ABSTRACT:
The first protein structures were determined in the 1950s. In the decades that followed, development of new methods for sample preparation, crystallization, data collection, and structure analysis yielded tens of thousands of biomolecular structures. This short review highlights some of the major technical advances exemplified with selected structures. © 2013 Wiley Periodicals, Inc. Biopolymers 99: 817–824, 2013.

Keywords: Protein Data Bank; protein structure; nucleic acid structure

INTRODUCTION

As this special issue of Biopolymers goes to press, the holdings of the Protein Data Bank archive (PDB) are approaching 100,000 entries (wwpdb.org).¹ Examination of the archive reveals atomic structures for biomolecules spanning the range of structure and function, from small peptide hormones to entire ribosomes, and from compact digestive enzymes to molecular motors with rotating axles and articulated chemically driven engines.

The richness of this central archive of knowledge builds on six decades of scientific innovation. As determination of the first protein structure in the 1950s, the field of structural molecular biology has grown and advanced, developing new methods with each decade to confront ever larger and more elusive targets. Herein, we review this trend by highlighting a signature structure from each decade and pertinent enabling technological advances. Please note that this exercise is by no means comprehensive, and, given space limitations, arguably seminal contributions, including some of the structures determined using NMR spectroscopy, are not described.

1950–1959

Technologies for biomolecular structure determination first emerged in the 1950s. At the time, two key pieces of information were known: first, earlier work with small molecules showed that single crystals diffract X-rays and may be used to determine atomic resolution structures; and second, proteins have defined three-dimensional (3D) structures and may be crystallized. In this era, diffraction patterns were recorded on film, which, due to radiation damage, required multiple crystals to provide complete data. A major challenge remained: how to determine the phases of each reflection in the diffraction pattern. To overcome this obstacle, Max Perutz developed the method of multiple isomorphous replacement (MIR).² His remarkable insight involved creating several modified versions of the protein of interest, by attaching one or a few electron-dense atoms at specific locations. Then, by systematically comparing diffraction measurements of these different crystalline forms, the locations of the heavy atoms could be determined...
and the phase angle for each reflection estimated with remarkable accuracy.

Early in the development of the MIR method, it was not clear how large a measurable diffraction difference would be required so several huge complexes of very heavy metals such as uranium were developed. For the globins, soaking crystals with moderately heavy metals (e.g., mercury, platinum) was sufficient.

The first protein structure, of myoglobin (Figure 1), was revealed in tantalizing installments; first, a low resolution structure that revealed the locations of alpha helices and the heme group, and then at progressively higher resolution, culminating in full a atomic rendition. John Kendrew’s structure revolutionized the study of protein structure and function by revealing many of the atomic aspects of protein architecture we now take for granted. Macroscopically, the structure proved consistent with the oil drop model,4 with most of its hydrophobic residues located within the core of the folded protein. Also, evident was the hierarchical nature of protein structure: alpha helical secondary structure elements, predicted in 1950 by Linus Pauling,5 were present, and the complex tertiary topology of the globular structure embracing the heme group was revealed. The myoglobin structure also exemplified the manifold noncovalent interactions that stabilize the structure, and precisely position the heme cofactor for oxygen binding and release.

The structure of hemoglobin followed closely, extending many of the architectural principles seen in myoglobin and adding a “quaternary” level to our understanding of protein structural hierarchy. Max Perutz’s work also revealed the atomic basis of one of the great mysteries of hemoglobin function: cooperative binding of oxygen via allosteric motion of protein subunits. Since these first two protein structures were obtained, more than 500 hemoglobin and more than 300 myoglobin structures have been determined from a variety of organisms, making it possible to create phylogenetic trees and understand protein evolution at the level of macromolecular structure.9

The 1960s saw a veritable explosion of effort in protein crystallography, greatly expanding and improving methods for structure determination. Researchers struggled with the complexity of the targets they were attacking, and created new tools to address many of the unresolved technical challenges. Data from X-ray measurements from protein crystals yield estimates of electron density for each point within the crystal in 3D, which were often displayed using stacks of transparent plastic sheets bearing contours showing electron rich volume elements corresponding to alpha helices and beta sheets (which are not present in the globins). In parallel, precisely engineered metal components reflecting the primary amino acid sequence were constructed to model the atomic structure of the protein. In the late 1960s, Fred Richards created a device known as a Richards’ Box or (less charitably) Fred’s Folly, wherein these two depictions of protein structure could be viewed simultaneously with a half-silvered mirror, allowing building of the atomic model into the contoured electron density map. In time, the very same process was migrated to computers using interactive computer graphics to manipulate the atomic model to fit the experimental electron density. Today, much of this model building can be accomplished in silico with little if any human intervention.

