OASIS4.0—a new version of the program OASIS for phasing protein diffraction data*

Zhang Tao(张 涛)[a)b)], Gu Yuan-Xin(古元新)[b)†],
Zheng Chao-De(郑朝德)[b)], and Fan Hai-Fu(范海福)[b)‡]

[a)] *Research Institute of Magnetic Materials, School of Physical Sciences and Technology, Lanzhou University, Lanzhou* 730000, *China*

[b)] *Beijing National Laboratory for Condensed Matter Physics and Key Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing* 100190, *China*

(Received 21 April 2010; revised manuscript received 29 April 2010)

The program OASIS4.0 has been released. Apart from the improved single-wavelength anomalous diffraction (SAD) phasing algorithm described in a separate paper, an important new feature in this version is the automation of the iterative phasing and model-building process in solving protein structures. A new graphical user's interface (GUI) is provided for controlling and real-time monitoring the dual-space iterative process. The GUI is discussed in detail in the present paper.

Keywords: program OASIS4.0, automatic dual-space fragment extension, proteins
PACC: 6110M, 8715

1. Introduction

The program OASIS is for direct-method phasing of single-wavelength anomalous diffraction (SAD) or single isomorphous replacement (SIR) data from proteins and for reciprocal-space extension of partial protein-structure models. There are four different versions of OASIS so far released. The first version of OASIS was released in 2000.[1] Only single runs of direct-method SAD/SIR phasing can be performed with this version. The second version of OASIS was released in 2004.[2,3] The direct-method reciprocal-space fragment extension in the presence of SAD/SIR information was first introduced into this version. A combination of OASIS-2004 with density-modification program (DM[4−6] or RESOLVE[7]) and model-building program (RESOLVE[8,9] and/or ARP/wARP[10]) leads to the dual-space iterative SAD/SIR phasing and fragment extension, which dramatically enhances the efficiency of direct-method SAD/SIR phasing.[11,12] The third version of OASIS was released in 2006.[13] Introduced into this version was the direct-method reciprocal-space fragment extension in the absence of SAD/SIR information. Combination of OASIS-2006 with density-modification program and model-building program leads to the dual-space iterative fragment extension in the absence of SAD/SIR information, resulting in significant improvement of the molecular replacement (MR) model completion.[14] OASIS4.0 is the fourth version of OASIS. New features of this version include the improvement of SAD-phasing algorithm and the automation of dual-space fragment extension with or without the help of SAD/SIR information. The former has been discussed in a separate paper,[15] while the latter is described in the present paper.

2. Automation of dual-space iteration

A dual-space phasing method has been used by Wang in the "solvent flattening" process[16,17] with great success in improving the quality of electron-density maps of proteins. The direct-method dual-space phasing procedure by Hauptman and colleagues[18] enhanced the power of *ab initio* phasing of diffraction data at atomic resolution by an order of magnitude. The preliminary test of iterative dual-space SAD phasing of protein data has been re-

*Project supported by the Innovation Foundation of the Chinese Academy of Sciences, and the National Basic Research Program of China (Grant No. 2002CB713801).
†Corresponding author. E-mail: gu@cryst.iphy.ac.cn
‡Corresponding author. E-mail: fanhf@cryst.iphy.ac.cn

ported by Gu *et al.*[19] The iterative dual-space SAD phasing and fragment extension has dramatically enhanced the SAD/SIR phasing power for protein data far below the atomic resolution.[3,11,20] Applications of the dual-space fragment extension have been extended to protein data without SAD/SIR signals.[14] This enhances significantly the efficiency of MR-model completion in the case where SAD/SIR signals are unavailable. While the solvent flattening procedure and the dual-space *ab initio* direct-method phasing procedure have been well automated, no efforts have been made for the program OASIS to automate the dual-space iteration before the version OASIS4.0.[21]

The dual-space iteration in OASIS consists of the following steps:

(i) phasing by OASIS4.0[21] in one of the following cases:

(a) SAD/SIR data with known heavy-atom substructure;

(b) SAD/SIR data with known heavy-atom substructure and a partial (maybe fragmental) protein structure model;

(c) diffraction data without SAD/SIR signals, but with a partial (maybe fragmental) protein structure model;

(ii) the density modification by DM[4−6] or RESOVE[7] to improve the electron density map obtained from (i) and output a set of improved phases and figures of merit. This step is optional and can be skipped by the user.

(iii) model building and structure refinement based on (ii) or (i) by RESOLVE,[8,9] ARP/wARP[10] and REFMAC[22] or by the program AutoBuild in PHENIX;[23]

(iv) ending process or going back to (i).

