

MRC LABORATORY OF MOLECULAR BIOLOGY



PIONEERS Kendrew (left) and Perutz, the scientists who determined the crystal structures of myoglobin and hemoglobin, with a model of myoglobin in the 1960s.

UNDERSTANDING *of* THE WORKINGS LIFE

Pinpointing three-dimensional arrangements of proteins has been crucial for understanding the chemistry of living cells

SARAH EVERTS, C&EN BERLIN

IN 1937, A YEAR INTO HIS DOCTORAL RESEARCH, Max F. Perutz was bored. The young Austrian scientist was at Cambridge University, using X-ray crystallography to study mining waste. He was working to solve the crystal structures of mineral fragments in slag heaps, and the samples were “horrible things to work with,” Perutz told Vega Institute historians in a video shot in 2001, a year before he died.

In the summer of 1937, Perutz took a trip to Prague to visit relatives. In a conversation with a biochemist cousin, Perutz talked about finding a better, more stimulating project. The cousin then made a suggestion that would change Perutz’ life and establish the field of protein structural biology. “He said why don’t [you] take on hemoglobin?”

At the time, top-notch crystallographers were studying the structures of vitamins, amino acids, and peptides, molecules containing hundreds of atoms. Hemoglobin had nearly 10,000 atoms. Nobody had solved the structure of such a large molecule.

But Perutz was already convinced that “if we didn’t understand the structure of proteins we would never understand how life works.” At the time “there was just one method, X-ray crystallography, which offered any hope—but even that was doubtful,” he said.

Perutz’ attempt to solve the structure of hemoglobin would take 22 years.

Before he finally succeeded, in 1959, Perutz’ Cambridge colleagues James D. Watson and Francis H.

C. Crick would begin and end their work to solve the structure of DNA, with essential data from Rosalind Franklin. Another colleague, John Kendrew, would start a project to solve the protein structure of myoglobin and finish it in 1957, beating Perutz to the punch by two

years, in part by using some of the essential methodology Perutz had developed.

Yet if Perutz ever despaired that his efforts might be in vain—and he did—many thousands of researchers today are grateful for his perseverance as well as for the methods he developed. Visualizing the structures of proteins at atomic resolution underpins the molecular understanding of biology. It is vital for drug research, and it is crucial for engineering enzymes to do industrial work, from fuel refinement to pulp and paper production.

When Perutz finally caught his first glimpse of hemoglobin at 5.5-Å resolution (*Nature* 1960, DOI: 10.1038/185416a0), he was awestruck. “I finally saw this thing I had been working on for 22 years,” Perutz recalled. “It was like reaching the top of a mountain after a very hard climb and falling in love at the same time,” Perutz said. “That intensity of joy—maybe you find it only in science, when nature reveals its great secrets.”

In contrast, the famous structure of DNA was solved in about a year and a half because the molecule’s

basic architecture is simple and symmetrical. “The DNA structure was a pen and pencil operation,” explains Wayne Hendrickson, a structural biology pioneer at Columbia University. At the time it was solved, researchers were proposing theoretical structures for DNA and calculating the X-ray diffraction patterns that would result from crystals of those proposed structures. The theoretical diffraction patterns were then compared with real diffraction data, Hendrickson says. “If you were close, you knew you had the right structure,” he says, which is how Watson and Crick succeeded.

Proteins were far more complicated. “It was doomed to think that you could just think your way through the problem to predict the protein’s diffraction patterns,” Hendrickson says.

Perutz, Kendrew, and their teams needed to use computers to mathematically convert X-ray diffraction patterns, using Fourier transforms, into electron-density maps of three-dimensional proteins. From these electron-density maps, researchers could then identify atoms in the structures.

IN THE MID-1950s, Kendrew’s and Perutz’ teams began using commercially available Edsat-1 and Edsat-2 computers to analyze protein diffraction patterns. These computers’ paper tape inputs predated IBM punch card computers, says Michael G. Rossmann, a structural biologist at Purdue University who helped Perutz solve the first structure of hemoglobin.

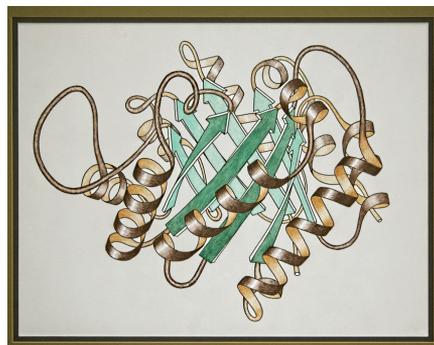
But computers weren’t the only advance necessary to solve protein structures from diffraction experiments. Sixteen years into the hemoglobin project—an era he called the dark years—Perutz was still stymied by the so-called phase problem. To mathematically convert diffraction data into electron-density data, researchers needed two important pieces of information: the intensity of diffracted light and the phase of that light. Perutz had figured out how to quantify the intensity of the diffraction pattern spots. But computing the phase of incident X-rays foiled Perutz until 1953.

