ISSN 1674-1056
http://cpb.iphy.ac.cn
http://iopscience.iop.org/cpb
中国物理 B
Chinese Physics
中国物理 Chinese Physics B

Development of “Parameter space screening”-based single-wavelength anomalous diffraction phasing and structure determination pipeline

Wei Ding(丁玮), Xiao-Ting Wang(王小婷), Yang-Yang Yi(易阳旸)
Citation: Chin. Phys. B . 2019, 28(11): 116101 . **doi:** 10.1088/1674-1056/ab43bd
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Effects of helium irradiation dose and temperature on the damage evolution of Ti_3SiC_2 ceramic

Hua-Hai Shen(申华海), Xia Xiang(向霞), Hai-Bin Zhang(张海斌), Xiao-Song Zhou(周晓松), Hong-Xiang Deng(邓洪祥), Xiao-Tao Zu(祖小涛)
Chin. Phys. B . 2019, 28(7): 076104 . **doi:** 10.1088/1674-1056/28/7/076104

Enhancement of MAD/MIR phasing at low resolution and a new procedure for automatic phase extension

Pu Han(韩普), Yuan-Xin Gu(古元新), Wei Ding(丁玮), Hai-Fu Fan(范海福)
Chin. Phys. B . 2019, 28(7): 076108 . **doi:** 10.1088/1674-1056/28/7/076108

Microstructure and lateral conductivity control of hydrogenated nanocrystalline silicon oxide and its application in a-Si:H/a-SiGe:H tandem solar cells

Tian-Tian Li(李天天), Tie Yang(杨铁), Jia Fang(方家), De-Kun Zhang(张德坤), Jian Sun(孙建), Chang-Chun Wei(魏长春), Sheng-Zhi Xu(许盛之), Guang-Cai Wang(王广才), Cai-Chi Liu(刘彩池), Ying Zhao(赵颖), Xiao-Dan Zhang(张晓丹)
Chin. Phys. B . 2016, 25(4): 046101 . **doi:** 10.1088/1674-1056/25/4/046101

FT-Raman spectroscopy study of solvent-in-salt electrolytes

Liumin Suo(索鎏敏), Zheng Fang(方铮), Yong-Sheng Hu(胡勇胜), Liquan Chen(陈立泉)
Chin. Phys. B . 2016, 25(1): 016101 . **doi:** 10.1088/1674-1056/25/1/016101

Complementary method to locate atomic coordinates by combined searching method of structure-sensitive indexes based on bond valence method

Song Zhen, Liu Xiao-Lang, He Li-Zhu, Xia Zhi-Guo, Liu Quan-Lin
Chin. Phys. B . 2015, 24(10): 106101 . **doi:** 10.1088/1674-1056/24/10/106101

Development of "Parameter space screening"-based single-wavelength anomalous diffraction phasing and structure determination pipeline

Wei Ding(丁玮)[1,2,†,‡], Xiao-Ting Wang(王小婷)[2,†], and Yang-Yang Yi(易阳旸)[3]

[1] *Key Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing* 100190, *China*
[2] *School of Environment, Tsinghua University, Beijing* 100084, *China*
[3] *School of Biomedical Sciences, University of Hong Kong,* 21 *Sassoon Road, Hong Kong, China*

(Received 12 August 2019; revised manuscript received 3 September 2019; published online 16 October 2019)

In this paper, we present a highly efficient structure determination pipeline software suite (X^2DF) that is based on the "Parameter space screening" method, by combining the popular crystallographic structure determination programs and high-performance parallel computing. The phasing method employed in X^2DF is based on the single-wavelength anomalous diffraction (SAD) theory. In the X^2DF, the choice of crystallographic software, the input parameters to this software and the results display layout, are all parameters which users can select and screen automatically. Users may submit multiple structure determination jobs each time, and each job uses a slightly different set of input parameters or programs. Upon completion, the results of the calculation performed can be displayed, harvested, and analyzed by using the graphical user interface (GUI) of the system. We have applied the X^2DF successfully to many cases including the cases that the structure solutions fail to be yielded by using manual approaches.

Keywords: single-wavelength anomalous diffraction, parameter space screening method, structure determination pipeline

PACS: 61.05.–a

DOI: 10.1088/1674-1056/ab43bd

1. Introduction

It has been reported that single-wavelength anomalous diffraction (SAD) is an indispensable technique in x-ray crystallography for the structure determination of biological macromolecules. Heavy-atom sites searching and phasing, density modification, and model building are the main steps in SAD structure solution. Many programs are available for automatic structure determinations, for instance, Phenix.autosol[1] and SHELXC/D/E[2,3] are the most commonly used programs for heavy-atom sites searching and phasing based on automatic interpretation of Patterson maps or by direct method (DM);[4] Parrot[5] is widely used for modifying the initial electron density maps using solvent flattening,[6] histogram matching, and non-crystallographic symmetry averaging methods; ARP/wARP,[7] Buccaneer,[8] and Phenix.autobuild[9,10] are highly automated tools for building the iterative model based on the density-weighted score function, statistical chain tracing method, automatic templates matching algorithm, *etc*. Further, CRANK,[11] Auto-Rickshaw,[12,13] and IPCAS[14–17] are highly automated and widely used pipelines for the SAD method. All these methods have been widely employed with great success, however, for some difficult cases, it may be necessary to fine-tune the settings or adjustable parameters and run more trials. Moreover, in some special cases, only a specific combination of parameters can realize an accurate structure solution, and there exists even no significant correlation between the combination of parameters and the final result. Therefore, relying solely on experience and trial and error analysis is time-consuming and has no guarantee to obtain ideal results.

