Introduction to single crystal X-ray analysis

IV. Data collection and processing

Akihito Yamano* and Mikio Yamasaki*

1. Introduction

Data collection and processing have a significant impact on the structure analysis step. Considering the power of current direct method programs, quality data are nearly equal in importance to obtaining the initial structure when crystallographic difficulties such as an ambiguous space group and twining are not involved. This article will describe problems and measures in obtaining diffraction data using two-dimensional detectors: a CCD and an IP detector.

2. Data collection

2.1. Crystal quality

Even a perfect-looking crystal under a microscope can seriously be flawed because of hidden defects. The real nature of the crystal can only be assessed by irradiating the crystal with X-rays and inspecting the quality of diffraction.

Examples of poorly diffracting crystals and measures to overcome those problems are discussed below.

a. Absence of high-angle reflections

Diffraction images on an IP and a CCD detector are respectively shown in Figs. 1 and 2. Neither of the images have diffraction spots in the high-angle region. The first approach to this problem is to extend the exposure time. If there is no significant improvement, the next choice will be to mount a new crystal. When no improvement is observed after checking several crystals, one has to go back to the crystallization step and explore different crystallization conditions.

b. Diffraction resembling a powder pattern

Figure 3 shows a diffraction image from the CCD

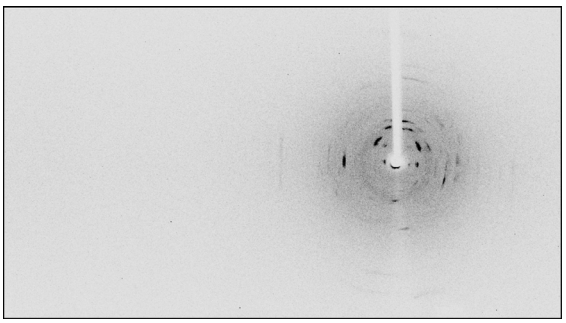

Fig. 1. A diffraction image from the IP detector. An example of the absence of high angle reflections.

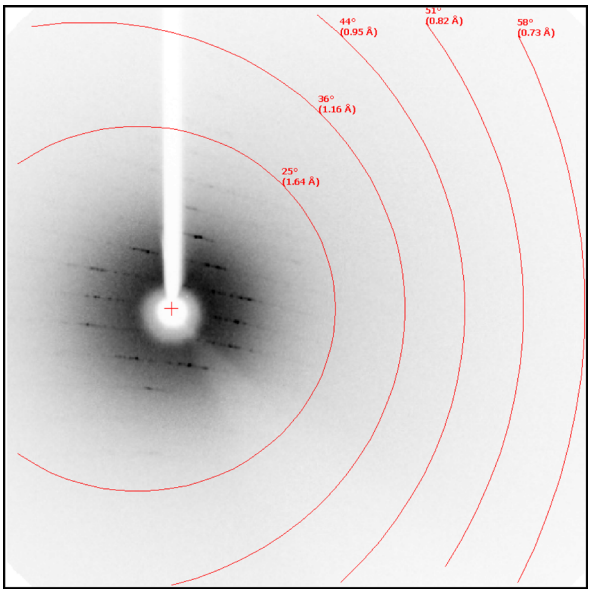

Fig. 2. A diffraction image from the CCD detector. Additional example of the absence of high angle reflections.

