% of the whole structure, respectively. The 40-residue model contains two highly conserved α-helices that form the core of the molecule. The results of the test are summarized in Table 2. The models automatically built by *ARP/wARP* and plotted using *PyMOL* (DeLano, 2002) at various stages of model completion based on the three starting models are shown in Figs. 2, 3 and 4, respectively. For the 63-residue starting model, a default run of *ARP/wARP* automatically led to a model of 119 residues (96% of the whole structure) all docked into the sequence. In this case neither *DM* nor *OASIS* was necessary. For the 51-residue starting model, a default run of *ARP/wARP* did not improve the model. However, either iteration of *DM–ARP/wARP* (bypassing *OASIS* in the flowchart of Fig. 1) or iteration of *OASIS–DM–ARP/wARP* led to a model of 121 residues (98% of the whole structure) all docked into the sequence. *DM–ARP/wARP* needed

Table 4
Summary of model completion of E7_C–Im7_C.

The number of residues found is given, with the number of residues docked in sequence given in parentheses. A, single run of *ARP/wARP* based on the MR model from *Phaser*; D-A, *DM–ARP/wARP* iteration; O-D-A, *OASIS–DM–ARP/wARP* iteration.

Protocol	A	D-A	O-D-A
Cycle 0	46 (13)		
Cycle 1		14 (0)	62 (13)
Cycle 2		0 (0)	68 (0)
Cycle 3			94 (50)
Cycle 4			81 (61)
Cycle 5			151 (111)
Cycle 6			191 (161)
Cycle 7			201 (201)

four cycles for model completion, while *OASIS–DM–ARP/wARP* only needed two cycles to obtain an equivalent model with a slightly smaller overall average phase error. For the 40-residue starting model, *DM–ARP/wARP* iteration failed to extend the model and was completely corrupted in the third cycle, resulting in a set of dummy atoms that bore no secondary-structure information. However, *OASIS–DM–ARP/wARP* iteration led to a 121-residue model in five cycles. The variations of overall average phase errors during model completion are plotted in Fig. 5. *DM–ARP/wARP* iteration is plotted as a black curve, while *OASIS–DM–ARP/wARP* iteration is plotted as a cyan curve. As can be seen from the latter, *ARP/wARP* is crucial for the final cycles. However, its success depends on whether the overall average phase error of the input model is lower than a certain value. In this example and those in our previous study (Yao *et al.*, 2006), this value is within the range 45–55°. Comparing the black curve with the cyan curve, it is clear that a combination of *OASIS* and *DM* is better than *DM* alone for improving the *ARP/wARP* phases. The property of *OASIS* described in §2.3 is validated in Table 3, which shows the correlation between input phase errors and output signs for $|\Delta\varphi_{\mathbf{h}}|$ by *OASIS*. Reflections are grouped into shells according to their input phase errors as

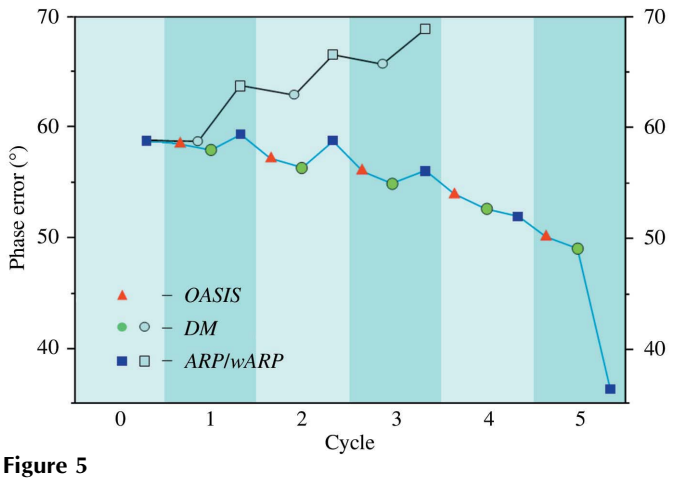

Figure 5
Variation of overall average phase errors during model completion of acidic phospholipase A_2 with the 40-residue starting model.