To solve this problem, we develop a highly efficient structure determination pipeline software suit, X^2DF. The X^2DF is an automated SAD phasing, density modification, and model building pipeline. The X^2DF is constructed based on the "Parameter space screening" method proposed by Liu *et al.*[18] The X^2DF can be used to screen dozens or hundreds of different combinations of input parameters and programs such as high-resolution limits for the heavy-atom sites searching and phasing, the number of heavy-atom sites to be searched, the space groups, *etc*. The X^2DF then spawns the multiple jobs in parallel on a Linux cluster by using various combinations of programs and input-parameter values. The X^2DF is successfully applied to many cases including some special and difficult cases which will be presented in Section 3 and Section 4.

2. Method

2.1. GUI of X^2DF

The X^2DF has a stand-alone interface written in Perl/tk. The X^2DF GUI window (Fig. 1) can be opened by running the "x2dfi.pl" in the installation folder in the command line.

†These authors contributed equally to this work.
‡Corresponding author. E-mail: dingwei@iphy.ac.cn

© 2019 Chinese Physical Society and IOP Publishing Ltd

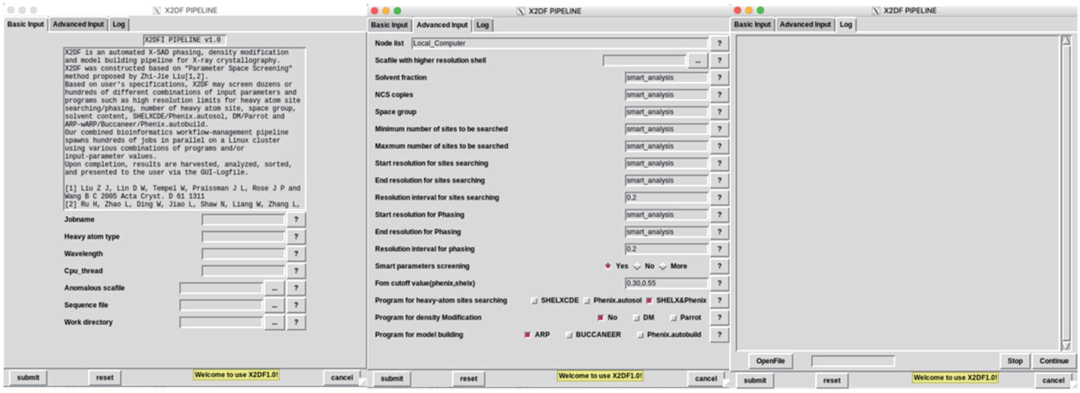


Fig. 1. X^2DF control panel. Control panel contains three panels, *i.e.* "Basic Input Panel", "Advanced Input Panel", and the "Log Panel" from left to right. Required parameters such as heavy-atom type, wavelength, CPU_thread, anomalous diffraction data, sequence file, and work directory are listed in "Basic Input Panel." Optional parameters are listed in "Advance Input Panel", and users can use default values of these parameters or adjust some of them manually. "Log Panel" is used for displaying progress and result of running jobs.

The X^2DF GUI contains three Panel bars, *i.e.*, "Basic Input Panel", "Advanced Input Panel", and "Log Panel". The "Basic Input Panel" can be used to fill in the required information and upload the necessary data, including "Heavy-atom type", "Wavelength", "CPU_thread", "Anomalous diffraction data", "Sequence file", and "Work directory".

Other parameters and experimental data can be input or uploaded by using an "Advanced Input Panel". In this panel, "Node list" item can be used for assigning the node name of the high-performance computing cluster. "Scafile with higher resolution shell" item can be used for uploading the high-resolution diffraction data. The "Minimum number of sites to be searched" and "Maximum number of sites to be searched" items are used to specify the search range of heavy-atom sites number. The "Start resolution for sites searching", "End resolution for sites searching", and "Resolution interval for sites searching" items are used for specifying the resolution screening range and interval for the heavy-atom sites searching. Similarly, The "Start resolution for phasing", "End resolution for phasing", and "Resolution interval for phasing" can be used to set the resolution screening range and interval for initial phases calculation. The "Program for heavy-atom sites searching" term, "Program for density modification" term, and "Program for model building" item list the candidate programs to be used in structure determination. The "figure-of-merit (FOM) cutoff value" is a constrained parameter used to limit the number of jobs. If the FOM value of the initial phases is less than the "FOM cutoff value," this job will be killed and deleted. The user may change any default control parameter in this panel, but in most cases using the default parameters will produce the desired results.

The "Log Panel" allows the user to monitor the progress of a running job and to analyze the final results.

