The Three-dimensional Structure of a Protein Molecule

The way in which the chain of amino acid units in a protein molecule is coiled and folded in space has been worked out for the first time. The protein is myoglobin, the molecule of which contains 2,600 atoms

by John C. Kendrew

When the early explorers of America made their first landfall, they had the unforgettable experience of glimpsing a New World that no European had seen before them. Moments such as this—first visions of new worlds—are one of the main attractions of exploration. From time to time scientists are privileged to share excitements of the same kind. Such a moment arrived for my colleagues and me one Sunday morning in 1957, when we looked at something no one before us had seen: a three-dimensional picture of a protein molecule in all its complexity. This first picture was a crude one, and two years later we had an almost equally exciting experience, extending over many days that were spent feeding data to a fast computing machine, of building up by degrees a far sharper picture of this same molecule. The protein was myoglobin, and our new picture was sharp enough to enable us to deduce the actual arrangement in space of nearly all of its 2,600 atoms.

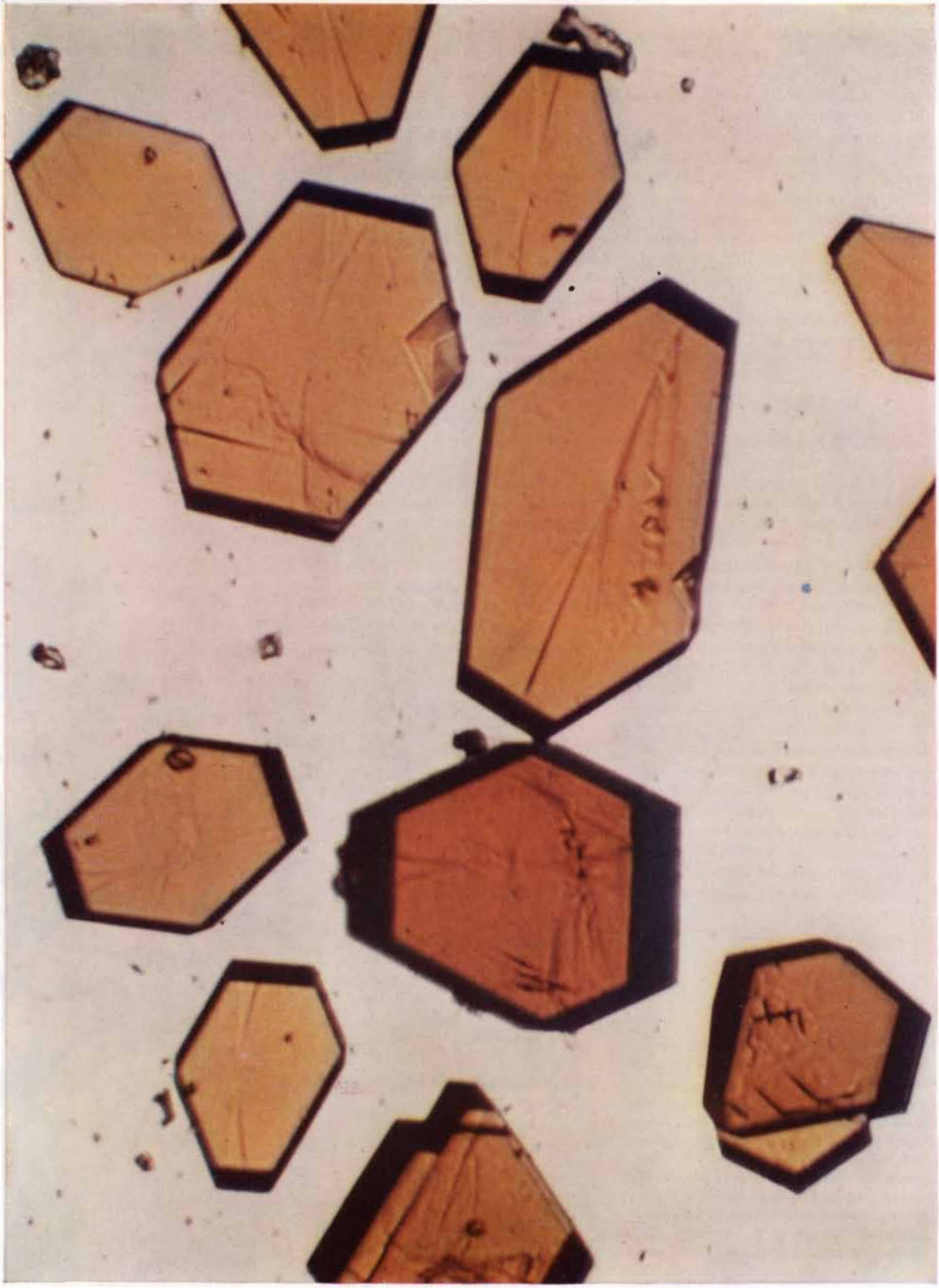
We had chosen myoglobin for our first attempt because, complex though it is, it is one of the smallest and presumably the simplest of protein molecules, some of which are 10 or even 100 times larger. The purpose of this article is to indicate some of the reasons why we thought it important to elucidate the three-dimensional architecture of a protein, to explain something of the methods we used and to describe our results.

In a real sense proteins are the "works" of living cells. Almost all chemical reactions that take place in cells are catalyzed by enzymes, and all known enzymes are proteins; an individual cell contains perhaps 1,000 different kinds of enzyme, each catalyzing a different and specific reaction. Proteins have many other important functions, being constituents of bone, muscle and tendon, of blood, of hair and skin and membranes. In addition to all this it is now evident that the hereditary information, transmitted from generation to generation in the nucleic acid of the chromosomes, finds its expression in the characteristic types of protein molecule synthesized by each cell. Clearly to understand the behavior of a living cell it is necessary first to find out how so wide a variety of functions can be assumed by molecules all made up for the most part of the same few basic units.

These units are amino acids, about 20 in number, joined together to form the long molecular chains known as polypeptides. Each link in a chain consists of the group —CO—CHR—NH—, where C, O, N and H represent atoms of carbon, oxygen, nitrogen and hydrogen respectively, and R represents any of the various groups of atoms in a side chain that differs for each of the 20 amino acids. All protein molecules contain polypeptide chains, and some of them contain no other constituents; in others there is an additional group of a different kind. For example, the hemoglobin in red blood corpuscles contains four polypeptide chains and four so-called heme groups: flat assemblages of atoms with an iron atom at the center. The function of the heme group is to combine reversibly with a molecule of oxygen, which is then carried by the blood from the lungs to the tissues. Myoglobin is, as it were, a junior relative of hemoglobin, being a quarter its size and consisting of a single polypeptide chain of about 150 amino acid units together with a single heme group. Myoglobin is contained within the cells of the tissues, and it acts as a temporary storehouse for the oxygen brought by the hemoglobin in the blood.

Following the classic researches on the insulin molecule by Frederick Sanger at the University of Cambridge, several groups of investigators have been able to discover the order in which the amino acids are arranged in the polypeptide chains of a number of proteins [see "The Chemical Structure of Proteins," by William H. Stein and Stanford Moore; SCIENTIFIC AMERICAN, February]. This laborious task does not, however, provide the whole story. A polypeptide chain of perhaps hundreds of links could be arranged in space in an almost infinite number of ways. Chemical methods give only the order of the links; equally important is their arrangement in space, the way in which particular side chains form crosslinks to bind the whole structure together into a nearly spherical object (as most proteins are known to be). Also of equal importance is the way in which certain key amino acid units, perhaps lying far apart in the sequence, are brought together by the three-dimensional folding to form a particular constellation of precise configuration—the so-called active site of the molecule—that enables the protein to perform its special functions. How is it possible to discover the three-dimensional arrangement of a molecule as complicated as a protein?

The key to the problem is that many proteins can be persuaded to crystallize, and often their crystals are as regular and as nearly perfect in shape as the crystals of simpler compounds. The fact that pro-



CRYSTALS OF MYOGLOBIN prepared from sperm-whale muscle are enlarged some 50 diameters. In them the molecules of myoglobin are stacked in regular array. By directing a beam of X rays at a single crystal and analyzing the pattern of the reflected rays, the author and his colleagues were able to plot the density of electrons in the molecule and thereby to locate the atoms in it.

teins crystallize is interesting in itself, for crystallization implies a regular three-dimensional array of identical molecules. If all the molecules did not have the same detailed shape, they could not form the repeating arrays that are necessary if the aggregate is to possess the regular external shape of a crystal. Therefore it appears that all protein molecules of a given type are identical—that is, they are not simply "colloidal" aggregates of indefinite shape. The existence of protein crystals means, in fact, that proteins do have a definite three-dimensional structure to solve. And the most powerful techniques for studying the structures of crystals are those of X-ray crystallography.