Table 5
Correlation between input phase errors and output signs of $\Delta\varphi_{\mathbf{h}}$ from *OASIS* during model completion of E7_C–Im7_C.

Nr, number of reflections with input phase error in the range specified in the first column; %, percentage of reflections with a plus sign for $|\Delta\varphi_{\mathbf{h}}|$ resulting from *OASIS*.

	Cycle 1		Cycle 3		Cycle 5		Cycle 7	
Range of phase errors (°)	Nr	%	Nr	%	Nr	%	Nr	%
0–30	3337	54.7	3761	60.0	4197	65.3	5733	76.0
30–60	2856	47.5	2874	49.1	3049	51.7	2469	43.2
60–90	2635	37.5	2359	38.0	2279	34.6	1852	17.5
90–120	2281	31.2	2189	27.3	1975	24.4	1702	6.7
120–150	2175	23.5	2097	21.1	1858	15.7	1813	5.6
150–180	2064	21.6	2071	17.3	1991	12.9	1779	4.8

listed in the first column. It is seen in each cycle that smaller input phase errors lead to larger percentages of reflections receiving a plus sign for $|\Delta\varphi_{\mathbf{h}}|$. This means that input phases with smaller errors have a higher possibility of being kept unchanged by *OASIS*, while those with larger errors have a higher possibility of receiving a large phase shift. Comparing the statistics of cycle 1 to those of cycle 5, it is found that the number of reflections with the smallest input phase errors increases and the percentage of these reflections that receive a plus sign for $|\Delta\varphi_{\mathbf{h}}|$ also increases after phase refinement. The opposite is found for the reflections with the largest input phase errors. This signifies that the model completion is progressing satisfactorily.

4.2. Model completion of the protein–protein complex E7_C–Im7_C

This test was designed to simulate a difficult situation. A protein complex consisting of two different molecules *A* and *B* was selected as the test sample. The MR model was derived from a search model that only bears structural similarity to molecule *A*. This MR model was then extended by dual-space iteration to solve the whole structure. The complex of colicin E7 DNAse and the IM7 immunity protein (E7_C–Im7_C; PDB code 1ujz) is a 215-residue protein containing two molecules: *A* of 87 residues and *B* of 128 residues. Molecule *A* of a similar complex (PDB code 1bxi) bears 60% sequence identity to molecule *A* of E7_C–Im7_C. The root-mean-square deviation between the two *A* molecules is 1.38 Å. Molecule *A* of 1bxi was pruned using the program *CHAINSAW* (Schwarzenbacher *et al.*, 2004) to provide a search model for *Phaser* (Read, 2001) to implement molecular replacement. Based on the MR model obtained from *Phaser*, *ARP/wARP* automatically built a model of 46 residues, none of which were within the region of molecule *B* of E7_C–Im7_C and of which only 13 have side chains. This 46-residue model, amounting to ∼20% of the structure of E7_C–Im7_C, was used as the starting point for *OASIS–DM–ARP/wARP* iteration in the test. The results are summarized in Table 4. Models at different stages automatically built by *ARP/wARP* and plotted using *PyMOL* are shown in Fig. 6. It was found in this case that iteration of *DM–ARP/wARP* was unable to obtain a solution and was corrupted completely in cycle 2,

