JOHN C. KENDREW

Myoglobin and the structure of proteins

Nobel Lecture, December 11, 1962

When I first became interested in the question of solving the structure of proteins, during the latter part of the Second world War, I had no doubt that this problem above all others deserved the attention of anyone concerned with fundamental aspects of biology. Had my interests been awakened a few years later I would, no doubt, have recognized that there were in fact two such basic unanswered questions, the structure of proteins and the structure of nucleic acids. As events have turned out, the second question was posed later and answered sooner. For me in the early 1940's, however, there seemed to be only one question uniquely qualified to engage the interest of anyone wishing to apply the disciplines of physics and chemistry to the problems of biology. It also seemed that the only technique offering any chance of success in determining the structures of molecules so large and complex as proteins was that of X-ray crystallography. Looking back on that time it occurs to me that my own almost total ignorance of this method was fortunate, in that it concealed from me the extent to which contemporary X-ray crystallographic techniques fell short of what was needed to solve the structures of molecules containing thousands of atoms; it was indeed a case of ignorance being bliss. For a number of years, this situation persisted - many roads were explored, but none of them seemed to offer real hope of a definitive solution - until my colleague Dr. Max Perutz showed that the method of isomorphous replacement, until then applied rather rarely in crystallography generally, and never in the field under discussion, was in fact ideally suited to the protein problem. His first successful application of this method to the haemoglobin structure provided the basis of all subsequent work in the field, my own included. Perutz has included an account of the method in his lecture, and in the present discussion I shall therefore refer to questions of methodology only in so far as they have special relevance to my own work.

As I have indicated, my choice of problem and of method seemed straightforward. The choice of material was not so simple. One looked for a protein of low molecular weight, easily prepared in quantity, readily crystallized,

and not already being studied by X-ray methods elsewhere. Myoglobin seemed to satisfy these criteria, and had the additional advantages of being closely related to haemoglobin, already the object of Perutz's attention for many years, and of sharing with haemoglobin a most important and interesting biological function, that of reversible combination with oxygen. As emerged more clearly later, myoglobin consists of a single polypeptide chain of about 150 amino acid residues, associated with a single haem group; its one-to-four relationship with haemoglobin already suggested in early days by a comparison of molecular weights, turned out to be not coincidental but a fundamental structural relationship, as has now been shown by comparing the molecular models of the two proteins. At the beginning, however, one was more concerned with practical problems which took a number of years to solve, than with hypothetical structural relationships.

First of all it was necessary to find some species whose myoglobin formed crystals suitable, both morphologically and structurally, to the purpose in hand; the search for this took us far and wide, through the world and through the animal kingdom, and eventually led us to the choice of the sperm whale, *Physeter catodon,* our material coming from Peru or from the Antarctic, with some close runners-up including the myoglobin of the common seal, whose structure is now being studied by Dr. Helen Scouloudi at the Royal Institution in London. Once the method of isomorphous replacement had been shown to be capable in principle of solving the structure, one was faced with the task of attaching a small number of very heavy atoms at well-defined sites to each protein molecule in the crystal. Myoglobin lacks the sulphydryl groups whose presence in haemoglobin was so successfully exploited by Perutz and Ingram for the attachment of mercurial reagents; we had to look for other ways, and our attempts to use the unique haem group for the attachment of ligands which contained heavy atoms having proved for the most part unsuccessful (our ligands were always rapidly ejected by even very small traces of oxygen which were almost impossible to exclude), we were thrown back to a more empirical approach. This consisted in crystallizing myoglobin in the presence of metallic ions and then seeing whether any changes in the X-ray pattern could be detected; further analysis was required to determine whether, as we desired, substitution had taken place at a single site. In the absence of any sound foundation of theory, it was necessary to examine a very large number of possible ligands - several hundreds - before two or three were found which satisfied all the rather rigid criteria. Such laboriously empirical procedures are still forced upon all

workers in this field, and very drastically limit the exploitation of the isomorphous replacement method. A rational and generally applicable solution to this problem still awaits discovery, and would do more than any other single factor to open up the field.

General strategy of the analysis

Turning now to the strategy actually adopted for the solution of the structure, we may remember that Perutz's first application of the isomorphous replacement method in haemoglobin, as well as our own in myoglobin, had been to produce a two-dimensional projection of the structure. For such a projection the number of X-ray reflexions required is fairly small, and the solution of the phase problem is simple; even with a single isomorphous replacement the results are unambiguous. But the amount of structural information which could be derived from a projection was almost nil, owing to the high degree of overlapping of the elements of so complex a structure. It was immediately clear that to exploit the method it had to be applied in three dimensions, to produce a spatial representation of the electron density throughout the crystal. This involved the study of a much larger number of reflexions and the calculation of general phases, and required, for anunambiguous solution, the comparison of several heavy-atom derivatives substituted in different parts of the molecule.