Iterative jobs for the automatic dual-space phasing and fragment extension are organized and monitored via the graphical user's interface (GUI) of OASIS4.0, which will be described in detail below.

3. GUI

The GUI of OASIS4.0 consists of the control panel and the monitoring board. The control panel is invoked via the CCP4 GUI (CCP4i, as shown in Fig. 1). Clicking the item "Oasis—direct methods phasing" in the CCP4 GUI will bring up a blank OASIS control panel (Fig. 2), which is used for setting up conditions/parameters for either a single run or an iterative job. This will be described in detail later. Starting an OASIS iterative process will bring up the monitoring board (Fig. 3), which is a graphical monitor for OASIS jobs. If for some reasons the board does not appear, or when the job had completed and the board is closed, the user can recall the board at: (work directory) /auto_oasis_data/0_html/index.htm.

On the monitoring board, there is a Gnuplot graph (Web site at http://www.gnuplot.info/) displaying the progress of the iteration. Besides, the user can click "details" at bottom left to obtain a detailed list for all finished cycles (see Fig. 4) or click "Sequence file" to check the file contents. By clicking the PyMOL logo on top right, the current-available "best model" (stored in a PDB file) will be opened and can be manipulated by PyMOL (Web site at http://www.pymol.org/) (Fig. 5).

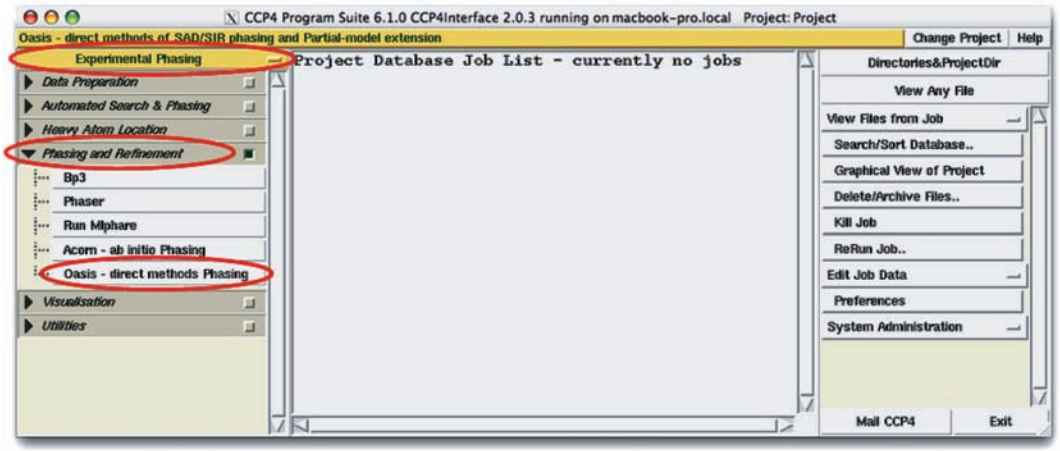

Fig. 1. CCP4 GUI.