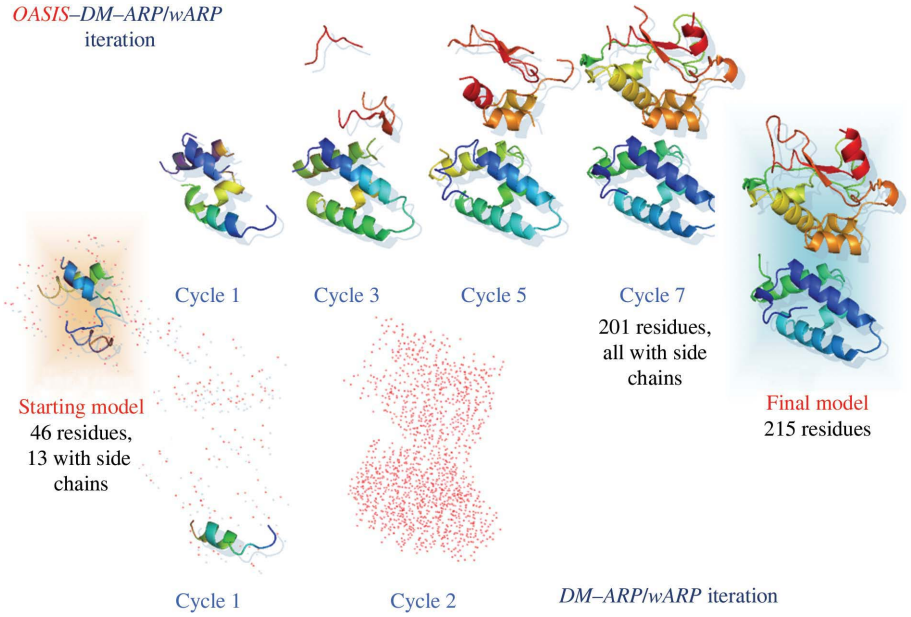

Figure 6
Model completion of E7_C–Im7_C with the 46-residue starting model. Upper row, models from the first, third, fifth and seventh cycles of *OASIS–DM–ARP/wARP* iteration; lower row, models from two cycles of *DM–ARP/wARP* iteration (bypassing *OASIS* in the flowchart in Fig. 1). The starting model and refined (final) model are shown on the left and right, respectively.

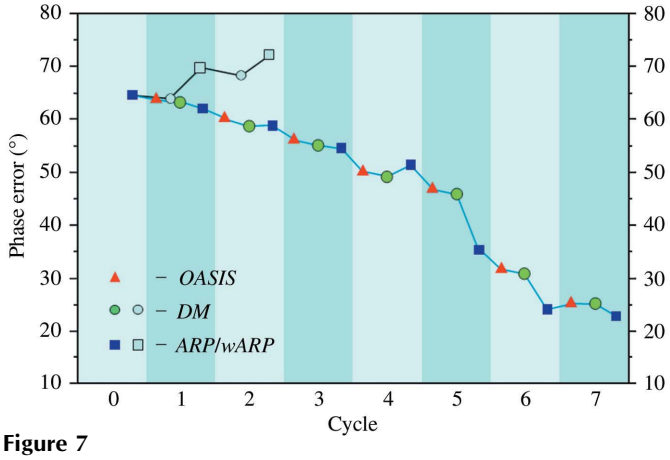

Figure 7
Variation of overall average phase errors during model completion of E7_C–Im7_C.

leading to a set of dummy atoms that contained no secondary-structure information. However, seven cycles of *OASIS–DM–ARP/wARP* iteration managed to produce a model of 201 residues (∼94% of the whole structure), all docked into the sequence. The variations of overall average phase errors during model completion are plotted in Fig. 7. The black curve shows *DM–ARP/wARP* iteration, while the cyan curve shows *OASIS–DM–ARP/wARP* iteration. The phasing property of *OASIS* described in §2.3 is validated in Table 5. Here, it is again seen in each cycle that smaller input phase errors lead to a larger percentage of reflections receiving a plus sign for $|\Delta\varphi_{\mathbf{h}}|$. On the other hand, from cycle 1 to cycle 7 the number of reflections having smallest input phase errors increases and their percentage of receiving a plus sign for $|\Delta\varphi_{\mathbf{h}}|$ also increases. The opposite is found for the reflections with the largest input phase errors. This implies a successful phase refinement.

