doi:10.1016/j.jmb.2009.05.024 J. Mol. Biol. (2009) 392, 2–32

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

50 Years of Protein Structure Analysis

Bror Strandberg[1], Richard E. Dickerson[2]* and Michael G. Rossmann[3]

[1]*Department of Cell and Molecular Biology, Uppsala University, Box 596, SE-751 24 Uppsala, Sweden*
E-mail address: bror.strandberg@telia.com

[2]*Molecular Biology Institute, Boyer Hall, University of California at Los Angeles, Los Angeles, CA 90095-1570, USA*
E-mail address: red@mbi.ucla.edu

[3]*Hockmeyer Hall of Structural Biology, Purdue University, 249 South Martin Jischke Drive, West Lafayette, IN 47907-1971, USA*
E-mail address: mr@purdue.edu

Received 7 May 2009;
accepted 8 May 2009

Fifty years ago, Max Perutz and John Kendrew at Cambridge University achieved something that many people at the time considered impossible: they were the first to use x-ray crystallography to decipher the molecular structures of proteins: haemoglobin and myoglobin. They found that both molecules were built from Linus Pauling's alpha helices, but folded and packed together in a complicated manner that never could have been deciphered by any other technique. With structure information in hand they could then explain how haemoglobin in the bloodstream binds and releases oxygen on cue, how it passes its cargo on to the related storage protein myoglobin, and how a single amino acid mutation can produce the catastrophe known as sickle-cell anemia. Perutz and Kendrew also observed that the folding of helices was identical in myoglobin and the two chains of haemoglobin, and this along with the simultaneously evolving new technique of amino acid sequence analysis established for the first time the concept of *molecular evolution*.

The crystallographic puzzle was "cracked" by Perutz when he demonstrated that the binding of only two heavy metal atoms to horse haemoglobin changed the x-ray pattern enough to allow him to solve the "phase problem" and circumvent the main obstacle to protein crystal structure analysis. Because myoglobin has a single chain whereas haemoglobin has four, Kendrew's work with myoglobin progressed more rapidly; a low resolution structure appeared in 1956 and the high resolution structure in 1959. That same year saw the low resolution picture of haemoglobin, and the high resolution structure followed shortly thereafter.

Much of the work in structure analysis was carried out by visiting postdoctoral fellows and technicians, under the watchful eye of Perutz and Kendrew. This celebratory review has been written by three of those former postdoctorals: Strandberg and Dickerson from the myoglobin project, and Rossmann from the haemoglobin.

Chapter 1: Building the Ground for the First Two Protein Structures: Myoglobin and Haemoglobin

Bror Strandberg

Introduction

Just 50 years ago, two Cambridge scientists, Max Perutz and John Kendrew, achieved a goal that had long been considered completely impossible: they solved the molecular structures of two related proteins—myoglobin and haemoglobin—by X-ray crystallography. In these days of automatic diffractometers; large high-speed computers; sophisticated methods for purifying, crystallizing, and labeling proteins with heavy atoms; and elegant computer programs for model building of protein structures, it is easy to forget how challenging the task was in the absence of all the aforementioned techniques. James Watson, who coparticipated a few years earlier in the structure analysis of DNA, once remarked sardonically that, "In some circles, an interest in the history of one's field is regarded as a sign of declining powers." But George Santayana warns us that, "Those who cannot remember the past are condemned to

*Corresponding author.

repeat it." So on the 50th year anniversary of such a remarkable achievement, it is worthwhile taking a few moments to consider the first time that anyone discovered what a protein molecule actually looked like.

Over the years, Kendrew and Perutz were assisted by many young collaborators from different parts of the world. Three of them were the authors of this memorial article. We were very privileged, since we had the good fortune to take part in the final stages of the structure determinations and could therefore experience the great moments of discovery when, for the first time, it could be seen how protein molecules were constructed. In the following two chapters, my colleagues Richard E. Dickerson and Michael G. Rossmann describe what happened when the structures of myoglobin and haemoglobin were determined at 2.0 Å and 5.5 Å resolution, respectively.

The First Steps

For a full appreciation of how these important results could be obtained, it is useful to have a brief background on how the basic method—X-ray crystallography—was invented and developed, in particular so that it could be used to determine the structures of such huge macromolecules. A handful of people played key roles in this development, and to follow their main discoveries is, of course, of special interest.

After the discovery of X-ray radiation by Wilhelm Röntgen in 1895, it was clear that this new radiation technique could have important medical applications. The use of X-rays in other fields was not so obvious. However, this situation changed through a discovery by the German physicist Max von Laue and his colleagues.

The phenomenon of interference was well known in optics using ordinary light. Light waves hitting regularly spaced objects (e.g., equally spaced scratches in a glass plate) enhance each other in certain directions and work against each other in other directions (here more or less entirely reducing each other's effect. In this manner, an interference pattern or diffraction pattern is produced. To obtain interference, the size of the wavelength has to be of the same order of magnitude as the distance between the interfering objects.

Since binding distances between atoms vary between about 1 Å and 3 Å and since X-ray wavelengths are on the order of 0.05–100 Å, there are many X-ray wavelengths that are of the same order as the binding distances between atoms. This fact led von Laue to predict that a crystal, which is a regular repetition of molecules and atoms in space, ought to function as a three-dimensional interference lattice for X-ray radiation. This prediction was verified experimentally by von Laue and his colleagues Friedrich and Knipping when they obtained the first X-ray diffraction photograph from a crystal, that of zinc blende ZnS, in 1912.[1]

The atoms (or, more correctly, electrons) in the crystals produce the diffraction of X-rays. Therefore, the next important development was to use the information in X-ray photographs to determine the structures of compounds built up by the atoms. The pioneer of this work was a very young scientist, W. L. Bragg, collaborating in part with his father, W. H. Bragg.

W. Lawrence Bragg was born in 1890 in Adelaide, Australia, where he got his basic education. His father, W. H. Bragg, was professor in physics there until 1909, when he was appointed as professor in physics at the University of Leeds. Lawrence Bragg then started as an undergraduate in mathematics and physics at Cambridge and had just graduated when the work by von Laue and his colleagues was published.

What Lawrence Bragg did was to calculate the intensities of all reflections (interference points) in an X-ray diffraction photograph and to compare these values with the observed intensities. The short description below will suggest how these calculations could be performed.

Usually a crystal contains several atoms of different kinds in the smallest volume (the unit cell), which is translated in three dimensions to build up the crystal. If there are n atoms in the unit cell, we can describe the crystal as being constructed by n congruent three-dimensional lattices inserted into one another. When X-ray radiation strikes the crystal, we get diffraction from all lattices in the same directions, but since the lattices are shifted relative to each other, the diffracted waves coming from the different atoms in the unit cell are not in phase with one another. If the atom n has the coordinates x_n, y_n, z_n (values in the range 0–1), the phase difference (the phase angle) for the nth lattice relative to the origin is:

$$\alpha_n = 2\pi(hx_n + ky_n + lz_n)$$

for the diffraction order h,k,l (which are the coordinates in diffraction space).

Thus, when calculating the sum of contributions from different atoms to every diffraction order (reflection), we have to sum up the waves with different starting points and different amplitudes f_n (proportional to the number of electrons in atom n). The result of this summation is called the structure factor $\mathbf{F(hkl)}$ for reflection h,k,l, and its expression below can easily be deduced:

$$\mathbf{F(hkl)} = \sum_n f_n \cdot e^{i2\pi(hx_n+ky_n+lz_n)} = \sum_n f_n e^{i\alpha_n}$$

The summation can be shown graphically to be a sum in the complex plane, and this is viewed for $n=3$ in Fig. 2.1b of Chapter 2 of this work. The heavy arrow in Fig. 2.1b is the structure factor, which represents the scattering for that reflection from the entire $n=3$ structure. Figure 2.1c shows the angle between the structure factor and the horizontal or real axis in the complex plane. This angle $\alpha(hkl)$ is a key parameter in structure determination by X-ray crystallography, as we will see further on.

When the atoms are correctly placed in the crystal unit cell, the calculated value of the structure factor $|\mathbf{F}(\mathbf{hkl})|_{calc}$ should compare well with the observed value $|\mathbf{F}(\mathbf{hkl})|_{obs}$, which can be obtained from the formula $|\mathbf{F}(\mathbf{hkl})|_{obs}=k\sqrt{I}$, where k is a constant and I is the intensity of the diffracted (or scattered) X-ray beam for reflection *hkl*.

Lawrence Bragg calculated structure factors and compared these with the corresponding observed values for several compounds such as sodium chloride (NaCl), potassium chloride (KCl), fluorspar (CaF), and calcite ($CaCO_3$).[2] From the starting positions of the atoms (based mainly on packing considerations), the atoms were moved until the calculated and observed structure factors were in the best possible agreement. This "trial-and-error" method for structure determination worked because the compounds were simple and the calculated structure factors were very sensitive to the positions of the atoms.

In 1915, W. L. and W. H. Bragg shared the Nobel Prize in Physics—W. L. Bragg (Fig. 1.1), then only 25 years old, for his work on diffraction and crystal structures, and his father, W. H. Bragg, for studies on the origin and properties of X-rays.

Lawrence Bragg continued to work on the structure of minerals, but the problems increased when the compounds became larger. In his work on the silicate mineral diopside $CaMg(SiO_3)_2$, he had great difficulties finding the positions, especially of the light oxygen atoms. He finally found a solution,[3] but asked himself an important question: "Is there an easier way?" Yes there was, and help came from work performed by the mathematician Fourier, who in the 1850s had already shown that a reasonably behaving periodic function could be expressed as a series of exponential terms. If such a function was combined with the expression for the structure factor given above, it was possible to obtain values for the electron density (a well-behaving function) at any point *x,y,z* in the crystal unit cell. Below, I show the expression deduced for the electron density, mainly to point out to the reader the two important parameters needed: the observed structure factor and the phase angle:

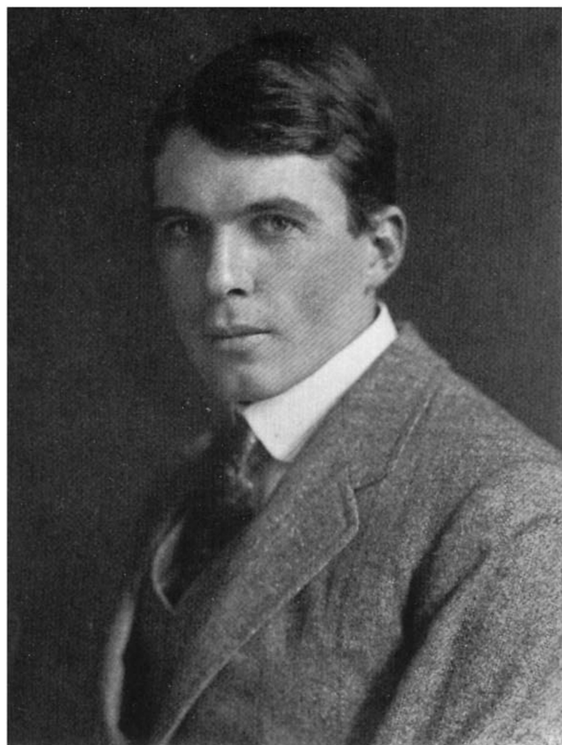


Fig. 1.1. W. Lawrence Bragg photographed in 1913 when he was 23 years old. Reprinted with permission from The Royal Society, London.

$$\rho(xyz) = 1/V \cdot \sum\sum_{h,k,l=-\infty}^{+\infty}\sum |\mathbf{F}(\mathbf{hkl})| e^{i\alpha(hkl)} \cdot e^{-i2\pi(hx+ky+lz)}$$

where V is the volume of the unit cell; $|\mathbf{F}(\mathbf{hkl})|$ is the observed structure factor; *h,k,l* are the diffraction orders; and $\alpha(hkl)$ is just the phase angle as described above and shown in Fig. 2.1c of Chapter 2.

If the crystal does not contain any atom that scatters in an anomalous way (and this is valid in most cases), then the expression above can be formulated as follows:

$$\rho(xyz) = 2/V \cdot \sum_{h=0}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} |\mathbf{F}(\mathbf{hkl})| \cdot \cos\left[2\pi(hx+ky+lz) - \alpha_{(hkl)}\right]$$

where it is somewhat easier to recognize $\alpha(hkl)$ as an angle in the expression. The electron density at each single point (*x,y,z*) in the unit cell is thus a summation of contributions from all the diffraction orders (*h,k,l*) in the diffraction photograph.

The American physicist R. J. Havighurst had, in 1927, used this expression to calculate the electron density distribution in a crystal of sodium chloride along cubic edges and cube diagonals.[4] Bragg extended the use of this method to two dimensions for the structure determination of diopside and obtained a clear solution for the position of oxygen atoms.[5]

Now one absolutely necessary piece was built into the ground for determining the structures of larger molecules, especially of molecules as large as proteins. However, two other important pieces remained to be found: the diffraction data from protein crystals and the phase angles, $|\mathbf{F}(\mathbf{hkl})|_{obs}$ and $\alpha(hkl)$, respectively, in the above expression for the electron density.

The Continued Road towards Protein Structures

W. L. Bragg had begun his very successful X-ray analysis of crystal structures in Cambridge. Therefore, it was logical that a post as lecturer in structural crystallography was created there, in the Department of Mineralogy, in 1926. The person who was appointed to this position was a young and talented scientist, J. D. Bernal. After studies in mathematics, mineralogy, and geology in Cambridge, he began research in 1923 in the laboratory of W. H. Bragg

(the father), who now had moved to the Royal Institution in London. Bernal became fascinated with crystallography, but his great interest was to use these X-ray techniques on biologically interesting molecules. Thus, in the Department of Mineralogy, he studied some minerals, but his main interest was to investigate molecules of biological importance, even some proteins. How reasonable was that, when so far it had not even been possible to obtain any diffraction data from protein crystals?

However, Bernal was a visionary and optimist. Furthermore, laboratory facilities were most likely improved when crystallography was transferred to the famous Cavendish Laboratory in 1931. Then in 1934, Bernal's group happened to obtain some very nice crystals of the protein pepsin. They were brought from Prof. Svedberg's laboratory at Uppsala University, Sweden (my home university). Svedberg had developed the ultracentrifuge, which was an excellent instrument for separation of macromolecules. A visiting scientist from Oxford came to Uppsala with a sample containing pepsin molecules in order to prepare a pure sample of the protein. When the separation had been performed, he left the pure pepsin solution in a test tube, where very nice crystals grew after some time.

A friend of Bernal who shortly afterwards visited Svedberg's laboratory and saw the crystals was sure that Bernal would love to look at them, and he got some of them to bring to Cambridge. When Bernal exposed these crystals to X-rays, they gave no reflections, only a vague blur—the same negative result that other people had obtained using protein crystals. However, Bernal did not give up. He examined the crystals under a microscope and discovered that when they were removed from their mother liquor, they deteriorated, became white, and lost birefringence. Bernal realized that the crystals had to be kept in their mother liquor; therefore, he mounted a crystal wet in a thin-walled glass capillary. It remained clear and, when exposed to X-rays, produced an excellent diffraction pattern; the film was covered with a mass of reflections. A young visiting scientist from Oxford, Dorothy Crowfoot (later Hodgkin), helped Bernal with the characterization of X-ray films by determining unit cell dimensions, space groups, etc. In May 1934, they published in *Nature* the first X-ray photograph from a protein crystal.[6] When this work was finished, Dorothy Crowfoot returned to Oxford, where she made an outstanding contribution to structure biology by solving the X-ray crystal structures of cholesterol, penicillin, vitamin B_{12}, and, finally, insulin. Thus, together with her colleagues in Cambridge (see Chapters 2 and 3), she became one of the founders of protein crystallography.

Now only one vital piece of information—the phase angles—remained missing in the process of making protein structure determinations possible. Finding a solution to the phase-angle problem in protein crystallography required almost another 20 years.

Perutz and Kendrew, and Their Work to Reach the Goal

In 1936, Max Perutz, a young research student from Austria, joined Bernal's group. Perutz was born in Vienna in 1914 and studied chemistry at a university there. During a lecture in organic chemistry, Max heard about work on vitamins and enzymes going on in Cambridge (in the Department of Biochemistry), and he became certain that Cambridge was where he wanted to work for his doctorate. As it so happened, Max never started his research studies in biochemistry, but instead in Bernal's group at the Cavendish Laboratory. There, biologically important molecules were studied, attracting Max even though the method was crystallography about which Max knew nothing.

When Max started to work in Bernal's group, there was no useful biological specimen available in the laboratory. Therefore, he had to begin with work on some minerals, which was somewhat disappointing. However, Max was still happy to work in the laboratory with his visionary leader who always was optimistic about the power of X-ray methods. The biological specimen that Max so much wanted to work on was finally obtained by his own action. A relative who was a physical chemist recommended that he should begin structural studies on haemoglobin and also suggested somebody in Cambridge who might be able to supply him with crystals of haemoglobin. This worked out well, and Max got beautiful crystals of the protein that colleagues in the group helped him to mount and start X-ray studies on. Max determined the crystals' basic parameters, unit cell dimensions, and space groups. Furthermore, the high quality of the X-ray pattern clearly indicated that the structure of the protein molecules in the crystals was well defined, but how to proceed further was, of course, not obvious.

In 1937, Bernal accepted a position as professor at Birkbeck College in London, and the members of his group were invited to go with him there. However, Max decided to stay since he liked both Cambridge and the Cavendish Laboratory very much. At the same time, the old Cavendish Professor, Rutherford, died, and his successor happened to be W. Lawrence Bragg. Max had, of course, learned that Bragg was the inventor of X-ray crystallography, and he was both nervous and hopeful when he called on Bragg to describe what he worked on and to show the X-ray photographs from the haemoglobin crystals. Bragg became very glad and enthusiastic. He immediately saw the great possibility that the method he had invented could one day be used to solve important problems in biology and medicine. Shortly after this meeting, Bragg managed to obtain a grant from the Rockefeller Foundation, and Max was appointed as his research assistant. Max had really found the ideal supporter and mentor. Bragg was highly respected and knew extremely

well both the powers and the difficulties of X-ray analysis.