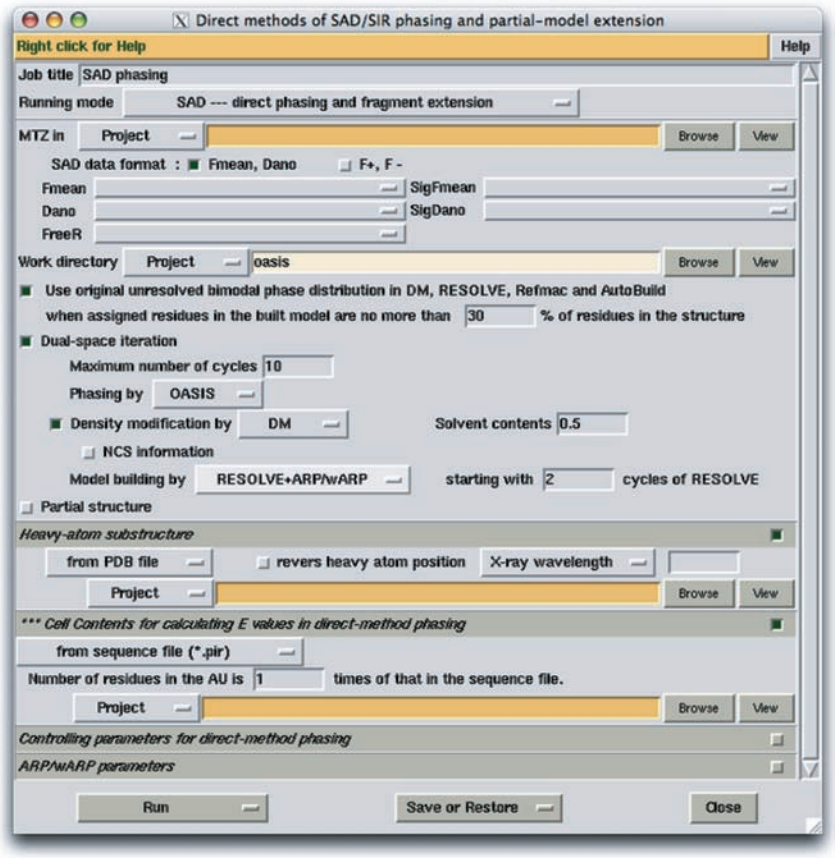

Fig. 2. Control panel of the OASIS4.0 GUI.

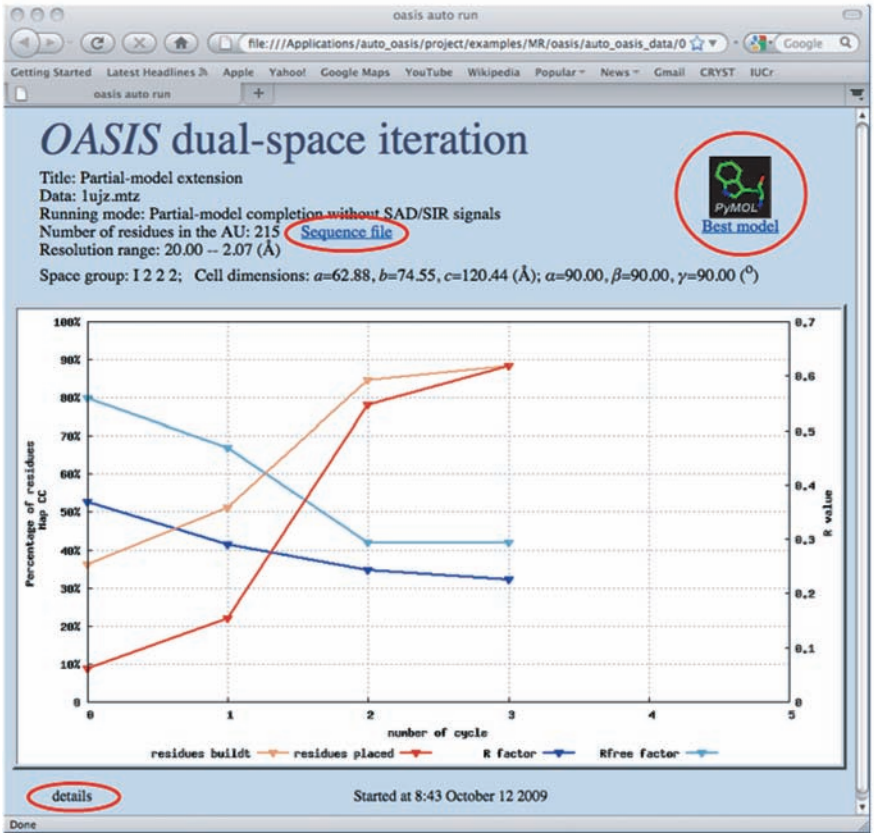

Fig. 3. Monitoring board of the OASIS4.0 GUI.

Fig. 4. Detailed list of results for finished cycles.