5. Concluding remarks

The diffraction data used in the present study were at about 2 Å resolution. For data at lower resolution, the program *RESOLVE* (build only; Terwilliger, 2003*a*,*b*) can be used to replace *ARP/wARP* and manual adjustment to model building may sometimes be inevitable during the iteration. Dual-space procedures involving direct methods have proved to be very efficient for *ab initio* solution of protein structures using diffraction data at atomic resolution (Weeks *et al.*, 1993; Sheldrick & Gould, 1995; Foadi *et al.*, 2000), for locating heavy-atom sites in proteins (Weeks *et al.*, 1993; Sheldrick & Gould, 1995), for SAD/SIR phasing (Gu, Chang *et al.*, 2002; Gu, Jiang *et al.*, 2002) and for model completion in the presence of SAD/SIR signals (Yao *et al.*, 2006). It is shown in the present study that they are also very efficient in MR model completion in the absence of SAD/SIR information. Direct methods thus have an important influence on nearly all aspects of single-crystal diffraction analysis of proteins.

H-FF would like to thank Professor D. C. Liang for helpful discussions. This work was supported by the Innovation Project of the Chinese Academy of Sciences and by the 973 Project (grant No. 2002CB713801) of the Ministry of Science and Technology of China.

References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, San Carlos, CA, USA.
Fan, H.-F. & Gu, Y.-X. (1985). *Acta Cryst.* A**41**, 280–284.
Fan, H.-F., Han, F.-S. & Qian, J.-Z. (1984). *Acta Cryst.* A**40**, 495–498.
Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Yao, J.-X. & Zheng, C.-D. (2000). *Acta Cryst.* D**56**, 1137–1147.
Gu, Y. X., Chang, W. R., Jiang, T., Zheng, C. D. & Fan, H. F. (2002). *Acta Cryst.* A**58**, 547–551.
Gu, Y. X., Jiang, F., Sha, B. D. & Fan, H. F. (2002). *Z. Kristallogr.* **217**, 710–714.
Karle, J. (1968). *Acta Cryst.* B**24**, 182–186.

Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004). *Nature Struct. Mol. Biol.* **11**, 371–379.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.
Schwarzenbacher, R., Godzik, A.., Grzechnik, S. K. & Jaroszewski, L. (2004). *Acta Cryst.* D**60**, 1229–1236.
Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* B**51**, 423–431.
Terwilliger, T. C. (2003*a*). *Acta Cryst.* D**59**, 38–44.
Terwilliger, T. C. (2003*b*). *Acta Cryst.* D**59**, 45–49.
Wang, J. W., Chen, J. R., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2004). *Acta Cryst.* D**60**, 1991–1996.
Wang, J. W., Chen, J. R., Gu, Y. X., Zheng, C. D., Jiang, F. & Fan, H. F. (2004). *Acta Cryst.* D**60**, 1987–1990.
Wang, X.-Q., Yang, J., Gui, L.-L., Lin, Z.-J., Chen, Y.-C. & Zhou, Y.-C. (1996). *J. Mol. Biol.* **255**, 669–676.
Weeks, C. M., DeTitta, G. T., Miller, R. & Hauptman, H. A. (1993). *Acta Cryst.* D**49**, 179–181.
Xu, S., Gu, L., Jiang, T., Zhou, Y. & Lin, Z. (2003). *Biochem. Biophys. Res. Commun.* **300**, 271–277.
Yao, D.-Q., Huang, S., Wang, J.-W., Gu, Y.-X., Zheng, C.-D., Fan, H.-F., Watanabe, N. & Tanaka, I. (2006). *Acta Cryst.* D**62**, 883–890.
Zhang, T., He, Y., Gu, Y.-X., Zheng, C.-D., Hao, Q., Wang, J.-W. & Fan, H.-F. (2007). *OASIS*06 – *A Direct-Method Program for SAD/SIR Phasing and Reciprocal-Space Fragment Extension.* Institute of Physics, Chinese Academy of Sciences, Beijing, People's Republic of China.