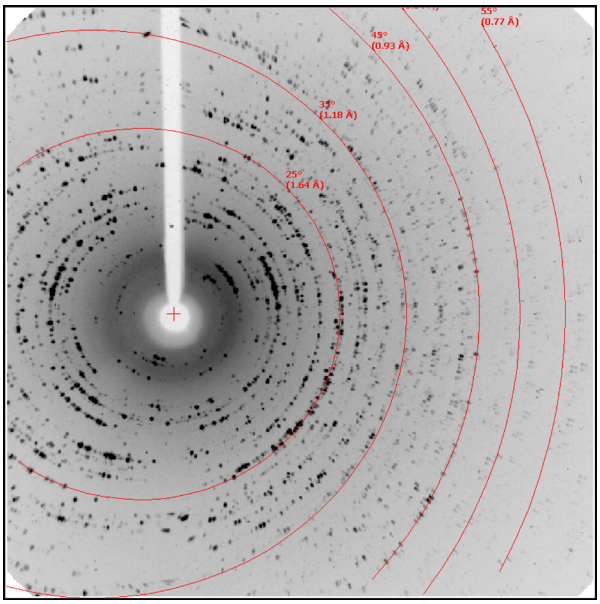

Fig. 3. A diffraction image from a crystal poor in crystallinity. Diffraction spots are nearly forming Debye rings, implying this crystal is a cluster of multiple crystallites.

* Application Laboratories, Rigaku Corporation.

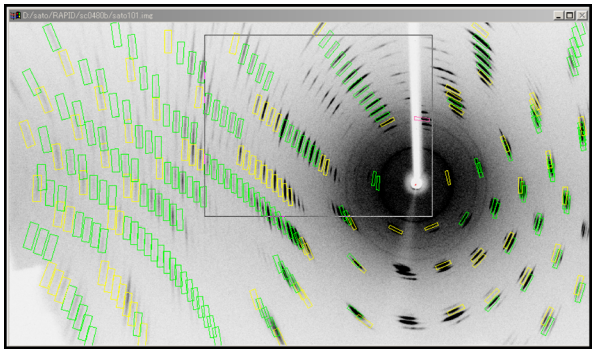

Fig. 4. A diffraction image from a crystal with high mosaicity. All diffraction spots are elongated severely in the circular direction. The arc-shaped integration boxes of the "tangential" mode of RAPID-AUTO are also shown.

created by overlaying multiple frames equivalent to 6°. The diffraction spots are aligned along Debye rings implying the sample is not a single crystal but a cluster of numerous crystallites. If this is the case, one must replace the crystal because there is no chance even to determine the unit cell.

d. Diffraction spots elongated in the circular direction

Figure 4 shows an image from the IP detector. The diffraction spot is elongated in the circular direction and an accurate structure cannot be derived by a regular intensity integration method. The first choice is to change the crystal but if the number of crystals is limited, or there seems to be little chance of finding a good crystal, it may be worth proceeding to data collection. RAPID-AUTO, a data collection and processing software originally written for the curved IP system, has an option to generate arc-shaped integration boxes that can cover the elongated reflections.

2.2. Redundancy

Generally speaking, the accuracy of the integrated intensity of each reflection is lower with IP and CCD two-dimensional detectors than with a scintillation counter, a photon counting detector. With a four circle diffractometer that equipped with a scintillation counter the measurement is repeated until a certain signal-to-noise ratio is satisfied. In contrast, a two-dimensional detector achieves comparable accuracy by averaging multiple measurements. This is possible because of the size of the two-dimensional detector allows recording of numerous reflections in one exposure.

Redundancy in single crystal analysis is an index of the average multiplicity of measurements of equivalent and identical reflections. To improve the data quality, one should plan the experiment to maximize the redundancy in a limited data collection time.

Table 1 summarizes the results of structure analyses differing in redundancy. A sucrose crystal was used throughout the experiment. When the redundancy was 1.67, the data collection time was 1 hour and the R1 was 3.26%. When the redundancy was raised to 3.04, the R1 became 2.77%, but the data collection time was

Table 1. Results of structure analysis of a sucrose crystal.