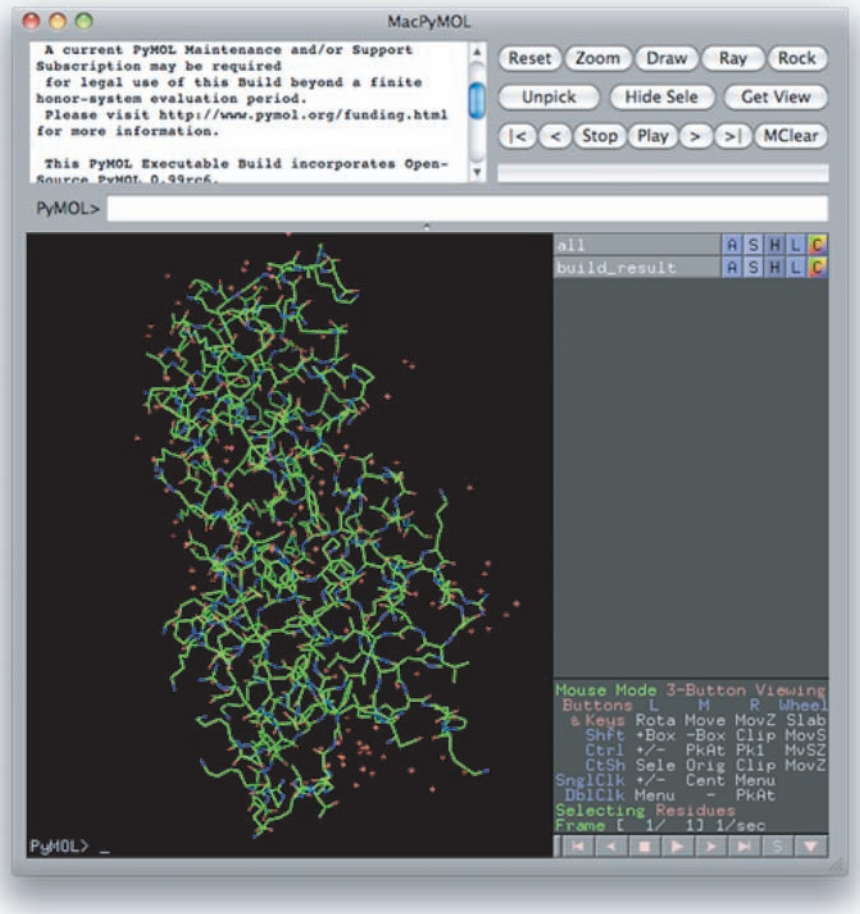

Fig. 5. Real-time PyMOL display of the best structure model.

3.1. Running mode and Job title; message windows – on-line help

The first thing to do on the control panel is to select the "Running mode". Click the button next to "Running mode" and select from the pulldown menu what you would like the job to do. Three choices are available (see Fig. 6). Suppose that you take "SAD—direct phasing and fragment extension", then the default

"Job title" will be "SAD phasing". You can change the contents of "Job title" by putting any character string into the adjacent slot. Two items are seen at the bottom of Fig. 6. They are related to some features of OASIS4.0 and are by default selected. Right click the first item "Use original unresolved $\cdots$" a message window will appear giving explanations for that item (Fig. 7). From the item "Use original unresolved $\cdots$" downward, the user can right click an item (including buttons, slots and widgets) on the control panel and obtain some explanations (on-line help) for that item. Such messages make up an important part of the tutorials of OASIS4.0.

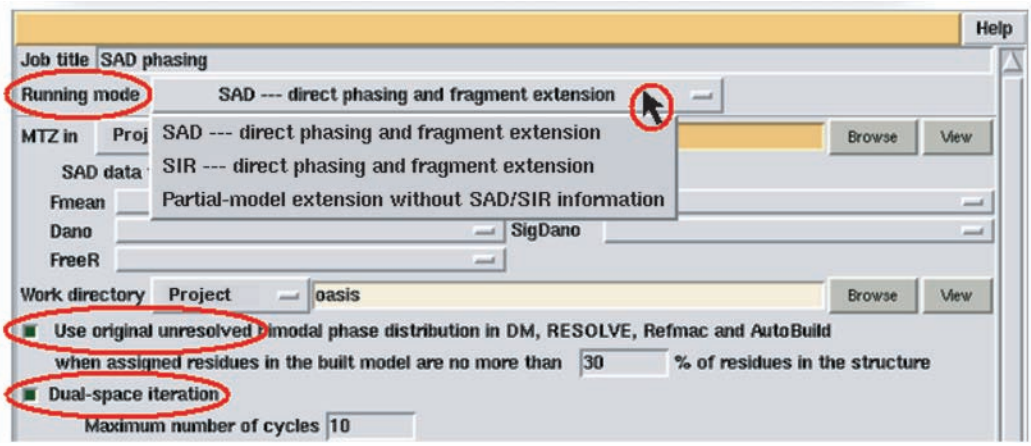

Fig. 6. Running mode selection.

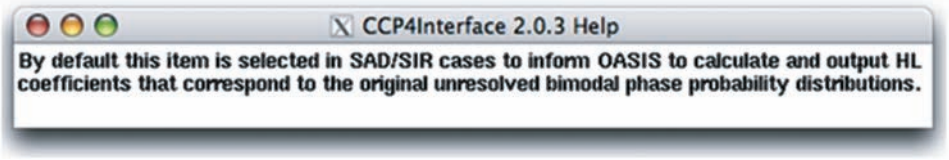

Fig. 7. On-line help—message widow.