2.2. Workflow of X^2DF

The workflow chart of the X^2DF (shown in Fig. 2) is as follows.

Step 1 Preparation of the x-ray data

For SAD phasing, the SAD data need inputting in reflection intensity SCA format. Other required parameters are the sequence, wavelength, and heavy-atom type. The optional parameters described in Subsection 2.1 are optimized with extensive test dataset or analyzed by using some crystallography tools. For instance, space group, solvent context, and site number are automatically analyzed by using Phenix.autosol; screening resolution for heavy-atom sites searching and phasing starts from the highest resolution shell of the anomalous diffraction data and ends in a range of 4.00 Å, with a resolution screening interval of 0.20 Å; and the default programs for heavy-atom sites searching/phasing, density modification, and model building are SHELXC/D/E + DM + ARP/wARP.

Step 2 Parameter space screening and job submission

When the users complete data entry and click the "submit" button, the number of jobs will be calculated by using a parameter space screening method;

$$\mathrm{Job}_{\mathrm{Number}} = HR * PR * SP * HA * SC * HP * DM * MB, \quad (1)$$

where HR represents the number of resolutions used for heavy-atom sites searching, PR the number of resolutions used for phasing, SP the number of candidate space groups, HA the number of possible heavy-atom sites, SC the quantity of possible solvent content, HP the number of programs used for heavy-atom sites searching and phasing, DM the number of programs used for density modification, MB the number of programs used for model building.

Then, the X^2DF can be used to create a series of independent jobs with the combination of unified parameter and program and to submit these jobs to the Linux cluster for further processing involved in structure determination.

Step 3 Substructure determination

Substructure information is essential for SAD phasing. The SHELXC/D or/and Phenix.autosol can be used for heavy-atom sites searching with different resolutions and the number of the expected sites, which have been screened as discussed in Step 2.

Step 4 Initial phases calculation

Once the substructure is determined, the initial phases of the anomalous data can be calculated by using SHELXE or/and Phenix.autosol, and a set of the initial phases with FOM is written in a new MTZ file.

Step 5 Job-analysis

If the FOM value of the initial phases is less than the "FOM cutoff value" (input using the "Advanced Input Panel" and the default value for SHELXE is 0.55, for Phenix.autosol, it is 0.35), the jobs are terminated, and the outputted files are deleted.

Step 6 Density modification

After Job-analysis, the remaining electron density map calculated in Step 4 can be further refined by DM and/or Parrot. New MTZ files with a set of improved phases will be created. This step is optional and can be skipped by selecting "No" in "Density modification" item in the "Advanced Input Panel."

Step 7 Model building

Based on the selected programs in "Model building" item, the model building step is carried out by using ARP/wARP, Buccaneer or/and Phenix.autobuild. Further, if the user provides high-resolution diffraction data in Step 1, its amplitude is used for building high-resolution model in this step.

Step 8 Outcome

After model building, the final coordinate and MTZ files are stored in the "Result" folder, and the structure information of the outputted models ($R_{\text{free}}/R_{\text{work}}$/Residues Built/Residues Placed) is written in a log file, which can be presented in the "Log Panel" in real-time.