The work on haemoglobin was interrupted by the war. Max was first interned in Canada as an "enemy alien" (from Austria) and then was invited to join a secret war project that aimed at producing a floating airbase from ice, a work that never came to any practical application. (Max, an avid skier, had published scientific papers on the structures of ice.) At the end of the war, Max returned to Cambridge and, in 1946, was joined by John Kendrew.

Kendrew was born in Oxford in 1917. After basic studies there, he went on to Trinity College in Cambridge, where he graduated in Chemistry in 1939. He then started research in physical chemistry (reaction kinetics) but, after a few months, was called up for military service, where he, among other things, worked on radar. During military service in Southeast Asia, Kendrew happened to meet J. D. Bernal, who explained to him the great and challenging possibility of determining protein structures by means of X-ray analysis. Bernal's description of his research no doubt had a big influence on Kendrew, and this interest in protein crystallography was further amplified during a visit to California, where John met Linus Pauling. It was natural that John wanted to finish his Trinity scholarship in Cambridge, where he was accepted to work on haemoglobin together with Max. It was decided that John should not work on exactly the same form of haemoglobin as Max (the protein from horse), but instead compare the protein forms from adult and foetal sheep.

Thus, there was a good plan for the scientific work. However, the financial situation was more difficult. Both John and Max had grants and awards lasting only another 2 years. Bragg had tried to get a university lectureship for Max, but the decision on this request seemed to last extremely long. Max's remark on this was that he probably was considered a very odd scientist—a chemist in a Physics Department working on a biological problem. Today, we would, of course, say that this is an ideal situation for somebody working in "molecular biology."

The Medical Research Council Unit

The financial difficulties of Max and John were solved in an interesting way. John and Max had been given bench space in a parasitology institute in Cambridge, the Molteno Institute, headed by David Keilin. Max told Keilin about the difficult financial situation, and this opened the way towards a solution. Keilin was a good friend of Sir Edward Mellanby, the Executive Head of the Medical Research Council (MRC), and Keilin suggested that Bragg should discuss with Mellanby the possibility of getting governmental financial support from the MRC. After a meeting with Bragg, Mellanby presented a written suggestion to the Council and, in their meeting in October 1947, the establishment of the "Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems" in Cavendish Laboratory was decided. The grant was £2550 per year to support Max and John and two research assistants for 5 years. This was the start of a research laboratory that would grow into a very famous research institute. The interesting thing about this important decision was that the physicist Bragg managed to convince the MRC that this research could lead to important results of great biological and medical value.

The Final Solution: The Phase-Angle Problem Solved

After Kendrew had finished his doctorate, he decided to change project from haemoglobin to the four-times-smaller protein myoglobin. The two molecules were closely related. Haemoglobin, with its four protein chains and haem groups, has the role of transporting oxygen to cells, while the single-chain myoglobin, with only one haem group, stored oxygen in muscles (serving as a reserve for oxygen). Kendrew chose this new project most probably because it would be easier to determine the structure of this smaller molecule and, with their functions being so closely related, there naturally were good possibilities for interesting comparisons of the two structures.

It was, of course, a great help that the financial situation was secure, especially since the initial years turned out to be scientifically difficult. Of the many methods used in trying to get structural results, probably the best was to vary the salt concentration in haemoglobin crystals. This led to some information on phase angles in case these were 0° or 180°. That made it possible to calculate a projection of the haemoglobin molecule, which, however, was impossible to interpret.

Then in 1953, Max started to work along a line that had been discussed and even used many years earlier. The method was to incorporate a heavy atom into the molecules, crystallize the modified molecules, and thus obtain somewhat changed X-ray reflection intensities. This method had been used successfully on smaller molecules (e.g., by M. Robertson in Glasgow, where Michael Rossmann, the author of Chapter 3 of this paper, did his PhD work). However, nobody thought that it should at all be possible to apply this method to such huge molecules as proteins. Even a heavy atom with, for example, 80 electrons was not expected to produce a measurable change in the diffraction pattern from a molecule containing a thousand or more lighter atoms. However, Max did not give up. Instead, he made a very important experiment. He measured on haemoglobin crystals the so-called absolute intensities (i.e., the intensities of diffracted X-rays relative to the intensities of the incident beam). The result was that these absolute intensities were surprisingly weak. The scatterings from the many light atoms apparently tended to cancel each other to a great extent (see Fig. 2.1c in Chapter 2 of this work).

Furthermore, Max obtained interesting information from a biochemist at Harvard who investigated normal and sickle cell haemoglobin to look for differences. During these studies, he could see that it was possible to attach mercury atoms to sulphydryl groups on haemoglobin without any change in the oxygen uptake. This indicated that the structure of the molecules most likely had not been altered, a good sign for Max. Vernon Ingram, a good chemist in the group, now prepared a mercury compound that he attached to the sulphydryl groups of haemoglobin. Max succeeded in crystallizing this haemoglobin "derivative" and exposed one of these crystals to X-rays. When he developed the photograph, he could, even when the film was still wet, see clear changes in reflection intensities compared to the intensities from crystals of the unmodified protein. He immediately rushed to Bragg's office and brought him to the dark room. They looked at the two films and at each other, and, almost without saying anything, they both understood that this was **the solution**.

Fifteen years earlier, they had together looked at one film. Now they had two. The present "derivative" film itself was not enough. Mercury atoms did not dominate but merely altered the intensities on the "unmodified" film, producing small but measurable differences that could be used for the determination of phase angles. In fact, two different derivatives were needed (even theoretically) to calculate unique phase angles. These derivatives were quite possible to prepare even if some effort was needed.

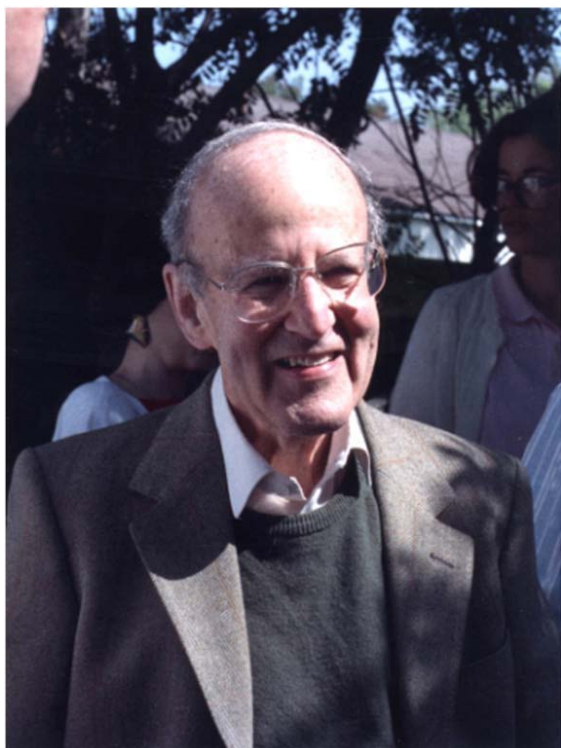


Fig. 1.2. Max Perutz photographed by Dan Anderson in 1991 during a visit to the University of California at Los Angeles.

Fig. 1.3. John Kendrew photographed *ca* 1962, the year he shared the Nobel Prize in Chemistry with Max Perutz.

The way to determine the phase angles by means of two or more derivatives is well explained in the text and in Figs. 2.2 and 2.3 (Chapter 2) of this work. Since the crystals of the derivatives and the unmodified protein must be isomorphous with one another, the method was called *isomorphous replacement*.

Max and his colleagues published this important result in 1954 in the *Proceedings of the Royal Society*,[7] and now Max (Fig. 1.2) and John (Fig. 1.3) began to "see the light at the end of the tunnel."

The First Picture of a Protein Molecule: Myoglobin at 6 Å Resolution

The way towards a first low-resolution structure was not surprisingly easier and faster for the smaller myoglobin. The type of myoglobin specimen used by Kendrew, from sperm whale, could be obtained in large quantities and gave large and robust crystals. A low-resolution analysis was tried first, since this involved a reasonable amount of data to be handled in this first step. The resolution of 6 Å was chosen because, at this level, it should be possible to recognize polypeptide chains having a compact configuration such as a helix. (In 1951, Pauling *et al.*[8] had predicted the existence of α-helix configuration in protein molecules, and Max had later that year produced experimental support for this.)

Preparation of heavy-atom derivatives for myoglobin, which had no free sulphydryl groups, was

somewhat more difficult than that for haemoglobin. The two chemists Gerhard Bodo and Howard Dintzis in John's group prepared and tested a large number of derivatives and succeeded in finally obtaining five compounds whose crystals gave reasonable intensity changes compared to crystals from the unmodified protein. The positions of the heavy atoms were determined by using a function derived by Patterson.[9] A "Patterson map," for which knowledge of phase angles was not needed, gave information on the relative positions of the heavy atoms in the different derivatives, but not on the atom positions themselves. For myoglobin, an acceptable but nonoptimal positioning of the heavy atoms was obtained in this way. (During the work on haemoglobin to 5.5 Å resolution, Michael Rossmann, the author of Chapter 3 of this work, derived a perfect solution to this problem.) Now all information for determination of the phase angles was available, and this was performed graphically with hand-drawn phase-circle diagrams as shown and explained in Fig. 2.4 of Chapter 2. Finally, the electron density was calculated using the 400 measured intensity data from unmodified crystals and the corresponding phase angles. These calculations, which took a total of 70 min, were performed on the Cambridge University EDSAC I computer.

From the calculated electron density maps, a model of the myoglobin molecule to 6 Å resolution was constructed (Fig. 1.4). The model contained a number of dense rod-like features that had the dimensions of α-helices and made up the bulk of the polypeptide chain. The rods were joined by corners where the electron density, in most cases, was somewhat lower or even broken. The chain was folded in an irregular manner, certainly much more complicated than had been anticipated. In a pocket in the molecule (having approximate dimensions of 45 Å×35 Å×25 Å), there was a dense flattened disc, which presumably was the position of the heme group with its central iron atom.

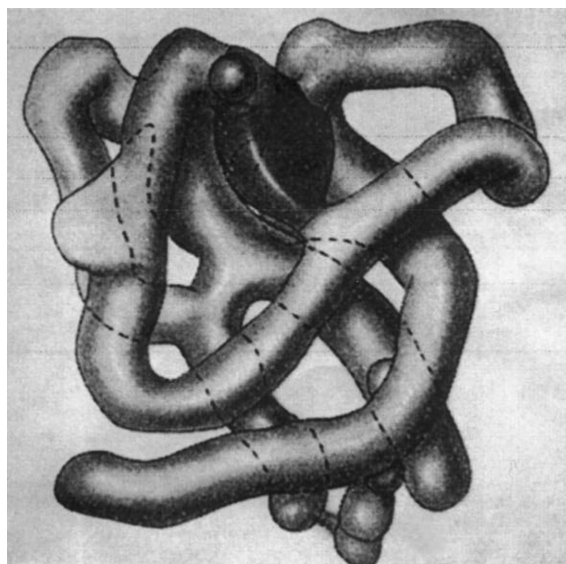


Fig. 1.4. The molecular model of myoglobin at 6 Å resolution. The model contains a number of dense rod-like features that have the dimensions of α-helices and make up the bulk of the polypeptide chain. The chain is folded in an irregular manner, more complicated than had been anticipated. In a pocket in the molecule is a dense flattened disc, presumably the position of the heme group. Reprinted with permission from The Royal Society, London.

This result was ready by the beginning of 1957 and was published in *Nature* in the following year.[10] It was the very first picture of a protein molecule, and it was followed by the wonderful structure determinations of myoglobin to 2.0 Å and of haemoglobin to 5.5 Å, performed in 1959 and described in Chapters 2 and 3, respectively.

Working on a Protein Molecule in Cambridge: A Privilege and a Great Experience

I started research in the Department of General and Inorganic Chemistry at Uppsala University in 1955. Since the professor at the department, Gunnar Hägg, was among the three or four people who introduced X-ray crystallography in Sweden, this was the key method at the institute. My supervisor Ingvar Lindqvist, who was interested in metal complexes, gave me the compound $AgNH_4(SCN)_2$ as my first structural problem. I thought this compound was not so exciting, but a nice experience during this work was to use a machine for the calculation of Fourier series, recently constructed by Gunnar Hägg himself. Our building was part of a big complex that included all branches of chemistry. When we went for coffee in the morning, we passed by the Physical Chemistry Department (from where Bernal's pepsin crystals, produced with the help of Svedberg's ultracentrifuge, came in 1934) and almost reached the Biochemistry Department, where Arne Tiselius, the man who developed the electrophoresis method, was the professor. Coffee times were therefore excellent occasions to learn about problems in other branches of chemistry and resulted in my second and third structural problems being related to biochemistry. These were the metal complexes $Cu(glycylglycine)_2$ and $Zn(imidazole)_2$, with the latter being a model compound for the active center of the enzyme carbonic anhydrase, which was studied in the Biochemistry Department. This was something quite different and, during work on these structures, I also attended a course in biochemistry in Tiselius' department.

Thus, it was very exciting when Gunnar Hägg in May 1957 received a letter from Max Perutz, asking if somebody in Hägg's department might be interested in coming to Cambridge and taking part in continued work on myoglobin or haemoglobin. I was, of course, immediately very interested, but there was a small problem. I had not yet finished my PhD work. On the other hand, such an opportunity would probably not come again, so I was given the right to make a break in my PhD studies. Hägg

wrote back to Max and explained my background and my big interest in joining Max and John. Thus, it was decided that I should go to Cambridge, but since our first child was recently born, it was first in June 1958 that I traveled to England together with my wife Karin and my little son Peter.

When I arrived, Max asked me which of the two problems I wanted to work on. Both were, of course, very interesting. Taking part in work on the first high-resolution structure of a protein molecule seemed an irresistible challenge, so I started to work in John's group and became a close colleague of Dick Dickerson, who already had been working for some months on the 2Å structure of myoglobin. The laboratory, "The Hut," was much smaller than the Department of Chemistry from which I had come, but the atmosphere was wonderful. I came to share a small office with Dick; Michael Rossmann, who arrived a few months later and started to work on haemoglobin; and Larry Steinrauf, who was a postdoctoral fellow from Linus Pauling's laboratory. It was really a great privilege to share an office with these wonderful colleagues. All the people in the laboratory—such as Mary Pinkerton, who was the head of the staff measuring the large amount of X-ray data; Ann Cullis, who was assistant to Max; Hilary Muirhead, who was a research student in the haemoglobin group; and many others—were very kind and helpful. The leaders Max and John were excellent examples for us, as were Francis Crick and Sidney Brenner, who shared an office across the corridor from our office. Francis was the great leader during the 11 o'clock coffee break, when all possible subjects were discussed. Figure 1.5 is a group picture from the autumn of 1958, with some of the people working in The Hut gathered outside the entrance door of the small laboratory building.

To me, three things were special with this laboratory and with the people working there:

(1) Everybody was always keen to solve all upcoming problems at any stage of the work, and this was, of course, very natural with leading scientists such as Max and John.
(2) The equipment was top class for that time. X-ray generators with rotating anodes (constructed by the laboratory engineer Tony Broad); a microdensitometer for measuring the vast number of X-ray reflections; and the EDSAC II computer at the nearby Mathematics Laboratory (probably the fastest computer in the world then) were all facilities

Fig. 1.5. Group picture of some of the people working in The Hut (autumn of 1958). From left to right, back row: Larry Steinrauf, Richard Dickerson, Hillary Muirhead, Michael Rossmann, Philip ___?___ (face obscured), Ann Cullis, Bror Strandberg, Wibeke, and an unidentified person. Front row: Leslie Barnett, Mary Pinkerton, and Max Perutz.

that strongly contributed to making the work possible.

(3) Finally, the many distinguished visitors certainly stimulated us and made us feel that we were working on very important problems.

My biggest memory from this time was no doubt a sunny morning in the beginning of August 1959, after a whole night of calculating the electron density map of myoglobin. In Chapter 2, Dick has given a very nice description of that event.

After taking part in the model building of myoglobin, I went back to Uppsala in December 1959 and started about a year later to work on the three-dimensional structure of the enzyme carbonic anhydrase, which was solved in 1969.

References

1. Friedrich, W., Knipping, P. & Laue, M. (1912). Interferenz-Erscheinungen bei Röntgenstrahlen. *Sitzungsber. Kgl. Bayrischen Akad. Wiss.* 303–322.
2. Bragg, W. L. (1913). The structure of some crystals as indicated by their diffraction of X-rays. *Proc. R. Soc. London Ser. A*, **89**, 248–277.
3. Bragg, W. L. & Warren, B. (1928). The structure of diopside, $CaMg(SiO_3)_2$. *Z. Kristallogr.* **69**, 168–193.
4. Havighurst, R. J. (1927). Electron distribution in the atoms of crystals. Sodium chloride, and lithium, sodium and calcium fluoride. *Phys. Rev.* **29**, 1.
5. Bragg, W. L. (1929). The determinations of parameters in crystal structures by means of Fourier series. *Proc. R. Soc. London Ser. A*, **123**, 537–559.
6. Bernal, J. D. & Crowfoot, D. (1934). X-ray photographs of crystalline pepsin. *Nature*, **133**, 794–795.
7. Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). Structure of haemoglobin: IV. Sign determination by the isomorphous replacement method. *Proc. R. Soc. London Ser. A*, **225**, 287–307.
8. Pauling, L., Corey, R. B. & Branson, H. R. (1951). The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl Acad. Sci. USA*, **37**, 205–211.
9. Patterson, A. L. (1935). A direct method for the determination of the components of interatomic distances in crystals. *Z. Kristallogr.* **90**, 517–542.
10. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature (London)*, **181**, 662–666.