3.2. Input data and output files

After choosing the "Running mode", the user should specify the location of the diffraction data file (in MTZ format) in the slot next to the label "MTZ in" (see Fig. 8). The user should also specify a "Work directory", which by default will be a subdirectory named "oasis" under user's default project directory. The "Work directory" will keep all intermediate files for the whole iterative process.

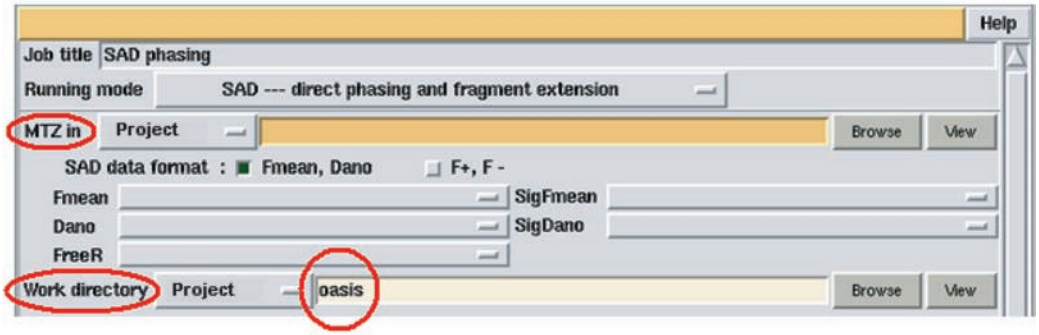

Fig. 8. Input and output specification (a).

If, instead of an iterative process, you wish to make a single run of OASIS, then you should unselect the item "Dual-space iteration" (see Fig. 9). In this case you should specify the location of the output file in the

slot next to the label "MTZ out". Notice that in this case the control panel will not provide the function of automatic model building.

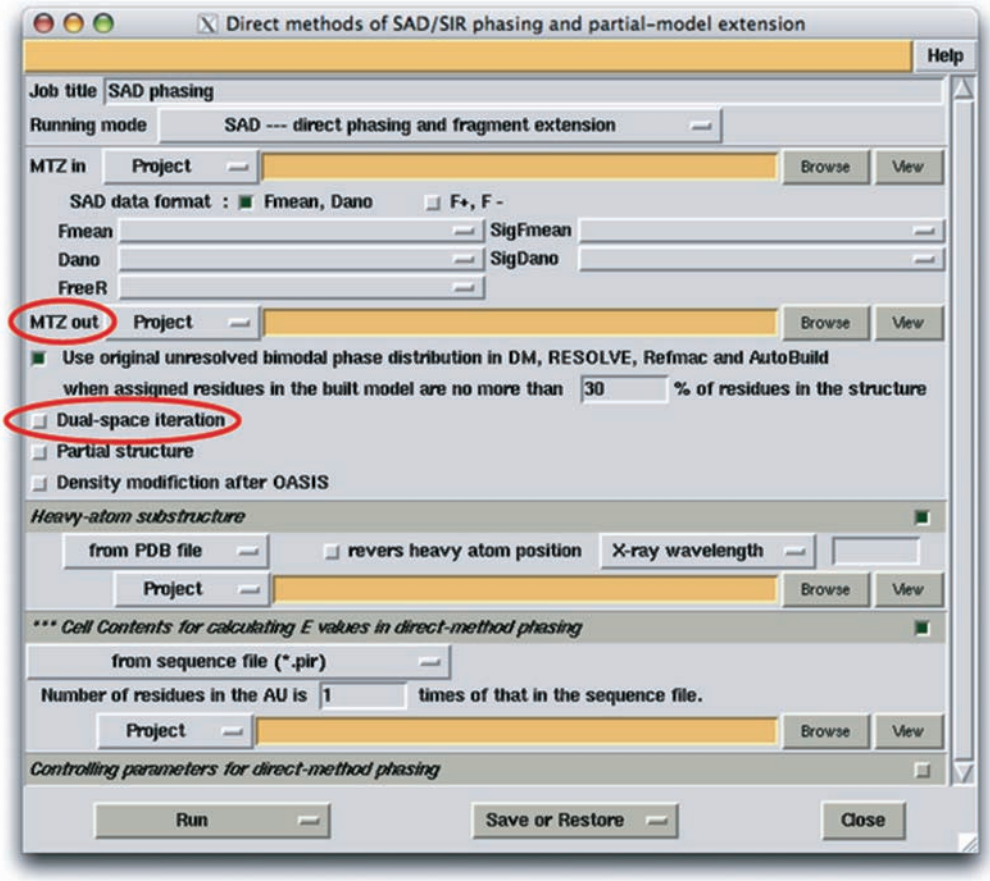

Fig. 9. Input and output specification (b).

