International Tables for Crystallography (2006). Vol. F, Chapter 1.1, pp. 1–3.

1. INTRODUCTION

1.1. Overview

BY E. ARNOLD AND M. G. ROSSMANN

As the first *International Tables* volume devoted to the crystallography of large biological molecules, Volume F is intended to complement existing volumes of *International Tables for Crystallography*. A background history of the subject is followed by a concise introduction to the basic theory of X-ray diffraction and other requirements for the practice of crystallography. Basic crystallographic theory is presented in considerably greater depth in other volumes of *International Tables*. Much of the information in the latter portion of this volume is more specifically related to macromolecular structure. This chapter is intended to serve as a basic guide to the contents of this book and to how the information herein relates to material in the other *International Tables* volumes.

Chapter 1.2 presents a brief history of the field of macromolecular crystallography. This is followed by an article describing many of the connections of crystallography with the field of medicine and providing an exciting look into the future possibilities of structure-based design of drugs, vaccines and other agents. Chapter 1.4 provides some personal perspectives on the future of science and crystallography, and is followed by a complementary response suggesting how crystallography could play a central role in unifying diverse scientific fields in the future.

Chapter 2.1 introduces diffraction theory and fundamentals of crystallography, including concepts of real and reciprocal space, unit-cell geometry, and symmetry. It is shown how scattering from electron density and atoms leads to the formulation of structure factors. The phase problem is introduced, as well as the basic theory behind some of the more common methods for its solution. All of the existing *International Tables* volumes are central references for basic crystallography.

Molecular biology has had a major impact in terms of accelerating progress in structural biology, and remains a rapidly developing area. Chapter 3.1 is a primer on modern molecular-biology techniques for producing materials for crystallographic studies. Since large amounts of highly purified materials are required, emphasis is placed on approaches for efficiently and economically yielding samples of biological macromolecules suitable for crystallization. This is complemented by Chapter 4.3, which describes molecular-engineering approaches for enhancing the likelihood of obtaining high-quality crystals of biological macromolecules.

The basic theory and practice of macromolecular crystallization are described in Chapters 4.1 and 4.2. This, too, is a rapidly evolving area, with continual advances in theory and practice. It is remarkable to consider the macromolecules that have been crystallized. We expect macromolecular engineering to play a central role in coaxing more macromolecules to form crystals suitable for structure determination in the future. The material in Part 4 is complemented by Part 5, which summarizes traditional properties of and methods for handling macromolecular crystals, as well as how to measure crystal density.

Part 6 provides a brief introduction to the theory and practice of generating X-rays and neutrons for diffraction experiments. Chapter 6.1 describes the basic theory of X-ray production from both conventional and synchrotron X-ray sources, as well as methods for defining the energy spectrum and geometry of X-ray beams. Numerous excellent articles in other volumes of *International Tables* go into more depth in these areas and the reader is referred in particular to Volume C, Chapter 4.2. Chapter 6.2 describes the generation and definition of neutron beams; related articles in other *International Tables* volumes include those in Volume C, Chapter 4.4.

Part 7 describes common methods for detecting X-rays, with a focus on detection devices that are currently most frequently used, including storage phosphor image plate and CCD detectors. This has been another rapidly developing area, particularly in the past two decades. A further article describing X-ray detector theory and practice is *International Tables* Volume C, Chapter 7.1.

Synchrotron-radiation sources have played a prominent role in advancing the frontiers of macromolecular structure determination in terms of size, quality and throughput. The extremely high intensity, tunable wavelength characteristics and pulsed time structure of synchrotron beams have enabled many novel experiments. Some of the unique characteristics of synchrotron radiation are being harnessed to help solve the phase problem using anomalous scattering measurements, *e.g.* in multiwavelength anomalous diffraction (MAD) experiments (see Chapter 14.2). The quality of synchrotron-radiation facilities for macromolecular studies has also been increasing rapidly, partly in response to the perceived value of the structures being determined. Many synchrotron beamlines have been designed to meet the needs of macromolecular experiments. Chapter 8.1 surveys many of the roles that synchrotron radiation plays in modern macromolecular structure determination. Chapter 8.2 summarizes applications of the age-old Laue crystallography technique, which has seen a revival in the study of macromolecular crystal structures using portions of the white spectrum of synchrotron X-radiation. Chapter 4.2 of *International Tables* Volume C is also a useful reference for understanding synchrotron radiation.