The eventual solution, called isomorphous replacement, involved inserting a heavy atom into the protein crystals. This

MRC LABORATORY OF MOLECULAR BIOLOGY



TIMBER TROPHY Perutz carved this balsa wood model of hemoglobin in 1959.



JANE RICHARDS

OLD-SCHOOL TALENT Hand-drawn structure of triose phosphate isomerase by Richards, who developed the ribbon model of secondary structure.

modified diffraction patterns so that the phase of the incident light could be inferred. Kendrew used the technique to solve myoglobin; hemoglobin yielded shortly thereafter.

And thus began a new field of research.

Throughout the 1960s, the structures of carboxypeptidase, papain, chymotrypsin, and ribonuclease began to appear from old and new protein structure labs, while Perutz and Kendrew obtained increasingly better resolution structures of hemoglobin and myoglobin.

“Every new structure was a surprise,” says Robert Huber of the Max Planck Institute for Biochemistry, in Munich, whose thesis supervisor, Walter Hoppe, asked him to solve the structure of an oxygen-carrying protein in insects during the

MAGIC BOX The original Richards’ box, built by Frederic Richards at Oxford in 1968 to help construct accurate atomic models of proteins.



FREDERIC RICHARDS

1960s. After getting his feet wet in the field, Huber went on to win the 1988 Nobel Prize in Chemistry for solving the first structure of a membrane protein.

Yet the field still faced many challenges, particularly in finding ways to visually represent these complicated protein structures. “It really helped to have some artist’s skills,” says Helen M. Berman, the head of the Research Collaborating for Structural Bioinformatics’ Protein Data Bank at Rutgers University, part of a worldwide repository for protein structure coordinates. Sometimes people would draw structures of the complicated proteins by hand to make stereodiagrams, whereas others carved structures out of balsa wood.

Models of proteins containing thousands of atoms were labors of love. To construct them, researchers would plot two-dimensional sections of the protein’s electron-density data onto transparent plastic sheets. Stacking these plastic sheets together, researchers could get a 3-D representation of the protein’s electron density, from which they would then painstakingly

& MORE ONLINE

Look at proteins, from hand-drawn masterpieces to electron-density map sculptures, at <http://cenm.ag/prot>.

construct atomic models of the protein using wires and screws. To improve proper alignment of the models with the electron-density maps, researchers would build a contraption called a Richards' box. This box used half-silvered mirrors to superimpose the atomic models being built with the electron-density maps. The whole process "took months and months," says Ian A. Wilson, a pioneering structural biologist at Scripps Research Institute California who solved the structure of the glycolysis enzyme triose phosphate isomerase as a graduate student at Oxford University.

Once you were done building the model, you began the exhausting process of figuring out the x - y - z coordinates of the protein. Prior to computers, this required rulers and plumb lines.

TO ILLUSTRATE proteins for a journal publication, early structural biologists often hired professional artists. One of the few people who had both scientific know-how and artistic skills was Duke University's Jane S. Richardson. She invented the now ubiquitous ribbon diagram to simplify representations of secondary structures such as α -helices and β -sheets. These types of representations are still found in most articles that include protein structure images. In 2004, Richardson told a biographer that she was staggered that "a whole generation of scientists sees protein structure through my eyes" (*Biol. Physicist* 2004, 4, 5).

In the mid- to late 1970s, computers and their displays got powerful enough that researchers could use them to visualize protein structures in minutes instead of months. Yet programmers never found a better representation for α -helices and β -sheets than Richardson's ribbon diagrams, so they were incorporated into visualization software as well.

In addition to advances in computer visualization, the 1970s saw the establishment of the Protein Data Bank (PDB), a tool universally used by structural biologists today. But at the time, the repository of 3-D details was extremely controversial.

Several scientists believed that "those who generated the first structure of an important protein should have the privilege to work on it without competition, at least for a certain time, and to determine protein-ligand complexes, particularly for proteins with relevance in pharma," Huber says. PDB, in contrast, made all of this information available to anyone

who asked—if the bank had the structure.