3.3. Partial structure

Partial-structure information is compulsory in "Partial-model extension" mode, but it is optional in SAD and SIR modes. Partial-structure information should be provided with a file in either PDB format or OASIS format. The user would right click the button "from PDB file" and the following slot (see Fig. 10) for details.

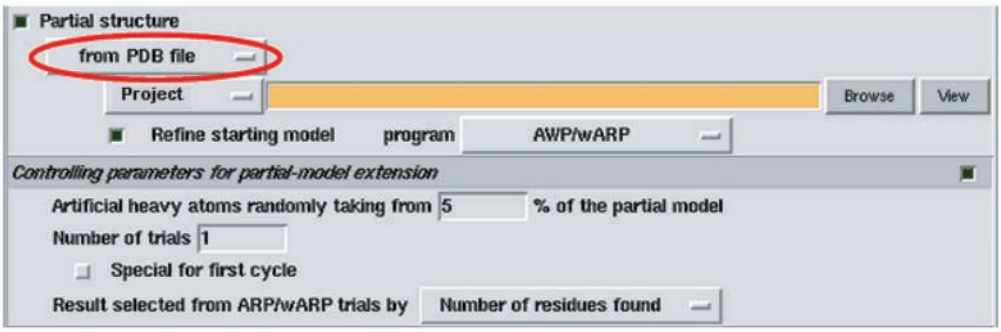

Fig. 10. Partial structure input.

Default controlling parameters for partial-model extension are set on the control panel as shown on the lower-half part of Fig. 10. In most cases, there is no need to change the default settings. However if you want to change something, before you do it, please right click the corresponding item to learn more.

3.4. Heavy-atom substructure

Information of the heavy-atom substructure is compulsory for SAD/SIR phasing. This can be provided by specifying the location of a file in the pink slot (see Fig. 11), or by entering the necessary atomic parameters on the control panel (see the lower-half part of Fig. 11). The item "x-ray wavelength" is only necessary for SAD phasing, which is used to calculate the imaginary-part correction (f″) of atomic scattering factors. Click the button "x-ray wavelength", you may alternatively select "input f″" instead of "x-ray wavelength".

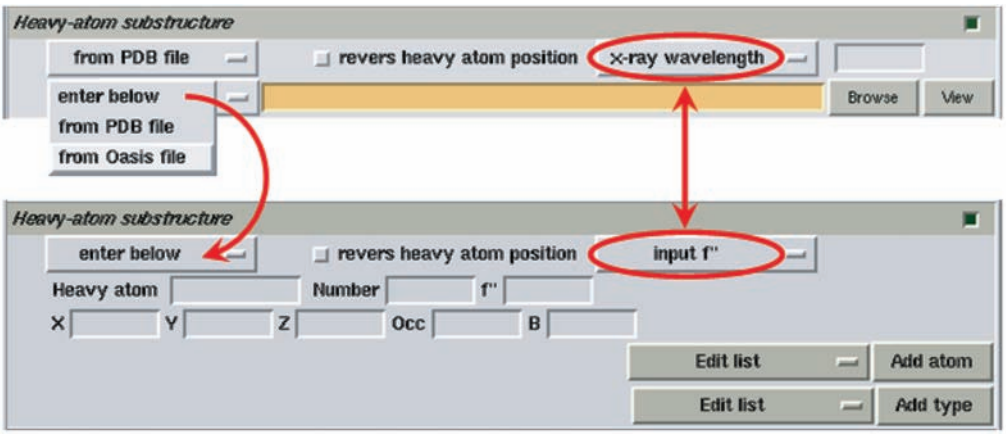

Fig. 11. Heavy-atom substructure, and x-ray wavelength or f″ input.

In the "partial-model extension" mode, the item "Heavy-atom substructure" will not appear on the control panel. Instead, there will be a widget "Heavy atom" next to the "Running mode" button as shown in Fig. 12. If (not compulsory) the user wants to input the heavy-atom substructure, please select the widget to bring up the item "Heavy-atom substructure".

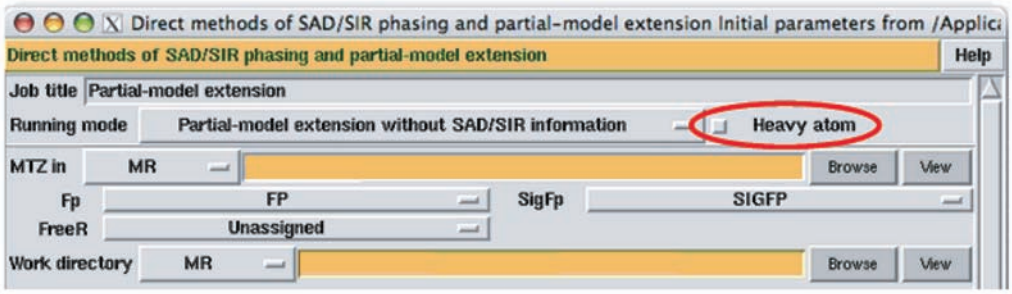

Fig. 12. Heavy atom option in partial-model extension mode.