Chapter 2: Myoglobin: A Whale of a Structure!

Richard E. Dickerson

Introduction

How is a crystal put together? The picture in many people's minds is of atoms or ions packed tightly together like tennis balls in a large cardboard box. Some crystals have only one type of ball, such as carbon atoms in diamond. Other crystals have two different kinds, such as sodium and chlorine ions in NaCl. But crystalline *proteins* or crystalline *DNA*? Ridiculous!

In 1953, I joined the laboratory of Bill Lipscomb at the University of Minnesota to work on crystal structures as a step toward a doctorate in physical chemistry. When I arrived, on the "Colonel's" (Lipscomb was from Kentucky) desk was a weird skeleton of linked triangles forming a fragment of an icosahedron, with one half open like a football helmet. This was the structure of a boron hydride, the Colonel told me, with boron atoms at the corners of the triangles and hydrogen atoms attached in unexpected places. I remember being astounded that such an awkward and ungainly entity could ever be induced to form a crystal. Little did I realize that in Cambridge, England, two colleagues named Perutz and Kendrew were, at that very moment, well on their way toward solving the crystal structures of molecules of vastly greater complexity than mere boron hydrides.

Things were much simpler in those distant days. I earned my doctorate with the aid of 375 X-ray reflections by solving the crystal structure of B_9H_{15}.[1–3] The project had been begun by a visiting professor from Leeds University, Peter Wheatley, and I inherited it when he returned to England. Data were collected on film with a precession camera, spot intensity was gauged by visual comparison with a standard scale, and calculations were carried out on a desk calculator with the help of an IBM 407 punched-card machine, which could perform simple arithmetic operations on data entered via punched cards. IBM 407 was "programmed" by connecting wires on a plugboard, rather like a primitive telephone switchboard. For the production of electron density maps, punched-card data were sent to Ray Pepinsky at Penn State University, who had developed the X-Ray Analogue Computer or X-RAC specifically for crystal structure analysis†.[3]

† Ray Pepinsky, by all accounts, was a tyrant in the laboratory and a rugged taskmaster. A postdoctoral fellow from the Netherlands whom I knew in Lipscomb's laboratory had moved from Penn State to Minneapolis after a nasty row in which Pepinsky had threatened to have his student visa revoked and to send him back home again. A few years later, as a new faculty member at Caltech, I mentioned Ray Pepinsky to a faculty colleague, Eddie Hughes. Hughes was the most courteous, gentlemanly, and mild-spoken academic whom I can recall. I never heard Eddie say a bad or spiteful word about anyone, no matter what the provocation. I told Eddie that I had heard that Pepinsky had just moved from Penn State to Nova University in Florida, and that Penn State had allowed him to take with him the vast X-RAC facility that he had built up. When I moved from the University of Illinois to Caltech in 1963, I brought almost no equipment with me. "How did Pepinsky manage to get Penn State's permission to walk off with many thousands of dollars of computer equipment?", I asked. Eddie surprised me with a calm reply, "Well, I suppose they figured that was the price they had to pay."

Things improved vastly when the Colonel persuaded Honeywell Corporation in nearby St. Paul to let us use their Remington Rand UNIVAC 1103A computer after office hours. The University of Minnesota had no digital computers of any kind at that time. They had considered purchasing an IBM 650 magnetic drum computer but decided that there was too little demand for its services to justify spending US$200,000–400,000 on such a machine. The UNIVAC 1103A was a "huge" computer, with 1024 thirty-six-bit words of electronic memory and 16,384 words on a magnetic drum. Compare that to 2 GB or 2,000,000,000 bytes of storage on a removable flash drive memory stick for a modern laptop computer. One byte in modern usage corresponds roughly to 8 bits, so today's 2-GB laptop memory stick has more storage capacity than 25,000 UNIVAC 1103A computers! As I was finishing my degree, a young postdoctoral fellow from Glasgow, Michael Rossmann, showed up on the scene. He and Audrey, and my wife Lola and I became good friends, but when I left Minneapolis in February 1957, I never dreamed that, in only a couple of years, we would be working together again, a third of the way around the world.

I did not go straight to Cambridge. In fact, the Colonel told me that Oxford and Cambridge were old hat, and that I would be better off at Leeds University, which offered such ground-breaking crystallographers as Peter Wheatley, Durward Cruickshank, and George Jeffrey. So one cold day in February of 1957, Lola and I disembarked at Southampton and took a train north to Leeds, where Peter Wheatley met us at the station. But things were not what they might have been. Jeffrey had finally moved permanently to the University of Pittsburgh where, over the years, he built up a large and productive Graduate Crystallography Department, the only one of its kind in the United States. Durward Cruickshank had moved to Manchester, which had a digital computer (the Ferranti Mark I) that Leeds University lacked. (When Leeds obtained its first computer a few years later, the university housed it in a former Methodist church building adjacent to the campus. In a bizarre fit of inspiration, its official name was the "*L*eeds *U*niversity *C*omputer and *I*ntegrator, *Fer*ranti" or LUCIFER.)

At Leeds, I collected precession data on a now mercifully forgotten dimethylsulfoximine, had a Patterson function calculated on the Manchester computer, and then solved the three projections of the structure using Beevers–Lipson strips and a desk calculator. But during the summer of 1957, a crisis arose. As I have related before,[4,5] Wheatley announced that he could no longer support a wife and two children on his annual academic salary of £900 and was leaving Leeds to head up a new Monsanto crystallographic laboratory in Zurich.

By a stroke of extraordinarily good fortune, at that very moment, Max Perutz and John Kendrew in Cambridge were advertising for one or more postdoctoral fellows to work with them on the high-resolution crystal structure of myoglobin and the low-resolution structure of the larger haemoglobin. Perhaps in a fit of conscience, Bill Lipscomb proposed my name and provided a very positive recommendation. Peter Wheatley also suggested me, and the result was that I received a phone call from Max, inviting me down to spend a day talking and looking around the laboratory. During the day, we all visited the Royal Institution in London to talk with David Phillips, who was collaborating with the Cambridge team. It was there that I first met a wonderful person, Sir Lawrence Bragg, the previous director of the Cavendish Laboratory at Cambridge and now the director of the Royal Institution. It was Bragg who, together with Max Perutz, had built the Medical Research Council (MRC) Laboratory in Cambridge and attracted to it people such as John Kendrew, Francis Crick, and Jim Watson. In 1915, the 25-year-old Bragg had shared the Nobel Prize in Physics with his father, W. H. Bragg, for essentially inventing X-ray crystal structure analysis. I looked at the man who had created Bragg's Law and suppressed an urge to exclaim, "I assumed you were dead!"

There was a meeting of the minds, and I accepted their invitation to join the MRC group at Cambridge in the fall of 1957. I chose to work with John Kendrew in part because the idea of getting a structure at atomic or near-atomic resolution appealed to a small-molecule crystallographer.

Getting Started

The MRC Laboratory of Molecular Biology occupied a long, one-story temporary building that had been built as a metallurgical laboratory during World War II. It was known irreverently as "The Hut" and sat in a courtyard of the Cavendish Physics Laboratory. The Hut contained a wet laboratory and offices, but the X-ray machines themselves were mercifully housed in the basement of the adjacent and more sturdily constructed Cavendish Laboratory. The wet laboratory was at one end of The Hut. One side of the central corridor had offices for Max Perutz and for Francis Crick and Sidney Brenner. On the other side were the offices for John Kendrew and for postdoctoral fellows. When I arrived in the fall of 1957, I shared this office with Erwin Alver and Roger Hart. Alver, from the University of Oslo, was an inorganic crystallographer like myself. Roger Hart was somewhat more qualified: although originally from the United States, he had first worked with Rosalind Franklin, and then with Aaron Klug after her death.

Alver only stayed for 1 year and returned to Oslo in the summer of 1958, shortly after the arrival of a third colleague, Bror Strandberg, from Uppsala. Alver worked with Kendrew like myself, and Hart worked with Perutz. Prior to Bror's arrival, Alver and I joked that we were the only two people in the MRC Laboratory who had ever completely solved a crystal structure, although mine had only nine boron atoms and Alver's was comparably insigni-

ficant. Proteins constituted an entirely new universe for both of us.

Kendrew had initially studied horse myoglobin, but soon reasoned correctly that the oxygen-storing molecule in tissues should be especially plentiful in aquatic mammals that dived and remained under water for appreciable time periods. Finback-whale and sperm-whale myoglobins were isolated and tried, and the latter was selected because its myoglobin concentration was so high that the meat was almost black.[6] Sperm-whale meat was shipped from Peru and stored frozen in large garbage cans in a deep-freeze room of the Molteno Institute, down the road a piece from the MRC Laboratory. Myoglobin was isolated and purified by laboratory aides, and crystals were grown in 2- or 4-ml screw-top vials. Crystals grew so large that they had to be cut down to size with a scalpel blade before being inserted and sealed in glass capillaries for X-ray photography.

Roger Hart was unexpectedly an important source of inspiration to me in these early days. John Kendrew carefully demonstrated to me every step of the crystal-mounting process: choosing a good crystal, cutting it down to size, inserting it into the capillary and drying away excess solvent from around it, adding short plugs of solvent to both ends of the capillary to maintain a moist atmosphere, sealing both ends of the capillary with wax, and mounting it on a goniometer head for placement on a precession camera. I found this procedure incredibly difficult and became rather intimidated by the entire process. I began to wonder whether I would ever master the trick. But Roger gave me pause. His hands were twisted and crippled by polio, yet he was taking survey picture after survey picture of derivatives of haemoglobin for Max Perutz. I said to myself, "By God, if Roger can mount crystals with those hands, I most certainly should be able to." And so it was, after hours of trial and error. Only later did I learn that Roger did not actually mount the crystals that he was photographing. Perutz came in after hours and mounted the next day's supply of haemoglobin crystals!

Bror Strandberg arrived from Uppsala in June 1958, and the project really got going. The third member of our myoglobin team was David Phillips in London. He was to collect data for one of the heavy-atom derivatives at the Royal Institution. John's original plan had been to follow his 6-Å low-resolution analysis with a higher-resolution analysis at 2.5 Å. The 6-Å analysis required ca 400 independent X-ray reflections from parent myoglobin crystals and from each of the five heavy-atom derivatives. Increasing the resolution from 6 Å to 2.5 Å would necessitate collection of $(6/2.5)^3 \times 400 = 5500$ reflections from each crystal type. But in planning discussions, someone pointed out that, at 2.5 Å, we were not using the full power of the precession camera. The greatest resolution that we could obtain on our 5 in. × 5 in. photographic plates was 2.0 Å, for which the reflections would extend all the way out to the edges of the plate. Why not go for broke and collect 2.0-Å data instead? We decided to do so and to collect and measure ca $(6/2)^3 \times 400 = 10{,}800$ reflections per data set. This would give us twice the information as the originally planned 2.5-Å analysis. (In the end, after allowing for space group extinct reflections and other problems, a dependable set of 9600 reflections was used in the 2-Å analysis.)

Data were collected from four heavy-atom derivatives: *para*-chloromercuribenzene sulphonate, mercury diamine ($HgAm_2$), gold chloride, and the double derivative *para*-chloromercuribenzene sulphonate/mercury. Each data set required 22 different precession camera photographs; for each photograph, a newly mounted myoglobin crystal was used to minimize radiation damage. In precession cameras, X-ray reflections or spots are arranged in a series of parallel lines. This meant that one could use a Joyce–Loebl microdensitomer to scan along each of these lines, yielding peaks that then could be measured with a millimeter scale. Max and John had a crew of technicians who were available to carry out this process, and we ended with a long list of (h,k,l) reflection indices and their corresponding peak intensities. These then were typed on punched paper tape for input to the Electronic Delay Storage Automatic Calculator (EDSAC) II computer in the nearby Cambridge Mathematical Laboratory.

EDSAC II was considered a major step up from the EDSAC I that Kendrew had used for low-resolution myoglobin. Kendrew and his colleague Bennett in 1952 were actually the first people to program a digital computer or "digital electronic calculating machine," the EDSAC I, to carry out a three-dimensional Fourier synthesis.[7] Bennett subsequently moved to the Ferranti Corporation in Manchester, the place that also lured Durward Cruikshank away from Leeds. This new EDSAC II computer had 2000 words of high-speed core storage and two magnetic tape drives. The core was far too small to permit shuffling and reordering of 10,000 X-ray reflections from each of five sets of data: a parent protein and four derivatives. Merging of data had to be accomplished by thumb-tacking strips of punched paper tape containing intensity data to a bulletin board, shuffling the strips manually to merge the data sets, and then running everything back into the computer and emitting one grand master tape. Figure 3.7 of Michael Rossmann's chapter shows Bror and me walking our data bulletin board back from the computing center. A good strong wind might have made a major change in our analysis.

Input to EDSAC II was performed by punched paper tape using standard British telegraph readers. Little metal fingers in the reader pushed up through the holes in the tape, reading the information in the pattern of holes. If an input tape was mispunched during preparation, it was ruined and could only be corrected by making a new tape. In desperation, the custom of patching the occasional mispunched hole with Scotch tape was developed. This worked fine, unless the tape reader fingers pushed off the patch, causing a misreading and sometimes a tape jam. The

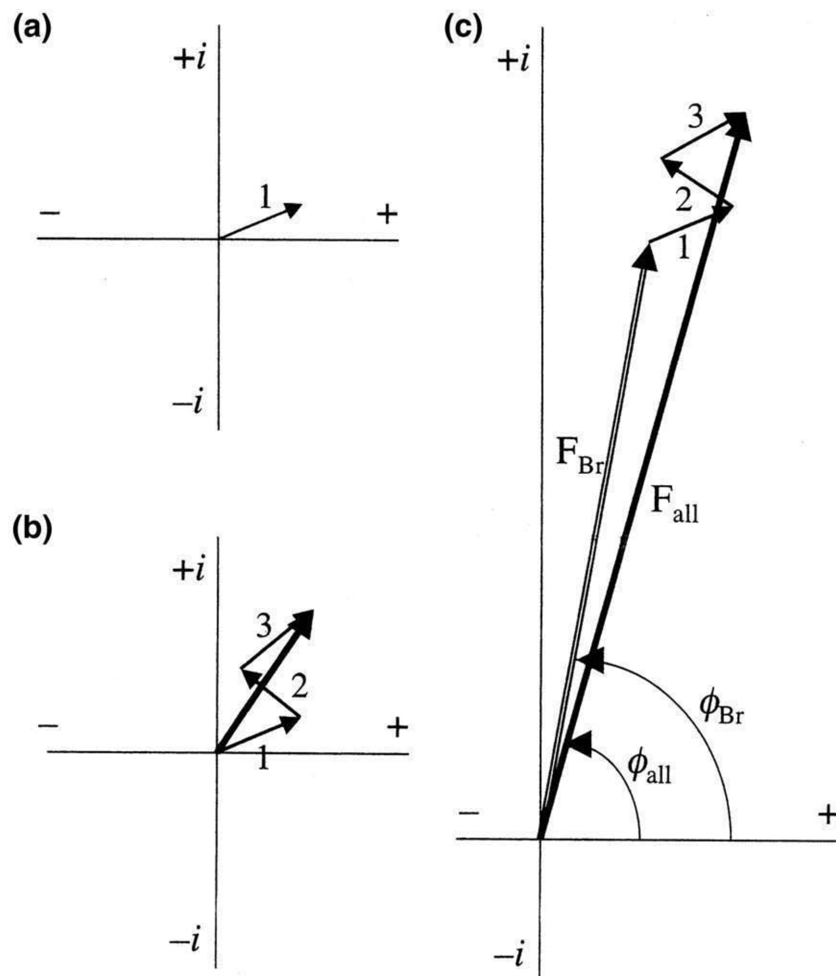


Fig. 2.1. For a given X-ray reflection, addition of waves from different atoms in one molecule. (a) A single atom X. (b) A three-atom X_3 molecule, with similar atoms in different locations. The heavy arrow represents scattering for that reflection from the entire molecule. (c) Result of tagging the three-atom X_3 molecule with a bromine atom. Because the heavy atom has so many electrons, it dominates scattering, and its phase is very similar to that for the entire X_3Br molecule.[5]

computing center came down on this with a hammer, threatening that any tape found to be patched in this manner would be confiscated. (I never recall this threat actually being carried out.)

The heart of EDSAC II was an array of long racks of vacuum tubes, with each rack being capable of being pulled out individually by its handle when one of its tubes blew. Wiring diagrams for the computer were in pencil in a large stack of circuit diagrams. When a change was made, the diagrams were erased and corrected. Whether permanent or even inked drawings of the computer existed, I was never aware of them.

There were three grades of access to EDSAC II: users, partially authorized users, and fully authorized users. Users could only have normal access to the computer during the day while it was manned by computer center staff. Partially authorized users were allowed to run the computer as late as they chose and to turn it off when finished by pulling a set of wall switches in a defined order. Fully authorized users were permitted to turn the computer on and off. I eventually qualified as a partially authorized user. I do not know if any member of the protein crystal structure group ever rose to the rank of a fully authorized user, but if so, it probably was Michael Rossmann. His programs for difference Patterson and difference Fourier refinements were so complex that they were sometimes used for the morning computer check, instead of the standard system-checking routines.

I Don't Recall the Name, But the Phase Is Familiar

We have not yet considered why five different sets of data, each of roughly 10,000 reflections, were

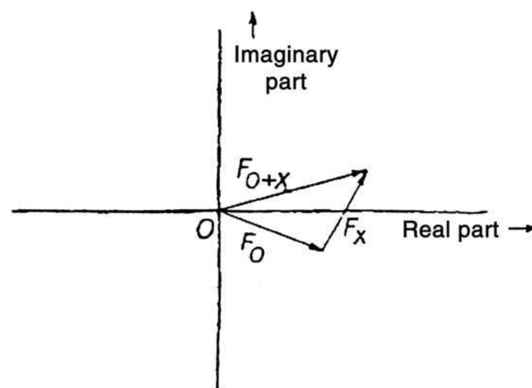


Fig. 2.2. For one X-ray reflection, addition of scattering vectors from a parent molecule $\mathbf{F_o}$ and an added heavy atom $\mathbf{F_x}$ to yield the observed experimental scattering from the heavy-atom derivative $\mathbf{F_{o+x}}$.[9] Reprinted with permission from the International Union of Crystallography (http://journals.iucr.org).