Chapter 9.1 summarizes many aspects of data collection from single crystals using monochromatic X-ray beams. Common camera-geometry and coordinate-system-definition schemes are given. Because most macromolecular data collection is carried out using the oscillation (or rotation) method, strategies related to this technique are emphasized. A variety of articles in Volume C of *International Tables* serve as additional references.

The use of cryogenic cooling of macromolecular crystals for data collection ('cryocrystallography') has become the most frequently used method of crystal handling for data collection. Part 10 summarizes the theory and practice of cryocrystallography. Among its advantages are enhanced crystal lifetime and improved resolution. Most current experiments in cryocrystallography use liquid-nitrogen-cooled gas streams, though some attempts have been made to use liquid-helium-cooled gas streams. Just a decade ago, it was still widely believed that many macromolecular crystals could not be studied successfully using cryocrystallography, or that the practice would be troublesome or would lead to inferior results. Now, crystallographers routinely screen for suitable cryoprotective conditions for data collection even in initial experiments, and often crystal diffraction quality is no longer assessed except using cryogenic cooling. However, some crystals have resisted attempts to cool successfully to cryogenic temperatures. Thus, data collection using ambient conditions, or moderate cooling (from approximately −40 °C to a few degrees below ambient temperature), are not likely to become obsolete in the near future.

Part 11 describes the processing of X-ray diffraction data from macromolecular crystals. Special associated problems concern

dealing with large numbers of observations, large unit cells (hence crowded reciprocal lattices) and diverse factors related to crystal imperfection (large and often anisotropic mosaicity, variability of unit-cell dimensions *etc*.). Various camera geometries have been used in macromolecular crystallography, including precession, Weissenberg, three- and four-circle diffractometry, and oscillation or rotation. The majority of diffraction data sets are collected now *via* the oscillation method and using a variety of detectors. Among the topics covered in Part 11 are autoindexing, integration, space-group assignment, scaling and post refinement.

Part 12 describes the theory and practice of the isomorphous replacement method, and begins the portion of Volume F that addresses how the phase problem in macromolecular crystallography can be solved. The isomorphous replacement method was the first technique used for solving macromolecular crystal structures, and will continue to play a central role for the foreseeable future. Chapter 12.1 describes the basic practice of isomorphous replacement, including the selection of heavy-metal reagents as candidate derivatives and crystal-derivatization procedures. Chapter 12.2 surveys some of the techniques used in isomorphous replacement calculations, including the location of heavy-atom sites and use of that information in phasing. Readers are also referred to Chapter 2.4 of *International Tables* Volume B for additional information about the isomorphous replacement method.

Part 13 describes the molecular replacement method and many of its uses in solving macromolecular crystal structures. This part covers general definitions of noncrystallographic symmetry, the use of rotation and translation functions, and phase improvement and extension *via* noncrystallographic symmetry. The molecular replacement method is very commonly used to solve macromolecular crystal structures where redundant information is present either in a given crystal lattice or among different crystals. In some cases, phase information is obtained by averaging noncrystallographically redundant electron density either within a single crystal lattice or among multiple crystal lattices. In other cases, atomic models from known structures can be used to help phase unknown crystal structures containing related structures. Molecular replacement phasing is often used in conjunction with other phasing methods, including isomorphous replacement and density modification methods. *International Tables* Volume B, Chapter 2.3 is also a useful reference for molecular replacement techniques.

Anomalous-dispersion measurements have played an increasingly important role in solving the phase problem for macromolecular crystals. Anomalous dispersion has been long recognized as a source of experimental phase information; for more than three decades, macromolecular crystallographers have been exploiting anomalous-dispersion measurements from crystals containing heavy metals, using even conventional X-ray sources. In the past two decades, synchrotron sources have permitted optimized anomalous-scattering experiments, where the X-ray energy is selected to be near an absorption edge of a scattering element. Chapter 14.1 summarizes applications of anomalous scattering using single wavelengths for macromolecular crystal structure determination. The multiwavelength anomalous diffraction (MAD) technique, in particular, is used to solve the phase problem for a broad array of macromolecular crystal structures. In the MAD experiment, intensities measured from a crystal at a number of wavelengths permit direct solution of the phase problem, frequently yielding easily interpretable electron-density maps. The theory and practice of the MAD technique are described in Chapter 14.2.