3.5. Unit-cell contents

Unit-cell contents are necessary for calculating normalized structure amplitudes $E(h,k,l)$ in direct-method phasing. For this purpose the user should specify the sequence file in the pink slot in Fig. 13. If the sequence file does not have the suffix "pir", then please click the button "from sequence file (∗.pir)" and take the option "from other sequence file (∗.∗)". If your protein consists of n identical molecules, the sequence file should only contain residues of one molecule and, the number "1" in the widget above the pink slot should be replaced by the number "n".

In case you do not have the sequence file, the unit-cell contents can also be calculated from the "number of residues in asymmetric unit". However, this option is only available for a single run, i.e., you should turn off (unselect) the item "Dual-space iteration" on the control panel.

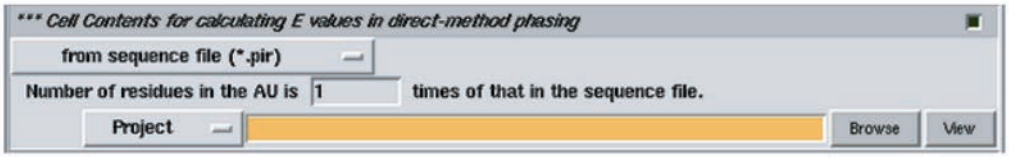

Fig. 13. Unit-cell contents input.

3.6. Default settings of direct methods

The default controlling parameters for direct-method phasing are provided near the bottom of the control panel as shown in Fig. 14. Values defining the resolution range are detected automatically from the input data. In most cases, there is no need to change any of the default settings. For details of each setting, please right click the corresponding item on the control panel.

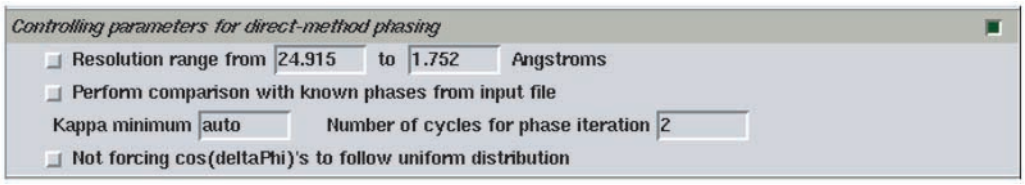

Fig. 14. Default settings of direct methods.

3.7. Limited control on ARP/wARP

Limited items for controlling ARP/wARP are available on the control panel (see Fig. 15). The user is referred to ARP/wARP documents for more details. In our experience, when the model-building result is unsatisfactory, unselecting the item "Use dipeptide distribution $\cdots$" may give chance to obtain a better result. The item "Include dummy atoms $\cdots$" is a special function of OASIS dual-space iteration. By default it is selected in SAD/SIR phasing, but unselected in Partial-model extension. Right click this item to have more explanations.

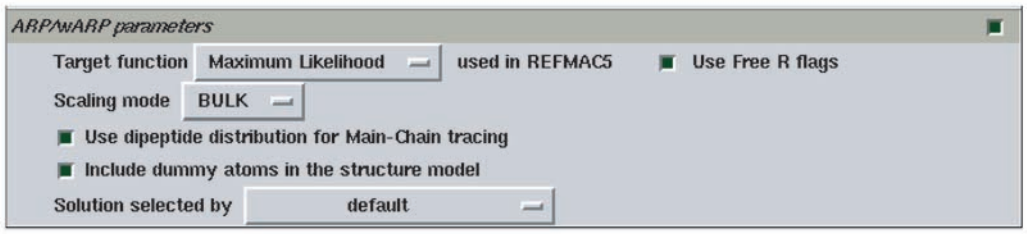

Fig. 15. Limited control of ARP/wARP.

3.8. Limited control on PHENIX

Limited items for controlling PHENIX.AutoBuild are available on the control panel (see Fig. 16). The user is referred to PHENIX documents for more details. The item "Secondary structure only" is important in model-building on electron-density maps at low resolution. By default it is selected for model building at lower than 3 Å (1 Å=0.1 nm) resolution.