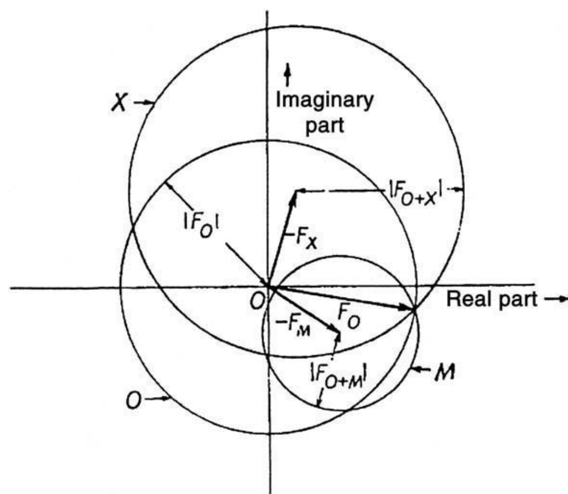


Fig. 2.3. Phase-circle diagram for one X-ray reflection, using experimental amplitudes from the parent molecule O and the two heavy-atom derivatives X and M. Each derivative phase circle crosses the parent circle O at two places. The intersection that is common to the two derivatives is the correct phase angle.[9] Reprinted with permission from the International Union of Crystallography (http://journals.iucr.org).

necessary. This brings us to the phase problem, which is the key to the structure analysis of macromolecules. One can generate an electron density map of the protein by a process of Fourier synthesis. Each of the thousands of reflections in the X-ray pattern contributes one wave to the image of the protein. The direction of each wave is determined by the position of its reflection in the X-ray pattern, and the wavelength of the wave is inversely proportional to the distance of the reflection from the center of the X-ray pattern. The amplitude of each wave is the square root of the measured intensity of the X-ray reflection.

So far, so good. This is nothing that a computer cannot handle. But one more—and very serious—complication remains. Each one of these thousands of waves has a phase—the fraction of a 360° cycle by which it is shifted compared to the other waves before they are all added together. To calculate the image of a protein, one must know both the amplitude and the phase of each of the 10,000 reflections. How does one find out the phase of each and every reflection in the X-ray pattern?

Each wave corresponding to one X-ray reflection actually is the sum of waves from all of the atoms in the scattering molecule. This sum, again, is a complex number. The amplitude of scattering from each atom depends on the number of electrons in the atom, and its phase depends on the position of the atom in the molecule. Figure 2.1a represents scattering from a single atom for one particular X-ray reflection. The length of the vector is the amplitude of the wave, and the angle of the vector from the horizontal or real axis is its phase. Figure 2.1b shows how scattering vectors from three atoms with similar numbers of electrons but different locations add together to produce an amplitude and a phase for the three-atom molecule as a whole.

One classic phase-solving approach for small molecules is the heavy-atom method. Tag the molecule with a heavy atom having so many electrons that it dominates scattering, as in Fig. 2.1c. First, locate the heavy atoms in the unit cell. (See Michael Rossmann's Chapter 3 for more on this issue.) Compute an electron density map using the phases of the heavy atom itself, which you can calculate if you know where the heavy atoms are located in the cell. If you are lucky, this map will show the heavy atoms plus at least a partial image of the rest of the molecule. Add more of these atoms as they become visible in the maps, and iterate until you have the picture of the complete molecule.

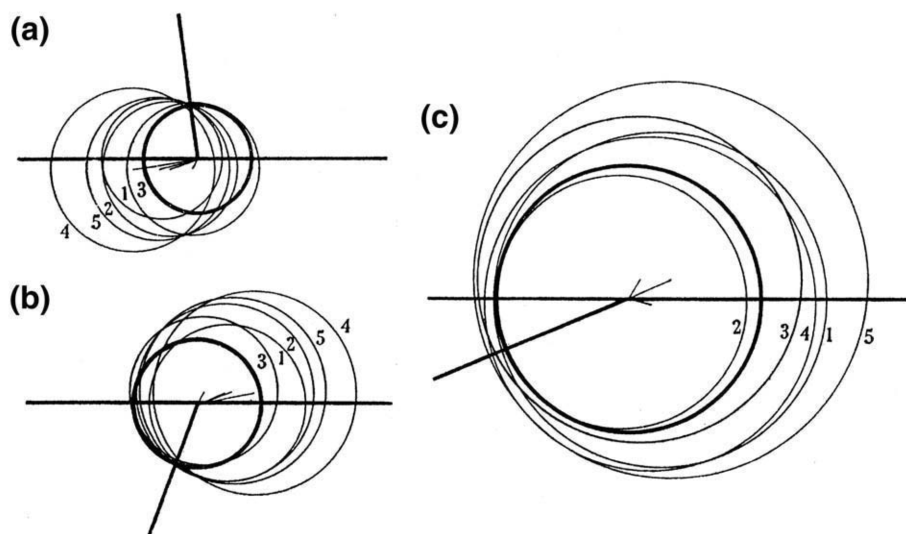


Fig. 2.4. Phase circles for three X-ray reflections from myoglobin at 6 Å resolution, using parent myoglobin and five different heavy-atom derivatives. (a) An excellent, unambiguous phase determination. (b) A somewhat less well-defined phase. (c) A poorly defined phase angle.[11]

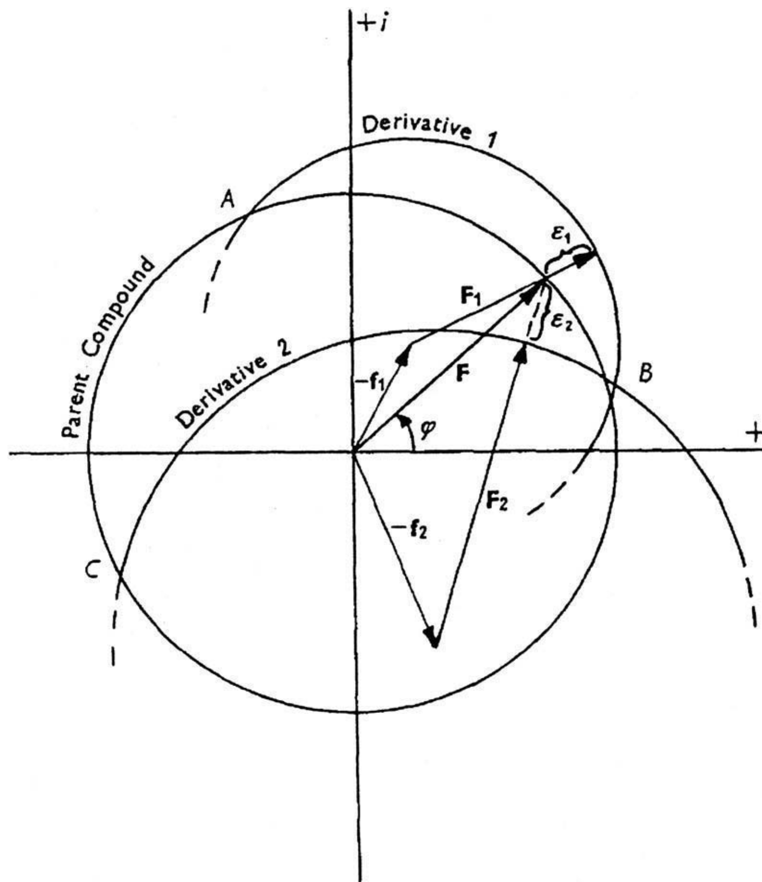


Fig. 2.5. Lack-of-closure errors in phase analysis. In this diagram, the amplitudes of the protein and its two heavy-atom derivatives are represented by **F**, $\mathbf{F}_1$, and $\mathbf{F}_2$. Scattering from heavy atoms alone is shown as vectors $\mathbf{f_1}$ and $\mathbf{f_2}$. For derivative 1, the requirement that $\mathbf{F_1} = \mathbf{F} + \mathbf{f_1}$ is satisfied at points A and B around the parent phase circle; for derivative 2, the requirement $\mathbf{F_2} = \mathbf{F} + \mathbf{f_2}$ is met at points B and C. Hence, the true phase angle is found at the shared point B. But for the general phase angle ϕ, the measured $|\mathbf{F}_1|$ from derivative 1 is too long by distance ε_1, and that for derivative 2 is too short by distance ε_2. These lack-of-closure errors can be used to calculate the probability of correctness of any angle. At the correct point B, this probability rises, in principle, to unity.[12] Reprinted with permission from the International Union of Crystallography (http://journals.iucr.org).

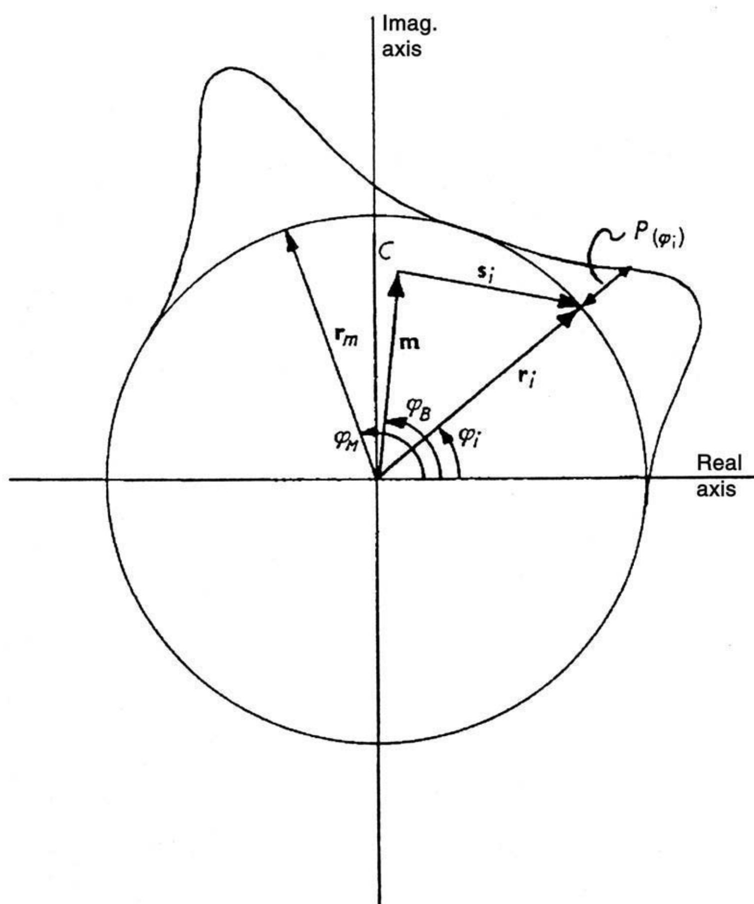


Fig. 2.6. Phase-angle probabilities plotted around a 360° phase circle of unit radius. This particular example exhibits two Gaussian-like peaks of probability, one roughly twice the other. Vector **m** marks the center of gravity of the probability curve.[12] Reprinted with permission from the International Union of Crystallography (http://journals.iucr.org).

But it was not at all clear whether this approach would work with proteins. Haemoglobin, with a molecular weight of 68,000, would have something on the order of 34,000 electrons around its atoms. How could one imagine that one or two mercury atoms of 80 electrons each could dominate the phasing? Indeed, could one expect to see any change in X-ray intensities at all when 80 or 160 electrons are added to a molecule that already contains 34,000? In 1954, Max Perutz provided the Rosetta Stone for protein crystal structure analysis when he demonstrated that, although heavy atoms did not dom-

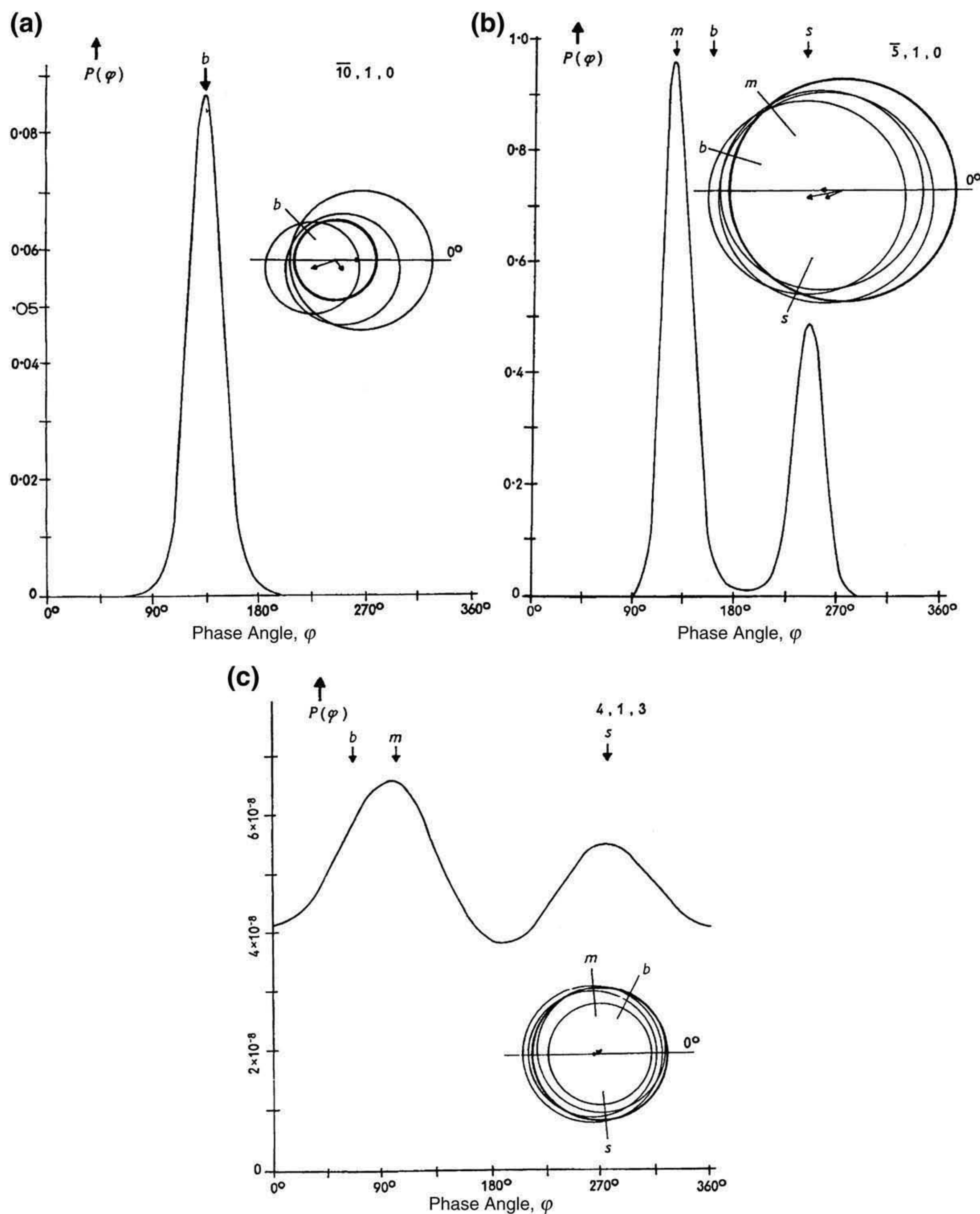


Fig. 2.7. Examples of phase determination for three X-ray reflections in the 2-Å phase analysis. The figures of merit are 0.96 (a), 0.59 (b), and 0.05 (c), which was the most poorly determined phase in the entire 2-Å myoglobin analysis. Position **m** in each diagram marks the most probable phase, and position **b** is the "best" phase using the centroid of the full probability circle. In the single-peak function (a), the most probable and best phases are the same. In the two bimodal functions, the best phase is shifted away from **m** somewhat toward the secondary peak.[12] Reprinted with permission from the International Union of Crystallography (http://journals.iucr.org).

inate phases from a protein, they did at least alter the intensities of the reflections in a manner that might allow one to deduce the phase angles of reflections from the parent protein.[8] Classical heavy-atom phasing methods would not work, but isomorphous replacement phase analysis might.

The man who provided the methodology for this kind of phase analysis was David Harker, who in 1956 was trying to solve the structure of the protein ribonuclease at the Brooklyn Polytechnic Institute.[9] Figure 2.2, reproduced from his work, shows how the scattering vector from the protein plus heavy atom ($\mathbf{F}_{o+x}$ in his notation) is the vector sum of scattering from the protein alone $\mathbf{F}_o$ and the heavy atom alone $\mathbf{F}_x$. (Boldface represents vectors in phase space, and regular font is used for vector amplitudes). The Harker phase-circle diagram in Fig. 2.3 for two heavy-atom derivatives, X and M, is produced by:

(1) Drawing a circle of radius $\mathbf{F}_o$ around the origin
(2) Drawing a circle of radius $\mathbf{F}_{o+x}$ around a point located at $-\mathbf{F}_x$
(3) Drawing a circle of radius $\mathbf{F}_{o+M}$ around a point located at $-\mathbf{F}_M$.

$\mathbf{F}_o$, $\mathbf{F}_{o+x}$, and $\mathbf{F}_{o+M}$ are the measured amplitudes of X-ray scattering for that reflection from the parent protein and the two heavy-atom derivatives. $\mathbf{F}_X$ and $\mathbf{F}_M$ are the scattering vectors for heavy atoms X and M alone, which you can calculate, in both phase and amplitude, once you have located the positions of the heavy atoms in the unit cell. The location on the $\mathbf{F}_o$ circle where all three circles intersect marks the correct phase angle for the parent protein.