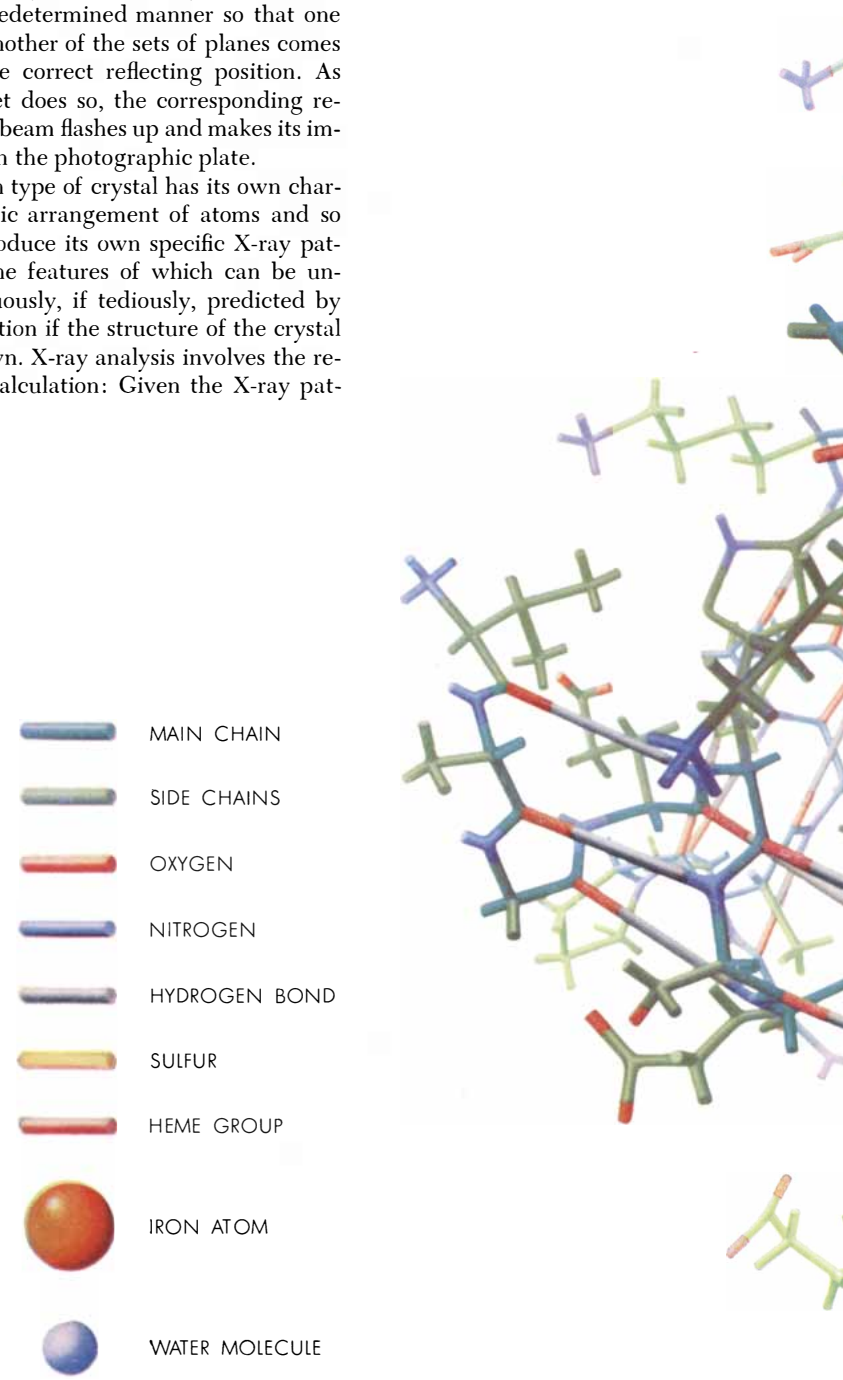
The X-Ray Approach

In 1912 Walter Friedrich, C. M. Paul Knipping and Max von Laue discovered that if a crystal is turned in various directions while a beam of X rays is sent through it, some of the X rays do not travel in a straight line. When the transmitted rays fall on a photographic plate, they produce not only a dark central spot but also a pattern of fainter spots around it. The reason for this diffraction pattern is that X rays are scattered, or reflected, by the electrons that form the outer part of each atom in the crystal.

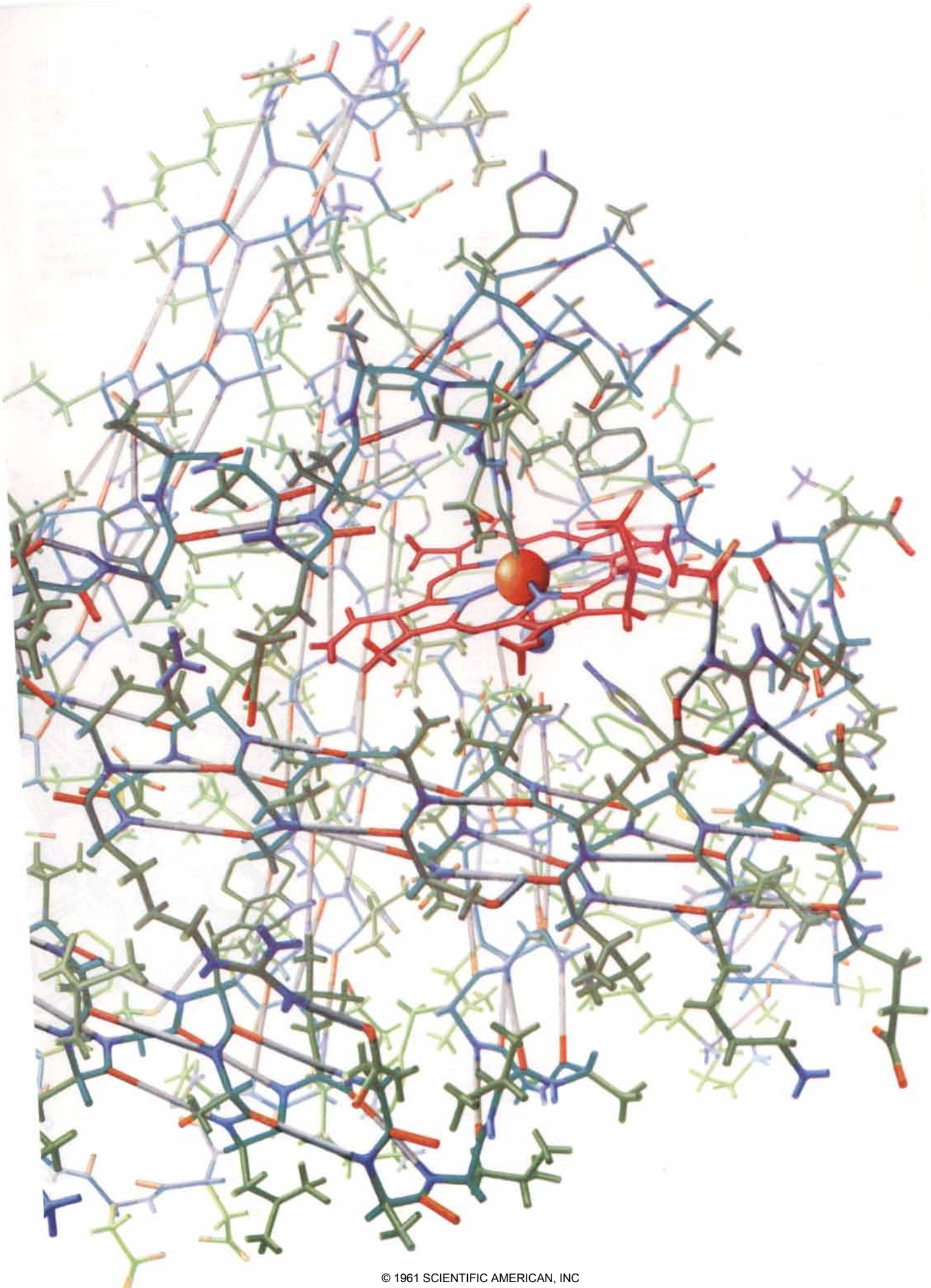
The atoms are arranged in an orderly array, something like the trees in a regularly planted orchard. As one drives past an orchard in an automobile and looks into it along different directions, one sees one set after another of lines of trees coming into view end on. Similarly, if one could look at the atoms of a crystal, one would see different planes of atoms in different directions. The X-ray beam is reflected by these sets of planes much as light is reflected from the surface of a mirror; that is to say, the angle of reflection is equal to the angle of incidence. But it can be shown that because the reflection is a set of parallel planes rather than a single surface, as in a mirror, the reflected beam will "flash up" only at a particular angle of incidence between incident beam and planes, this angle becoming greater the closer together the planes of the set are. Thus each spot in the X-ray diffraction pattern corresponds to a particular set of planes; and the spots farthest out in the pattern (those made by X rays diffracted through the biggest angles) correspond to the most closely spaced sets of planes. In an X-ray camera the crystal is rotated in a predetermined manner so that one after another of the sets of planes comes into the correct reflecting position. As each set does so, the corresponding reflected beam flashes up and makes its imprint on the photographic plate.

Each type of crystal has its own characteristic arrangement of atoms and so will produce its own specific X-ray pattern, the features of which can be unambiguously, if tediously, predicted by calculation if the structure of the crystal is known. X-ray analysis involves the reverse calculation: Given the X-ray pattern, what is the crystal structure that must have produced it?

In analyzing complex crystals the calculation is carried out by applying a method known as Fourier synthesis to the repeating, three-dimensional configuration. To understand what this involves, consider first a one-dimensional analogy: a musical note. Physically, a steady musical note is a repeating sequence of rarefactions and condensations in the air between the listener and the instrument producing the note. If the den-



THREE-DIMENSIONAL MODEL of the myoglobin molecule is depicted in this painting by Irving Geis. The key to the model is at the left side of the painting. The molecule consists of some 150 amino acid units strung together in a single chain with a heme group attached to it. At the center of the heme group is a single atom of iron. Most of the amino acid units are arranged in helical sections such as the one running diagonally across the bottom of the painting. Each amino acid unit in the model is identified in the illustration on the following two pages. The model is the result of work by the author, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, V. C. Shore and H. C. Watson.



sity of the air is plotted along the path, the graph is a complex but perfectly repetitive wave form. More than 150 years ago the French physicist Jean Baptiste Fourier discovered that any such wave form can be decomposed, or analyzed, into a set of harmonics that are pure sine waves of shorter and shorter wavelength and thus of higher and higher frequency [see "The Reproduction of Sound," by Edward E. David, SCIENTIFIC AMERICAN, August]. The reverse process—Fourier synthesis—consists in combining a series of pure sine waves of the proper relative amplitude in the proper relative phases (that is, in or out of step with one another to the correct extent) so as to reproduce the original wave form, or note. In practice it is not necessary to use all the components to obtain a reasonably faithful reproduction. The greater the number of higher harmonics that are included, however, the more nearly perfect is the rendering of the note.