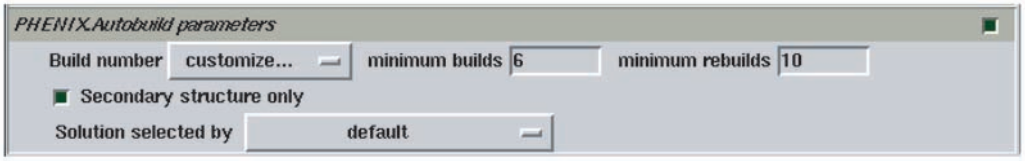

Fig. 16. Limited control of PHENIX.

4. Installation

OASIS4.0 is freely available for academic users from the Web site http://cryst.iphy.ac.cn. Both the Linux version and the MAC OS X version are available. Please read the "Readme file" first after unpacked the downloaded package. Then by executing the script install.csh all the necessary OASIS4.0 files will be put into the existent CCP4 (available at http://www.ccp4.ac.uk/) directories in the local computer. In order to use the OASIS4.0 GUI, the user should have also ARP/wARP (available from http://www.embl-hamburg.de/ARP/) and PHENIX (available from http://www.phenix-online.org/) preinstalled in the local computer. People, who used OASIS4.0 in their work, please cite the present paper in the relevant publications.

References

[1] Hao Q, Gu Y X, Zheng C D and Fan H F 2000 *J. Appl. Cryst.* **33** 980

[2] Wang J W, Chen J R, Gu Y X, Zheng C D, Jiang F and Fan H F 2004 *Acta Cryst.* D **60** 1987

[3] Wang J W, Chen J R, Gu Y X, Zheng C D and Fan H F 2004 *Acta Cryst.* **60** 1991

[4] Cowtan K D and Main P 1993 *Acta Cryst.* D **49** 148

[5] Cowtan K 1994 *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31** 34

[6] Collaborative Computational Project No. 4 1994 *Acta Cryst.* D **50** 760

[7] Terwilliger T C 2000 *Acta Cryst.* D **56** 965

[8] Terwilliger T C 2003 *Acta Cryst.* D **59** 38

[9] Terwilliger T C 2003 *Acta Cryst.* D **59** 45

[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] He Y, Gu Y X, Lin Z J, Zheng C D and Fan H F 2007 *Chin. Phys.* **16** 3022

[13] Zhang T, He Y, Gu Y X, Zheng C D, Hao Q, Wang J W and Fan H F 2007 *OASIS-2006: a direct-methods program for SAD/SIR phasing and reciprocal-space fragment extension* Institute of Physics, Chinese Academy of Sciences, China (available at http://cryst.iphy.ac.cn)

[14] He Y, Yao D Q, Gu Y X, Lin Z J, Zheng C D and Fan H F 2007 *Acta Cryst.* D **63** 793

[15] Wu L J, Zhang T, Gu Y X, Zheng C D and Fan H F 2009 *Acta Cryst.* D **65** 1213

[16] Wang B C 1981 *Acta Cryst.* A **37** Suppl. C-11

[17] Wang B C 1985 *Methods in Enzymology* **115** 90

[18] DeTitta G T, Weeks C M, Thuman P, Miller R and Hauptman H A 1994 *Acta Cryst.* A **50** 203

[19] Gu Y X, Zheng C D, Zhao Y D, Ke H M and Fan H F 1997 *Acta Cryst.* D **53** 792

[20] Yao D Q, Li H, Chen Q, Gu Y X, Zheng C D, Lin Z J and Fan H F 2008 *Chin. Phys.* B **17** 1

[21] Zhang T, Wu L J, He Y, Wang J W, Zheng C D, Hao Q, Gu Y X and Fan H F 2009 *OASIS4.0: a direct-methods program for SAD/SIR phasing and reciprocal-space fragment extension with a GUI for controlling and monitoring iterative processes* Institute of Physics, Chinese Academy of Sciences, China (available at http://cryst.iphy.ac.cn)

[22] Vagin A A, Steiner R S, Lebedev A A, Potterton L, McNicholas S, Long F and Murshudov G N 2004 *Acta Cryst.* D **60** 2284

[23] Adams P D, Afonine P V, Bunkóczi G, Chen V B, Davis I W, Echols N, Headd J J, Hung L W, Kapral G J, Grosse-Kunstleve R W, McCoy A J, Moriarty N W, Oeffner R, Read R J, Richardson D C, Richardson J S, Terwilliger T C and Zwart P H 2010 *Acta Cryst.* D **66** 213