Of course, nothing is ever as simple as it first appears. Errors in intensity measurements produce errors in the three phase-circle radii, which in turn can make phase-angle determination inexact. The solution is to collect data from three, four, or five heavy-atom derivatives instead of the minimal two, and to draw the best phase angle through their cloud of intersections. Because myoglobin is so much smaller than haemoglobin, it was actually the first protein structure to be solved in three dimensions by Harker's isomorphous replacement method.[10,11] In the Kendrew Archives at the Bodelian Library in Oxford, there resides a three-ring notebook in which John Kendrew drew a set of four hundred 6-Å Harker phase-circle diagrams with a ruler and a compass, using different-colored pencils for each derivative. When we began preparing for 2-Å high-resolution phase analysis in 1958, John handed me this notebook as a guide to what I should incorporate into a computer program. Today, the notebook sits jealously guarded in the Kendrew Archives at the Bodelian, and the photocopies that I obtained recently are not of publication quality. I wish that I had had access to a really good color photocopy machine half a century ago. But Figure 2.4, from the published work, will have to do.

The most serious problem with phase analysis at 2 Å was how to obtain information from poor phase determinations such as Fig. 2.4c. Several approaches were tried.[12,13] Michael Rossmann proposed and Hilary Muirhead programmed a "least scatter" approach, choosing as a phase the radial line along which intersections with phase circles lay most closely together. Roger Hart used a function to calculate the probability of any particular phase angle, and then chose the direction that led to maximum probability. But the optimal solution was developed by Blow and Crick.[14] Using lack-of-closure errors as defined in Fig. 2.5, they calculated the phase-angle probability all around the 360° phase circle (Fig. 2.6), and then chose the centroid or center of gravity of that probability distribution as the "best" phase angle. The vector to this centroid, **m** in Fig. 2.6, then yields the best phase ϕ_B for that reflection in the Fourier synthesis and the optimal weighting factor **m**. The length of that vector **m** is 1 for a perfect phase determination and 0 for a hopelessly bad determination. The length **m** itself is then a measure of the dependability of that particular phase angle and is called the "figure of merit."[12,13] The mean figure of merit provided a way of assessing the quality of the entire phase analysis.

I wrote a program for EDSAC II that calculated phase probabilities and defined the centroid vector. Figure 2.7 exhibits three examples from the high-resolution myoglobin analysis and shows the usefulness of the figure of merit **m** as a measure of

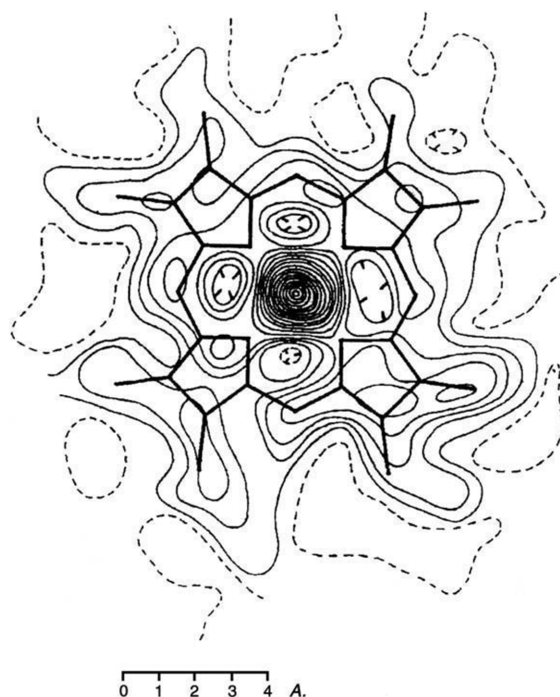


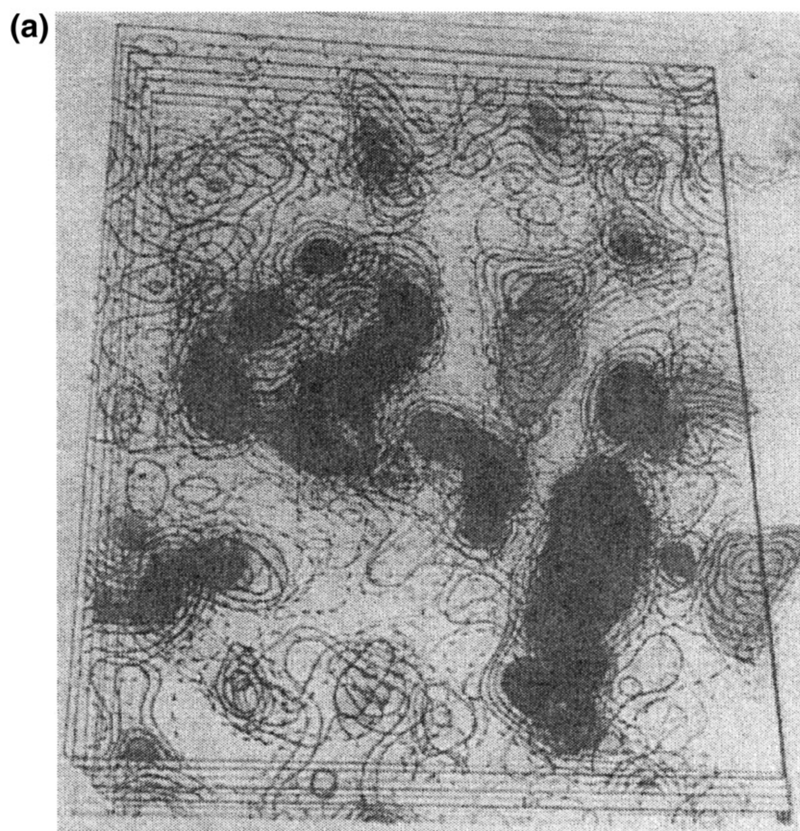
Fig. 2.8. Section through the electron density of the haem group in the 2-Å structure analysis, with an ideal framework superimposed. Positions of the "ears" extending from the four five-membered rings are clearly established.[15]

the quality of each phase determination. Calculation of phase angles was a slow process by modern standards:

> "With four heavy-atom compounds, the time required for one phase determination on EDSAC II was 3.5 sec, of which about 1 sec was occupied in punching results. The full set of 9600 reflections required 10 hr."[12]

The mean figure of merit **m** ranged from 0.90 at the center of the X-ray pattern (lowest resolution) to 0.45 at the high-resolution 2-Å limit. We experimented with various exponential sharpening functions. One approach was simply to accept the figure-of-merit results, with its falloff of amplitudes in the high-resolution region of the diffraction pattern. But this damped down the calculated image too much and smeared out fine structure features. The best solution was to apply a sharpening function so that the *average* falloff of the "best" data, including figures of merit, matched the average falloff of the original experimental data set. This had the result of keeping the effective resolution the same, but

(a)



(b)

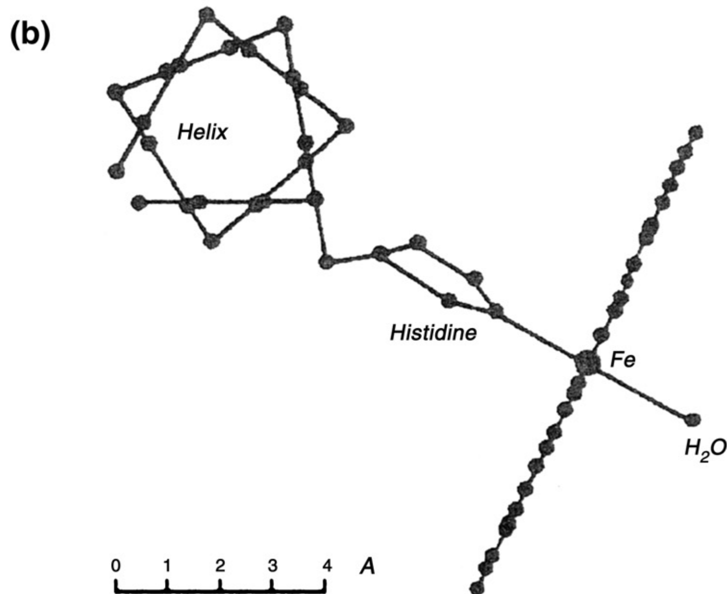


Fig. 2.9. View through stacked Plexiglas sheets showing the haem group seen on edge and a view down the hollow F-helix that is attached to the haem by a His side chain. Map, above; interpretation, below.[15]

emphasizing well-established phases and damping down those that were badly underdetermined.

The All-Night Party

For the last stages of myoglobin analysis—phase refinement, electron density calculation, and interpretation of results—myoglobin had a three-man team. David Davies from the National Institutes of Health in Washington, DC, joined us in April or May 1959 and, like Bror, remained for a time after I departed for the University of Illinois in September 1959. Once the weighted phase angles had been calculated, the next step was to calculate the 2-Å electron density map of myoglobin and see what new information could be learned. The main question was whether those sausage-like rods seen in the 6-Å map actually were α-helices. There was a strong presumption of this at 6 Å, but no proof. Moreover, the connections between "sausages" in the 6-Å map were unclear in places, especially where they appeared to navigate a sharp corner.

Calculation of the three-dimensional electron density map of myoglobin, involving 9600 reflections, was an all-night affair at the EDSAC II computer. John Kendrew, Bror Strandberg, David Davies, and I, as well as our London collaborator David Phillips and some of his coworkers, were all on hand. Ninety-six sections were calculated through the crystals, with each section being a grid of density values spaced roughly 2/3 Å apart.[15] The complete electron density map required 12 h of computer time. In contrast, calculation of the earlier 6-Å map with 400 X-ray reflections had taken only 72 min on EDSAC I.[11]

As the night dragged on and as completed electron density sections began appearing, the sections were scanned for information that the low-resolution map did not have. The haem group was easily identified (Figures 2.8 and 2.9) and, to everyone's satisfaction, the 6-Å "sausages" did indeed turn out to be α-helices. Side chains were clearly visible, and even C=O groups within the helix were clear, thereby establishing the direction of the backbone chain (Fig. 2.10). The overall structure of the sperm-whale myoglobin molecule is shown in Fig. 2.11.

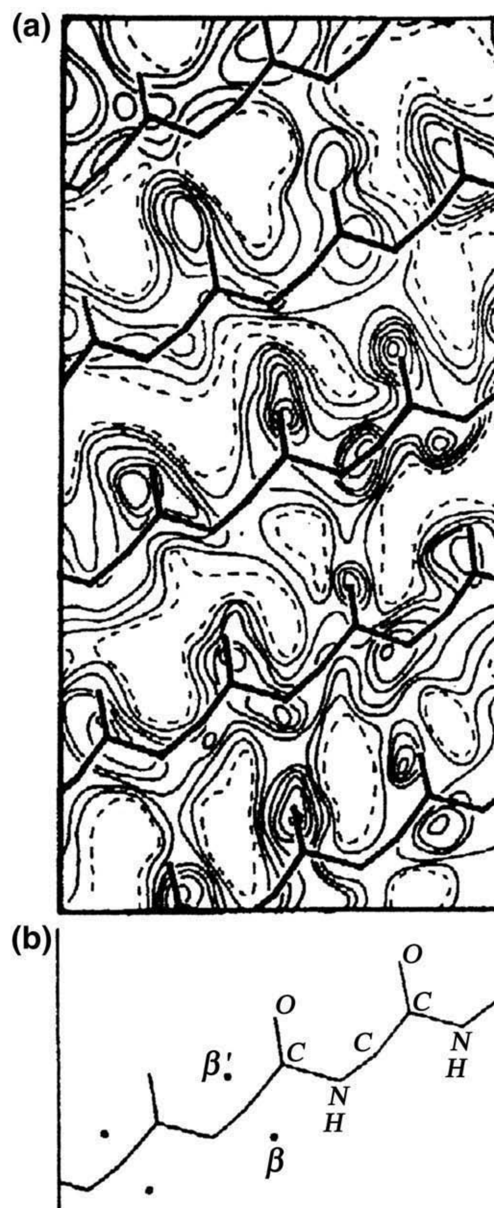


Fig. 2.10. Density in an unrolled α-helix at 2 Å. Not only are side chains clearly visible above and below the plane of this unrolled cylinder, the C=O groups along the backbone are so well defined that they establish the direction in which α-chain runs. The 6-Å "sausages" are clearly Pauling/Corey α-helices.[15]

To celebrate the new structure, a garden party was held at dusk on the lawn of Peterhouse. Contoured Plexiglas sheets of the 2-Å myoglobin map were stacked on a large light box that was the center of attraction. One of my fondest memories of that party is that of Sir Lawrence Bragg, who had pushed for protein structure analysis for more than a quarter of a century, dragging one party attendee after another over to the light box, pointing down the E-helix, and saying in exultation, "Look! It's hollow! It's hollow!"

Shortly thereafter, my wife Lola, my infant son Ian (a Cambridge citizen by birth), and I dashed down to Southampton to catch a ship home so I could take up my duties as a new faculty member at the University of Illinois. Bror and David Davies remained some months longer, helping John construct a display of electron density built of colored clips on a forest of steel rods (Fig. 2.12), into which one could build an accurate atomic structure using wire models invented by John Kendrew. (Kendrew model parts are now a collector's item.) The model was built in the basement of the Cavendish Laboratory and was put on proud display the following autumn, when the International Union of Crystallography held its biennial congress in Cambridge. For the congress, the new myoglobin and haemoglobin results were displayed in a room

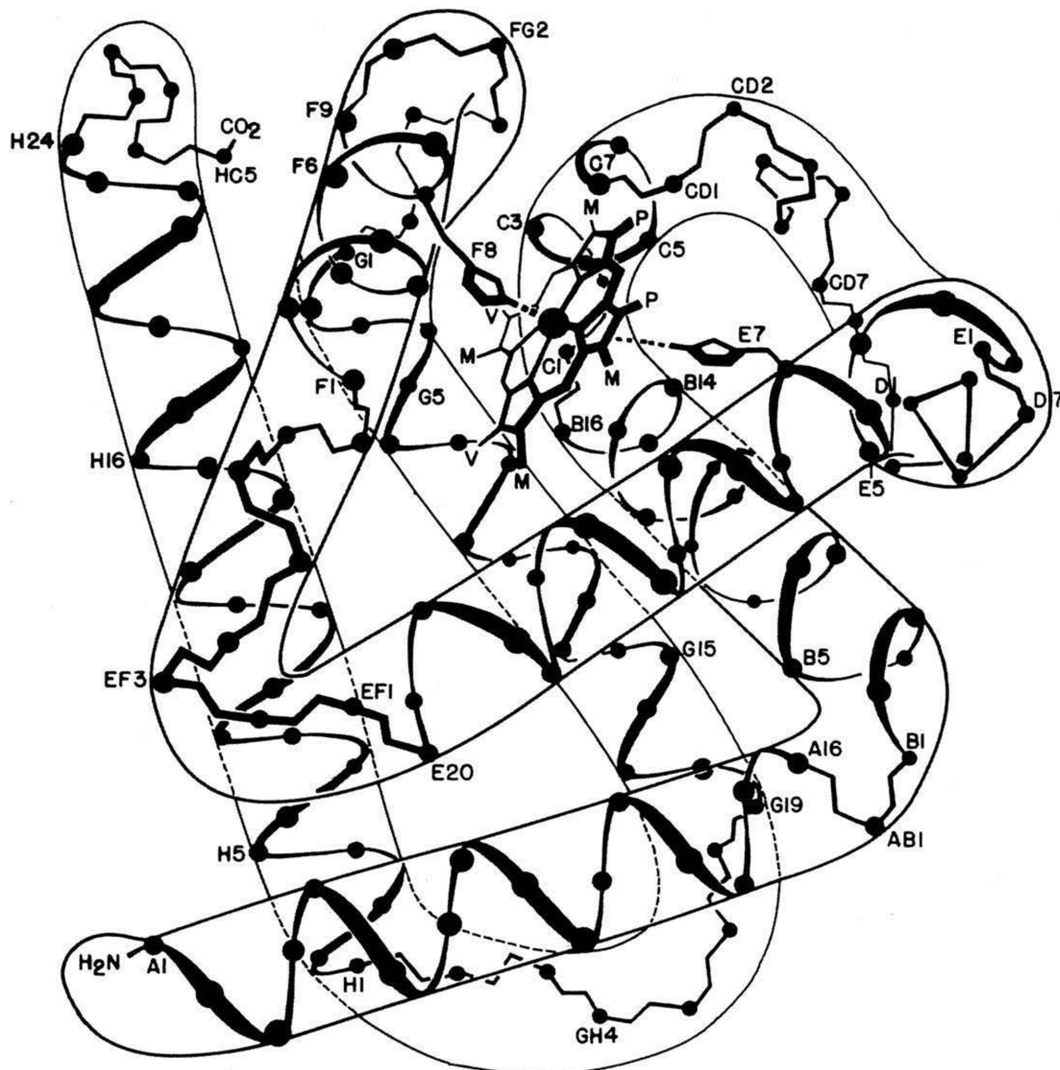


Fig. 2.11. An accurate but simplified diagram of the folding of α-helices in sperm-whale myoglobin from the 2-Å analysis. Each black circle along a helix marks the α-carbon of an amino acid. Zigzag lines follow the chain backbone in nonhelical regions such as bends between helices. The haem group and its attachments also are shown. Drawn by the author to advertise a 1962 lecture by John Kendrew at the University of Illinois.[5]

that belonged to the Electrical Engineering Department. Permission for use of the room was granted, but John and Max were told in no uncertain terms to get "all of that stuff" out of the room quickly before the fall term began. The models now reside in the Kensington Science Museum in London.

Not long after the low-resolution 6-Å myoglobin structure was published, someone in Howard Schachman's laboratory at Berkeley drew the cartoon seen in Fig. 2.13. John Kendrew stands uncertainly to the right of the monster that he has created, with computer output strewn on the floor. Schachman sent the drawing to me in 1965, but could not recall just which of his students had been the artist. The molecule clearly has the same overall structure as Fig. 1.4 of Chapter 1, complete with the breaks between helices and the superfluous connections that were relics of the low-resolution structure.