The first atomic structures of enzymes were determined in the 1960s. The structure of lysozyme (Figure 2) from David Phillips’ laboratory was rapidly followed by that of ribonuclease from the laboratories of Fred Richards and Gopinath Kartha. During this era, most enzyme structures were those of hydrolases, which are quite stable and relatively easy to purify and crystallize. This impressive body of work inspired many hypotheses regarding enzyme action, including the
mechanics of transition state stabilization and induced fit of enzyme active sites. Brian Matthews’ later work on T4 bacteriophage lysozyme showed that substrates also undergo induced fit, when he described a sugar ring distortion that favors transition state formation during the cleavage reaction. The very same lysozyme was later used as a structural biology “lab rat” in studies of hundreds of mutant variants of the enzyme, exploring relationships between amino acid sequence and structural stability (e.g., Ref. 16). As the first two structures of hydrolases were deposited to the PDB, more than 50,000 enzymes have followed, including representatives of every known enzyme class.

1970–1979
In the 1970s, the purview of structural biology expanded from proteins to nucleic acids. The structure of the phenylalanine transfer RNA (tRNA) was determined independently by competing research groups in the United States and United Kingdom. Multiple isomorphous replacement with osmium, samarium, and platinum derivatives was used to determine phases and data were painstakingly collected one reflection at a time using a single crystal X-ray diffractometer. Because phosphorus is such an electron dense atom, they appear as prominent features in electron density maps and proved easily recognizable landmarks when these competing structures were modeled.

Transfer RNA (Figure 3) in 3D brought many surprises. The cloverleaf secondary structure, which was correctly proposed on the basis of nucleotide sequence, was shown to fold into an L-shaped tertiary structure, with the three anticodon bases (magenta) at one end, and the site of amino acid attachment (red) more than 75 Å apart. The structure of tRNA also revealed modified bases, which continue to puzzle researchers studying RNA function and stability. The tRNA structures were also among the earliest structures for which crystallographic refinement techniques were used to improve the fit of the atomic model to the data with constrained-restrained methods as efficient alternatives to the full-matrix least-squares refinement approaches used by Keith Watenpaugh in his pioneering work on rubredoxin. During this same period, important advances in nucleic acid chemistry lead to deeper understanding of biological information transfer. Methods were developed for synthesis...
and crystallization of short DNA oligonucleotides of defined sequence. The first DNA crystal structure appeared in 1979,\textsuperscript{25} which strangely was left-handed. The first structure of right-handed B-DNA soon followed.\textsuperscript{26} There are more than 1400 oligonucleotide structures in the PDB.

1980–1989
Starting in the late 1970s and continuing into the 1980s, biomolecular crystallography moved on to tackle larger and larger targets. The phase problem, however, remained an enormous stumbling block. When determining the first structure of a virus,\textsuperscript{27} Harrison took advantage of the symmetrical nature of the tomato bushy stunt virus (TBSV) capsid to expedite phase estimation. With the benefit of certain assumptions about the noncrystallographic symmetry of the capsid, experimental phases could be determined quite accurately. Concurrently, Rossmann determined the structure of southern bean mosaic virus.\textsuperscript{28} As with many of the protein structures, these first virus structures were determined using X-rays produced by a rotating anode source with data collected on photographic film with an oscillation camera. By the time the rhinovirus cause of the common cold was crystallized much later, the increased intensity of X-ray beams produced by synchrotron radiation sources could be put to good advantage. The “American” method of data collection (shoot first, ask questions later) was invented during this era to maximize the amount of data that can be collected from a single crystal.\textsuperscript{29}

The structure of tomato bushy stunt virus (Figure 4) was by far the largest structure of its day, shedding light on puzzling aspects of capsid symmetry. In 1962, Don Caspar and Aaron Klug proposed the principle of quasisymmetry to explain the fact that some viruses form capsids containing large numbers of protein subunits incompatible with true icosahedral symmetry.\textsuperscript{30} TBSV represents one such case: a perfectly symmetrical capsid would be composed of 60 identical subunits, but the TBSV capsid contains 180 copies of a single protein. Harrison’s structure revealed that the larger capsid is assembled by allowing small local deformations of subunits, yielding a “quasisymmetric” capsid that is not perfectly icosahedral. There are now in excess of 400 virus structures in the PDB, with more than 100 determined using cryo-electron microscopy methods.

1990–1999
The next decade opened with the atomic structure of bacteriorhodopsin,\textsuperscript{32} the culmination of decades of biochemical and structural study, which paved the way for many additional structures of integral membrane proteins. During this same period, researchers began to crystallize proteins recognizing defined sequence elements within DNA double helices. Crystallization of these complexes was (and remains) something of an art, often requiring careful tuning of the exact length of the oligonucleotide and the presence of overhanging ends.\textsuperscript{33}

The 1990s also witnessed a decisive turn in the battle against the phase problem—multiple wavelength anomalous dispersion (or MAD) phasing. Wayne Hendrickson pioneered use of selenium incorporation in place of methionine sulfur atoms to produce protein crystals that can be used directly for structure determination with tunable sources of synchrotron radiation. His technological breakthrough realized at long last the promise evident in Johannes Bijvoet’s observations regarding anomalous diffraction published in 1949. Anomalous scattering information, obtained from just one protein crystal, allows atomic resolution structure determination without the need for heavy-metal derivatives. Hendrickson’s approach revolutionized protein crystallography. Indeed, it is the preferred strategy today, and most novel X-ray structures deposited to the PDB are the result of anomalous dispersion phasing.