To an X-ray beam a crystal is an extended electron cloud, the density of

ALANINE	ALA
ARGININE	ARG
ASPARTIC ACID OR ASPARAGINE	ASP
GLUTAMIC ACID OR GLUTAMINE	GLU
GLUTAMIC ACID	GLU. C
GLYCINE	GLY
HISTIDINE	HIS
ISOLEUCINE	ILEU
LEUCINE	LEU
LYSINE	LYS
METHIONINE	MET
PHENYLALANINE	PHE
PROLINE	PRO
SERINE	SER
THREONINE	THR
TYROSINE	TYR
VALINE	VAL

A- 1 VAL (AMINO END)
2 ALA
3 GLY
4 GLU
5 TRY
6 SER
7 GLU
8 ILEU
9 LEU
10 LYS
11 (NOT GLY)
12 TRY
13 (NOT GLY)
14 LEU
15 LEU
16 GLU

AB- 1 (NOT GLY)

B- 1 LEU
2 VAL OR THR
3 ALA
4 GLY
5 HIS
6 GLY
7 LYS
8 LEU
9 THR
10 LEU
11 ILEU
12 SER
13 LEU
14 PHE
15 LYS
16 SER

C- 1 HIS
2 PRO
3 GLU.C
4 THR
5 LEU
6 GLU
7 LYS

CD- 1 PHE
2 ASP
3 ARG
4 PHE
5 LYS
6 HIS
7 LEU
8 LYS

D- 1 THR
2 GLU.C
3 ALA
4 GLU.C
5 MET
6 LYS
7 ALA

E- 1 SER
2 GLU.C
3 ASP
4 LEU
5 LYS
6 VAL
7 HIS
8 GLY
9 ILEU
10 GLU
11 VAL
12 ASP
13 (NOT ALA, GLY)
14 ALA
15 LEU
16 GLY
17 ALA
18 ILEU
19 ASP
20 ARG

EF- 1 LYS
2 LYS
3 GLY
4 LEU
5 HIS
6 (NOT GLY)
7 (NOT GLY)
8 GLU

F- 1 GLU
2 ALA
3 PRO
4 THR
5 ALA
6 HIS
7 SER
8 HIS
9 ALA

FG- 1 (NOT GLY)
2 (NOT GLY)
3 PHE
4 (NOT ALA)
5 ILEU

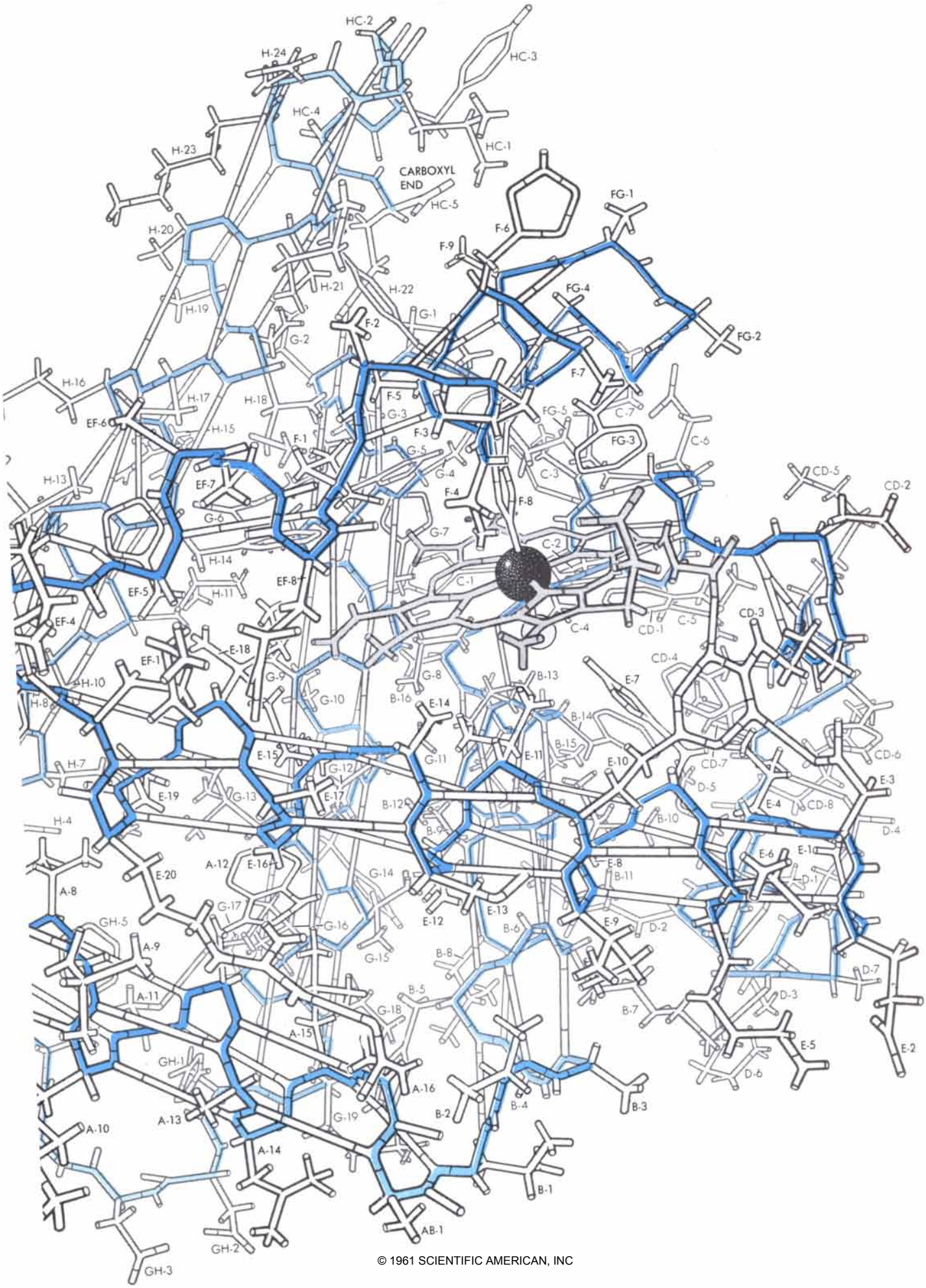
G- 1 PRO
2 ILEU
3 LYS
4 TYR
5 (NOT ALA, GLY)
6 GLU
7 HIS
8 LEU
9 SER
10 (NOT GLY, ALA)
11 ALA
12 VAL OR THR
13 ILEU
14 HIS
15 VAL
16 ARG
17 ALA
18 THR
19 LYS

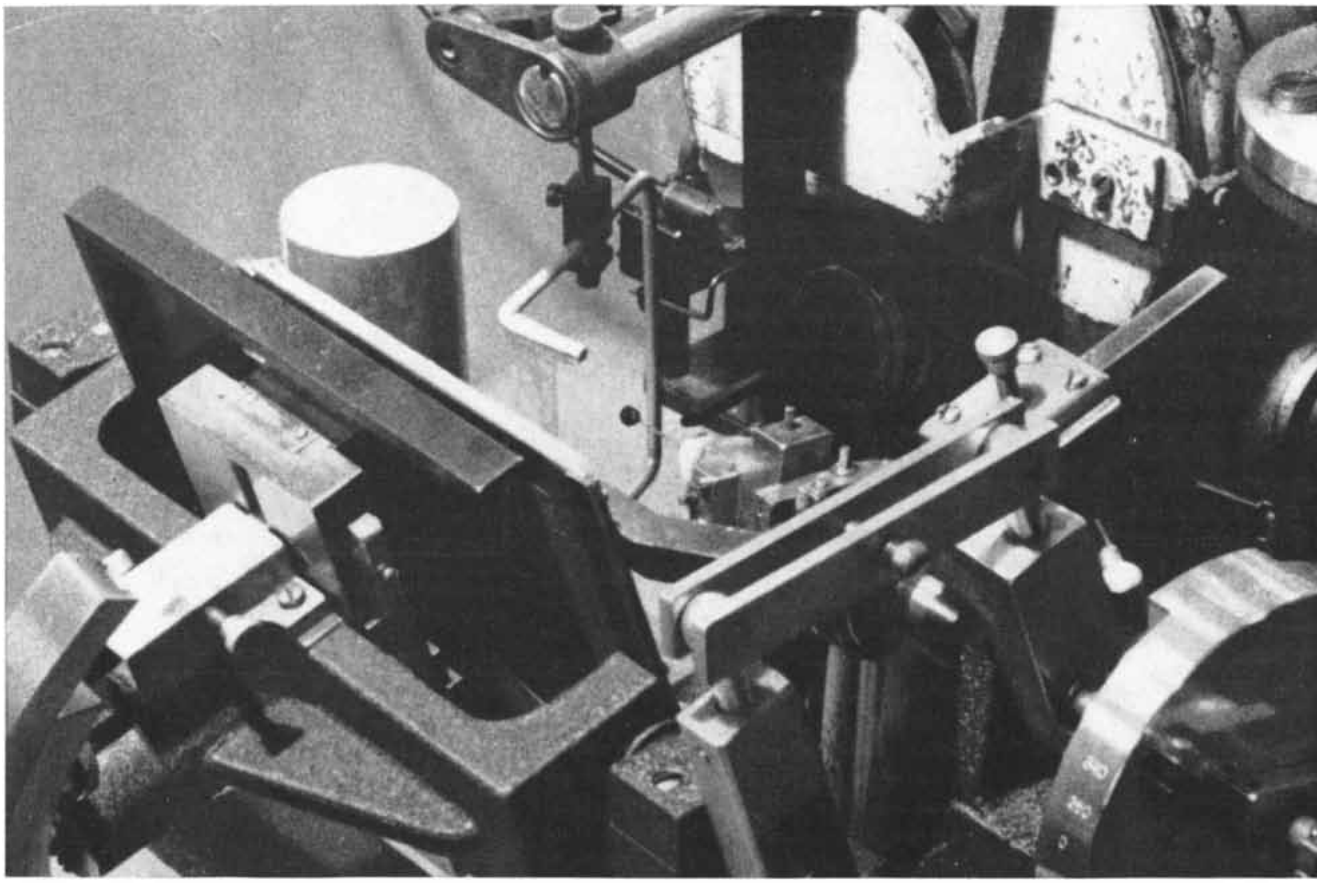
GH- 1 HIS
2 ASP
3 ASP
4 GLU
5 PHE
6 GLY

H- 1 ALA
2 PRO
3 ALA
4 ASP
5 GLY
6 ALA
7 MET
8 GLY
9 LYS
10 ALA
11 LEU
12 GLU.C
13 LEU
14 PHE
15 ARG
16 LYS
17 ASP.C
18 ILEU
19 ALA
20 ALA
21 LYS
22 TYR
23 LYS
24 GLU.C