Afterthoughts

John Kendrew and Max Perutz had quite different styles in running a research laboratory. John was the mentor, guide, and organizer. He was simultaneously involved in several other activities outside the laboratory: renovation of the art in Peterhouse and renovation of the electrical system that (as I remember) Rutherford had first installed, science advisor to the British government, liaison with the US government over the Polaris missile program, and other useful activities. He would come into the laboratory for a group meeting several mornings per week and would ask us, "How is progress? What has been happening? What do you plan to do next? What do you need?" He was valuable to the group as a resource and inspiration. In a sense, he was an excellent mentor for someone who aspired running an independent research group some day. We learned how to run our own show. Myoglobin, however,

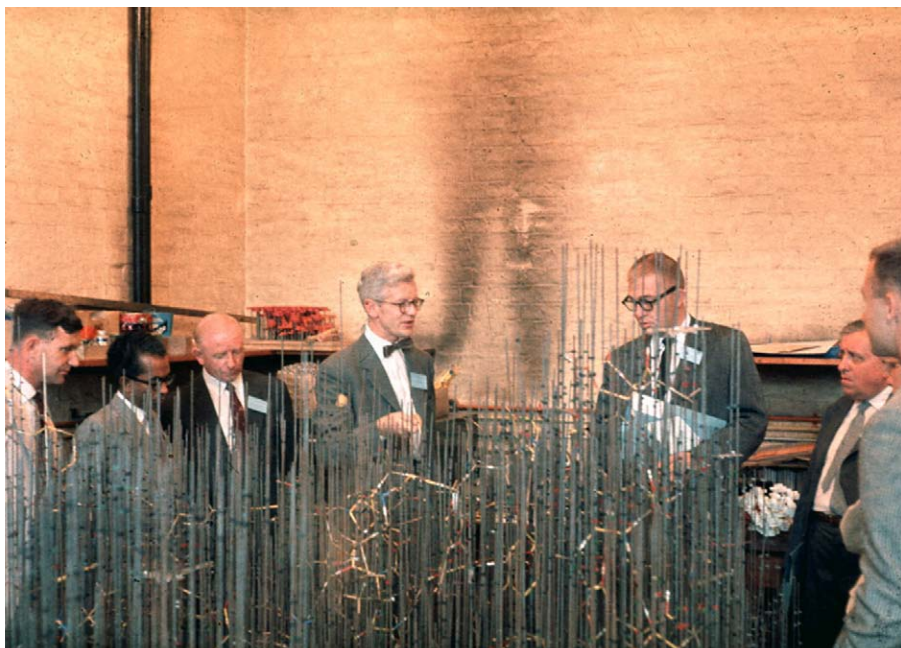


Fig. 2.12. The 2-Å electron density displayed as colored clips in a forest of steel rods mounted on a plywood base. The detailed backbone of the polypeptide chain can be seen snaking through this forest. John Kendrew (center left) explains myoglobin features to attendees at the 1960 meeting of the International Union of Crystallographers in Cambridge.[5]

marked the end of John's scientific research career. In 1974, after the conclusion of the final 1.5-Å-resolution analysis of myoglobin by Herman Watson and coworkers, John left Cambridge to become the first director of the European Molecular Biology Organization Laboratory in Heidelberg. By all accounts, he did a superb job. When he retired from there in 1981, he accepted a position, appropriately enough, as president or master of St. John's College at Oxford. (This is why all his files eventually wound up at Oxford instead of Cambridge.) I can fantasize some awe-stricken freshman coming up to the distinguished, white-haired President and asking timidly, "Are you the St. John for whom the college is named?"

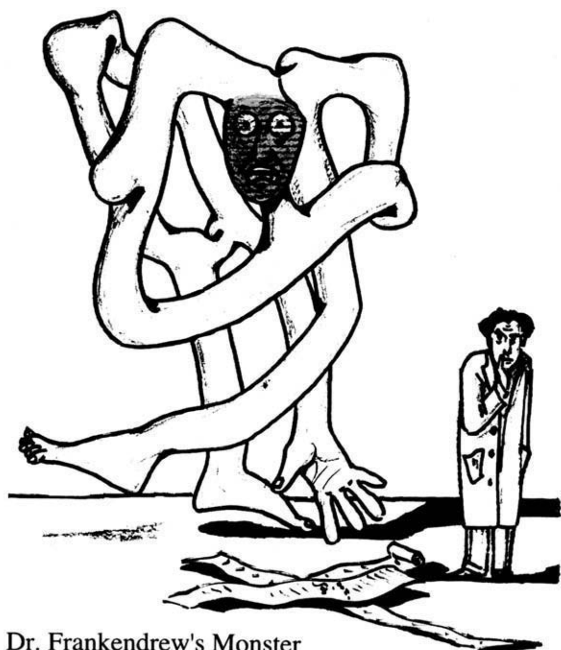


Fig. 2.13. Dr. Frankendrew's monster. The 6-Å model, caricatured around 1958 by a now-forgotten member of Howard Schachman's laboratory at UC Berkeley. The caricature was sent to me by Schachman in 1965. Note the breaks and false connections between helices that were cleared up by the 2-Å analysis.[5]

In contrast, Max was a hands-on bench biochemist whose center of gravity was always the laboratory itself. On many occasions, he would involve himself directly in the ongoing structure analysis as, for example, when he came back to the laboratory in the evenings to mount crystals for Roger Hart's next-day precession camera surveys. He became director of the MRC Laboratory and served for many years, but this did not stop his daily involvement with the laboratory. Both styles had their merits: one learned from John, but one learned with Max. Max's lifelong commitment to research is acknowledged in Fig. 2.14, which depicts him as the modern Noah of the MRC, bringing proteins safely through the deluge on an ark that he himself constructed. The approaching dove at right carries the reward that Max and John earned for their efforts.

In 1996, the International Union of Crystallography held its congress in Seattle, WA. A panel of crystallographic Nobel laureates was invited to speak about their life in structure analysis. One question asked by the moderator was: "If you had to live your life over again, would you do again just what you actually did?" My graduate mentor Bill Lipscomb and most other Nobel laureates replied, "Of course." But John Kendrew had a different take on things: over the years, he had come to appreciate how much he enjoyed scientific administration

Fig. 2.14. Max as Noah, bringing his proteins safely through the deluge. Drawn by editorial cartoonist Dave Granlund after my rough sketches for the cover of *Present at the Flood: How Structural Molecular Biology Came About*.[5] Note, to be rigorously fair, that the myoglobin molecule leading off at the prow of the Ark more properly belongs to John Kendrew, not to Max. But it was indeed first. For more superb cartoons by Granlund, see his Web site [www.davegranlund.com/cartoons].

at the European Molecular Biology Organization Laboratory. If he had to do it over, he said, he probably would not have spent so many years on research, but would have moved into administration much earlier. But there is an interesting flaw in this argument. He was presumably offered the European Molecular Biology Organization position because he was both a capable administrator and a Nobel laureate. Had he not solved the myoglobin structure or carried out some comparable notable scientific achievement, it was unlikely that he would have been chosen as the founding leader of the European Molecular Biology Organization Laboratory. But all things turned out for the best for both John and Max. *De mortuis nil nisi bonum*.

References

1. Dickerson, R. E., Wheatley, P. J., Howell, P. A. & Lipscomb, W. N. (1956). Boron arrangement in a B_9 hydride. *J. Chem. Phys.* **25**, 606–607.
2. Dickerson, R. E., Wheatley, P. J., Howell, P. A. & Lipscomb, W. N. (1957). Crystal and molecular structure of B_9H_{15}. *J. Chem. Phys.* **27**, *200–209.*
3. Dickerson, R. E. & Lipscomb, W. N. (1957). Semitopological approach to boron-hydride structures. *J. Chem. Phys.* **27**, 212–217.
4. Dickerson, R. E. (1992). A little ancient history. *Protein Sci.* **1**, 182–186 (Reproduced in Reference 5).
5. Dickerson, R. E. (2005). *Present at the Flood: How Structural Molecular Biology Came About.* Sinauer Associates, Sunderland, MA.
6. Kendrew, J. C. & Parrish, R. G. (1956). The crystal structure of myoglobin: III. Sperm-whale myoglobin. *Proc. R. Soc. London Ser. A*, **238**, 305–324.
7. Bennett, J. M. & Kendrew, J. C. (1952). The computation of Fourier syntheses with a digital electronic calculating machine. *Acta Crystallogr.* **5**, 109–116.
8. Green, D. W., Ingram, M. & Perutz, M. F. (1954). The structure of haemoglobin: IV. Sign determination by the isomorphous replacement method. *Proc. R. Soc. London Ser. A*, **225**, 287–307.
9. Harker, D. (1956). The determination of the phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. *Acta Crystallogr.* **9**, 1–9.
10. Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature (London)*, **181**, 662–666 (in Ref. 5).
11. Bodo, G., Dintzis, H. M., Kendrew, J. C. & Wyckoff, H. W. (1959). The crystal structure of myoglobin: V. A low-resolution three-dimensional Fourier synthesis of sperm-whale myoglobin crystals. *Proc. R. Soc. London Ser. A*, **253**, 70–102.
12. Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961). The crystal structure of myoglobin: phase determination to a resolution of 2 Å by the method of isomorphous replacement. *Acta Crystallogr.* **14**, 1188–1195.
13. Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961). The phase problem and isomorphous replacement methods in protein structures. *Computing Methods and the Phase Problem in X-ray Crystal Analysis*. pp. 236–251, Pergamon Press, New York.
14. Blow, D. M. & Crick, F. H. C. (1959). The treatment of errors in the isomorphous replacement method. *Acta Crystallogr.* **12**, 794–802.

15. Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). Structure of myoglobin: a three-dimensional Fourier synthesis at 2 Å resolution. *Nature (London)*, **185**, 422–427 (in Ref. 5).

Chapter 3: Recollection of the Events Leading to the Discovery of the Structure of haemoglobin

Michael G. Rossmann

Introduction

My PhD work in the laboratory of Monteath Robertson at the University of Glasgow in Scotland required lengthy calculations of Fourier series summations to determine electron densities. Our computational equipment consisted of desk calculators powered by electric motors and Beevers–Lipson strips representing sine and cosine curves with different amplitudes. It used to take me about one week to calculate one two-dimensional projection for the flat aromatic molecules I was analyzing. It was, therefore, a great delight for me to encounter at the University of Minnesota in Minneapolis one of the first commercial computers, the UNIVAC 1103, when I joined the laboratory of Bill Lipscomb (the "Colonel") as a young postdoctoral fellow. The Colonel had arranged for a few hours per week of computing time for his group on the UNIVAC computer at Honeywell in neighboring St. Paul. I overlapped with Dick Dickerson for a couple of months as he finished his doctorate on the structure of boron hydrides, a subject that gained the Nobel Prize for the Colonel about 15 years later. Dick taught me my first lessons on how to program a computer. Before I left Minnesota two years later, Bob Jacobson, a graduate student with the Colonel, and I had written a fairly complete crystallographic computing package that continued to be in use for many years.

While in Minnesota, I attended the International Union of Crystallography meeting in Montreal and listened to a fascinating lecture by Dorothy Hodgkin[1] on the structure of myoglobin being tackled by John Kendrew in Cambridge. Work on the structure determination of proteins seemed, to me, to be the best kind of project for my interests in methods for solving the crystallographic phase problem and the use of modern computational techniques. Fortunately, Max Perutz, the director of the Cambridge Laboratory that also included John Kendrew, quickly replied positively to my request for a job.

When my family and I arrived in Cambridge, I found a kind of environment different from what I had known either in Glasgow or Minneapolis. We worked in the little "Hut" (still standing even today; Fig. 3.1) in the New Museum Site off Free School Lane, close to the central market square. I gathered that there was insufficient space in the Cavendish Laboratory, which was world famous for the work of Thompson, Rutherford, and others on the structure of the atom, to also accommodate Max Perutz and members of his group who were studying biological molecules. Thus, we were all crowded into The Hut, next to the Austin Wing of the Cavendish. I shared a very small office with Bror Strandberg, a Swedish graduate student; Dick Dickerson, who had arrived via a circuitous route after a short stay at the University of Leeds; and Larry Steinrauf, a former student of Linus Pauling. The office opposite to ours housed Francis Crick and Sydney Brenner. Max and John each had their own office. There was one office for the "computer girls" who helped with manual calculations, another office for Ann Cullis (assistant to Max Perutz) and Mary Pinkerton (assistant to John Kendrew), an office for a graduate student (Hilary Muirhead) and assorted postdoctoral fellows who worked for Sydney and Francis, plus a very crowded biochemistry laboratory. Everyone packed tightly into the central corridor of The Hut at 11 o'clock each morning, reaching for the coffee pot brewed by Leslie Barnett, Sydney Brenner's assistant, while debating everything from the relation between church and science in the then new Churchill College to the best ways of finding heavy-atom positions in protein crystals. Many a morning the conversations were dominated by Francis's loud but appealing voice and provocative ideas. This was, for me, an intellectual environment more stimulating than I had ever previously encountered. I started to realize that biology was a more interesting subject than I had thought and to recognize how ignorant I had remained in my narrow undergraduate physics and mathematics education. No longer would the phase problem be my only challenge, as I became aware that the most significant problems were those of the way the universe and Earth had evolved to make the present day.

Fig. 3.1. The Medical Research Council Hut at the New Museum Site off Free School Lane, Cambridge, photographed in 2003, still almost unchanged after nearly 50 years. However, in 2003, it was all locked up, whereas in 1959, it was full of activity day and night.

Protein Structure in Cambridge before 1958

The likelihood of ever solving the three-dimensional structure of a protein seemed extremely small even in the 1950s. There were too many data to be collected, no reasonable way of solving the phase problem, and no way of computing the results to make an interpretation of the collected data. But maybe even more daunting was the nagging question of whether each protein molecule had a unique structure. To some extent, these problems were starting to be solved. Bernal and Dorothy Crowfoot (later Hodgkin) had shown that pepsin could be induced to grow crystals that could diffract to near-atomic resolution,[2] implying that all the molecules in the crystal had the same structure. Max encouraged Tony Broad to build a powerful rotating-anode X-ray generator. Max had also overseen the development of an accurate instrument for measuring the intensity differences between the X-ray reflections from isomorphous heavy-atom derivatives of haemoglobin crystals. Punched-card-driven devices, which were first used in the English textile industry and later became standard items of office equipment as the forerunners of computers, were starting to be used for mathematical computations. Furthermore, Max had been creating a set of very original techniques for solving the phase problem. One of these was the isomorphous replacement technique that had been the basis of the first crystal structures ever solved by X-ray diffraction, namely, those of NaCl and KCl.[3] It had also been explored by my former mentor Monteath Robertson as a tool for determining the structure of some haem-like compounds,[4] but most crystallographers at the time thought that the incorporation of even the heaviest atoms would have only a negligible impact on the X-ray diffraction of a protein crystal. Nevertheless, in 1953, Max had shown that, using heavy-metal isomorphous replacements, it was possible to determine the phases of the ($h0l$) reflections of his monoclinic haemoglobin crystals.[5] Admittedly, these reflections were of a centric zone and therefore limited to be either 0° or 180°, but this was enough to give hope that a structure determination might be possible.

John Kendrew, a former student of Max, had initiated a study of the smaller myoglobin molecule. Myoglobin is a carrier of oxygen in muscles, much like haemoglobin is a carrier of oxygen in blood. However, myoglobin has a molecular weight of only 17 kDa and consists of a single polypeptide chain, whereas haemoglobin consists of four polypeptide chains: two α- and two β-chains each of about 17 kDa molecular weight. With his smaller protein molecule, John had made more rapid advances than Max, using the techniques Max had developed. In 1957, John had been able to obtain a 6-Å-resolution electron density map of sperm-whale myoglobin that seemed to represent a series of bent helices,[6] probably α-helices as had been proposed by Linus Pauling.[7] Models of John's myoglobin structure in a variety of different scales were everywhere in The Hut. This structure had been the big event in the year before I arrived in Cambridge and had been heralded as being imminent by Dorothy when I heard her speak in Montreal.

Autumn of 1958

By the time I arrived in Cambridge in September 1958, Max and Ann Cullis had regained their ability to grow good crystals of monoclinic oxygenated haemoglobin and were preparing a series of heavy-atom derivatives based on the substitution of a couple of Cys amino acids with Hg derivatives. Max had determined the x and z coordinates using the ($h0l$) projection data he had so carefully explored over the previous six years or more. The unresolved problem was the determination of the relative y coordinates of the heavy-atom sites. This problem existed because there is no centric projection perpendicular to the unique 2-fold axis in a monoclinic space group and no defined origin as is provided by the crystallographic 2-fold axis in the ($h0l$) projection. John Kendrew had encountered this problem in the determination of the 6-Å-resolution map of myoglobin. Perutz[8] and Blow[9] had made suggestions as to how this problem might be solved, but none of the proposed functions gave unambiguous solutions.

A few steps from The Hut was an old, ugly, brick building that housed the very new Mathematics Laboratory. In this building, Maurice Wilkes and his "boffin" friends had put together a machine from leftover wartime radar equipment that was capable of being programmed to do all manner of fast computations. This Electronic Delay Storage Automatic Calculator (EDSAC) was arguably the first automatic stored-program electronic computer ever built. John Kendrew had used this machine to calculate the 6-Å-resolution map of myoglobin. However, by the time I arrived in Cambridge, the original EDSAC had been replaced by a new, much faster, much more powerful EDSAC II computer that had just opened its doors to other selected users (meaning friends and acquaintances of college "dons"). EDSAC II was the brainchild of David Wheeler. It was quickly apparent to me that EDSAC II was far more sophisticated than the machine I had used in Minnesota. I read the well-written manual and started putting some of my crystallographic experience into this wonderful new machine. However, I did not have a Cambridge background and was not aware of the conventions assumed by in-group users. As a result, I was treated with some suspicion that was not dispelled until I was able to show that my programs produced correct results in an acceptably short time frame. In addition, there was a large group of more conventional crystallographers at Cambridge who distrusted the new computers, while others, like myself, advocated their use. The prevailing thought was that the human mind should be more powerful than any machine.