Composition	$C_{12}H_{22}O_{11}$	
Crystal size (mm)	0.43×0.38×0.14	
Space group	$P2_1$(#4)	
# of observed reflections	7707	3592
# of independent reflections (Friedel pairs)	2535 (1149)	2157 (807)
Redundancy	3.04	1.67
Rmerge (%)	3.35	3.23
R1 [I>2.0σ(I)] (%)	2.77	3.26
Rw [all data] (%)	7.10	9.75
Flack parameter	−0.04 (13)	−0.1(2)
Data collection time (min.)	133	62

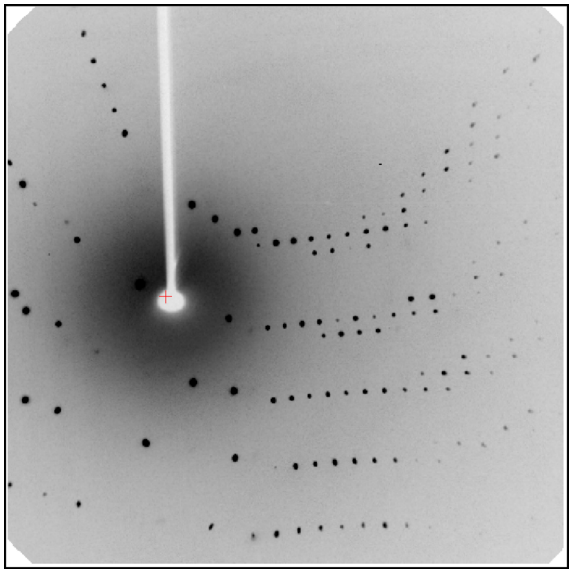

Fig. 5. A diffraction image collected with an oscillation angle of 0.5° and with 32 sec. exposure time. Images corresponding to 5°, that is 10 frames are overlaid.

doubled. Additionally the Flack parameter, an indicator of correctness of the absolute structure, has a smaller uncertainty.

For a regular structure determination, the former suffices but for a reliable determination of the absolute structure, the latter is desirable.

2.3. Oscillation angle

Apart from the crystallinity, a smaller oscillation angle usually gives a better signal-to-noise ratio. Figures 5 and 6 show diffraction images differing in oscillation angle range. The diffraction image in Fig. 5 was taken with a 0.5° oscillation and 64 second exposure and that in Fig. 6 was done with a 0.25° oscillation and 32 second exposure. Both images are created by overlaying images corresponding to 5°. Even though the exposure time per 1° is consistent in the two data sets, Figure 6

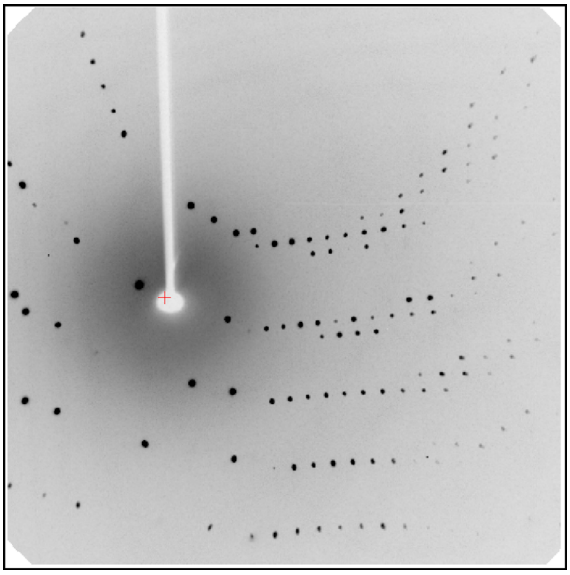

Fig. 6. A diffraction image collected with an oscillation angle of 0.25° and with 64 sec. exposure time. Images corresponding to 5°, that is 20 frames are overlaid.

shows lower background especially in the vicinity of the direct beam stop. This tendency can be explained qualitatively by assuming a crystal having a mosaic spread of 0.3°, a typical value for a small molecule crystal. On one hand, with the 0.25° oscillation angle, the intensity is recorded throughout the 0.25° oscillation when the centroid of the diffraction coincides with the middle of the oscillation angle. On the other hand, with the 0.5° oscillation angle, only 0.3° contributes to the intensity and remaining 0.2° contributes to the background. However, an exceedingly small oscillation angle often brings in negative effects such as a long data collection time, X-ray damage of the sample, etc.