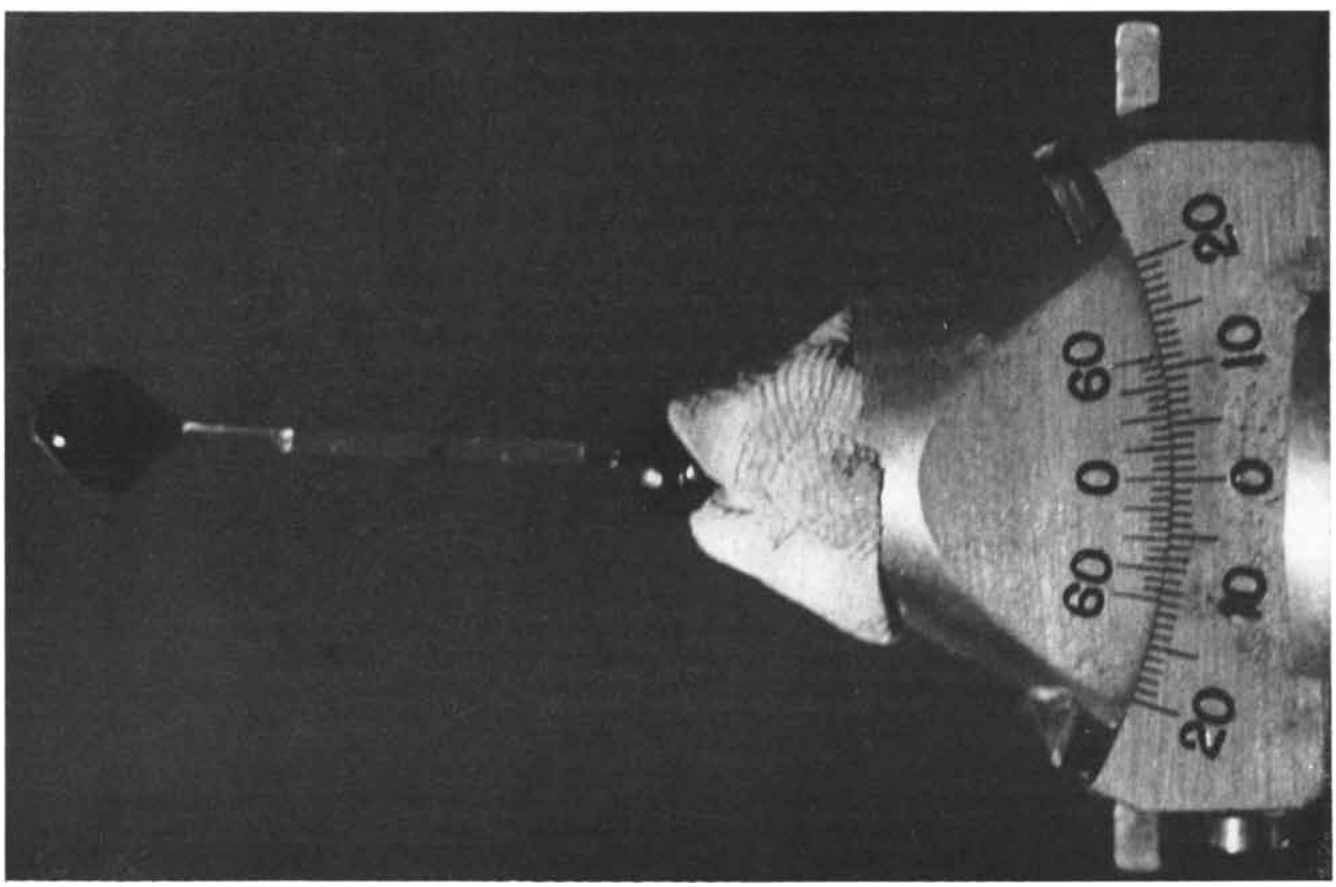
HC- 1 LEU
2 GLY
3 TYR
4 GLY
5 GLU.C (CARBOXYL END)

SEQUENCE OF AMINO ACID UNITS in the model of myoglobin is indicated by the letters and numbers in the illustration on these two pages. The amino acid unit represented by each symbol is given in the table above; the key to the abbreviations is at top left in the table. The brackets in the table indicate those amino acid units which form a helical section. The direction of the main chain is traced in color in the illustration; the chain begins at far left (*amino end*) and ends near the top (*carboxyl end*). Here the heme group is indicated in gray. Not all the amino acid units in the model have been positively identified. In some cases it has only been determined that they cannot be certain units. The over-all configuration of the molecule, however, is known with a considerable degree of confidence.





SPECIAL X-RAY CAMERA is used to make X-ray diffraction photographs of the myoglobin crystal. The crystal is contained in the thin glass tube exactly at the center of the photograph. The beam of X rays comes out of metal tube just to the right of glass one.



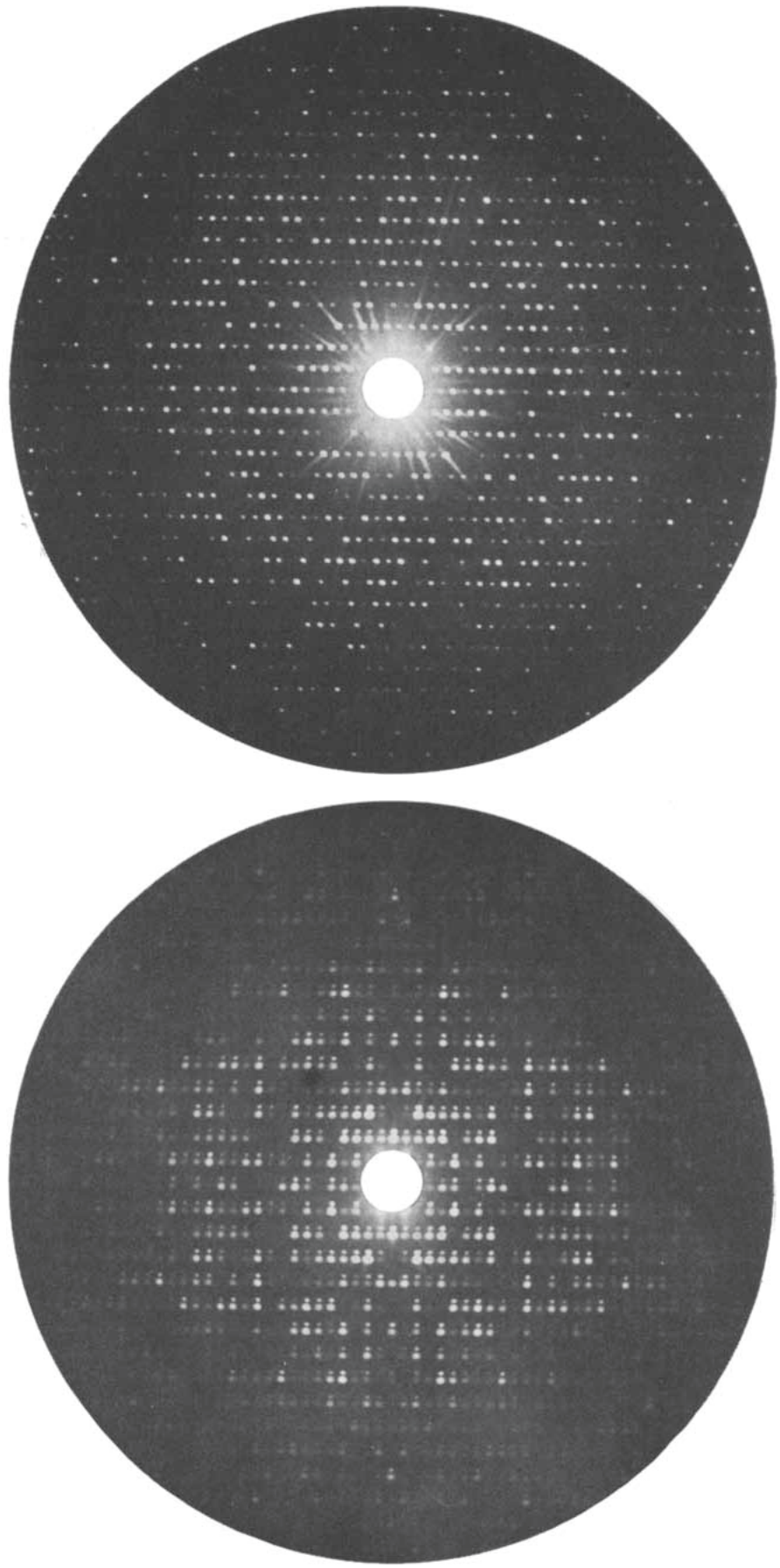
CRYSTAL OF MYOGLOBIN is the dark speck in the middle of the glass tube in this close-up. As the X-ray exposure is made the crystal is rotated so that the X rays reflected from planes of atoms in the crystal "flash up" to make spots on a photographic plate.

which varies from place to place in three dimensions but in a regular, repeating way. (The density at any point depends on the types of atom in the neighborhood and their spatial arrangement.) The crystal can therefore be thought of as a kind of three-dimensional sound wave consisting of rarefactions and condensations of electrons rather than of air particles. This wave too can be decomposed into harmonics—that is, simple, repetitive patterns of density variation; along any single direction the density of each harmonic varies sinusoidally.

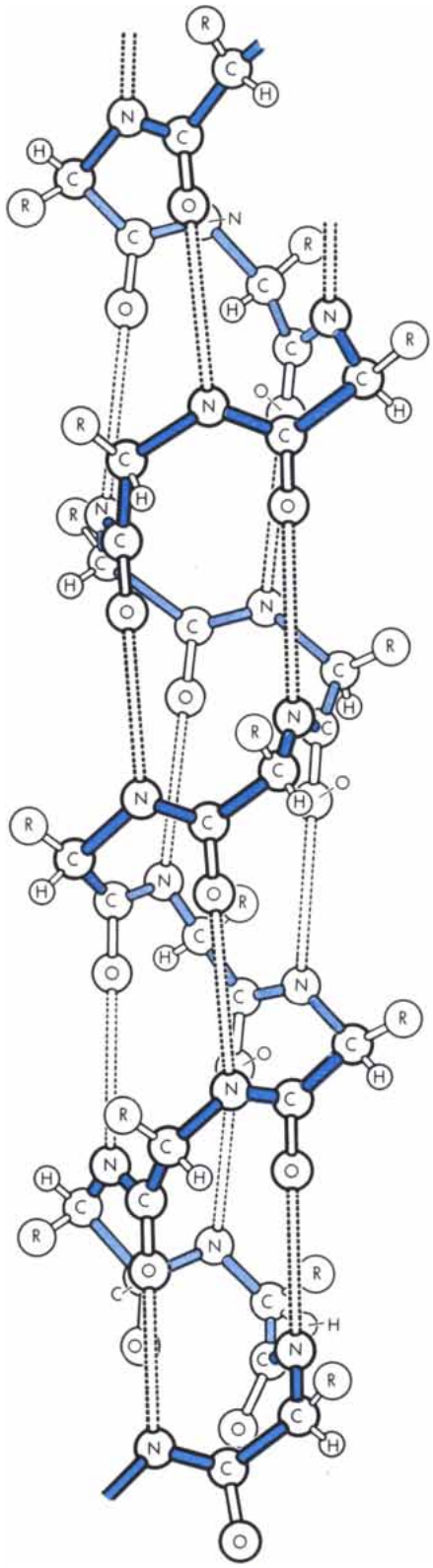
It turns out that each harmonic corresponds to one particular spot in the X-ray diffraction pattern. The reason is that each set of possible atomic planes in the crystal constitutes one element, so to speak, of the over-all periodic structure. That is just what a harmonic is: a component of the over-all periodic structure. From the position and darkness of a spot, the "wavelength" (the spacing between high-density peaks) and "amplitude" (the value of the density at the peaks) of the corresponding harmonic can be computed. So the problem is reduced to calculating the harmonics from the spot pattern and then adding them together to arrive at the total structure.

