
Preface

The determination of the first protein crystal structure took Max Perutz 22 years of titanic work, and his beloved hemoglobin was in fact not the winner: Perutz was scooped by his colleague, John Kendrew (both awarded the Nobel Prize in 1962), who determined two years earlier, in 1957, the structure of the four times smaller myoglobin. Fifteen years later, when the Protein Data Bank (PDB) was created in 1971, there were only seven protein structures deposited there, whereas today the PDB holds over 130,000 experimental macromolecular structures. The overwhelming majority (90%) were determined by crystallography. Since about 6% of the PDB crystal structures contain nucleic acids, we should properly refer to this research area as macromolecular crystallography, but the historically sanctioned term, protein crystallography, is still used. After nearly three decades of slow trickle of structures, in the mid-1990s the PDB received a tremendous boost, entering an exponential growth phase. The main factors were (1) advances in computer and information technology, providing the much needed computer power for complex calculations, but also better algorithms and means of experiment automation; (2) introduction of genetic engineering for easy production of practically any protein in bacterial cell “factories”; and (3) widespread use of powerful synchrotron sources of X-rays. A strong impetus was provided by several structural proteomics projects, which, in the wake of the genomic era, set the ambitious goal of inferring the function of all proteins encoded by the sequenced genomes from their structure.

With the use of third-generation synchrotron sources and ultra-fast pixel area detectors (APD), the data collection time has been reduced to seconds, with concomitant reduction of the crystal size (to microns) and improvement of data quality. The speed of data collection and the routine use of cryogenic temperatures (100 K) have led to increased popularity of “mail-in crystallography,” where cryopreserved samples are shipped to a robot-operated beamline, and the data collection is conducted remotely. Another possibility is offered by polychromatic Laue diffraction, where structural transformations within a protein crystal (e.g., during a millisecond enzymatic reaction) are mapped using a series of nanosecond snapshots of complete datasets.

This is not the last word, however, because the emerging X-ray free electron lasers (XFELs) are offering beams more than 10 orders of magnitude brighter than even the most powerful synchrotrons. With such bright pulses, crystallites as small as 100 nm are injected to the beam, and a series of still diffraction images (from objects that are destroyed femtoseconds later) are used to reconstruct the complete diffraction pattern. The next step in this direction opens the possibility of studying the structure of single macromolecules injected into the XFEL beam. Inspired by the XFEL solutions, synchrotron beamlines are also turning towards serial crystallography (SSX).

Essentially all the steps of the crystallographic process have undergone a tremendous transformation in the last 10–20 years, and the progress has changed completely the way structural biology is practiced. Although the crystallization process is still based on the familiar phase diagram with oversaturation achieved by vapor diffusion, it is nowadays handled by crystallization robots capable of reproducible setting of thousands of trials in nanoliter volumes, and allowing for remote inspection of the crystallization process. Progress in

crystallization techniques is also noted in attacking the challenge of membrane protein crystallization. Nonexistent in the PDB earlier, membrane proteins started appearing in the PDB with increasing frequency from the late-1980s, led by the Nobel Prize-winning structure of the photosynthetic reaction center. The original method based on the use of detergents has been largely replaced by crystallization in mesophases, called lipidic cubic phases (LCPs).

Progress has been noted in all three basic methods for the solution of the phase problem in macromolecular crystallography. The method of multiple isomorphous replacement (MIR) is still occasionally used for novel protein structures but with new density modification algorithms it is more readily applicable in the single-wavelength SIR version, especially in combination with anomalous scattering. Additionally, the introduction of quick halide soaks has made it possible to avoid the complications and dangers of heavy metals. However, a more convenient tackling of novel structures uses the approach of multiwavelength anomalous diffraction (MAD) worked out by Wayne Hendrickson, or its single-wavelength variant (SAD). It is based on scattering of tunable synchrotron radiation by anomalous atoms such as selenium, which can be introduced into recombinant proteins in the form of Se-Met. Owing to the presence of many possible search models in the PDB, the most successful method of choice for homologous proteins is molecular replacement (MR), originally proposed by Michael Rossmann and David Blow, now available in a number of powerful algorithms, including those based on maximum likelihood (ML). The constantly improving data resolution and quality make it possible to solve protein structures using the weak anomalous signal of the natural sulfur atoms or even by direct methods.

Also the stage of structure refinement has advanced beyond recognition from the early simplistic and tedious algorithms. It is now possible to refine within minutes models with hundreds of thousands of parameters using millions of reflections. ML is usually the algorithm of choice. It uses a different probabilistic approach to model parameter optimization, asking for such a model that maximizes the probability of the concrete data set at hand. It easily incorporates prior knowledge in the form of stereochemical restraints but requires rigorous information about data statistics.