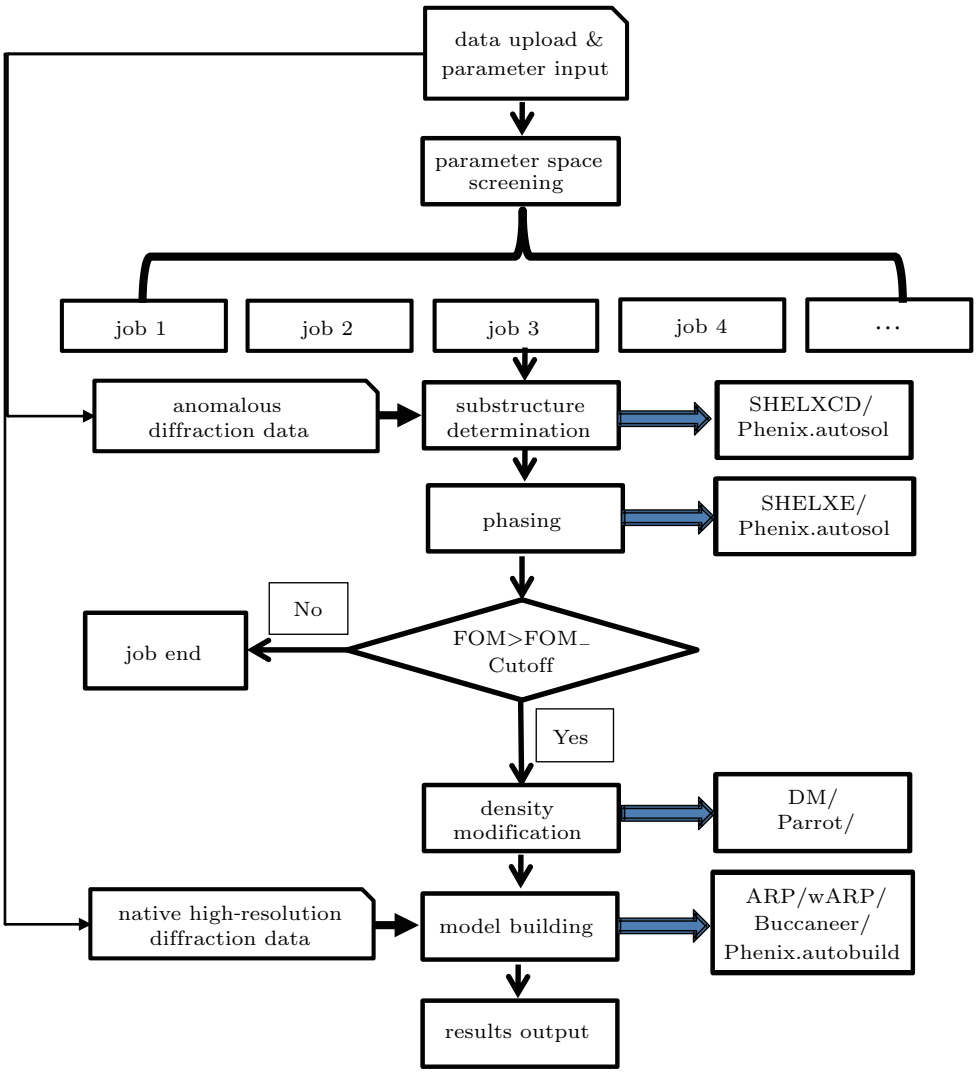


Fig. 2. Workflow chart of X^2DF. Snip single corner single rectangle denotes input data, rectangle represents process steps, pipeline refers to steps from data preparation through output results and can be run without user intervention.

2.3. Installation and runtime environment

The X^{2}DF is freely available to academic users and has been tested on Linux (Centos, Fedora, and Ubuntu) platforms. Users can download the latest version from the website: http://cryst.iphy.ac.cn/Download/X2DF/ and add the script folder to the environment variables of the local system. Further, CCP4,[19,20] PHENIX, ARP/wARP, Buccaneer, and SHELXC/D/E should be pre-installed and added to the environment variables too so that some subroutines of CCP4 and Phenix can be called by the X^{2}DF. The X^{2}DF is written in PERL/Tk language so that it can be run directly by using the "x2dfi.pl" script without needing to compile the program first. For more details, interested readers can refer to the "Readme.txt" in X^{2}DF package.

All calculations presented in this paper were performed on a T7910 (DELL) with a 3.40 GHz, 24 processors Intel Xeon E5-2643 v4 CPU, and 64 GB RAM. The versions of the supported programs are CCP4-7.0.075, PHENIX-1.15.2-3472, ARP/wARP-8.0, and cbuccaneer version 1.6.5, SHELXC/D/E-version 2016/1.

3. Results and discussion

The general applicability of the X^{2}DF is tested with many unknown structures by using various protocols; however, only four typical and intractable cases are presented in this paper. The resolutions of the test cases range from 2.70 Å to 3.21 Å. The quantity of residues ranges from 164 to 952. The types of heavy-atoms used are S and Se. The mean anomalous difference ranges from 0.0349 to 0.0716. Detailed information about the x-ray diffraction data is summarized in Table 1.

Table 1. Diffraction data used in the case studies. Four typical cases are selected. T1 denotes a typical low-resolution SeMet-SAD case used in revealing effects of resolution limit on final result; T2 denotes a typical case of low redundancy of SeMet-SAD and low crystal symmetry, which presents the influence of the combination of different programs; and T3 denotes a typical case of long-wavelength Sulphur SAD, showing the crucial role of screening heavy-atom sites. T4 denotes a typical case of long-wavelength native Sulphur SAD with high-resolution data.

Cases	T1	T2	T3	T4	
Crystal	STING	LEAV	DR6	DR6	DR6
PDB ID code	4EF5	4J1G	3U3S	3U3T[a)]	3U3P[b)]
Wavelength/Å	0.98	0.98	2.00	2.70	0.98
Space group	$C222_1$	P1	$P6_122$	$P6_122$	$P6_122$
a, b, c/Å	79.34	50.41	77.68	77.47	77.86
	89.34	74.62	77.68	77.47	77.86
	74.72	80.94	186.81	185.81	186.56
α, β, γ/(°)	90.00	69.46	90.00	90.00	90.00
	90.00	89.75	90.00	90.00	90.00
	90.00	104.12	120.00	120.00	120.00
Resolution range/Å	50.00–3.10	50.00–3.08	50.00–2.70	50.00–3.21	50.00–2.09
	(3.21–3.10)	(3.19–3.08)	(2.82–2.70)	(3.31–3.21)	(2.16–2.09)
Redundancy	9.8(10.2)	3.8(3.8)	39.3(40.3)	19.9(19.5)	20.4(20.3)
Heavy-atom type	Se	Se	S	S	S
Heavy-atom sites/ASU[c)]	5	24	20	20	20
No. molecules/ASU	1	4	1	1	1
No. of residues/ASU	265	952	164	164	164

[a)] Anomalous diffraction data for test case T4, [b)] High-resolution diffraction data for test case T4, [c)] ASU = asymmetry unit.

