

PART 9. X-RAY DATA COLLECTION

Chapter 9.1. Principles of monochromatic data collection

Z. DAUTER AND K. S. WILSON

9.1.1. Introduction

X-ray data collection is the central experiment in a crystal structure analysis. For small-molecule structures, the availability of intensity data to atomic, usually around 0.8 Å, resolution means that the phase problem can be solved directly and the atomic positions refined with a full anisotropic model. This results in a truly automatic structure solution for most small molecules.

Macromolecular crystals pose much greater problems with regard to data collection. The first arise from the size of the unit cell, resulting in lower average intensities of individual reflections coupled with a much greater number of reflections (Table 9.1.1.1). Secondly, the crystals usually contain considerable proportions of disordered aqueous solvent, giving further reduction in intensity at high resolution and, in the majority of cases, restricting the resolution to be much less than atomic. Thirdly, the crystals are sensitive to radiation damage (see Section 9.1.12). Such problems have severe implications for all subsequent steps in a structure analysis. Solution of the phase problem is generally not possible through direct methods, except for a small number of exceptionally well diffracting proteins. The refined models require the imposition of stereochemical constraints or restraints to maintain an acceptable geometry.

At modern synchrotron beamlines, cryogenic cooling and high-efficiency two-dimensional (2D) detectors have made data collection technically easier, but it remains a fundamental scientific procedure underpinning the whole structural analysis. Therefore, it is essential to take the greatest care over this key step. The aim of this chapter is to indicate procedures for optimizing data acquisition. Overviews on several issues related to this topic have been published (Carter & Sweet, 1997; Evans & Walsh, 2005; Dauter, 2005).

9.1.2. The components of a monochromatic X-ray experiment

To collect X-ray data from single crystals, the following elements are required:

- (1) a source of X-rays;
- (2) optical elements to focus the X-rays onto the sample;
- (3) a monochromator to select a single wavelength;
- (4) a collimator to produce a beam of defined dimension;
- (5) a shutter to limit the exposure of the sample to X-rays;
- (6) a goniostat with associated sample holder to allow rotation of the crystal;
- (7) the crystalline sample itself;
- (8) a cryogenic cooling device for vitrified crystals;
- (9) an efficient, generally 2D, detector system;
- (10) software to control the experiment and store and display the X-ray images;
- (11) data-processing software to extract intensities and associated standard uncertainties for the Bragg reflections in the images.

On a number of beamlines, automated procedures have been implemented for sample changing, automatic crystal centring, evaluating the diffraction and proposing a strategy for data collection. These allow rapid and more effective selection of the best sample and optimal parameters.

Many of these are discussed elsewhere in this volume. This chapter aims to provide guidance in those areas where choices are to be made by the experimenter and is concerned with the interrelations between parameters and how they conspire for or against different strategies of data collection.

9.1.3. Data completeness

The advantage of diffraction methods over spectroscopy is that they provide a full 3D view of the object. Diffraction methods are theoretically limited by the wavelength of the radiation used, but, in practice, every diffraction experiment is further limited by the aperture and quality of the lens. In the X-ray experiment, the aperture corresponds to the resolution limit and the quality of the 'lens' to the completeness and accuracy of the measured Bragg reflection intensities.

In this context, completeness has two components, the first of which is geometric and hence quantitative. It is necessary to rotate the crystal so that all unique reciprocal-lattice points pass through the Ewald sphere and the associated intensities are recorded on the detector. Ideally, the intensities of 100% of the unique Bragg reflections should be measured. The second component is qualitative and statistical: for each hkl , the intensity, I_{hkl} , should be significant, with its accuracy correctly estimated in the form of an associated standard uncertainty, $\sigma(I)$. The data should be significant in terms of the $I/\sigma(I)$ ratio throughout the resolution range. This point will be returned to below, but it is especially important that the data at low resolution are complete and not overloaded on the detector, and that there is not a significant fraction of essentially zero-level intensities in the higher-resolution shells.

9.1.4. X-ray sources

There are two principal sources of X-rays appropriate for macromolecular data collection: rotating anodes and synchrotron storage rings. These are not discussed in detail here as they are described in Chapters 6.1 and 8.1.

Table 9.1.1.1
Size of the unit cell and number of reflections

Compound	Unit cell		Reflections	Average intensity
	Edge (Å)	Volume (Å ³)		
Small organic	10	1000	2000	1
Supramolecule	30	25000	30000	1/25000
Protein	100	1000000	100000	1/1000000
Virus	400	100000000	1000000	1/100000000

9. X-RAY DATA COLLECTION

9.1.4.1. Conventional sources

Rotating anodes were initially developed for biological scattering experiments on muscle samples and have the advantage of higher intensity compared with sealed-tube generators. They usually have a copper target providing radiation at a fixed wavelength of 1.542 Å. Alternative targets, such as silver or molybdenum, provide lower intensities at short wavelengths, but have not found general applications to macromolecules. It is also possible to use a chromium target, giving a longer wavelength of 2.29 Å. Historically, rotating anodes were first used with nickel filters to give monochromatic Cu $K\alpha$ radiation. Current systems are equipped with either graphite monochromators, a focusing mirror or multilayer optics. The latter provide substantially enhanced intensity. Rotating anodes remain the source of choice in most structural biology laboratories. An important choice for the user is in the selection of optimal collimator aperture: this should roughly match the crystal sample dimensions. For large crystals, especially if the cell dimensions are also large, it may be preferable to use collimator settings smaller than the crystal in order to resolve the diffraction spots on the detector. Considerable progress has been made in the technologies for rotating anodes in recent years, as well as in the development of high-intensity sealed-tube sources for home laboratories (Bloomer & Arndt, 1999).