There is, however, a serious catch: the diffraction pattern provides information on the wavelength and amplitude of the harmonics but not on their relative phases. In the case of sound, phase is not particularly important in synthesizing a wave; the ear is rather insensitive to phase difference and hears very nearly the same note so long as the relative amplitudes of the harmonics are correct. On the other hand, the shape of the wave as seen by the eye varies greatly when the relative phases of the components are shifted.

In deriving crystal structure the correct shape of the three-dimensional "wave" is precisely what one is looking for. But the X-ray picture contains only half of the information required; it contains the amplitudes but not the phases. In simpler structures crystallographers get around the difficulty by a method of trial and error; from a plausible model structure they calculate the phases and use these in conjunction with the measured amplitudes to calculate a Fourier synthesis, that is to say, an enlarged picture of the distribution of the electrons (and hence of the atoms) in the structure. The result should be a good deal closer to the real structure than the original model, and from it crystallographers can calculate a new and improved set of phases. If the original model was good enough to put them on



X-RAY PHOTOGRAPHS of myoglobin are patterns of spots. At top is a photograph of a normal crystal. At bottom is a photograph of a different type in which patterns of normal crystal and one labeled with heavy atoms are superimposed slightly out of register. Differences in density between two sets made it possible to determine phase of X-ray reflections.



ALPHA HELIX of a protein molecule is a coiled chain of amino acid units. The backbones of the units form a repeating sequence of atoms of carbon (*C*), oxygen (*O*), hydrogen (*H*) and nitrogen (*N*). The *R* stands for the side chain that distinguishes one amino acid from another. The configuration of the helix is maintained by hydrogen bonds (*broken lines*). The hydrogen atom that participates in each of these bonds is not shown.

the right track, they gradually approach the true structure by a series of successive approximations, or refinements.

As in the case of the musical note, the greater the number of higher harmonics that are included, the sharper and more precise is the resulting picture. The higher harmonics of a musical note are, of course, the components of shortest wavelength (highest frequency); in a crystal structure the harmonics are correspondingly the reflections from the most closely spaced sets of planes. As has been mentioned, these reflections occur at the largest angles and show up as spots farthest out in the pattern. The resolution of the final picture—that is, the smallest scale of detail it can show—depends on the outer limit of the spots included in the analysis; the number of spots that have to be included goes up in proportion to the cube of the resolving power required.

The first X-ray photographs of protein crystals were made nearly 25 years ago, but for many years it was not possible even in principle to imagine how the structures of crystals so complex could be discovered. Their X-ray patterns contained many thousands of reflections, paralleling the complexity of the molecules themselves. There was no hope of proceeding by trial and error; the first model could never be good enough to provide a useful starting point. So although protein crystallographers discovered many interesting facts about protein crystals, they did not succeed in extracting much information bearing directly on the molecular structure.

In 1953 the whole prospect was transformed by a discovery of my University of Cambridge colleague Max F. Perutz, who had been studying hemoglobin crystals for many years. The hemoglobin molecule contains two free sulfhydryl groups (SH) of the amino acid cystine; by well-known reactions it is possible to attach atoms of mercury to these groups. Perutz found that if he made crystals of hemoglobin labeled with mercury, their X-ray pattern differed significantly from that of unlabeled crystals, even though the mass of a mercury atom is very small compared with that of a complete hemoglobin molecule. This made it possible to apply the so-called method of isomorphous replacement in the Fourier synthesis.

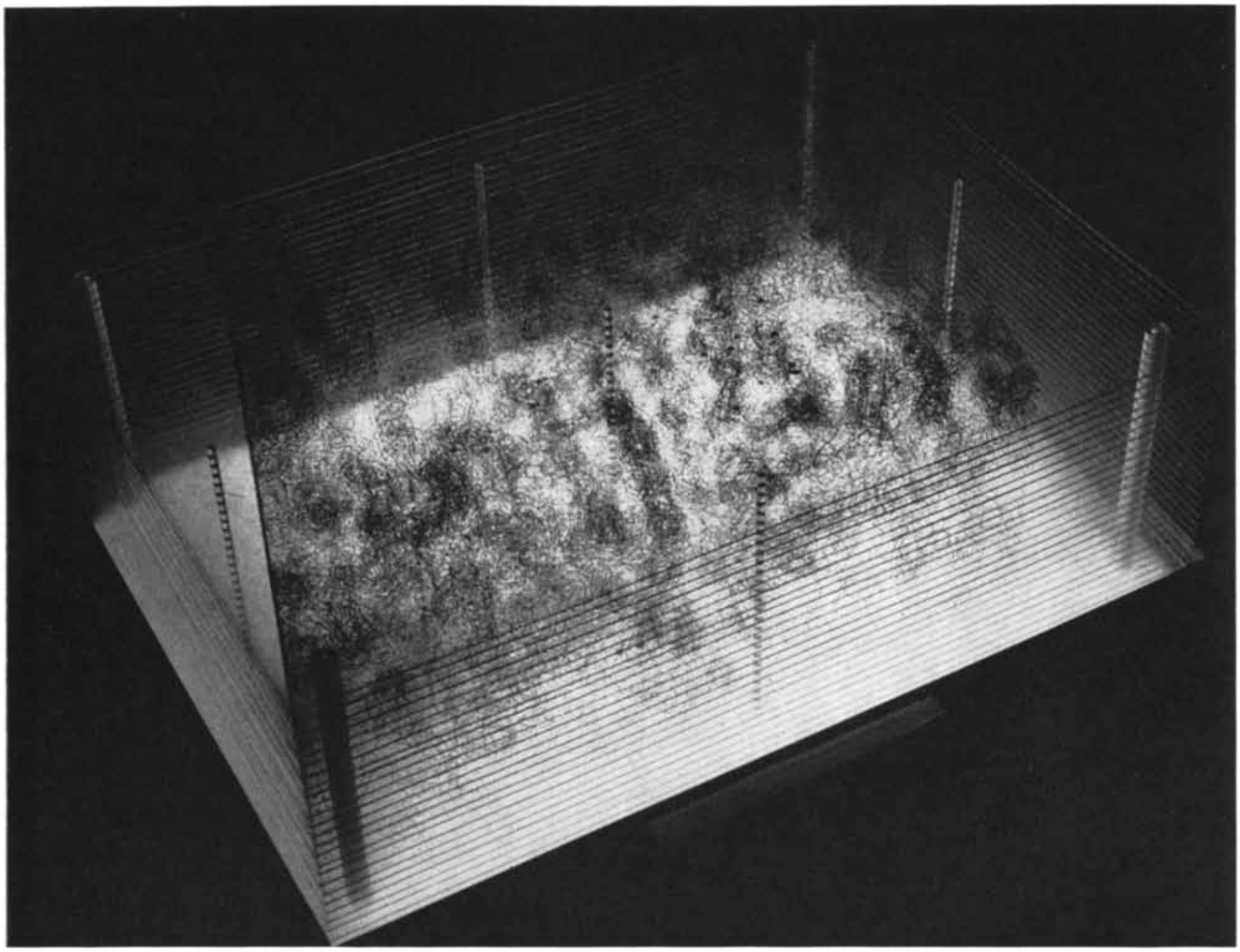
A full explanation of the method is beyond the scope of the present article. Suffice it to say that by comparing in detail the X-ray patterns of crystals with and without heavy atoms it is possible to deduce the phases of all the reflections, and this without any of the guesswork of the trial-and-error method. Thus Perutz' observation for the first time made it possible, in principle at least, to solve the complex X-ray pattern of a protein crystal and to produce a model of the structure of the molecule.

In our studies of myoglobin we could not follow Perutz' method for attaching mercury atoms, because myoglobin lacks free sulfhydryl groups. We were, however, successful in finding other ways to attach four or five kinds of heavy atom at different sites in the molecule, and we were then able to proceed to a study of the three-dimensional structure of the crystal. A complete solution would involve including all the reflections in the X-ray pattern in our calculation—some 25,000 in all. At the time this work began no computers in existence were fast enough or large enough to handle so great an amount of data; besides, we thought it better in the first instance to test the new method on a smaller scale.

