Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Data collection and processing

Johan Turkenburg[a] and Leo Brady[b]

[a]York Structural Biology Laboratory, Chemistry Department, University of York, York YO1 5DD, England, and [b]Department of Biochemistry, University of Bristol, Bristol BS8 1TD, England

The articles in this issue form the proceedings of the 1999 CCP4 Study Weekend on Data Collection and Processing. It is traditional for the CCP4 Study Weekends to ask the invited speakers to present both an overview and recent developments for their topic. As a result, this issue provides a comprehensive overview of nearly all aspects of a typical macromolecular X-ray diffraction experiment.

Data collection and processing is obviously a crucial step in the determination of any X-ray crystal structure, as it provides the observations which are used throughout the process of structure solution and refinement. Recent developments include third-generation synchrotron sources, new optics systems for home sources, the more widespread availability of image-plate and CCD detectors and the now universal application of cryogenic techniques. Together with advances in data processing and other crystallographic software, and the increased speed of computing, this has led to a considerably higher throughput in macromolecular crystallography.

The general introduction by M. Rossmann gives some of the history of the software currently in widespread usage and describes programs more recently developed in Purdue in collaboration with ADSC. This has resulted in autoindexing routines which are extensively described in the literature and now also form part of the *MOSFLM* suite, as described in a later presentation by H. Powell.

The widespread application of cryogenic techniques (E. Garman), which substantially reduce radiation sensitivity, has revolutionized macromolecular crystallography. It is now often possible to collect useful data on small weakly diffracting crystals where, until recently, limited crystal life-time would have been an insurmountable hurdle. The potential of cooling crystals to liquid-helium temperatures and the benefits this might bring are also discussed.

The next two papers discuss issues related to X-ray radiation sources. P. Lindley describes the nature of the ESRF, the third-generation synchrotron source at Grenoble. The impact on protein crystallography of modern synchrotron sources is illustrated with examples of data collected from small crystals, anomalous dispersion measurements, time-resolved studies and crystals with large unit cells. Each of these problems requires a careful match between the properties of the crystal and the characteristics of the X-ray source, the optics and the detector; with these requirements and considerations being reviewed in the paper by C. Nave.

The availability of state of the art X-ray equipment in the home laboratory is indispensable. This allows careful testing of pre-frozen crystals and screening of potential derivatives and complexes as well as routine data collection. Recently, there have been major developments in radiation sources, optics and detectors. The next three papers are concerned with these: experiences with a CCD detector mounted on a rotating-anode source (S. Muchmore), the development of the microfocus tube by U. Arndt and Bede Scientific (A. Bloomer) and the use of Yale mirrors and multilayer optics in combination with an image-plate detector and a rotating anode (J. Ferrara).

The two following papers discuss aspects of data processing. One of the challenges for data-processing packages is to keep apace with the speed of data collection at third-generation synchrotron sources. The first step is usually the indexing of one or more images and this needs to be fast and automatic. The implementation of one of the autoindexing algorithms in the data-processing program *MOSFLM* is described by H. Powell, while the integration of the diffraction data and the determination of standard uncertainties are covered by A. Leslie.

Fast indexing of initial images allows the experimentalist to plan a diffraction experiment rationally. The paper by Z. Dauter explains data-collection strategies with

great clarity. With the advent of very high intensity radiation sources and fast read-out detectors, data collection in fine slicing mode has become a viable alternative to the more traditional choice of oscillation width. The advantages of this approach and matching data-processing requirements are discussed by J. Pflugrath.

The next two papers concentrate on MAD and time-resolved studies, respectively. The application of cryogenic techniques and the availability of rapidly tunable high-intensity beamlines has greatly extended the use of multiple anomalous dispersion as a phasing tool. Results obtained at the Advanced Photon Source are described by M. Walsh. The need for specialized software when processing time-resolved Laue data is illustrated in the paper by D. Bourgeois. Algorithms developed can also be used to alleviate overlap problems in conventional data sets, and some examples are discussed.

A review of recent developments in electron microscopy techniques and the potential use of low-resolution phase information obtained in EM experiments for phasing in X-ray experiments is given by J. Grimes. The paper by N. Chandra describes various kinds of twinning as well as how to recognize them. In many cases it is possible to overcome the problems introduced by twinning, and a few examples are given. Outlier detection in non-redundant data often requires additional information if it is to be performed reliably. R. Read discusses how this can be performed more reliably using structure-factor statistics and shows how these ideas can be applied using his program *Outliar*.

A. Perrakis describes the development of a beamline set up specifically for the use of microcrystals at the ESRF. Crystallization is often the bottleneck in crystallographic studies, and in many cases only microcrystals are available. To take full advantage of the unique properties of beamline ID13 at the ESRF, and to make dealing with microcrystals practicable, elements such as the goniometer and backstop need to be carefully redesigned. Such a system is described in detail.

The last paper (P. Evans) summarizes essential points that should be considered when planning to perform proper data collection and processing.

Unfortunately, Z. Otwinowski, who spoke on data scaling, did not provide a written version of his talk.

The organization of this meeting would not have been possible without Pat Broadhurst (Daresbury) and the CCP4 staff, especially Alun Ashton and Sue Bailey. The preparation of these proceedings was greatly aided by David Brown (CCP4).