Density modification, discussed in Part 15, encompasses an array of techniques used to aid solution of the phase problem *via* electron-density-map modifications. Recognition of usual density-distribution patterns in macromolecular crystal structures permits the application of such techniques as solvent flattening (disordered solvent regions have lower density), histogram matching (normal distributions of density are expected) and skeletonization (owing to the long-chain nature of macromolecules such as proteins). Electron-density averaging, discussed in Chapter 13.4, can be thought of as a density-modification technique as well. Chapter 15.1 surveys the general problem and practice of density modification, including a discussion of solvent flattening, histogram matching, skeletonization and phase combination methodology. Chapter 15.2 discusses weighting of Fourier terms for calculation of electron-density maps in a more general sense, especially with respect to the problem of minimizing model bias in phase improvement. Electron-density modification techniques can often be implemented efficiently in reciprocal space, too.

Part 16 describes the use of direct methods in macromolecular crystallography. Some 30 years ago, direct methods revolutionized the practice of small-molecule crystallography by facilitating structure solution directly from intensity measurements. As a result, phase determination of most small-molecule crystal structures has become quite routine. In the meantime, many attempts have been made to apply direct methods to solving macromolecular crystal structures. Prospects in this area are improving, but success has been obtained in only a limited number of cases, often with extremely high resolution data measured from small proteins. Chapter 16.1 surveys progress in the application of direct methods to solve macromolecular crystal structures.

The use of computer graphics for building models of macromolecular structures has facilitated the efficiency of macromolecular structure solution and refinement immensely (Part 17). Until just a little more than 20 years ago, all models of macromolecular structures were built as physical models, with parts of appropriate dimensions scaled up to our size! Computer-graphics representations of structures have made macromolecular structure models more precise, especially when coupled with refinement methods, and have contributed to the rapid proliferation of new structural information. With continual improvement in computer hardware and software for three-dimensional visualization of molecules (the crystallographer's version of 'virtual reality'), continuing rapid progress and evolution in this area is likely. The availability of computer graphics has also contributed greatly to the magnificent illustration of crystal structures, one of the factors that has thrust structural biology into many prominent roles in modern life and chemical sciences. Chapter 17.2 surveys the field of computer visualization and animation of molecular structures, with a valuable historical perspective. Chapter 3.3 of *International Tables* Volume B is a useful reference for basics of computer-graphics visualization of molecules.

As in other areas of crystallography, refinement methods are used to obtain the most complete and precise structural information from macromolecular crystallographic data. The often limited resolution and other factors lead to underdetermination of structural parameters relative to small-molecule crystal structures. In addition to X-ray intensity observations, macromolecular refinement incorporates observations about the normal stereochemistry of molecules, thereby improving the data-to-parameter ratio. Whereas incorporation of geometrical restraints and constraints in macromolecular refinement was initially implemented about 30 years ago, it is now generally a publication prerequisite that this methodology be used in structure refinement. Basic principles of crystallographic refinement, including least-squares minimization, constrained refinement and restrained refinement, are described in Chapter 18.1. Simulated-annealing methods, discussed in Chapter 18.2, can accelerate convergence to a refined structure, and are now widely used in refining macromolecular crystal structures. Structure quality and target parameters for stereochemical constraints and restraints are discussed in Chapter 18.3. High-resolution refinement of macromolecular structures, including handling of hydrogen-atom

positions, is discussed in Chapter 18.4. Estimation of coordinate error in structure refinement is discussed in Chapter 18.5.

Part 19 is a collection of short reviews of alternative methods for studying macromolecular structure. Each can provide information complementary to that obtained from single-crystal X-ray diffraction methods. In fact, structural information obtained from nuclear magnetic resonance (NMR) spectroscopy or cryo-electron microscopy is now frequently used in initiating crystal structure solution *via* the molecular replacement method (Part 13). Neutron diffraction, discussed in Chapter 19.1, can be used to obtain high-precision information about hydrogen atoms in macromolecular structures. Electron diffraction studies of thin crystals are yielding structural information to increasingly high resolution, often for problems where obtaining three-dimensional crystals is challenging (Chapter 19.2). Small-angle X-ray (Chapter 19.3) and neutron (Chapter 19.4) scattering studies can be used to obtain information about shape and electron-density contrast even in noncrystalline materials and are especially informative in cases of large macromolecular assemblies (*e.g.* viruses and ribosomes). Fibre diffraction (Chapter 19.5) can be used to study the structure of fibrous biological molecules. Cryo-electron microscopy and high-resolution electron microscopy have been applied to the study of detailed structures of noncrystalline molecules of increasing complexity (Chapter 19.6). The combination of electron microscopy and crystallography is helping to bridge molecular structure and multi-molecular ultrastructure in living cells. NMR spectroscopy has become a central method in the determination of small and medium-sized protein structures (Chapter 19.7), and yields unique descriptions of molecular interactions and motion in solution. Continuing breakthroughs in NMR technology are expanding greatly the size range of structures that can be studied by NMR.