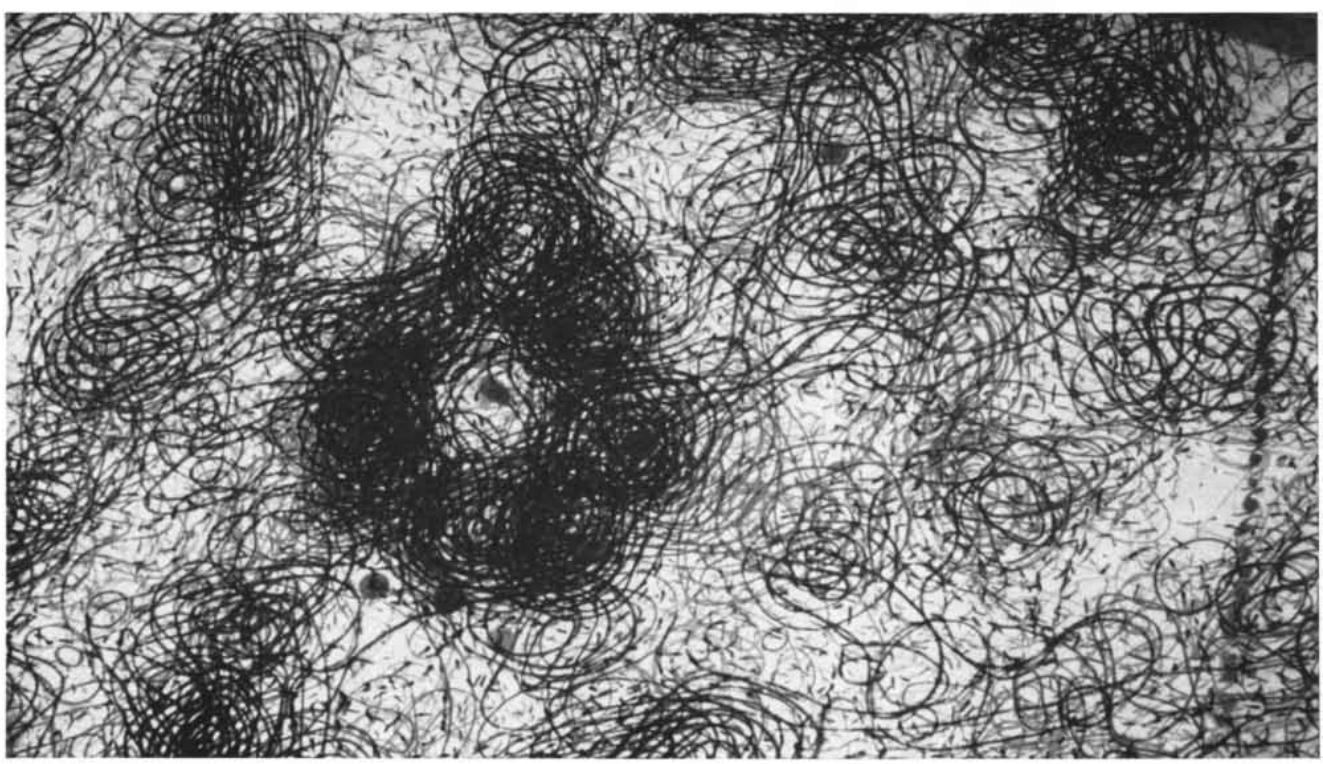
The Six-Angstrom-Unit Picture

As has already been indicated, if we include only the central reflections of the pattern, we obtain a low-resolution, or crude, representation of the structure. The higher the resolution that is desired, the farther out in the pattern must the reflections be measured. We decided that in the first stage of the project we would aim for a resolution of six angstrom units (an angstrom unit is one hundred-millionth of a centimeter). This would be sufficient to reveal the general arrangement of the polypeptide chains in the molecule, but not the configuration of the atoms within the chains or that of the side chains surrounding them. To achieve a six-angstrom resolution we had to measure 400 reflections from the unlabeled protein and from each of five types of crystal containing heavy atoms. Our calculations, which were completed in the summer of 1957, gave us the density of electrons at a large number of points in the crystal, a high electron density being found where many atoms are concentrated. Crystallographers usually represent a three-dimensional density distribution, or Fourier synthesis, by cutting an imaginary series of parallel sections through the structure. The density distribution in each section is represented by a series of density contours drawn on a lucite sheet. When all the sheets are stacked together, they give a representation in space of the density throughout the molecule.

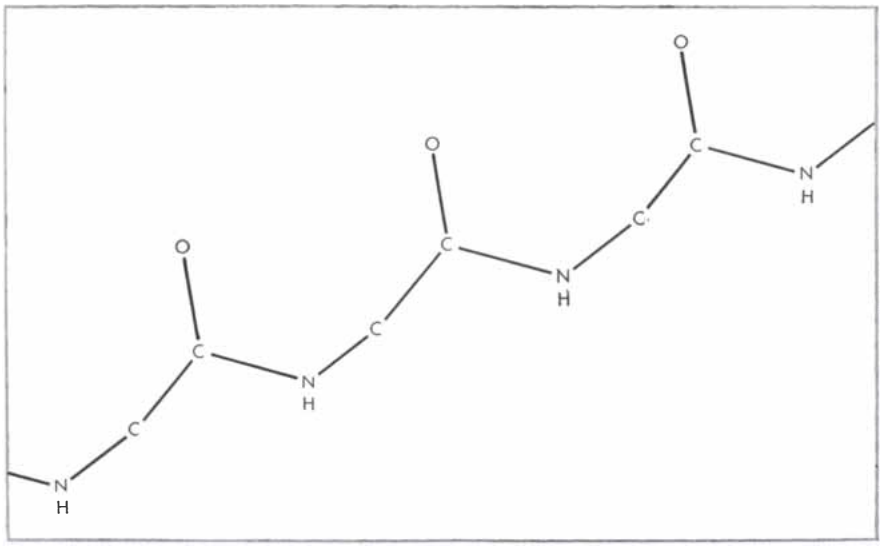
As soon as we had constructed our



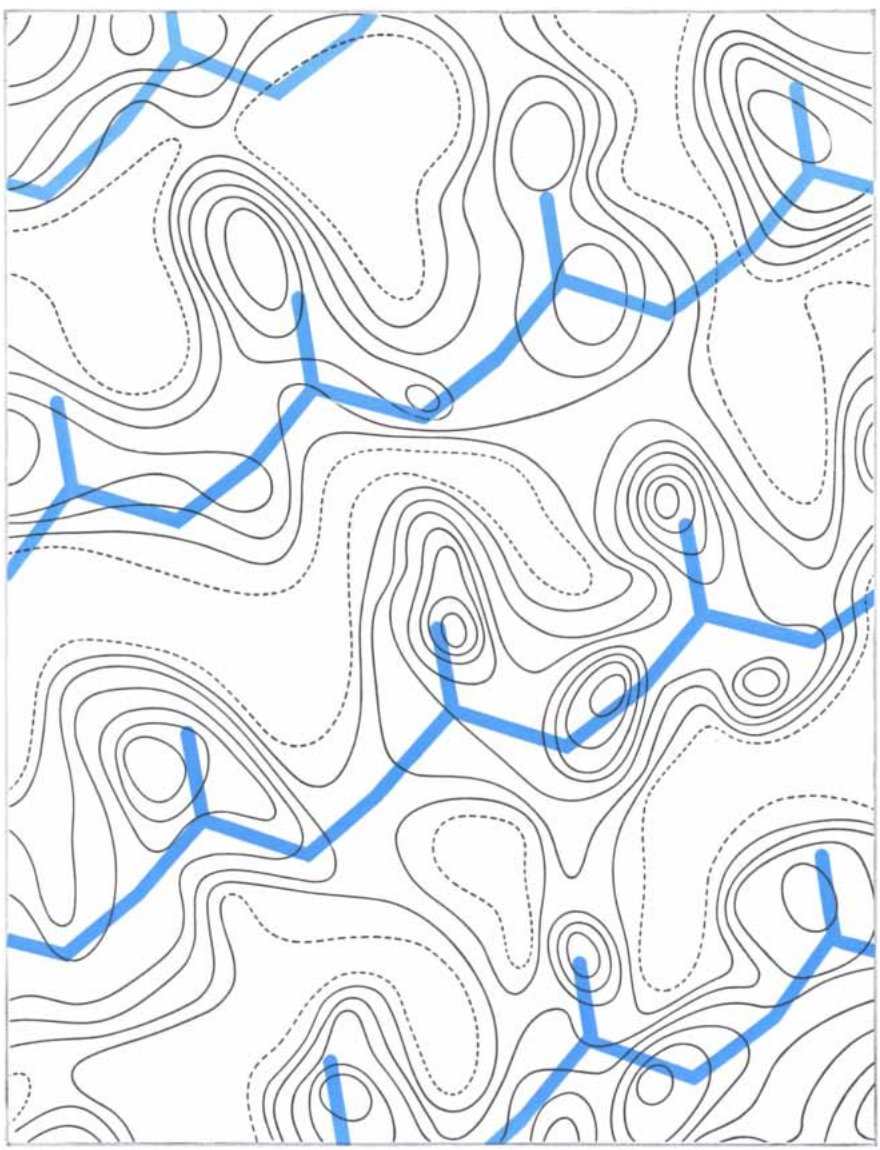
ELECTRON-DENSITY MAP of the myoglobin crystal is made up of lucite sheets, on each of which are traced the contours of electron density at that depth. The dark band in the middle is the heme group. This map was made at a resolution of two angstrom units.



MAP SHOWS ALPHA HELIX when it is seen from the appropriate angle. Here the camera looks through the contours on a series of lucite sheets. Alpha helix is the dark ring at left center. Thus it is seen along its axis, as though it were a cylinder seen from the end.



BACKBONE OF THE ALPHA HELIX is shown schematically in this diagram. The sequence of atoms in the helix is —CO—CHR—NH—. Here the HR attached to isolated C is omitted.



BACKBONE IS SUPERIMPOSED on contours made by plotting on a cylinder density along helix in crystal. Cylinder was then unrolled. Backbone thus appears to repeat.

first lucite density map of the myoglobin crystal, we could see at a glance that it contained the features we were looking for, namely a set of high-density rods of just the dimensions one would expect for a polypeptide chain. Closer examination showed that in fact it consisted of almost nothing but a complicated and intertwining set of these rods, sometimes going straight for a distance, then turning a corner and going off in a new direction. In addition to the rods we were able to see very dense peaks, which we took to be the heme groups themselves. The iron atom at the center of the heme group, being by far the heaviest atom in the molecule and therefore having the largest number of planetary electrons, would be expected to stand out as a prominent feature. It was not at all easy, however, to gain any impression of what the molecule was actually like, largely because the molecules are packed together in the crystal and it is hard to see where one begins and the next one ends.

Our next task was to dissect out a single molecule from the enlarged density map of the crystal so that we could look at it separately. Fortunately all protein crystals, including myoglobin, contain a good deal of liquid which fills up the interstices between neighboring molecules, and which at low resolution looks like a uniform sea of density, so that it can easily be distinguished from the irregular variations of density within the molecule itself. By looking for the liquid regions we were able to draw an outline surface around the molecule and so isolate it from its neighbors. Extracted in this way, the molecule stood forth as the complicated and asymmetrical object shown in the top illustration on page 109. The polypeptide chain winds irregularly around the structure, supporting the heme group in a kind of basket. For the most part the course of the polypeptide chain could be followed, but we could not be sure of its route everywhere, since its density became lower at the corners and it tended to fade into the background at those points. Our model did, however, give us a good general picture of the layout of the molecule, and it showed us that it was indeed much more complicated and irregular than most of the earlier theories of the structure of proteins had suggested.