Table 2. Input parameters and solutions of test cases. R1–R3 corresponding to the test cases T1–T3 in Table 1. R4 denotes solution of the test case T4 with anomalous diffraction data only; R5 denotes other solution of T4 by using both anomalous and high-resolution diffraction data. In R1–R3, multiple jobs are created by using screening resolution, screening program, and screening heavy-atom sites, separately. However, for each case, only one "Good Result" is obtained. In R4, no "Good Result" is found in 1152 jobs; but in R5, there are two "Good Results" out of 64 jobs.

	R1	R2	R3	R4	R5
Crystal	STING	LEAV	DR6	DR6	DR6
PDB ID code	4EF5	4J1G	3U3S	3U3T	3U3T & 3U3P
Resolution screening range for heavy-atom sites searching/Å	3.10–3.50	3.07	2.70	3.21–4.00	3.21–4.00
Resolution screening range for phasing/Å	3.10–3.50	3.07	2.70	3.21–4.00	3.21–4.00
The screening range of heavy-atom sites	1	24	6–21	9	9
Resolution screening interval for heavy-atom sites searching/Å	0.05	–	–	0.1	0.1
Resolution screening range for phasing/Å	0.05	–	–	0.1	0.1
Programs[a)]	Pas + D + B	S/Pas + D/P/N + Pab/A/B	S + A	S/Pas + D/P/N + Pab/A/B	S + A
Total job number	100	18	16	1152	64
Good result number	1	1	1	0	2
Resolution for heavy-atom site searching of the best job/Å	3.40	3.07	2.70	–	3.81
Resolution for phasing of the best job/Å	3.20	3.07	2.70	–	3.91
The number of heavy-atom sites of the best result	1	24	9	–	9
Program combination of the best result	Pas + D + B	Pas + D + Pab	S + A	–	S + A
R_{work}/R_{free} of the best result.	0.32/0.37	0.32/0.38	0.24/0.28	–	0.25/0.29
Time/h	6	4	2	136	8

[a)] S = SHELXC/D/E, Pas = Phenix.autosol, D = DM, P = Parrort, N = N/A (skip this step), A = RP/wARP, B = Buccaneer, Pab = Phenix.autobuild.

The quality of the output models is measured by using two indicators, R_{work} and R_{free}.[21–24] If the R_{free}/R_{work} is less than 0.40/0.40, it is regarded as a "Good Result" and the lowest R_{free}/R_{work} ratio is seen as the "Best Result." The resolution of the "Best Result" is considered as the "Best Resolution." An overview of the results of the test cases is presented in Table 2. Structure comparison between the X^{2}DF model and the structure deposited in PDB is shown in Fig. 3. Further, the accuracy of the experimental phases is measured by using figure-of-merit-weighted mean phase error (FOM-wMPE)[25] which is calculated against the final phases of the deposited structure.

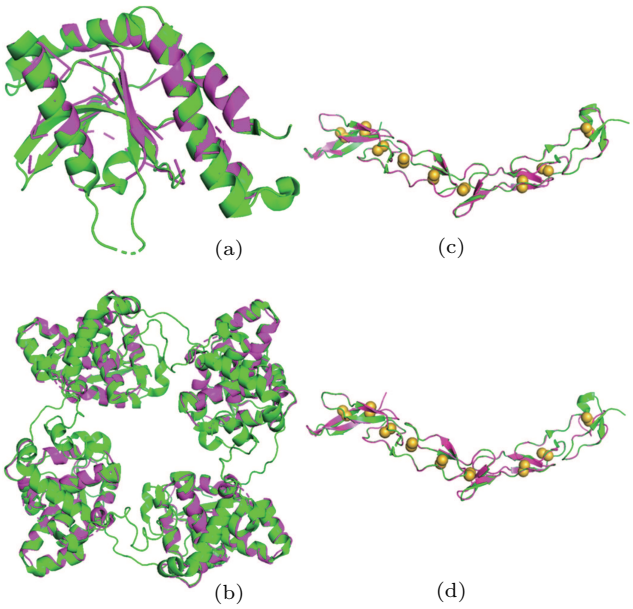


Fig. 3. Structure comparison. Models from X^{2}DF (in magenta) are superimposed on the structures deposited in PDB (in green), showing (a) X^{2}DF model of T1 VS 4EF5, (b) X^{2}DF model of T2 VS LEAV, (c) X^{2}DF model of T3 VS 3U3S, (d) X^{2}DF model of T4 VS 3U3P. In panels (c) and (d), disulphide bonds in deposited structures are shown as yellow spheres. There are nine disulfide bonds in 3U3S/3U3P, which belong to 18 cysteines.

3.1. Case study 1: resolution screening for low-resolution data

Test case 1 (T1) represents a typical case of SeMet-SAD with a resolution lower than 3.00 Å. It is the structure of STING adaptor protein with five SeMet sites and 265 amino acid residues per asymmetry unit (PDB entry 4EF5[26]). The anomalous diffraction data are collected at a wavelength of 0.98 Å and indexed, integrated, and scaled at 3.10 Å with the space group of $C222_1$ and the mean anomalous difference of 0.0716.