The whole diffraction pattern of a myoglobin crystal consists of at least 25,000 reflexions. In 1955, when the three-dimensional work began, no computers existed which were fast enough to calculate Fourier syntheses containing so many terms; besides, the method was unproved and it seemed advisable to test it first on a smaller sample of data.

We may regard a typical X-ray photograph of a myoglobin crystal (Fig. I) as a two-dimensional section through a three-dimensional array of reflexions; each reflexion corresponds to a single Fourier component, and the whole structure can be reconstructed by using all the components as terms of a Fourier synthesis. As Perutz has indicated in his lecture, the components of higher frequency (higher harmonics), which are responsible for filling in the fine details of the structure, lie to the outside of the pattern. Thus one can obtain a rendering of the molecule at low resolution by using simply those reflexions within a spherical surface at the centre of the pattern. By doubling the radius of the sphere (which now encloses eight times as many reflexions)

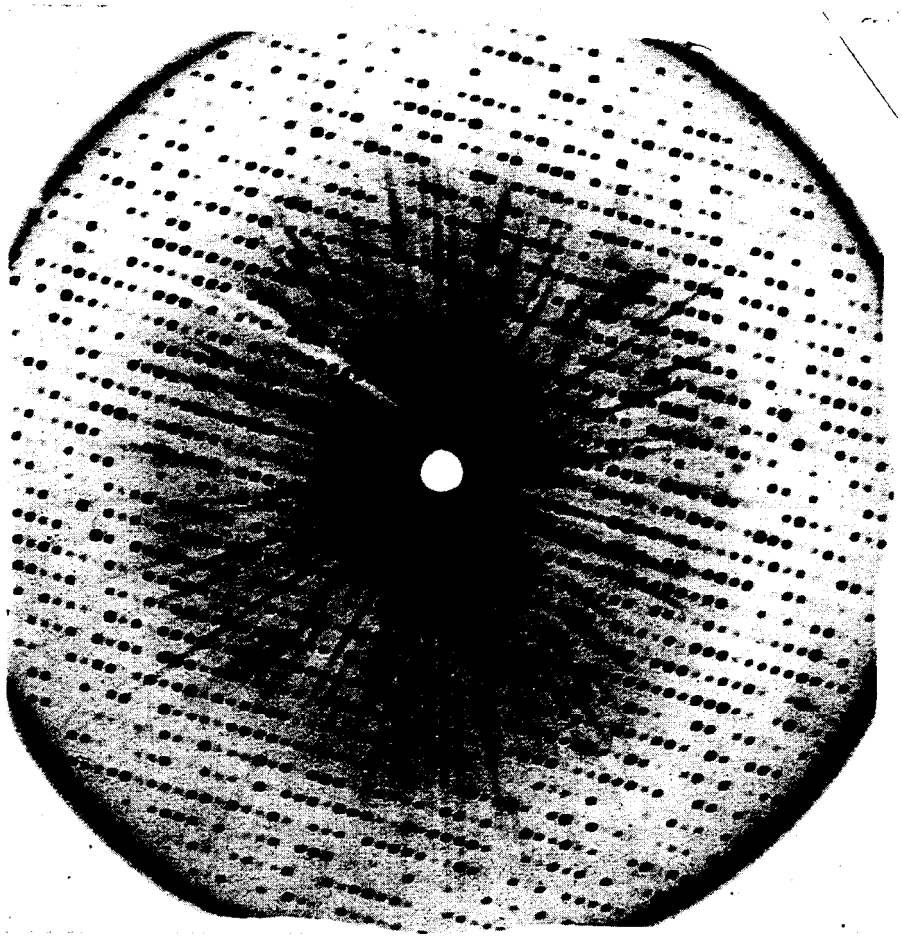

Fig. I. X-ray precession photograph of a myoglobin crystal.

