Guidelines for high-efficiency/accuracy data collection with multiple small-wedge data collection using the automatic data collection ZOO system

In protein crystallography, the data collection method is modified in accordance with the combination of crystal and X-ray beam sizes, such as in rotation data collection, helical data collection, small-wedge synchrotron crystallography (SWSX), and serial synchrotron rotation crystallography (SSROX) experiments [1]. In SWSX, a cryoloop with multiple crystals is raster-scanned with X-rays, 5–20° small-wedge data are collected from each crystal, and a large number of small-wedge datasets are merged to produce a highly complete dataset after individual data reduction [2]. SPring-8 **BL45XU** was designed to enable fully automatic measurements. All beamline components are controlled by the beamline control software BSS [3]. The ZOO system [4] has achieved automated data collection for all goniometer-based data collections in protein crystallography by communicating with BSS. Small crystals such as LCP crystals of membrane proteins (membrane protein crystals) are measured by SWSX combined with microbeams using the ZOO system at SPring-8. However, there has been no systematic study of the effect of the absorbed dose on the final merged dataset in SWSX.

We performed experiments to evaluate the optimal dose for SWSX using microcrystals of the same size and microbeams in order to efficiently collect highly accurate data. The optimal dose for obtaining highly accurate data was investigated by sulfur-SAD (S-SAD) phasing. The evaluation samples used were lysozyme crystals with a controlled size of approximately 20 μm. All SWSX datasets were automatically collected using the ZOO system at BL45XU. The beam size for data measurement was 18(H)×20(V) μm^2. 10° small-wedge data were collected from each crystal with 18 combinations of a wavelength set of 1.0, 1.4, and 1.7 Å, and doses of 1, 2, 5, 10, 20, and 40 MGy (Fig. 1). More than 400 small-wedge datasets were collected under each condition. The dose per crystal of the data measured using the ZOO system was estimated by RADDOSE 3D [5]. Data processing and merging for each small-wedge dataset were performed by XDS using KAMO [6]. The clustering process for data merging was BLEND [7]. For each merged dataset, S-SAD phasing was performed using SHELXC, SHELXD, and SHELXE in SHELX [8]. Among the six doses at a wavelength of 1.0 Å, only the dose of 5 MGy resulted in successful phasing by S-SAD (Table 1). At a wavelength of 1.4 Å, the doses of 1, 2, 5, 10, and 20 MGy resulted in successful phasing by S-SAD, but not the dose of 40 MGy. Experimental phases were successfully determined for all doses at a wavelength of 1.7 Å.

Furthermore, to determine the contribution of the number of merged datasets at wavelengths of 1.4 and 1.7 Å in phase determination, CCmap for each dose was plotted (Fig. 2). The number of merged datasets was set to eight patterns (25, 50, 75, 100, 125, 150, 175, and 200 sets), and 10 repetitions of random dataset extraction and merging were independently conducted for each number of merged datasets for two different wavelengths, 1.4 and 1.7 Å. For both wavelengths, CCmap increased with the number of merged datasets, except at a dose of 40 MGy at a wavelength of 1.4 Å. It was clearly shown that the higher the number of merged datasets, the easier the phase determination becomes at any dose.

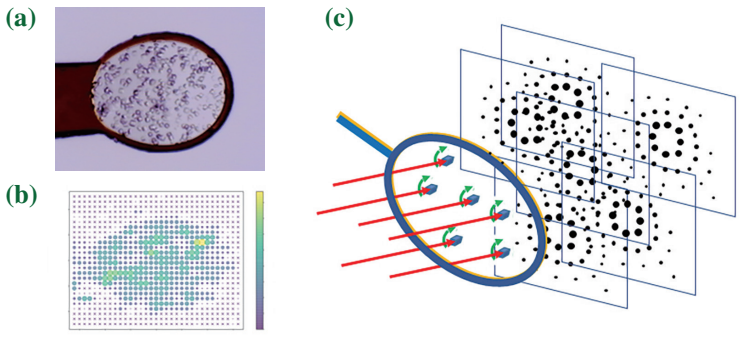

Fig. 1. Details of SWSX data collection. **(a)** Multiple crystals in cryoloop. **(b)** Results of an X-ray raster scan of multiple crystals in a cryoloop. **(c)** 5–20° small-wedge data are collected from each crystal.