If it didn't, "you had to write the author asking for coordinates," Rossmann says. "Some said no." Around 1981, Rossmann found a way to extract the data he wanted from a scientific paper, as long as its authors included a stereogram of the protein. "In desperation I wrote a computer program that could extract the protein's 3-D

coordinates from the stereogram," Rossmann says. "PDB supplied the program to anyone who asked."

According to Berman, everything changed again after 1989. That's when the International Union of Crystallography published guidelines for submitting structural data to PDB. "As soon as these guidelines were established, funding agen-

NMR with the resolution you need

Pharmaceutical and industrial labs now have an exciting analytical alternative to the expense and time required for dedicated NMR facilities. The new Thermo Scientific™ picoSpin™ 80 NMR spectrometer brings 80 MHz resolution, expected for most routine analysis, to virtually anywhere in your facility. With no liquid cryogenes and minimal shimming required, the small size and lightweight portability of the picoSpin 80 spectrometer provides the answers you need with less downtime and within an extremely affordable budget.

where you need it most

Thermo
SCIENTIFIC

- Find useful NMR resolution everywhere at thermoscientific.com/picospin



- More resolution with 80 MHz chemical shift dispersion.

cies and journals increasingly required PDB submission as part of receiving funds or getting published,” she says. The number of structures in the bank began to rise exponentially, from 507 in 1990 to 13,596 in 2000.

The growing number of PDB submissions set the stage for a better understanding of the rules of protein structure, says

“That intensity of joy—maybe you find it only in science, when nature reveals its great secrets.”

Thomas L. Blundell, a structural biologist at Cambridge University who solved the structure of insulin. Researchers such as Janet Thornton, now the director of the

European Bioinformatics Institute at the Wellcome Trust Genome Campus, in England, used PDB to compare, contrast, and statistically analyze the prevalence of certain protein folds or interactions such as salt bridges.

THESE EARLY structural bioinformaticians also began developing computer tools to superimpose the structures of seemingly unrelated proteins, allowing them to discover that nature sometimes reused useful topologies. The comparisons provided the first evidence that Darwinian evolutionary principles applied at the molecular scale in proteins. For example, some folds, such as α/β -barrels, cylinders of β -strands and α -helices, “kept popping up all over the place,” Thornton says.

In the 1980s, researchers increasingly noticed that evolution sometimes cut-and-pasted amino acid sequences from one protein to another and then slightly adjusted them to do different chemistry. This is a case of divergent evolution: Similar protein structures have related amino acid sequences, likely passed on from evolutionary protein ancestors. Other times, completely different amino acid sequences adopted the same 3-D fold, a case of convergent evolution, where a biomolecular structure or trait is so useful it evolves several times independently.

By the 1990s, protein structural biology had matured into a well-established analytical discipline. Researchers began using recombinant DNA to clone and express proteins at precise lengths, thus reducing the floppy ends that had stymied earlier crystallization attempts. Meanwhile, the use of X-rays produced by synchrotron sources improved the quality of diffraction data and its speed of acquisition. At the same time, Kurt Wüthrich and Richard R. Ernst were improving nuclear magnetic resonance techniques to solve protein structures, sometimes faster than with X-ray crystallization.

Throughout the 2000s, structural biology for simple and small proteins became more highly automated. High-throughput structural biology programs tackled previously unknown proteins uncovered from genome-sequencing projects. High-throughput technology also automated the production

How important are your APIs & Intermediates?

stra·te·gic
(strə-’tē-jik)

AFC offers unique competitive advantages creating long-term strategic benefits for our customers.

AFC is uniquely capable of paring one or more of our state-of-the-art technology offerings into the same synthetic sequence to develop innovative solutions for our customers, creating long-term strategic and competitive advantages.

Located in California and Texas, our facilities specialize in the following state-of-the-art technology platforms:

- **Commercial-scale production utilizing hazardous/energetic chemistries**
- **High-containment facilities** for the production of high-potency APIs (HPAPIs) and intermediates
- **Chromatographic separations** using simulated moving bed (SMB) technology
- **Continuous processing**
- **Controlled substances & specialized chemical capabilities**

 **AMPAC™**
FINE CHEMICALS

America's Leading Manufacturer of Custom Small Molecule APIs
+1 916 357 6880 | +1 800 311 9668 toll-free | AFCbusdev@apfc.com | ampacfinechemicals.com



of protein crystal samples. Meanwhile, researchers in the field set their sights on solving the structures of large, multidomain cellular machines like the ribosome and complex membrane proteins, such as ion channels. Finally, electron microscopy evolved as a third tool for determining the structure of proteins. In 2013, PDB topped 85,000 protein structures, many solved by a combination of X-ray crystallography, NMR, and electron microscopy.