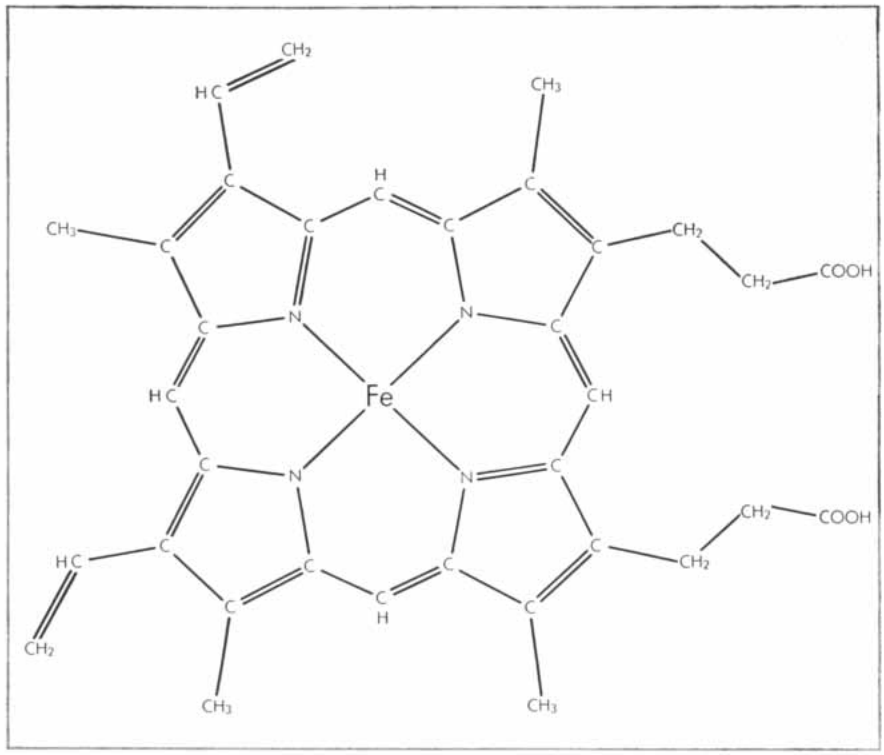
The Two-Angstrom-Unit Picture

At a resolution of six angstroms we could not expect to see any details of the polypeptide chain or of the side chains attached to it. To see all the atoms

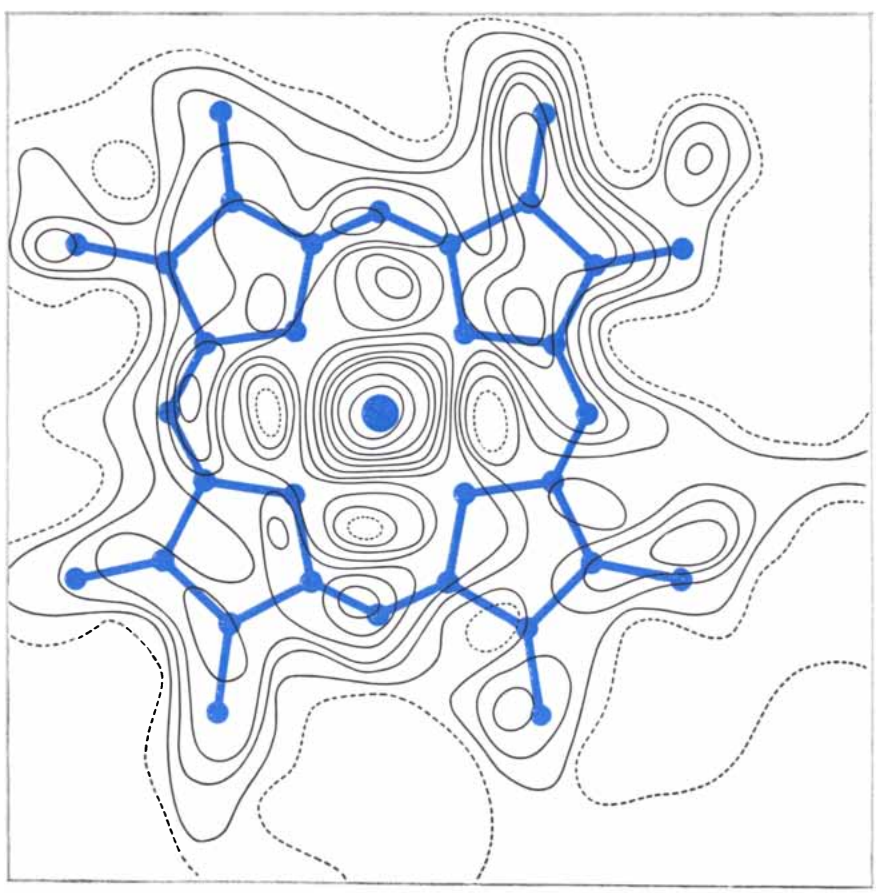
of a structure as separate blobs of density it would be necessary to work at a resolution higher than 1.5 angstroms, for neighboring atoms attached to one another by chemical bonds lie only from one to 1.5 angstroms apart. A Fourier synthesis of myoglobin at 1.5 angstroms resolution would involve 25,000 reflections; we decided in the second stage of the work to limit our ambitions to a resolution of two angstroms. Even this required that we include in our calculations nearly 10,000 reflections for the unsubstituted crystal and the same number for each of the heavy-atom derivatives. It was necessary to measure a formidable number of X-ray photographs, a task that took a team of six people many months to complete. At the end of this time the mass of data that we had accumulated was so great that it could be handled only by a truly fast computer. We were fortunate that a machine of this class—EDSAC Mark II —had recently come into service at the University of Cambridge, and we were able to use it for deducing the phases of all the 10,000 reflections and for the ensuing calculations of the Fourier synthesis itself. Fast though it is, we taxed the powers of EDSAC II to the utmost, and it was clear that any further improvement in resolution would demand the use of still more powerful machines.

Once more our results were plotted in the form of a three-dimensional contour map [*see top illustration on page 105*]. Since we were looking for finer details, it was necessary to cut sections through the structure at closer intervals than before; in fact, this time we had about 50 sections compared to 16 in the six-angstrom map. To construct the new map it was necessary to calculate the electron density at 100,000 points in the molecule. Indeed, the amount of information contained in the final synthesis was so great that drawing and building the density map was in itself a lengthy task, amounting to some six man-months of work. The result was a complicated set of dense and less dense regions that at first sight seemed completely irregular. Our first step was to see what we could learn from it about the configuration of the polypeptide chains, which in our earlier synthesis had appeared merely as solid rods.

Here I shall digress briefly to consider some earlier work on the fibrous protein of hair. In fibrous proteins the polypeptide chains probably run parallel to the axis of the fiber for considerable distances. Such protein fibers were among the earliest biological macromole-



SKELETON OF THE HEME GROUP is outlined in this diagram. At the center of the group is the iron atom (*Fe*). There are four such groups in hemoglobin and one in myoglobin.



SKELETON IS SUPERIMPOSED on another section of the electron-density map of the myoglobin crystal. Here bonds to the iron atom are omitted to show contours around atom.

cules to be examined by X-ray methods. W. T. Astbury, in his classical work carried out at the University of Leeds in the early 1930's, showed that a human hair gave a characteristic X-ray pattern, which on stretching changed reversibly into quite a different pattern. He was able to show that the pattern of stretched hair—the so-called beta form—corresponded to the polypeptide chains being almost fully extended; it followed that in the unstretched, or alpha, form the chains must assume some kind of folded configuration. For many years different workers proposed a succession of more or less unsatisfactory models of the folded chain in unstretched hair, but finally in 1951 Linus Pauling and Robert B. Corey of the California Institute of Technology found the definitive solution, showing that the chain actually took up a helical or spiral shape, the now famous alpha helix [*see illustration on page 104*]. In this configuration the successive turns of the helix are held together by weak hydrogen bonds between NH groups on one turn and CO groups opposite them on the next turn. The alpha helix turned out to be present in several other fibrous proteins besides hair; and although there was no definite proof of the fact, indirect evidence indicated that the alpha helix, or something like it, could exist in the molecule of globular proteins too.

The first thing we wanted to do when we finished our Fourier synthesis at two angstroms resolution was to see whether or not there was anything to the idea of helical structures in a globular protein. Accordingly we looked through the stack of lucite sheets in a direction corresponding to the axis of one of the rods we had seen at low resolution. We were delighted to find that the dense rod now revealed itself as a hollow cylinder running dead straight through the structure [*see bottom illustration on page 105*]. Closer examination showed that the density followed a spiral course along the surface of the cylinder, indicating that the polypeptide chain indeed assumed a helical shape. Detailed measurement of the spiral density showed that it followed precisely the dimensions of the alpha helix deduced by Pauling and Corey 10 years earlier. In fact, it turned out that about three-quarters of the polypeptide chain in the molecule took the form of straight lengths of alpha helix, the helical segments being joined by irregular regions at the corners. In all there were eight such segments, varying in length from seven to 24 amino acid units. In each segment it was possible to fit the alpha helix exactly to the observed density in such a way that we could be reasonably sure of the placing of each atom, even though at this resolution we had not secured full separation between one atom and its neighbors.

The next object of interest was the heme group. Looking at the map from the appropriate angles, we now saw this group as a flat object with a region of high density at the iron atom in the center. A section through the map through the plane of the flat object shows a variation in density that closely follows the known chemical structure of the system of rings in the heme group [*see illustrations on preceding page*].

The Three-dimensional Model

When we came to study our structure in detail, we soon felt the need of a better way to represent the three-dimensional density distribution. We wanted some method that would enable us to fit actual atomic models to the features we could see. Our solution was to erect a forest of steel rods on which we placed colored clips to represent the distribution in space of points of high density, different colors representing different values of the density [*see illustration on page 110*]. The scale of this model was five centimeters per angstrom, so that the whole model would fit in a cube about six feet on a side. Each helical segment of polypeptide chain could be seen as a spiral of colored clips passing through the model, and we were then able to insert actual alpha helices made of skeleton-type models (similar to the familiar ball-and-spoke models but with the balls omitted) and to show that they precisely followed the dense trail of clips. In this way we were able to trace the polypeptide chain from beginning to end, right through the molecule, and to establish its configuration in each of the irregular corners joining neighboring helices. Once the course of the main chain had been delineated with atomic models, we were able to see the side chains emerging from it at appropriate intervals as dense branches of various sizes. At first we thought it unlikely that we would be able to identify many of the side chains, but after some practice we found that in fact we were able to do so surprisingly often. As mentioned earlier, side chains in proteins are of only 20 kinds (in myoglobin only 17), and they are of very different shapes and sizes, ranging from the one in glycine, which is only a single hydrogen atom (invisible to the crystallographer), to the chain in tryptophan, with a double-ring system of 10 carbon and nitrogen atoms. Our problem was reduced to deciding among 17 possible side chains in each case.