Normally, the "Signal-to-Noise" and "Measurability" of the anomalous diffraction data increase with resolution decreasing,[27] while too low resolution will reduce the phase accuracy. Thus, in heavy-atom sites searching and phasing steps, a compromise selection of resolution limit is necessary. So, in this case, the resolution screening range for heavy-atom sites searching and phasing is set to be from 3.10 Å to 3.50 Å with an increasing interval of 0.05 Å. The 100 jobs are created by using the X^{2}DF, but yield only one "Good Result." The "Best resolution" is 3.40 Å/3.20 Å, not the highest resolution of the diffraction data (as shown in Table 2 (R1) and Fig. 4). This implies that for the extreme case the selection of the resolution limit is decisive in structure determination.

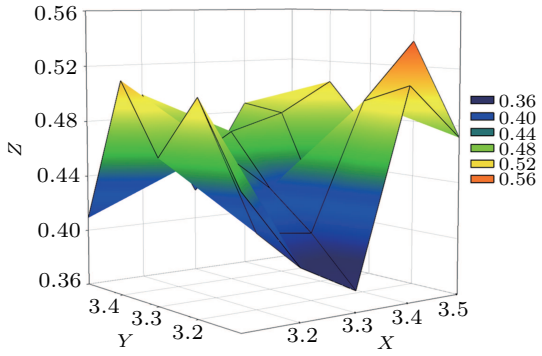


Fig. 4. Profile of parameter space screening results for T1 dataset. High-resolution limits for heavy-atom sites searching (X axis) and high-resolution limits for phasing (Y axis) are screened for solutions with lowest R_{free} (Z axis). Colors representing the respective R_{free} values are shown in a box on the right. For instance, blue color implies that R_{free} value is equal to 0.40 and that there is only one point in this color.

Similar resolution limits are recommended based on "Measurability" and "Anisotropy" analysis.[28] The "Measurability" can be defined as the fraction of Bijvoet related intensity difference, and it is a function of resolution. When its value is greater than 0.10, the anomalous signal will be strong enough for identifying the anomalous substructure sites. "Anisotropy" can be defined as the difference in resolution limit along the reciprocal space axes. In this case, when the "Measurability" is greater than 0.10, the resolution can extend to 3.43 Å. Further, there is a slight resolution anisotropy" in T1 data. The resolution limits along the reciprocal space axes (a^*, b^*, and c^*) are 3.17 Å, 3.19 Å, and 3.11 Å, respectively. The resolution limit (3.40 Å/3.20 Å) used in this case is evaluated, and can guarantee the strength of the anomalous signal for searching the substructure sites and also avoid influencing the building of the final model by crystal anisotropy.

3.2. Case study 2: program screening for low crystal symmetry data

Test case 2 (T2) represents a typical case of SeMet-SAD with low redundancy and low crystal symmetry. It is the structure of the Leanyer orthobunyavirus nucleoprotein-RNA complex, with 24 SeMet site and 952 amino acid residues per asymmetry unit (PDB entry 4J1G[29]). The anomalous diffraction data are collected at a wavelength of 0.98 Å and indexed, integrated, and scaled at 3.07 Å with the space group of $P1$, the redundancy of 3.8, and the mean anomalous difference of 0.0702. In this case, all the programs described in Subsection 2.2 (Steps 3, 4, 6, 7) are used for solving the structure.

In this case, 18 different combinations of programs are used for determining the structure, however, only one combination (Phenix.autosol + DM + Phenix.autobuild) yields "Good Result" (as shown in Table 2 (R2) and Table 3). It is not proven that which is stronger than the others because all of them are widely used for phasing, density modification, and model building programs, and have recorded several successes. Moreover, in other test cases, the best program combinations are different. Here in the present article, we show that in the main steps of the structure solution, the appropriate selection of the programs can significantly influence the final results.

Table 3. Screening program for T2 in heavy-atom sites searching/phasing, density modification, and model building steps. Eighteen jobs are created by using X^2DF with different program combinations, but only one combination yields "Good Result."

Program for heavy-atom sites searching/Phasing	Program for density modification[a)]	Program for model building	R_{work}/R_{free}
SHELXC/D/E	N/A	ARP/wARP	0.22/0.54
SHELXC/D/E	N/A	Buccaneer	0.51/0.55
SHELXC/D/E	N/A	Phenix.autobuild	0.42/0.48
SHELXC/D/E	DM	ARP/wARP	0.22/0.52
SHELXC/D/E	DM	Buccaneer	0.52/0.53
SHELXC/D/E	DM	Phenix.autobuild	0.42/0.47
SHELXC/D/E	Parrot	ARP/wARP	0.20/0.54
SHELXC/D/E	Parrot	Buccaneer	0.52/0.52
SHELXC/D/E	Parrot	Phenix.autobuild	0.42/0.47
Phenix.autosol	N/A	ARP/wARP	0.26/0.55
Phenix.autosol	N/A	Buccaneer	0.52/0.53
Phenix.autosol	N/A	Phenix.autobuild	0.41/0.48
Phenix.autosol	DM	ARP/wARP	0.52/0.61
Phenix.autosol	DM	Buccaneer	0.50/0.53
Phenix.autosol	**DM**	**Phenix.autobuild**	**0.32/0.37**
Phenix.autosol	Parrot	ARP/wARP	0.48/0.61
Phenix.autosol	Parrot	Buccaneer	0.50/0.52
Phenix.autosol	Parrot	Phenix.autobuild	0.35/0.48

[a)] N/A = skipping this step.