One of the first tasks that I set for myself was to write a Fourier synthesis program for the calculation of electron density and other crystallographic maps. I also proposed a new and simpler function to solve the "*y*-axis" problem mentioned above. Although program testing could be performed at certain specific times during the day, large-scale computations could be performed only on allocated nights. Max and John had been given Monday nights, to be shared with the radio astronomers of Martin Ryle's group. Thus, one Monday, maybe in November 1958, I used my new Fourier program to compute my new *y*-axis correlation functions. Alas, when I came to examine the results the next morning, I realized that the symmetry was wrong. Clearly, there was a mistake (maybe many) in the program. Max and Ann were very supportive and assured me that they could wait for me to try to fix my program. Fortunately, the problem was not serious and, two weeks later, I was back on another Monday evening repeating the same calculations with a corrected program.

My delight was indeed great the next morning. The maps showed exactly what I had hoped—positive peaks for vectors between atoms in the same compound and negative peaks for vectors between atoms in different compounds[10] (Fig. 3.2). Furthermore, the vectors were all self-consistent. Best of all, the peaks were sky high, leaving absolutely no doubt in differentiating between signal and noise. Now it was certain that we had the necessary heavy-atom derivatives, the data-processing procedures, and the technology to find the correct positions of the heavy atoms, all prerequisites for determining the structure of haemoglobin. Max was delighted. He brought in every person who would understand the significance of these results, starting with the local crystallographers such as Bill Cochran and then the outsiders such as Sir Lawrence Bragg, Dorothy, and Bernal (the "Sage"). Mostly, he would ask me for my folders containing the maps whenever he had a distinguished visitor, but it was obvious that he was happy to have had this progress. At the same time, I began to be recognized by those who had complained about my unconventional ways and techniques.

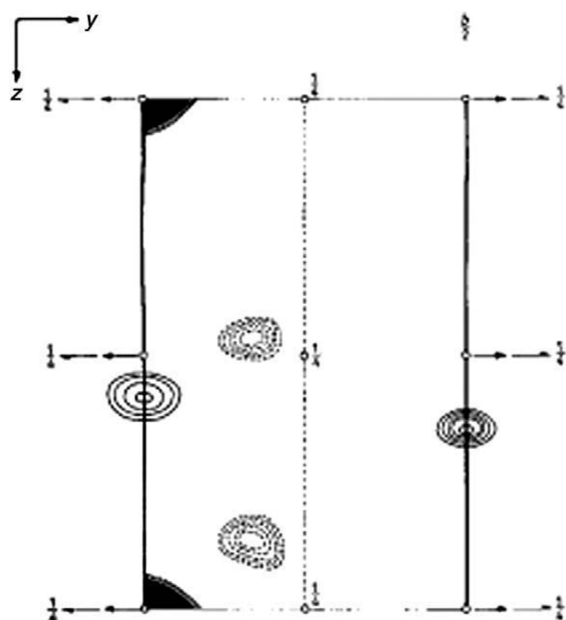


Fig. 3.2. Composite view of the $(F_{HgAc2}-F_{PCMB})^2$ correlation function between the $HgAc_2$ and *para*-chloromercurybenzoate (PCMB) heavy-atom derivatives. Continuous lines are positive contours, whereas broken lines are negative contours. The Harker sections at $v=0$ and $v=1/2$ have positive peaks relating the symmetry-related $HgAc_2$ sites and also the symmetry-related PCMB sites, whereas the $v=0.3$ section contains negative vectors between the $HgAc_2$ and the PCMB sites.[10]

Winter and Spring of 1959

Now we had to start thinking about phase determination given the heavy-atom sites. But we still needed to determine the parameters that defined the size of heavy-atom substitutions (occupancies and "temperature" factors). For this, I wrote a least squares program based on the *y*-axis correlation method.

John Kendrew and Sir Lawrence Bragg had used a manual technique for determining the phases of each reflection based on David Harker's diagram.[11] Each of the about 400 reflections used in the 6-Å-resolution map of myoglobin had been analyzed with a "Harker diagram." This was, in reality, an Argand diagram on which were drawn calculated vectors representing the structure factors of the heavy atoms given their parameters. Circles corresponding to the observed amplitudes of all the isomorphous compounds were then drawn. Where the circles intersected was the most probable phase. John and Sir Lawrence had simply guessed at what they thought was the best point of intersection of the phase circles and averaged their guesses. Now, however, we had five times as many reflections as John had had for his earlier work.

David Blow had been a graduate student of Max, although in 1958, he was in America as a postdoctoral fellow with Alex Rich, first at the National Institutes of Health and later at the Massachusetts Institute of Technology. David had spent his PhD time thinking with Francis about how to be more rigorous[12] in the selection of the best phase from the Harker diagram. Dick and Bror simplified the Blow–Crick ideas to a readily programmable algorithm.[13] I had helped to teach Hilary Muirhead to program EDSAC II and suggested that she should write a phasing program. I was, however, not entirely happy with Dick and Bror's algorithm because it assumed that all the phasing errors were in the measurement of heavy-atom derivative data, with no error in unsubstituted "native" data. Thus, Hilary and I somewhat modified the procedure[14]—a change that most probably aided in producing the clean map that we subsequently calculated in August 1959.

Max had briefly taught me how to mount the haemoglobin crystals into glass capillaries, how to

operate the Broad X-ray generators in the basement of the Austin Wing of the Cavendish Laboratory, and how to handle precession cameras with protein crystals. Among the man-made obstacles was the chief Cavendish Laboratory technician who could never understand why we needed so much film to collect our data. Persuading him to provide a new box of film was a major diplomatic feat that usually resulted in being given merely a few films at a time.

We now had all the tools and all the data I thought were necessary for producing a 5.5-Å-resolution map of haemoglobin. Max and Ann had collected about five heavy-atom derivative data sets, but Max wanted to collect one more. Although, in retrospect, that can now be seen to have been an unnecessary precaution, we had no idea what we would see in the map. Thus, when we eventually saw the electron density, there was no doubt about its significance and, considering the unexpected nature of the result, it might have been met with disbelief had the map been poorer.

August 1959

Everything was ready by August and Max agreed to the calculation of the haemoglobin electron density map. That is everything except for EDSAC II, which had had one of its frequent nervous breakdowns. We impatiently waited another two weeks when Max was invited to be the first user of the repaired computer. I persuaded Max to come with me to the Math lab that afternoon. I tried to teach Max how to feed the punched paper tapes into the computer, but he was too nervous, so eventually I started the computations myself. About an hour and a half later, the map was calculated, but it still needed plotting by hand, a tedious task that would consume our computer girls for the next few days. But I did notice that there were two sections with very high density, which presumably were the haem groups in the α- and β-chains of the molecule. (The space group was C2, with molecules sitting on crystallographic 2-fold axes.) Thus, I suggested that we should explore these positions with a finer grid interval. Max agreed, and I punched a couple of quick "jiffy tapes" to do this task. This time, Max was less nervous and succeeded. Unfortunately, seconds after Max had initiated the calculations, smoke started to pour out from behind the control panel, and everything closed down. Max made a comment, something like, "You see, I should not have operated the computer." EDSAC II was closed for another two weeks. Then John Kendrew, Dick Dickerson, Bror Strandberg, and David Phillips (their collaborator from the Royal Institution in London) had their turn in using EDSAC II to calculate the 2-Å-resolution map of myoglobin. I remember receiving a phone call late that evening from David Phillips asking me to come to the Math lab to sort out a problem with my Fourier program.

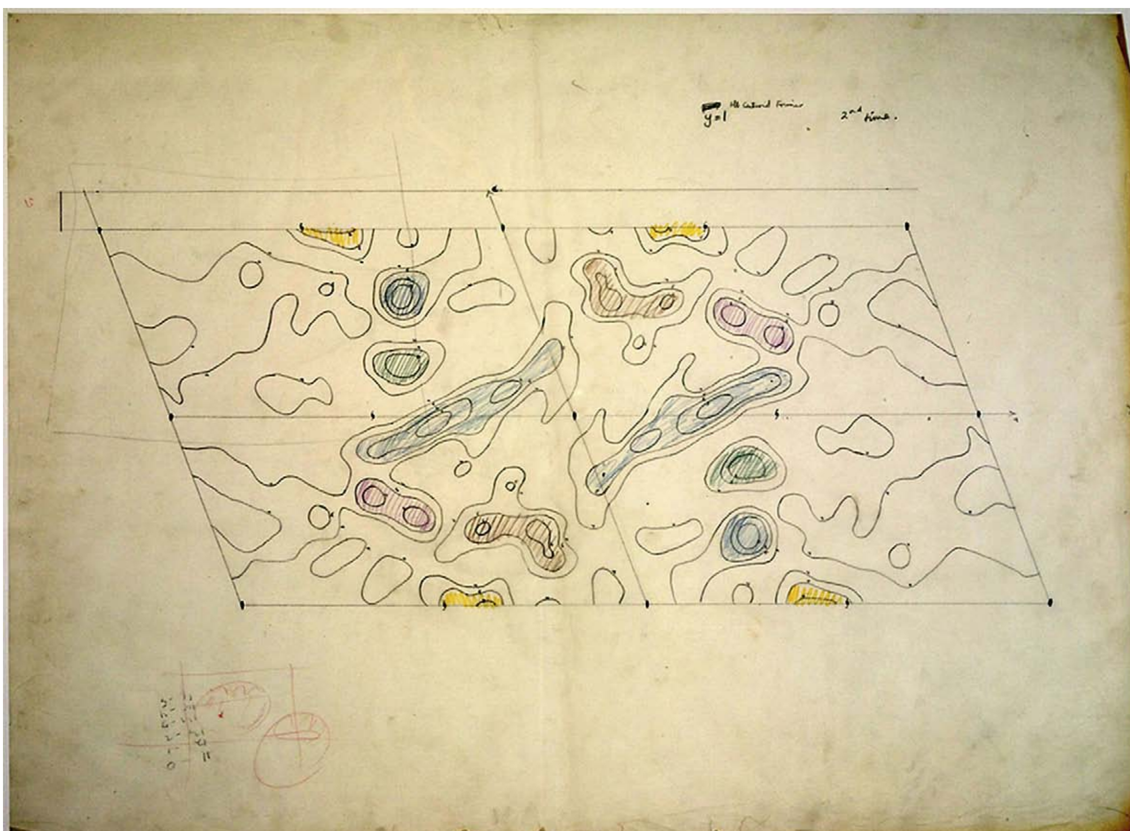


Fig. 3.3. The 5.5-Å-resolution haemoglobin map. Section $y=1$ showing long rods of density representing α-helices.

They then took most of the night to finish their calculation.

Once the haemoglobin maps had been plotted onto a semitransparent tracing paper (Figs. 3.3–3.5), Max tried to construct a cage made of Meccano parts in his office in order to display the map sections in a way that might help make the map more easily interpretable. Yet when I looked at the maps, I could see long, thin features that were surely α-helices. I really wanted to look at the maps and not wait on Max. However, maybe out of force of habit, Max started to prepare more heavy-atom derivatives in the laboratory. What happened next I kept to myself for many decades. I had been privileged to work on the haemoglobin project with Max, but it was also a project that Max had given his whole life to develop. In my enthusiasm to look at the results, I stole the final discovery from Max. When I realized what I had done, I felt terrible. I have never truly forgiven myself for my behavior. The excitement of discovery was completely shattered for me.

When Georgina Ferry was doing her research prior to writing Max's biography, she came to interview me at Purdue. I felt it right to tell her everything I knew about Max, which would have to include these moments on the summit of Mount Discovery as a background to understanding the character of her biographical study. Sometime after her visit, she sent me her presentation of those events with a request that she be allowed to use this passage in her upcoming book.[15] I found her writing to be sensitive, balanced, and accurate; thus, I gave her permission to use this passage in her book. For me, this was also a catharsis on a topic about which I had spent much anguish.

The helices in the maps were all connected by density permitting a trace of both the α and β polypeptide chains. I had once contemplated, long before the map was out, that there might be some similarity between myoglobin and haemoglobin because of the coincidences of molecular weights and functions. But I was quickly put down with disdain. Nevertheless, I was not overly cowed by these comments because, having traced the density chains, I immediately tried to compare the easily available three-dimensional 6-Å-resolution myoglobin model with my chain tracing. The result was obvious.

I found Max in the Biochemistry Laboratory. Only then did I realize how much I had hurt him. With the realization of what I had done, all desire to explore further was completely gone. Instead, Max used his outstanding experimental skills to build a model from thermally setting clay (Fig. 3.6), which I would never have had the skill or imagination to do myself. This was a perfect way of representing the structure, showing not only the "globin fold" but also indicating the level of error and the packing of helices. The model was beautiful! Nature was beautiful!

Notwithstanding the similarity of the myoglobin structure with the α- and β-chains of haemoglobin,

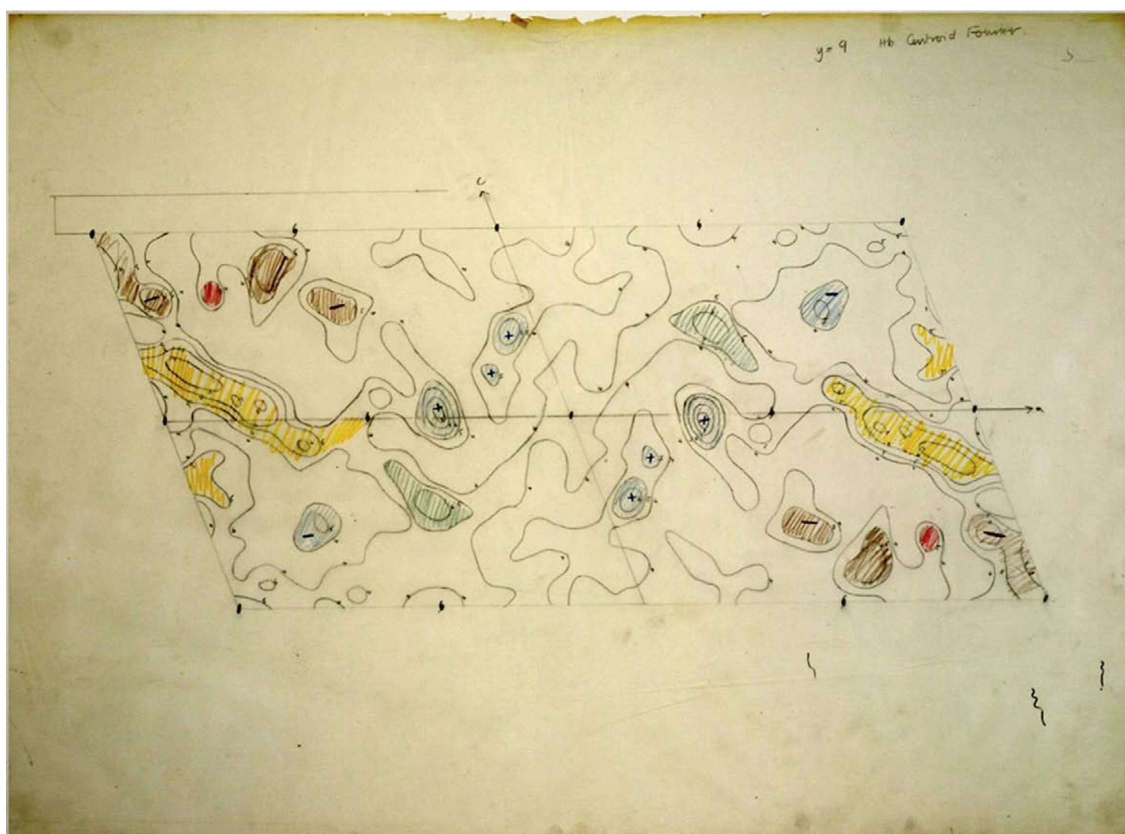


Fig. 3.4. The 5.5-Å-resolution haemoglobin map. Section $y = 9$ showing long rods of density representing α-helices.

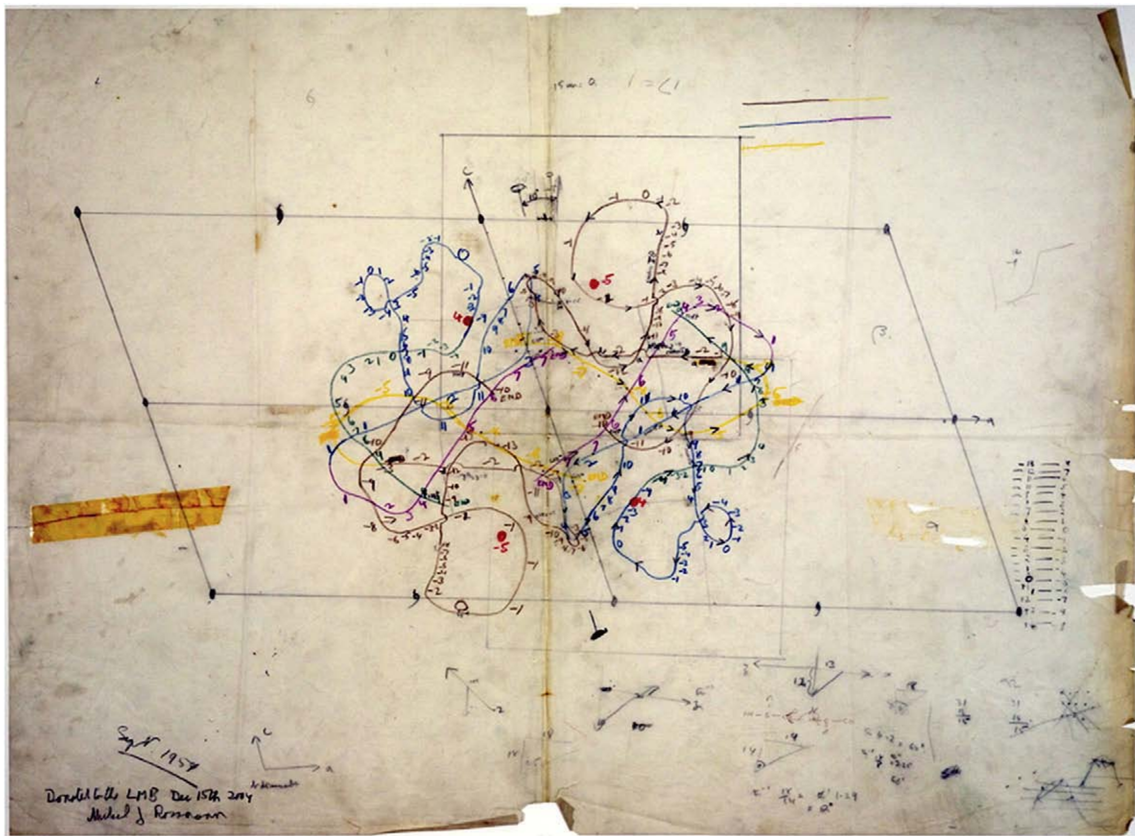


Fig. 3.5. Interpretation of the 5.5-Å-resolution haemoglobin map. I used different colors to trace regions of continuous density. A change in color represented uncertainty of the connectivity of densities between the colored traced sections. Numbers indicate the sections on which density was found. I could recognize the two independent, fully connected chains only when this tracing had been completed. In 2004 or 2005, I gave this map, plus my model of haemoglobin, for safekeeping to the Medical Research Council Laboratory of Molecular Biology historical archives.