9.1.4.2. Synchrotron storage rings

The radiation intensity available from rotating anodes is limited by the heat load per unit area on the target. In the early 1970s, it was realized that synchrotron storage rings produced X-radiation in the necessary spectral range for studies in structural molecular biology (Rosenbaum *et al.*, 1971), and the last four decades have seen great advances in their application to macromolecular crystallography (Helliwell, 2004; Hendrickson, 2000). Synchrotron radiation (SR) is now used for the great majority of newly determined protein-crystal structures.

The general advantages of SR are:

- (1) High intensity: third-generation sources provide more than 1000 times the intensity of a conventional source.
- (2) A highly parallel beam allowing the resolution of closely spaced spots from large unit cells.
- (3) Short wavelengths, less than 1 Å, essentially eliminating the problems of correcting for absorption.
- (4) Tunability of the wavelength, allowing its optimization for single- or multiple-wavelength applications; this is simply not possible with a conventional source.
- (5) The ability to use a white, non-monochromated beam, the so-called Laue technique discussed in Chapter 8.2.

SR beamlines take a number of forms. The source may be a bending magnet or an insertion device, such as a wiggler or an undulator. The properties of different beamlines thus vary considerably and it is vital to choose an appropriate beamline for any particular application. The beamline capabilities are, of course, affected by the detector as well as the source itself. As far as the user is concerned, the primary questions regard the intensity, the size of the focal spot, the wavelength tunability and the detector system.

The present consensus for new synchrotron beamlines for macromolecular crystallography is that they should be on sources with an energy of at least 3 GeV and should receive radiation from tunable undulators. Together, these provide high and tunable intensity over the range required for most crystal-

lographic experiments, including multiwavelength anomalous dispersion (MAD). The impact of free-electron lasers, which are currently under construction at a number of sites, is not yet possible to assess.

Present beamlines produce radiation of extremely high quality for macromolecular data collection. At third-generation sources complete data sets can be collected from cryogenically vitrified single crystals in minutes.

9.1.5. Goniostat geometry

9.1.5.1. Overview

The diffraction condition for a particular reflection is fulfilled when the corresponding reciprocal-lattice point lies on the surface of the Ewald sphere. If a stationary crystal is irradiated by the X-ray beam, only a few reflections will lie in the diffracting position. To record intensities of a larger number of reflections, either the size of the Ewald sphere or the crystal orientation has to be changed. The first option, with the use of non-monochromatic, or 'white', radiation, is the basis of the Laue method (Chapter 8.2). If the radiation is monochromatic, *i.e.* single-wavelength, the crystal has to be rotated during exposure to bring successive reflections into the diffraction condition.

9.1.5.2. The screenless rotation method and 2D detectors

In the early days of protein crystallography, a number of geometries were used for X-ray cameras, notably the Weissenberg and precession methods. In addition, single-counter diffractometers were used with four-circle goniostats. However, in practice only the screenless rotation geometry (Arndt & Wonacott, 1977) survives today. This requires a 2D detector, which was initially in the form of photographic film. It is of significance that typical film sizes were of the order of 10 × 10 cm with up to 2000 × 2000 scanned pixels; a similar effective area has proved effective for image plates and charge-coupled devices (CCDs).

However, automation of protein-data collection needed efficient 2D detectors (Part 7). The first were multiwire proportional counters, which found widespread use in the early 1980s (Hamlin, 1985). These finally proved to be limited by a combination of spatial resolution and dead time of the read-out. A major advance occurred in the late 1980s with the widespread introduction of imaging plates (Amemiya, 1995), scanned on-line both at synchrotron beamlines and on laboratory rotating-anode sources. This represented a revolution in macromolecular data collection, making it technically straightforward to save full 2D images with sufficient positional resolution and dynamic range to computer disk automatically. The limiting factor of the imaging plate proved to be the slow read-out time of the order of several seconds to minutes. At high-intensity sources in particular, *e.g.* third-generation SR sites, exposure times per image can fall to one second or less, and with an imaging plate the bulk of the time is spent reading the detector image rather than collecting data. Typical data-collection times with imaging plates remained of the order of several hours, even with the use of SR. This is a much smaller problem with rotating-anode sources, where exposure times dominate the duty cycle.

For high-intensity SR sites, the detector of choice is the CCD (Gruner & Ealick, 1995). The spatial resolution is comparable with that of imaging plates, but the read-out time can be as low as one to two seconds. This means that complete data can be