Structures of DNA-protein complexes revealed the molecular mechanisms underpinning reading and replicating genomic information. Repressor structures show how proteins can read nucleotide sequences, and structures of polymerases reveal the machinery of replication. The T7 DNA polymerase structure
depicted in Figure 5 caught the enzyme in the act, with a short template strand, a primer strand, and a nucleotide waiting to be added to the growing strand. Nucleic acid polymerases are similarly shaped with an overall fold shaped like a hand that grips the growing DNA and arrays the catalytic amino acids deep inside, in the “palm.” Some of these structures also explained high fidelity DNA replication, showing in atomic detail the proofreading site that detects and corrects most of the errors made in the polymerase active site. Today, the PDB contains more than 2,700 protein-DNA complexes, ranging from the lac repressor and its operator to entire nucleosomes.

The T7 bacteriophage also gave structural biologists another important polymerase. T7 RNA polymerase controlled by the lac repressor is now used almost universally for inducible expression of recombinant exogenous proteins in Escherichia coli. The days of purifying proteins for crystallization from natural sources are now (fortunately) long gone.

2000–2009

In the new millennium, X-ray crystallography reached another turning point with the determination of ribosomal subunit structures. Ribosomes had been studied for decades using every conceivable biophysical and structural technique, building up an impressive body of information. Key to the determination of ribosome structures were three important advances, including (i) using now well-established cryogenic methods to minimize crystal deterioration in the X-ray beam, (ii) using information from electron microscopy to determine the molecular envelope, and (iii) using tungsten heavy atom derivative clusters for phasing. These landmark structures then provided crystallographers with the all-important foot in the door, permitting use of molecular replacement to determine a profusion of follow-on structures of ribosomes complexed with antibiotic drugs, protein regulatory factors, tRNAs, and mRNAs (Figure 6). There are now more than 300 structures of ribosomes and ribosomal subunits in the PDB.

The structure of the large ribosomal subunit profoundly influenced our understanding of evolution of life on the Earth. The active site, where peptide bonds are made and nascent proteins formed, is composed entirely of RNA, with a key adenine base doing much of the work. Thus, the
ribosome represents the largest known ribozyme. This striking observation fostered much research and even more speculation into the origins of life, leading to suggestions that the earliest replicating molecules were nucleic acids, not proteins.

2010–2013
Just as in the 1950s, structural biology continues to explore new areas of biology with a wealth of exciting new technologies. Researchers are using intensely bright X-ray beams to analyze ever smaller crystals, some taking data with infinitesimally short pulses of X-radiation that yield diffraction patterns before radiation damage can set in. Teams of students and educators have been employed to crowd-source the difficult problem of protein structure prediction, and the resulting models have been used to determine some particularly recalcitrant structures. Perhaps most exciting, this decade is turning out to be the decade of integrative structural biology, wherein many diverse techniques working at multiple levels of spatial resolution are being combined to determine structures of particularly large or complex macromolecular machines. The intact proteasome represents a telling example of what is now possible (Figure 7). A combination of existing structures (from the PDB archive) and molecular dynamics calculations was used to build a near-atomic model based on experimental data coming from cryo-electron microscopy. This important advance revealed the molecular bases of recognition and regulated destruction of unwanted, obsolete proteins.

PROTEIN STRUCTURES AND THE PROTEIN DATA BANK
Very early in the evolution of structural biology as a discipline, the scientific community realized that this wealth of structural data and knowledge must be shared. The PDB archive was established in 1971 with just seven inaugural structures. Atomic coordinates were initially distributed on magnetic tapes, and then via the internet. Today, the archive is approaching 100,000 entries. Many of these structures are being used on a daily basis by researchers, students, educators, policy makers, and the general public.

As the discipline has matured, so has the PDB archive. Deposition of atomic coordinates and primary experimental data is now mandatory for publication in most prominent scientific journals. The Worldwide Protein Data Bank (wwPDB) was formed in 2003 by PDB data processing centers in the US, Europe, and Japan. The wwPDB deals with the continually enlarging scale of both the experimental data and the atomic coordinate files, and works to establish a comprehensive, consistent data structure and efficient procedures for capturing and validating these data.

Countless groundbreaking scientific advances can be attributed to the open access nature of the PDB resource. Biomedical research, in particular, has benefited enormously, allowing atomic-level understanding of proteins involved in health and disease, and supporting discovery of drugs to modify the action of target enzymes and the like. PDB structures have also fostered deep understanding of basic biomolecular processes, ranging from the central dogma of biology visualized at the atomic level to the complexities of immune recognition and control. With the advent of high throughput structural
genomics, our understanding of structure–function relationships is being extended to encompass entire organisms. PDB structures have also enhanced our understanding of basic principles of biomolecular structure and function, supporting rapid progress in biotechnology and bionanotechnology.

REFERENCES

FIGURE 7 Proteasome (PDB ID 4b4t and EMDatabank EMD-2165). A near-atomic structure of the entire proteasome was determined by fitting atomic structures of the components (rainbow coloring) to a 7.4 Å resolution electron density map, derived from single particle cryo-electron microscopy (gray contours).

Reviewing Editor: Alfred Wittinghofer