Energy and molecular-dynamics calculations already play an integral role in many approaches for refining macromolecular structures (Part 20). Simulation methods hold promise for greater understanding of the time course of macromolecular motion than can be obtained through painstaking experimental approaches. However, experimental structures are still the starting point for simulation methods, and the quality of simulations is judged relative to experimental observables. Chapters 20.1 and 20.2 present complementary surveys of the current field of energy and molecular-dynamics calculations.

Structure validation (Part 21) is an important part of macromolecular crystal structure determination. Owing in part to the low data-to-parameter ratio and to problems of model phase bias, it can be difficult to correct misinterpretations of structure that can occur at many stages of structure determination. Chapters 21.1, 21.2 and 21.3 present approaches to structure validation using a range of reference information about macromolecular structure, in addition to observed diffraction intensities. Structure-validation methods are especially important in cases where unusual or highly unexpected features are found in a new structure.

Part 22 presents a survey of many methods used in the analysis of macromolecular structure. Since macromolecular structures tend to be very complicated, it is essential to extract features, descriptions and representations that can simplify information in helpful ways. Calculations of molecular surface areas, volumes and solvent-accessible surface areas are discussed in Chapter 22.1. Useful generalizations relating surface areas buried at macromolecular interfaces and energies of association have emerged. Chapter 22.2 surveys the occurrence of hydrogen bonds in biological macromolecules. Electrostatic interactions in proteins are described in Chapter 22.3. The Cambridge Structural Database is the most complete compendium of small-molecule structural data; its role in assessing macromolecular crystal structures is discussed in Chapter 22.4.

Part 23 surveys current knowledge of protein and nucleic acid structures. Proliferation of structural data has created problems for classification schemes, which have been forced to co-evolve with new structural knowledge. Methods of protein structural classification are described in Chapter 23.1. Systematic aspects of ligand binding to macromolecules are discussed in Chapter 23.2. A survey of nucleic acid structure, geometry and classification schemes is presented in Chapter 23.3. Solvent structure in macromolecular crystals is reviewed in Chapter 23.4.

With the proliferation of macromolecular structures, it has been necessary to have databases as international resources for rapid access to, and archival of, primary structural data. The functioning of the former Brookhaven Protein Data Bank (PDB), which for almost thirty years was the depository for protein crystal (and later NMR) structures, is summarized in Chapter 24.1. Chapter 24.5 describes the organization and features of the new PDB, run by the Research Collaboratory for Structural Bioinformatics, which superseded the Brookhaven PDB in 1999. The PDB permits rapid access to the rapidly increasing store of macromolecular structural data *via* the internet, as well as rapid correlation of structural data with other key life sciences databases. The Nucleic Acid Database (NDB), containing nucleic acid structures with and without bound ligands and proteins, is described in Chapter 24.2. The Cambridge Structural Database (CSD), which is the central database for small-molecule structures, is described in Chapter 24.3. The Biological Macromolecule Crystallization Database (BMCD), a repository for macromolecular crystallization data, is described in Chapter 24.4.

Part 25 summarizes computer programs and packages in common use in macromolecular structure determination and analysis. Owing to constant changes in this area, the information in this chapter is expected to be more volatile than that in the remainder of the volume. Chapter 25.1 presents a survey of some of the most popular programs, with a brief description and references for further information. Specific programs and program systems summarized include *PHASES* (Section 25.2.1); *DM*/*DMMULTI* (Section 25.2.2); the *Crystallography & NMR System* or *CNS* (Section 25.2.3); the *TNT* refinement package (Section 25.2.4); *ARP* and *wARP* for automated model construction and refinement (Section 25.2.5); *PROCHECK* (Section 25.2.6); *MolScript* (Section 25.2.7); *MAGE*, *PROBE* and kinemages (Section 25.2.8); *XDS* (Section 25.2.9); and *SHELX* (Section 25.2.10).

Chapter 26.1 provides a detailed history of the structure determination of lysozyme, the first enzyme crystal structure to be solved. This chapter serves as a guide to the process by which the lysozyme structure was solved. Although the specific methods used to determine macromolecular structures have changed, the overall process is similar and the reader should find this account entertaining as well as instructive.