With the fast growing volume of deposits in the PDB, the problem of dubious or (very rarely) blatantly wrong models is becoming a major concern, especially with regard to complex structures, in which imagination or wishful thinking sometimes takes precedence over experimental data, particularly in ligand modeling. Such cases, however, stimulate continual development of validation tools and sensitize the community to the need of vigilance and maintenance of high standards. The absolute number of atomic resolution ($d_{\min} < 1.2 \text{ \AA}$) structures in the PDB is quite high (>3000), but their proportion has stayed at a less impressive level of $\sim 2\%$ for years. The fraction of ultrahigh resolution ($d_{\min} < 0.8 \text{ \AA}$) structures is dismally small (0.04%). These high-quality models in the PDB are, however, of paramount importance because together with data retrieved from the CSD (Cambridge Structural Database) they serve to define better standards for macromolecular structure refinement and validation. The recent developments explore the potential of machine learning and of conformation-dependent parametrization.

The most spectacular achievements of macromolecular crystallography, often crowned with Nobel Prizes, have significantly advanced our understanding of the molecular mechanisms of life as well as contributed to the development of successful medicines, therapies, or biotechnology tools. Crystallographic studies of virus structure have a long history, dating back to Stanley, Bawden, Pirie, Franklin, and Klug, are marked by two Nobel Prizes, and

have amassed several hundred models in the PDB. Crystallography has played a major role in dissecting the mechanisms of a number of viral pathogens. The most outstanding example is the battle with the HIV retrovirus. The prompt determination of the crystal structures of several key HIV proteins, most notably of the protease and reverse transcriptase, provided molecular targets for unprecedented structure-guided drug development success within just a few years. Indeed, the case of HIV protease set a new paradigm for rational drug design. Currently, this approach has been extended to fragment-based drug development, where crystallography is harnessed to identify molecular-cocktail components that can be stitched together to form drug molecules against specific macromolecular targets. In the recent outbreaks of viral infections, such as SARS, MERS, Ebola, or Zika, crystallographers have been in the front line of the battle, quickly providing dependable macromolecular structures for targeted drug development.

Perhaps the most iconic achievement of macromolecular crystallography in the recent years was the determination of the structure of the ribosome, which is a huge megadalton molecular machine responsible for the synthesis of all proteins in all living cells on our planet over several billion years. The structure explained that the ribosome is a ribozyme of catalytic RNA, as well as elucidated the mechanism of a number of antibiotics targeting the ribosomes of bacterial pathogens. The Nobel Prize to Venki Ramakrishnan, Tom Steitz, and Ada Yonath (2009) for the ribosome structure, which is a gene translation machine, followed the award to Roger Kornberg (2006) for the elucidation of the molecular mechanism of gene transcription.

The field of membrane-protein crystallography, initiated with the structure of the photosynthetic reaction center, is also growing very quickly. In the recent years, Nobel Prizes were awarded to Roderick MacKinnon (2003) for the determination of the structure of membrane channels and to Brian Kobilka and Robert Lefkowitz (2012) for the structure of the membrane-bound GPCR receptors. They sense diverse signals outside the cell (such as light, odor, hormone) and activate intracellular pathways by dissociating a subunit of the so-called G-protein that is coupled to the receptor (thus the name GPCR). There are ~800 different human GPCRs and they are the targets of ~50% of all modern drugs. It should be stressed that the first GPCR structure, determined by Krzysztof Palczewski for rhodopsin, explained the complicated molecular mechanism of our vision.

Also, recently crystallography has been used to explain the molecular mechanism of the promising versatile CRISPR-Cas9 genome-editing tool, adopted from the bacterial defense system based on clustered regularly interspaced short palindromic repeats (CRISPR) and coupled with a specific Cas nuclease.

The time is therefore ripe to describe what is currently available in the palette of methods and tools of contemporary macromolecular crystallography. The chapters included in this volume have been written by acclaimed specialists in each of the topics covered. It is hoped that this volume of *Methods in Molecular Biology* will help to acquaint the community of practicing and potential macromolecular crystallographers with the newest advances in the field and will inform them about the currently available tools.

*Frederick, MD, USA
Argonne, IL, USA
Poznan, Poland*

*Alexander Wlodawer
Zbigniew Dauter
Mariusz Jaskolski*

Protein Crystallography

Methods and Protocols

Wlodawer, A.; Dauter, Z.; Jaskolski, M. (Eds.)

2017, XIII, 672 p. 173 illus., 145 illus. in color.,

Hardcover

ISBN: 978-1-4939-6998-2

A product of Humana Press