As pioneers in the field reflect on the biggest difference between current protein structure projects and early ones, many mention the decreasing role of crystallographers and the increasing role of biologists in determining what proteins are studied.

“Initially, people solved the structures they could,” Huber says. Perutz, for instance, was aided back in the 1930s by a physiologist at Cambridge who had already crystallized horse hemoglobin. Now, says Huber, “You pose the biological question first. Then you determine which structures you need to solve in order to answer the question.”

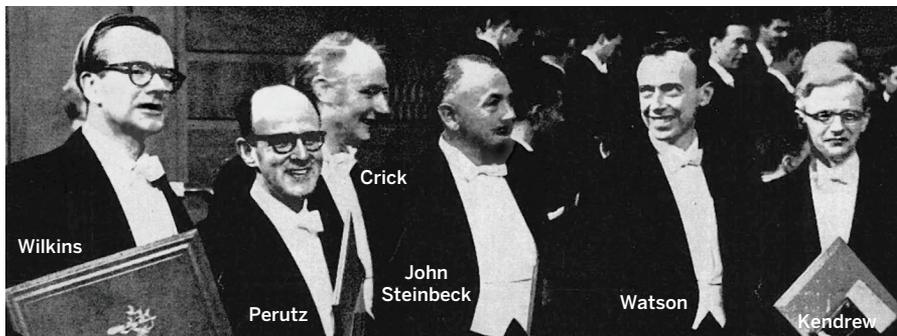
Blundell concurs: “A lot of the early protein structure people were physicists or computer people or chemists and didn’t know much about biology. The prerequisite was that you had crystals.” ■



From C&EN Archives

Protein Structure When Michael G. Rossmann solved the crystal structure of lactate dehydrogenase, C&EN visited the lab to photograph the pioneering structural biologist hard at work developing a stick atomic model of the 310-amino-acid protein.

Award Winning In 1962, the founders of structural biology won their share of Nobel Prizes. Max F. Perutz and John Kendrew shared the Nobel Prize in Chemistry for establishing tools to discover the structure of proteins, while James D. Watson, Francis H. C. Crick, and Maurice H. F. Wilkins won the Nobel Prize in Physiology or Medicine for their work on solving the structure of DNA.



C&EN ARCHIVES (BOTH)

Women's Work

Rosalind Franklin's X-ray diffraction data played a pivotal role in James D. Watson and Francis H. C. Crick's determination of DNA's three-dimensional structure. Their failure to properly acknowledge her data may tempt one to assume that the world of structural biology has been unwelcoming to women.

That would be an error. The first person to show that protein crystals could diffract was Dorothy Crowfoot Hodgkin of Oxford University, who went on to solve the structures of insulin and penicillin. In 1964, she won the Nobel Prize in Chemistry for her structure determination of vitamin B-12.

Kathleen Lonsdale, a crystallographer, showed benzene was flat. She was the first to use Fourier transforms to solve the X-ray crystal structure of a compound from diffraction data. She was also one of the

first female members of the Royal Society.

Many other women have high-profile positions in structural biology. Add to the list Helen M. Berman, director of the Research Collaborating for Structural Bioinformatics' Protein Data Bank, and Ada E. Yonath, who shared the 2009 Nobel Prize in Chemistry for her crystallographic studies of the ribosome, a mix of protein and RNA. “There do seem to be more women involved in crystallography than in other physical sciences,” Berman

says. The question is, why?

Many researchers point to Hodgkin. As a professor in an undergraduate college at Oxford, she mentored small groups of students. Since the colleges were gender-segregated in her era, she would have mentored only undergraduate women, many of whom were enticed to work in her lab as graduate students, says Cambridge University structural biologist Thomas L. Blundell, who solved the structure of insulin as her graduate student. “She had a big impact. She wasn’t somebody who was openly furthering the aims of women. She just did. Women just came to her lab. But men came as well.” Many of her students, such as Barbara Low, Marjorie Harding, and others, went on to set up

crystallography labs around the world.

And what was Hodgkin’s academic heritage? It can be traced back to the father of X-ray crystallography, William Bragg, who mentored both Lonsdale and John D. Bernal, who was Hodgkin’s doctoral adviser (as well as Max F. Perutz’ and John Kendrew’s). Bragg’s openness to having female doctoral students may have impressed Bernal when he himself was a student. Bernal certainly learned that his female graduate student peer—Lonsdale—could do excellent science, which may have inspired him to take on Hodgkin as a student. Later, Franklin moved to Bernal’s lab after her unpleasant experience with DNA, another sign that his lab welcomed women.