3. Data processing

3.1. High Rmerge

Rmerge is the primary index that should be checked at the end of data processing. It is the direct measure of the agreement among equivalent reflections, therefore the smaller the value is, the better the data quality. Since the Rmerge strongly correlates with the crystallographic R factor, a large Rmerge is often a sign of an unreliable structure. When the Rmerge is larger than 20%, the data collection and processing should closely be examined. Sometimes limiting 2θ to 50 improves Rmerge and results in an acceptable R1 value. Table 2 shows an example of such cases. The Rmerge at the highest resolution shell between 0.83 to 0.77 Å exceeds 30%, and the $F^2/\sigma(F^2)$ below 3.0 indicates poor data quality in this resolution range. When the structure is solved with data $2\theta_{max} \leqq 55°$, the R1 became 5.19% while it became 4.37% with $2\theta_{max} \leqq 50°$. One should remember that the number of reflections decreases as the 2θ is reduced.

3.2. Low completeness

Contrary to the estimated completeness of reflections predicted during the preliminary experiment, the actual completeness after data processing often becomes much lower than expected. There are a number of possibilities for this, but the primary suspect is large mosaicity. One can easily confirm this problem by the prediction procedure. If there are many excess reflections, this probably is the case. Manually inputting the mosaicity and sometimes excluding some reflections from the refinement may eliminate this problem.

The second possibility is that the detector is too close to the sample. This occurs mainly with a CCD detector because of the relatively large point spread function. With a large unit cell, the separation among diffraction spots may become insufficient and discarded during the data processing due to severe overlaps. This problem can be identified by checking the relevant log file. It should refer to the exclusion of a large number of reflections. If this is the case, data collection must be repeated with a longer crystal to detector distance and a smaller oscillation angle.

Table 2. Result of data processing with $2\theta \leq 55°$.

Resolution (Å)	Completeness	Rmerge	$F^2/\sigma(F^2)$
−1.66			
1.66−1.32	1.000	2.984	27.47
1.32−1.15	1.000	4.349	16.88
1.15−1.05	1.000	7.005	11.15
1.05−0.97	0.997	10.789	6.95
0.97−0.91	1.000	13.078	5.06
0.91−0.87	1.000	16.869	3.85
0.87−0.83	1.000	22.637	2.95
0.83−0.80	0.993	30.155	2.01
0.80−0.77	0.993	36.679	1.52
Total	0.996	3.963	18.65

Table 3. Result of data processing with $2\theta \leq 50°$.

Resolution (Å)	Completeness	Rmerge	$F^2/\sigma(F^2)$
−1.81	0.972	1.431	54.73
1.81−1.44	1.000	2.612	30.95
1.44−1.25	1.000	3.589	22.81
1.25−1.14	1.000	4.633	15.13
1.14−1.06	1.000	7.313	10.70
1.06−1.00	1.000	9.672	7.61
1.00−0.95	0.996	11.289	5.93
0.95−0.90	1.000	15.251	4.27
0.80−0.87	1.000	16.735	3.72
0.87−0.84	1.000	20.728	3.11
Total	0.997	3.462	21.48

4. Summary

The careful visual inspection of the crystal is the first step to the successful structure analysis. However, we often encounter a perfect looking crystal that diffracts X-rays poorly. Patience is the key to a successful screening and at least 5 to 10 crystals should be checked before abandoning the sample.

Generally speaking, a smaller Rmerge implies a good data set. We are often asked for advice concerning the difficulty of obtaining an initial structure.

A large percentage of the inquiries can immediately be answered by checking the Rmerge value. It is noted that the initial structure cannot be determined with the Rmerge beyond 20%. For direct methods, it is key to have high-resolution reflections with good accuracy. If this is the case, and the initial structure still can not be obtained, the data collection process must be inspected carefully.