3.3. Case study 3: heavy-atom sites screening for Sulphur SAD data

Test case 3 (T3) represents a typical case of long-wavelength native Sulphur SAD. It is the structure of the ectodomain of Death Receptor 6 consisting of 164 residues, of which 18 are cysteine and 3 are methionines (PDB entry 3U3S[30]). The anomalous diffraction data are collected at a wavelength of 2.00 Å and then indexed, integrated, and scaled at 2.70 Å resolution with the $P6_122$ space group and the mean anomalous difference of 0.0349. In general, the number of heavy-atom sites can be analyzed by using the protein sequence and the anomalous difference Patterson map. However, the anomalous signal of the sulphur atoms is always weaker than that of the mental atoms; therefore, it is difficult to distinguish the signal peak from the background noise in the anomalous difference Patterson map. In this case, the screening range for heavy-atom sites is set to be from 6 to 21.

It can be seen from the results that the optimal number of the heavy-atom sites is 9, which is not the same as the quantity of cysteine and methionines in the sequence. However, when we analyze the deposited structure of T3, we find nine disulfide bonds in the structure as shown in Fig. 3(d). Because these disulfide bonds cannot be distinguished at the current resolution, the strong anomalous scattering signal can be similarly derived from nine heavy-atom sites. So, when the number of searched heavy-atom sites is less than 9, some accurate sites can be lost, while when it is more than 9, there will appear multiple inaccurate sites in the heavy-atom substructure. These two scenarios have a negative influence on the calculation of the initial phases as shown by the FOM–wMPE curve in Fig. 5.

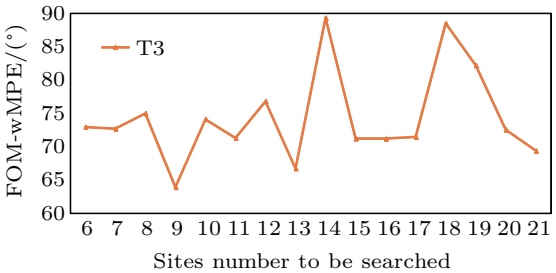


Fig. 5. Plot of figure-of-merit–weighted mean phase error (FOM–wMPE) for T3 datasets. Heavy-atom sites (horizontal axis) are screened for the best FOM–wMPE value (vertical axis). When the site number equals the number of disulphide bonds in the structure, FOM–wMPE produced the lowest value. The FOM–wMPE value is calculated by using initial phases (output by SHELXE) against the deposited structure.

In this curve, the difference in heavy-atom sites yields the drastic variation of the FOM–wMPE value, and when the number of the sites is 9, the FOM–wMPE has the lowest value, which is consistent with the previous analysis. Above all, an accurate estimation of heavy-atom sites number can play a crucial role in "Substructure determination" and the "Initial phases calculation" step.

3.4. Case study 4: high-resolution native data using for Sulphur SAD

Test case 4 (T4) represents a typical case of long-wavelength Sulphur SAD with high-resolution data. Two sets of data are used in this case. They are the same protein of T3 but from other two different crystals (PDB entry 3U3P, 3U3T[30]).

For Sulphur SAD method, long-wavelength x-ray is used to enhance the anomalous signal of the diffraction data; however, we observe that the longer the wavelength, the lower the resolution will be. Therefore, two kinds of diffraction data are used together in this case. The anomalous diffraction data of 3U3T are collected at a wavelength of 2.70 Å with 3.21 Å resolution and the mean anomalous difference of 0.0592. The native diffraction data of 3U3P are collected at a wavelength of 0.98 Å with 2.09 Å resolution.

The results of this case are shown in Table 2 (R4 and R5). In R4, more than 1000 jobs are created by the X^{2}DF through using the low-resolution anomalous data alone. The minimum FOM–wMPE value of the initial phases among these jobs is as low as 66.3°. However, there is not even one "Good Result" in the end due to the unsatisfactory resolution. But as shown in R5, when high-resolution data are provided in the model building step, two "Good Results" are obtained from 64 jobs. This case shows that the high-resolution data can be combined with the anomalous data to improve the quality of the final model in the X^{2}DF. Moreover, this case also provides the other treatment for SIRAS data processing, which involves the SAD phasing with the anomalous data and the model building by using native data.

4. Concluding remarks

As of July 2019, about 90% of the macromolecular structures deposited in the PDB (http://www.rcsb.org) have been obtained by using x-ray crystallography. The SAD method is the most commonly used method when the target molecule has no homologous structure. "Parameter space screening"-based SAD phasing and structure determination pipeline (X^{2}DF) provide a new strategy for the efficient implementation of the structure solution. First, the implementation of the parameter screening method can effectively improve the success rate of the SAD phasing and model building especially for the low resolution and weak anomalous scattering signal dataset. Second, the parallel operation mode dramatically improves the efficiency of the whole SAD structure determination processing. Some convincing examples can be found in cases T1–T4. Further, the parameter-input GUI of the X^{2}DF is designed to be intuitive and straightforward but with options for different structure determination protocols. The number of the required input parameters is minimized while the other optimal parameters are optimized through extensive tests. Thus, in most cases, the using of the default parameters can yield the desired results.