For both wavelengths of 1.4 and 1.7 Å, the dose of 5 MGy produced the least number of datasets and increased CCmap, and the second-best doses were 2 and 10 MGy. The number of merged datasets that increased CCmap was higher at a wavelength of 1.7 Å than at a wavelength of 1.4 Å. This is because the Bijvoet ratio of lysozyme in S-SAD is 1.76% (f"= 0.67) at a wavelength of 1.7 Å, which is larger than 1.23% (f"= 0.47) at a wavelength of 1.4 Å. These results indicate that a dose of approximately 5 MGy at any wavelength enables the phase determination of S-SAD at the lowest number of datasets, which can be measured in a shorter time. The higher the number of merged datasets, the easier the phase determination becomes. This shows that data accuracy improves with the number of merged datasets in SWSX. For high doses of 20 MGy at a wavelength of 1.4 Å and 20 and 40 MGy at a wavelength of 1.7 Å radiation damage was severe and the data accuracy decreased. S-SAD phasing was made possible by increasing the number of datasets to be merged and improving the accuracy. However, at a higher dose of 40 MGy at a wavelength of 1.4 Å, the degraded data accuracy due to radiation damage could not be recovered by the merging process.

On the other hand, at the lower doses of 1 and 2 MGy at both wavelengths of 1.7 and 1.4 Å, the signal was insufficient, but it was recovered by increasing the number of datasets to be merged. The general dose slicing strategy is to add the sliced data in a merging process to ensure accurate measurements. However, this is not efficient because it requires a long measurement time for a number of repetitions.

Table 1. Numbers of merged crystals and S-SAD phasing statistics of lysozyme crystals at different doses at a wavelengths of 1.0, 1.4, and 1.7 Å.

Dose (MGy)	1	2	5	10	20	40
Wavelength	1.0 Å					
Number of crystals	366	357	384	243	428	399
S-SAD Phasing	Failure	Failure	Success	Failure	Failure	Failure
Wavelength	1.4 Å					
Number of crystals	335	400	404	219	316	348
S-SAD Phasing	Success	Success	Success	Success	Success	Failure
Wavelength	1.7 Å					
Number of crystals	353	453	358	395	218	298
S-SAD Phasing	Success	Success	Success	Success	Success	Success

In SWSX, increasing the number of crystals in the merging process makes the phase determination easier. The dose of 5 MGy at both wavelengths of 1.7 and 1.4 Å was the most efficient in degrading the accuracy through radiation damage and in recovering the accuracy in accordance with the number of merges. In this study, we suggested that the optimal condition for high-efficiency/accuracy data collection condition with SWSX is to acquire one data set at a dose of 5 MGy per crystal.

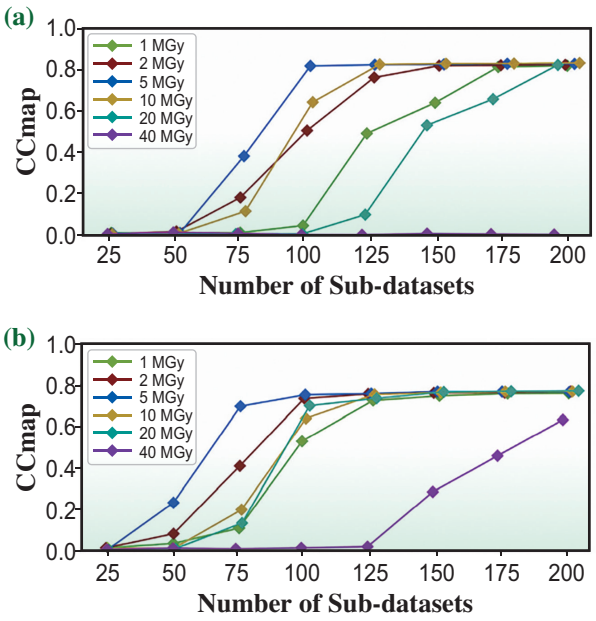

Fig. 2. Correlation between the number of datasets merged for each dose and CCmap for wavelength of **(a)** 1.4 Å and **(b)** 1.7 Å. The mean CCmap values derived from the phase determination for 10 randomly selected datasets are plotted against the number of datasets merged.

Seiki Baba

Japan Synchrotron Radiation Research Institute (JASRI)

Email: baba@spring8.or.jp

References

[1] K. Hasegawa *et al.*: J. Synchrotron Rad. **24** (2017) 29.
[2] V. Cherezov *et al.*: J. R. Soc. Interface. **6** (2009) S587.
[3] G. Ueno *et al.*: J. Synchrotron Rad. **12** (2005) 380.
[4] K. Hirata *et al.*: Acta Cryst. **D75** (2019) 138.
[5] O. B. Zeldin *et al.*: J. Appl. Crystallogr. **46** (2013) 1225.
[6] K. Yamashita *et al.*: Acta Cryst. **D74** (2018) 441.
[7] J. Foadi *et al.*: Acta Cryst. **D69** (2013) 1617.
[8] G. M. Sheldrick: Acta Cryst. **D66** (2010) 479.