Some of our identifications were definite, others were tentative. Fortunately an independent check on our conclusions lay at hand. For several years A. B. Edmundson and C. H. W. Hirs, working in Stein and Moore's laboratory at the Rockefeller Institute, had been trying to establish the amino acid sequence of myoglobin by chemical methods. Their work is still incomplete, but they have broken down the molecule into a set of short pieces, or peptides, the compositions—and in some cases the internal sequences—of which they have determined. The order in which the peptides are arranged in the intact molecule has yet to be established chemically. We have been able, however, to place almost every one of the peptides with certainty in its correct position along the chain by comparing its composition with our X-ray identifications of the side chains. There are virtually no gaps left, nor are there peptides unplaced. Once assigned to their correct positions, the peptides often help to confirm doubtful X-ray identifications, and by putting the two types of evidence together we arrive at a nearly complete amino acid sequence for the whole molecule.

Simply to determine the amino acid sequence was not our main aim in undertaking the X-ray analysis of myoglobin. We were much more concerned with the three-dimensional arrangement of the side chains in the molecule and with the interactions between them that produce and maintain the molecule's characteristic configuration. To study these interactions we undertook to make a model of the whole molecule, with every side chain in place [*see illustration on pages 98 and 99*]. The result was an object still more complex than the low-resolution model, although of course all the features of the latter are still apparent in the former. We can now discern many of the types of interaction that protein chemists have postulated on the basis of physicochemical studies. For example, positively charged basic groups such as those of lysine and arginine are held by electrostatic attraction close to negatively charged acid groups such as aspartic or glutamic acid; several types of hydrogen-bond interaction can be seen, among them NH groups in the main chain bonded to the oxygen atom of serine or threonine; and everywhere we find a close interlocking of hydrocarbon groups such as CH_2 or CH_3, giving rise to the so-called van der Waals' attraction. The structure is not yet sufficiently complete in all details to allow a

EARLY MODEL of the myoglobin molecule was made at a resolution of six angstrom units. This model has the same general configuration as that of the model depicted on pages 98 and 99, but it lacks detail. The heme group is the flat section at upper right.

CLOSE-UP OF CONTOURS of map on which the six-angstrom model was based shows that contours are coarser than those in two-angstrom map. Early model was based on work of author, G. Bodo, H. M. Dintzis, R. G. Parrish, H. W. Wyckoff and D. C. Phillips.

full analysis of the interactions, but at least we can now see the general pattern of forces that maintains the integrity of the molecule.

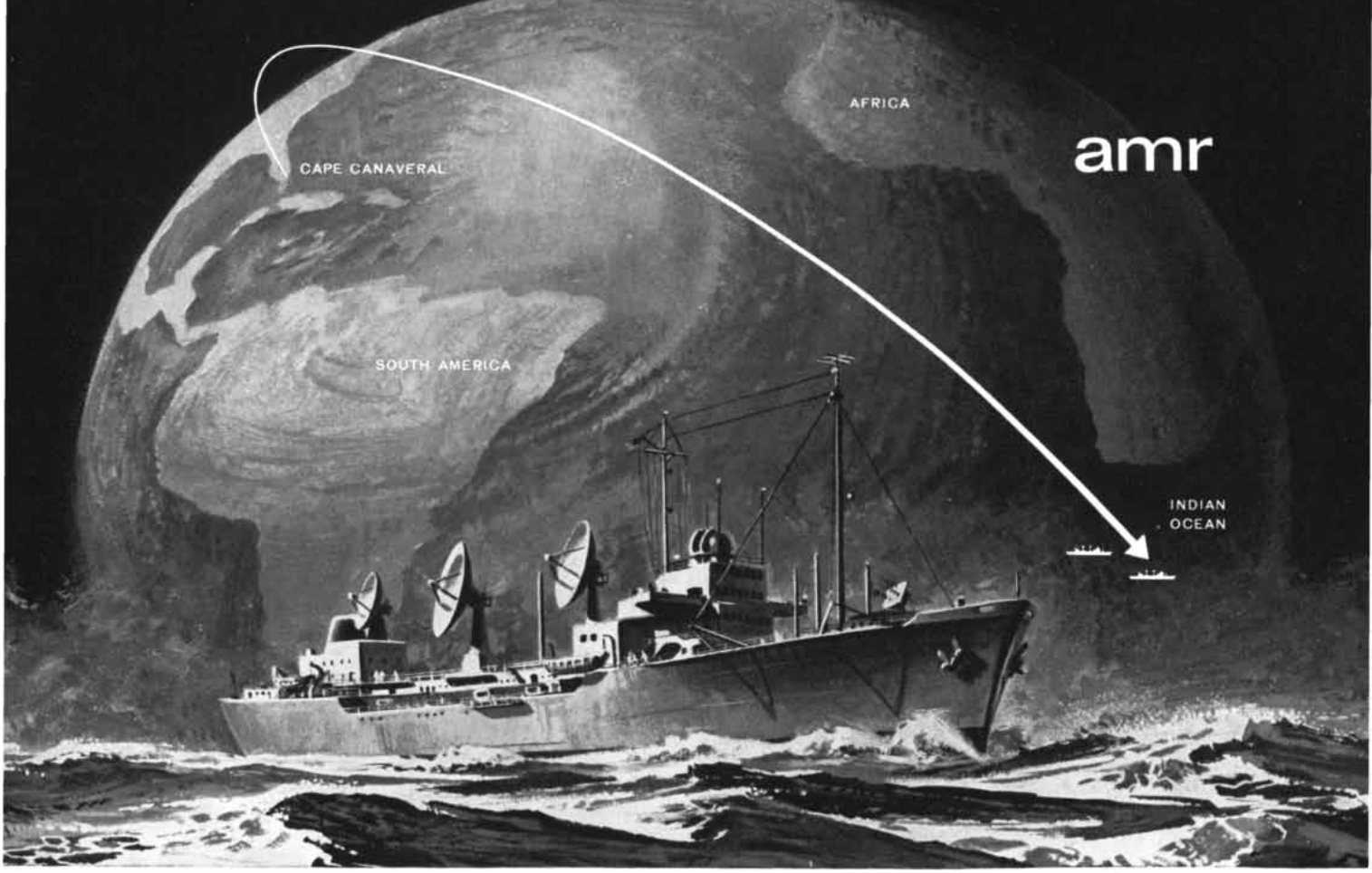
We can also often see why helical segments end at a particular place; in many instances proline side chains are found at the ends of helices and, as was pointed out several years ago, proline is bound to interfere with helix formation because of its peculiar shape, unlike that of any other naturally occurring amino acid. Finally, we can examine the way in which the heme group itself is attached to the rest of the molecule; the iron atom is attached to a nitrogen in a histidine side chain (as had been suggested years ago on the basis of chemical evidence), and the flat ring system is stabilized by hydrocarbon side chains, especially ring side chains, lying parallel to it.

In similar studies of the larger hemoglobin molecule Perutz and his collaborators have shown that, at least to the resolution of 5.5 angstroms that they have so far achieved, there is an astonishing similarity between the three-dimensional structure of myoglobin and the structure of each of the four subunits formed by the individual polypeptide chains of hemoglobin. This is a most remarkable result considering that we are dealing with two distinct proteins, one found in muscle and the other in red blood cells, one derived from sperm whale and the other from horse. Furthermore, the amino acid compositions of the two proteins are known to differ substantially.

The amino acid sequences of the hemoglobin chains have been completely determined. We have found that when we lay the hemoglobin sequences alongside those of myoglobin, making appropriate allowances for slight differences in their length, there are many correspondences, often just at those points where a study of the myoglobin molecule indicates that a crucial stabilizing reaction takes place. We can even begin to find chemical explanations for some of the peculiarities of the congenitally abnormal hemoglobins present in individuals suffering from certain rare blood diseases.

Even in the present incomplete state of our studies on myoglobin we are beginning to think of a protein molecule in terms of its three-dimensional chemical structure and hence to find rational explanations for its chemical behavior and physiological function, to understand its affinities with related proteins and to glimpse the problems involved in explaining the synthesis of proteins in living organisms and the nature of the malfunctions resulting from errors in this process. It is evident that today students of the living organism do indeed stand on the threshold of a new world. Analyses of many other proteins, and at still higher resolutions (such as we hope soon to achieve with myoglobin), will be needed before this new world can be fully invaded, and the manifold interactions between the giant molecules of living cells must be comprehended in terms of well-understood concepts of chemistry. Nevertheless, the prospect of establishing a firm basis for an understanding of the enormous complexities of structure, of biogenesis and of function of living organisms in health and disease is now distinctly in view.

FOREST OF RODS was used to build up the two-angstrom model of the myoglobin molecule from electron-density map. Densities were indicated by clips on rods, and model was based on position of clips. Outline of heme group is visible at upper left center.



Objective: increase capability of the Atlantic Missile Range – sponsored by the Missile Test Center of the Air Force Systems Command – to study missiles and space vehicles during the critical terminal phase of flight with a refinement never before possible. Prime requisite: *mobility* . . . ability to move to the most advantageous position in thousands of miles of ocean.

The answer: conversion of two large troop transports into Mobile Atlantic Range Stations, seagoing laboratories incorporating every proven technique of instrumentation, data acquisition, and instantaneous data transmission to Cape Canaveral, integrated with the most sophisticated long range tracking radars. System Manager: The Sperry Rand Systems Group. General Offices: Great Neck, N. Y.

The Mobile Atlantic Range Stations will be equipped with integrated radar, telemetry, infrared tracking, data handling, communications, inertial navigation, and sophisticated weather forecasting. Team members include Bethlehem Shipbuilding, Ford Instrument, Gibbs & Cox, IT&T, Remington Rand UNIVAC, Sperry Gyroscope Company.

SPERRY