Acknowledgment

The authors would like to thank professors S Y Ouyang, H Ru, and Dr. F F Niu for the data courtesy of STING, DR6, and LEAV.

References

[1] Terwilliger T C, Adams P D, Read R J, McCoy A J, Moriarty N W, Grosse-Kunstleve R W, Afonine P V, Zwart P H and Hung L W 2009 *Acta Cryst. D* **65** 582
[2] Sheldrick G M, Hauptman H A, Weeks C M, Miller R and Usón I 2001 *International Tables for Crystallography Volume F: Crystallography ofbiological macromolecules*, eds. Rossmann M G and Arnold E (Dordrecht: Springer Netherlands) pp. 333–345
[3] Uson I and Sheldrick G M 2018 *Acta Cryst. D* **74** 106
[4] Cowtan K D 1994 *dm: An automated procedure for phase improvement by density modification*, Vol. 31 (Warrington WA4 4AD, England: Daresbury Laboratory) pp. 34–38
[5] Cowtan K 2010 *Acta Cryst. D* **66** 470
[6] Wang B C 1985 *Methods Enzymol* **115** 90
[7] Morris R J, Perrakis A and Lamzin V S 2003 *Methods Enzymol.* **374** 229
[8] Cowtan K 2006 *Acta Cryst. D* **62** 1002
[9] Terwilliger T C, Grosse-Kunstleve R W, Afonine P V, Moriarty N W, Zwart P H, Hung L W, Read R J and Adams P D 2008 *Acta Cryst. D* **64** 61
[10] Adams P D, Afonine P V, Bunkoczi 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
[11] Ness S R, de Graaff R A, Abrahams J P and Pannu N S 2004 *Structure* **12** 1753
[12] Panjikar S, Parthasarathy V, Lamzin V S, Weiss M S and Tucker P A 2005 *Acta Cryst. D* **61** 449
[13] Panjikar S, Parthasarathy V, Lamzin V S, Weiss M S and Tucker P A 2009 *Acta Cryst. D* **65** 1089
[14] Zhang W Z, Zhang H M, Zhang T, Fan H F and Hao Q 2015 *Acta Cryst. D* **71** 1487
[15] Zhang T, Gu Y X, Zheng C D and Fan H F 2010 *Chin. Phys. B* **19** 086103
[16] Gu Y X, Zheng C D, Fan H F and Zhang T 2010 *Chin. Phys. B* **19** 086102
[17] Wu L J, Gu Y X, Zheng C D, Fan H F and Zhang T 2010 *Chin. Phys. B* **19** 096101
[18] Liu Z J, Lin D W, Tempel W, Praissman J L, Rose J P and Wang B C 2005 *Acta Cryst. D* **61** 1311

[19] Winn M D, Ballard C C, Cowtan K D, Dodson E J, Emsley P, Evans P R, Keegan R M, Krissinel E B, Leslie A G, McCoy A, McNicholas S J, Murshudov G N, Pannu N S, Potterton E A, Powell H R, Read R J, Vagin A and Wilson K S 2011 *Acta Cryst. D* **67** 235
[20] Vijayan M and Ramaseshan S 2006 *Isomorphous replacement and anomalous scattering* (Warrington WA4 4AD, England: Darebury Laboratory) pp. 80–85
[21] Strout G and Jensen L 1989 *X-ray structure determination*, Vol. Practical Guide, 2nd edn., (New York: John Wiley and Sons) pp. 343–378
[22] Brunger A T 1992 *Nature* **355** 472
[23] Tickle I J, Laskowski R A and Moss D S 1998 *Acta Cryst. D* **54** 547
[24] Tickle I J, Laskowski R A and Moss D S 2000 *Acta Cryst. D* **56** 442
[25] Lunin V Y and Woolfson M M 1993 *Acta Cryst. D* **49** 530
[26] Ouyang S, Song X, Wang Y, Ru H, Shaw N, Jiang Y, Niu F, Zhu Y, Qiu W, Parvatiyar K, Li Y, Zhang R, Cheng G and Liu Z J 2012 *Immunity* **36** 1073
[27] Dauter Z 2006 *Acta Cryst. D* **62** 867
[28] Zwart P H, Grosse-Kunstleve R W and Adams P D 2005 *CCP4 Newsl* **43** 27
[29] Niu F, Shaw N, Wang Y E, Jiao L, Ding W, Li X, Zhu P, Upur H, Ouyang S, Cheng G and Liu Z J 2013 *Proc. Natl. Acad. Sci. USA* **110** 9054
[30] Ru H, Zhao L, Ding W, Jiao L, Shaw N, Liang W, Zhang L, Hung L W, Matsugaki N, Wakatsuki S and Liu Z J 2012 *Acta Cryst. D* **68** 521