Max still felt it necessary to make further checks on the correctness of the structure, in particular of the site of the all-important heme groups. He noted that the Fe atoms in the center of the heme groups should give rise to a measurable anomalous X-ray diffraction effect. Thus, given the now-established Fe atom positions and the isomorphous replacement determination of the "native" phases, it would be possible to calculate the reflections that should have the largest anomalous effect and the sign of their Bijovet difference. In this way, I determined 32 reflections that should have the largest anomalous difference. Max then asked Tony North at the Royal Institution to measure the size and sign of the Bijovet difference for these reflections using the diffractometer that Uli Arndt and David Phillips had designed to help John Kendrew with the myoglobin data collection. Tony found that his experimental measurements agreed with my predictions for 28 of 32 listed reflections.[16] This confirmation also had a further benefit, namely, in determining the absolute hand of the haemoglobin map. That information could then be applied to establish the absolute hand of the 2-Å-resolution myoglobin map that Kendrew's group had just determined using the reasonable assumption that the myoglobin and haemoglobin structures have the same hand. This demonstrated that all the α-helices in the myoglobin map were right handed. This was also a satisfactory result, as there would be fewer steric clashes of the side chains of laevorotatory amino acids with main-chain atoms in a right-handed α-helix, given the correctness of Bijovet's determination of the absolute hand of amino acids being laevorotatory.[17]

The evolutionary significance of the haemoglobin structure[16,18] was not lost on anyone. Here was a verification of Darwin's ideas at the molecular—and therefore genetic—level. Francis Crick and Jim Watson had suggested the manner in which a gene determined the amino acid sequence of a protein, although the actual genetic code was still unknown at that time. Yet it would seem to be likely that there must have been an ancestral gene that coded for a primordial globin fold from which the present-day sperm-whale myoglobin and horse haemoglobin α- and β-chains had evolved by gene duplication and mutation. For me, this discovery has had more influence on my subsequent scientific direction than any other single event.

It was a fortuitous coincidence that, in the following year, the International Union of Crystallography had scheduled its triennial meeting in Cambridge. It had been three years since I had heard Dorothy talk

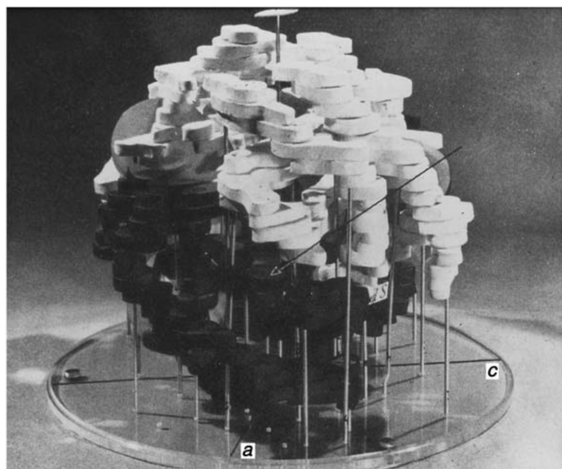


Fig. 3.6. Max's model of the 5.5-Å-resolution map of haemoglobin.[18] A thermally setting clay was molded by Max to correspond to the density on each section outlined by a chosen contour level. The different sections were then glued together. The photograph of the model is only available in black and white, but the α-chain was shown in white, the β-chain was shown in black, and the hemes were shown in red. The model referred to in the caption to Fig. 3.5 was made of wood, using a thread saw to cut along the contour lines, thereby isolating the higher-density regions that were subsequently glued together. The scale was the same as what Max had used to construct his model and the same as the original maps shown in Figs. 3.3–3.5.

at the previous International Union of Crystallography meeting in Montreal. What tumultuous years those turned out to be! The 1960 meeting in Cambridge was an opportunity to display our new discoveries. Max gave a great lecture in the local movie theater located in the market square, the only hall big enough to house the whole meeting. (In those days, movie theaters were bigger, and scientific conferences were smaller.)

After August 1959

Not long after those extraordinary days in 1959, my three officemates dispersed around the world. Dick and Larry returned to America, and Bror stayed on a few months to initiate the building of the myoglobin model in the cold basement of the Austin Wing before returning to Sweden. But the desk vacated by Larry was quickly occupied by David Blow on his return from the United States. Although he had missed two exciting years, he (as was I) was fascinated by the crystallographic implications of structure determinations that had happened in his absence. One of David's interests as a graduate student had been the use of anomalous dispersion to aid phase determination. I had also been thinking about the use of anomalous dispersion for finding the position of anomalous scatterers.[19] We found ourselves in daily stimulating discussions, producing, in the subsequent five years, a series of papers that have helped to form the foundations of modern structural biology. The topics we covered were single isomorphous replacement (SIR) and single anomalous dispersion (SID),[20] phase combination,[21] and molecular replacement (MR).[22–25] The genesis of these papers came largely from the haemoglobin structure determination. We asked ourselves how few derivatives would have really been necessary to solve the structure (SIR and SAD); how crystallographic structure factor calculations should be compared to multiple isomorphous replacement phasing; and whether it would have been possible to discover the relationship between the α- and β-chains of haemoglobin without actually determining the structure (MR).

David initiated a study on chymotrypsin, a study that we initially shared equally.[26] Not only were we able to use all our new ideas, but the data processing also benefited from our collaboration. Whereas Dick and Bror had used a pseudo-manual/computer sorting procedure captured on a now-famous photograph (Fig. 3.7), I developed a computer sorting system for the chymotrypsin studies that I later adapted and improved for use at Purdue University and in modern commercial software.[27]

Max made it clear to me that I should try to find a more permanent job somewhere else. Fortunately, there was no shortage of job offers once we had determined the structure of haemoglobin, but they were all from America. I accepted the offer from Purdue University because I felt that the head of the Biology Department was genuinely interested in

Fig. 3.7. Dick and Bror with the myoglobin sorting board. Each piece of punched tape represented a set of intensities from one line (constant h and k) of reflections on one precession photograph. The same line of reflections might have been recorded on a number of photographs, thus permitting the determination of scale factors between photographs.

helping me to work on the structure of proteins. My decision turned out to be right. Although for many years I missed the intense intellectual stimulus of Cambridge, challenges at Purdue allowed me to put a personal imprint on the development of my own laboratory, which might not have been possible in a more established environment. I remain enormously grateful to Max, Dorothy, David Phillips, and many other English colleagues who strongly supported my work at Purdue in numerous ways, including frequent visits and nominations to prestigious awards. It is significant that exactly 50 years since the discovery of the haemoglobin structure, we are about to move into the completely new "Hockmeyer Hall of Structural Biology" at Purdue University.

Acknowledgements

I thank Georgina Ferry for reading and correcting my original draft of this manuscript, and Sheryl Kelly for help in its preparation for publication.

1959 was both a year of joyful discovery and a year of domestic problems for me. I remain enormously grateful to Lola and Dick Dickerson for their extremely kind and generous help without which I would not have been able to do the work described in this article. Similarly, the concern of Max and Ann (Cullis) Kennedy for my welfare remains an overriding component of my memories. Although Max was not initially a fan of computerized technology, he was always willing to search for ways to satisfy my appetite for what I deemed necessary within reasonable financial limits. I also thank David Blow for the fun years we had together, as well as my two "computer girls" Jill (Collard) Dawes and Angela (Campbell) Mott for their outstanding technical help and lifelong friendship. Finally and most importantly, I thank my wife Audrey for her faithful loving support in the face of strong competition from science, my other love.

References

1. Ferry, G. (1998). *Dorothy Hodgkin: A Life.* Granta Books, London.
2. Bernal, J. D. & Crowfoot, D. (1934). X-ray photographs of crystalline pepsin. *Nature (London)*, **133**, 794–795.
3. Bragg, W. L. (1913). The structure of some crystals as indicated by their diffraction of X-rays. *Proc. R. Soc. London Ser. A*, **89**, 248–277.
4. Robertson, J. M. (1935). An X-ray study of the structure of phthalocyanines: Part I. The metal-free, nickel, copper, and platinum compounds. *J. Chem. Soc.* 615–621; 1935.
5. Green, D. W., Ingram, M. & Perutz, M. F. (1954). The structure of haemoglobin: IV. Sign determination by the isomorphous replacement method. *Proc. R. Soc. London Ser. A*, **225**, 287–307.
6. Bodo, G., Dintzis, H. M., Kendrew, J. C. & Wyckoff, H. W. (1959). The crystal structure of myoglobin: V. A low-resolution three-dimensional Fourier synthesis of sperm-whale myoglobin crystals. *Proc. R. Soc. London Ser. A*, **253**, 70–102.
7. Pauling, L., Corey, R. B. & Branson, H. R. (1951). The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl Acad. Sci. USA*, **37**, 205–211.
8. Perutz, M. F. (1956). Isomorphous replacement and phase determination in non-centrosymmetric space groups. *Acta Crystallogr.* **9**, 867–873.
9. Blow, D. M. (1958). The structure of haemoglobin: VII. Determination of phase angles in the non-centrosymmetric [100] zone. *Proc. R. Soc. London Ser. A*, **247**, 302–336.
10. Rossmann, M. G. (1960). The accurate determination of the position and shape of heavy-atom replacement groups in proteins. *Acta Crystallogr.* **13**, 221–226.
11. Harker, D. (1956). The determination of the phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. *Acta Crystallogr.* **9**, 1–9.
12. Blow, D. M. & Crick, F. H. C. (1959). The treatment of errors in the isomorphous replacement method. *Acta Crystallogr.* **12**, 794–802.
13. Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961). The crystal structure of myoglobin: phase determination to a resolution of 2 Å by the method of isomorphous replacement. *Acta Crystallogr.* **14**, 1188–1195.
14. Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1962). The structure of haemoglobin: VIII. A three-dimensional Fourier synthesis at 5.5 Å resolution: determination of the phase angles. *Proc. R. Soc. London Ser. A*, **265**, 15–38.
15. Ferry, G. (2007). *Max Perutz and the Secret of Life.* Chatto and Windus, London.
16. Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1962). The structure of haemoglobin: IX. A three-dimensional Fourier synthesis at 5.5 Å resolution: description of the structure. *Proc. R. Soc. London Ser. A*, **265**, 161–187.
17. Bijvoet, J. M., Peerdeman, A. F. & van Bommel, A. J. (1951). Determination of the absolute configuration of optically active compounds by means of X-rays. *Nature (London)*, **168**, 271–272.
18. Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G. & North, A. C. T. (1960). Structure of haemoglobin. A three-dimensional Fourier synthesis at 5.5-Å resolution, obtained by X-ray analysis. *Nature (London)*, **185**, 416–422.
19. Rossmann, M. G. (1961). The position of anomalous scatterers in protein crystals. *Acta Crystallogr.* **14**, 383–388.
20. Blow, D. M. & Rossmann, M. G. (1961). The single isomorphous replacement technique. *Acta Crystallogr.* **14**, 1195–1202.
21. Rossmann, M. G. & Blow, D. M. (1961). The refinement of structures partially determined by the isomorphous replacement method. *Acta Crystallogr.* **14**, 641–647.
22. Rossmann, M. G. & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.* **15**, 24–31.
23. Rossmann, M. G. & Blow, D. M. (1963). Determination of phases by the conditions of non-crystallographic symmetry. *Acta Crystallogr.* **16**, 39–45.
24. Rossmann, M. G., Blow, D. M., Harding, M. M. & Coller, E. (1964). The relative positions of independent molecules within the same asymmetric unit. *Acta Crystallogr.* **17**, 338–342.

25. Dodson, E., Harding, M. M., Hodgkin, D. C. & Rossmann, M. G. (1966). The crystal structure of insulin: III. Evidence for a 2-fold axis in rhombohedral zinc insulin. *J. Mol. Biol.* **16**, 227–241.
26. Blow, D. M., Rossmann, M. G. & Jeffery, B. A. (1964). The arrangement of α-chymotrypsin molecules in the monoclinic crystal form. *J. Mol. Biol.* **8**, 65–78.
27. Rossmann, M. G., Leslie, A. G. W., Abdel-Meguid, S. S. & Tsukihara, T. (1979). Processing and post-refinement of oscillation camera data. *J. Appl. Crystallogr.* **12**, 570–581.

Epilogue

John Kendrew met J. D. Bernal when they were both in military service in Southeast Asia during the Second World War. Bernal predicted to John that the great postwar challenge would be to determine the three-dimensional structure of a protein by means of X-ray crystallography. Bernal's final advice to Kendrew was: "When this is over, go home to England and solve the structure of a globular protein." Later, after the myoglobin structure had been solved, Kendrew's comment to this advice was, "And so I did."

Perhaps neither Bernal nor John realized, even after the structure of myoglobin had been determined, that this was the very beginning of a vast scientific revolution. The structures of myoglobin and haemoglobin (Fig. E1) were the distant rolls of thunder, unnoticed by most, of a tremendous storm that was gathering and that would completely change the biological landscape.

After that summer of 1959, nothing much happened in structural biology for several years. The general opinion was that perhaps there would be no further protein structures any time soon. Indeed, the world had to wait until 1967 for the structure of hen egg white lysozyme, determined by David Phillips' team at the Royal Institution in London and overseen by Sir Lawrence Bragg.[1,2] Lysozyme not only had elements of β-sheets as had been predicted by Pauling,[3] but it was also the first enzyme whose structure was determined. Furthermore, the structure was able to put onto a firm basis many of the previous suggestions on how an enzyme can catalyze a chemical reaction.

Fig. E1. John and Max standing behind the forest of rods into which the first sperm-whale myoglobin model was built in the basement of the Cavendish Laboratory. Reprinted with permission from the MRC Laboratory of Molecular Biology.

Over the subsequent few years, a few other enzymes, including carboxypeptidase, chymotrypsin, ribonuclease, papain, insulin, carbonic anhydrase, and lactate dehydrogenase, were determined. We realized that, suddenly, there were quite a lot of data that needed to be made accessible to all. At the Cold Spring Harbor meeting in the summer of 1971 there was a small meeting that included Max Perutz, Jan Drenth, Fred Richards, Walter Hamilton, and Michael Rossmann, which established the Protein Data Bank for the collection of all protein (and later nucleic acid) structures. The holdings of the Protein Data Bank grew only slowly at first, but the rate has increased exponentially to where about a total of 50,000 structures have now been deposited. The major factors that have made this possible are the advent of synchrotron radiation for X-ray diffraction studies, freezing of crystals to greatly reduce radiation damage, and automatic computerized data processing as well as improved methods for phase determination and phase refinement—all making it possible to solve crystal structures within hours of collecting the first diffraction data.

Biological revolution is not only the speed with which structures can now be solved, but also the increasing biological significance of the structures. Structures of viruses, ribosomes (barely known to exist in 1959), and ribozymes (not dreamed of in 1959) have become available. All these varied topics have been found to have a common thread in their evolution, as it has become apparent that many biologically essential functions are based on structures that have changed little since the appearance of life on Earth. The modern crystallographer knows little about space groups or structure factors, but does know a great deal about structure and biology. A new science, called "bioinformatics," has arisen—made necessary by the enormous growth of structural and gene sequence information. Indeed, structural biology has become the unifying factor of just about every aspect of biology.

In 1959, we were a small elite group of friends who had a partial and incomplete vision of what might be the future. Today, structural biologists are on every continent in every country. The biological knowledge exhibited by the newest generation of scientists, whether physicist, chemist or biologist, is very impressive, but he/she will probably know little, if anything, about what happened in 1959.

References

1. Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C., & Sarma, V. R. (1965). Structure of hen egg-white lysozyme. A three-

dimensional Fourier synthesis at 2 Å resolution. *Nature (London)* **206**, 757–761.
2. Blake, C. C. F., Fenn, R. H., Johnson, L. N., Koenig, D. F., Mair, G. A., North, A. C. T., Oldham, J. W. H., Phillips, D. C., Poljak, R. J., Sarma, V. R. & Vernon, C. A., (2001). Historical perspective. How the structure of lysozyme was actually determined. In *International Tables for Crystallography* (Rossmann M. G. & Arnold, E., eds), vol. F. Kluwer Academic Publishers, Dordrecht, pp. 744–772.
3. Pauling, L., & Corey, R. B. (1951). Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc. Natl Acad. Sci. USA.* **37**, 729–740.