we double the resolution of the density distribution. We actually decided to undertake the solution of the structure in three stages; the first, completed in 1957, involved 400 reflexions and gave a resolution of 6 Å; the second (1959) included nearly 10,000 reflexions and gave a resolution of 2 Å; the third (not yet complete) includes all the observable reflexions - about 25,000 - and gives a resolution of 1.4 Å. It may be recalled that polypeptide chains pack together at centre-to-centre distances of 5 to 10 Å; that atoms (other than hydrogen) of neighbouring groups in Van der Waals contact, or brought together by hydrogen bonds or charge interactions, lie 2.8 to 4 A apart; and that the separation between covalently bonded atoms is 1 to 1.5

Å. It follows that the three stages chosen would be expected to separate polypeptide chains, groups of atoms, and individual atoms, respectively. The third stage, with its resolving power of 1.4 Å, should only just distinguish neighbouring covalently bonded atoms, but this is as far as the analysis can go because beyond this point the diffraction pattern fades away. This limit represents a lower degree of order than is usual in crystals of molecules of low or moderate complexity; in fact myoglobin crystals possess a higher degree of order than do those of almost all other proteins, and this was an additional reason for my choice of this protein for analysis.

Before proceeding to describe the results of the three stages of the analysis, it will be convenient to revert to the question of computers. As will be evident from what follows, the amount of useful structural information obtainable increases rapidly with the resolving power. Indeed it seems probable that for most proteins the dividend obtained from high resolution would be even greater, for it has emerged that the helix content of myoglobin (75%) is a good deal higher than that of most other proteins, and the identification of structural features in myoglobin at less than atomic resolution was greatly dependent on the presence of many helical segments of polypeptide chain, readily identifiable even at 6 Å resolution and already at 2 Å resolution providing well-defined take-off points for side chains, often enabling these to be identified even though their individual atoms could not be distinguished. But, as already indicated, the amount of computation required increases very rapidly with the resolving power. Even at the first stage of the analysis we made use of an electronic computer, EDSAC I, which though small and slow by modern standards was at the time one of the very few such instruments in operation in the world; it is significant that these early Fourier syntheses of the myoglobin data were, to the best of my belief, the first crystallographic computations ever carried out on an electronic computer and initiated a practice which later (and incidentally after a time lag of several years) became universal among crystallographers. At each stage of the myoglobin analysis the computers employed were among the most rapid available at the time, and we are now using very fast and large computers such as EDSAC II and IBM 7090; most proteins are larger than myoglobin, and will need even bigger computers. There are also problems of data collection and data handling. In the myoglobin analysis the data for the 6 Å and 2 Å stages were mostly collected by conventional photographic methods; but at the 2 Å stage the solution of the phase problem for 9,600 reflexions involved the densitometry of some quarter of a million spots in all,

from different heavy-atom derivatives and exposures of different lengths. This represents something near the limit of the practicable, especially as we were aiming for, and achieved, a mean error of 2 to 4% in the determination of amplitude; personally I would not care to have to undertake such a task a second time. In any case, serious effects of radiation damage to the crystals make photographic techniques increasingly difficult if not impossible at the higher resolutions. Fortunately, the automatic diffractometer designed by my colleagues Drs. U. W. Arndt and D. C. Phillips became available just in time for the final stage of the work; with this apparatus the intensities of successive reflexions, measured with a proportional counter, are recorded on punched tape which can be fed direct into a high-speed computer. There is no doubt that automatic data-collecting equipment and very fast large computers will be highly desirable for all, and essential for most, X-ray studies of proteins.

Myoglobin at 6Å resolution

The three-dimensional electron density distribution in a crystal is most conveniently represented as a series of contour maps plotted on parallel transparent sheets; the function drawn in this way for myoglobin at 6 Å resolution is shown in Fig. 2. A cursory inspection of the map showed it to consist of a large number of rod-like segments, joined at the ends, and irregularly wandering through the structure; a single dense flattened disk in each molecule; and sundry connected regions of uniform density. These could be identified respectively with polypeptide chains, with the iron atom and its associated porphyrin ring, and with the liquid filling the interstices between neighbouring molecules. From the map it was possible to "dissect out" a single protein molecule, its boundaries being demarcated by the adjoining liquid; a scale model of this is shown in Fig. 3. For the most part the course of the single polypeptide chain could be followed as a continuous region of high density, but some ambiguities remained, especially at the irregular regions between two straight rods. The most striking features of the molecule were its irregularity and its total lack of symmetry; this made all the more remarkable the later finding by Perutz that each of the four sub-units of haemoglobin closely resembled the myoglobin molecule, in spite of wide differences in species and in amino acid composition.

As expected, it was not possible at 6 Å resolution to draw any conclusions

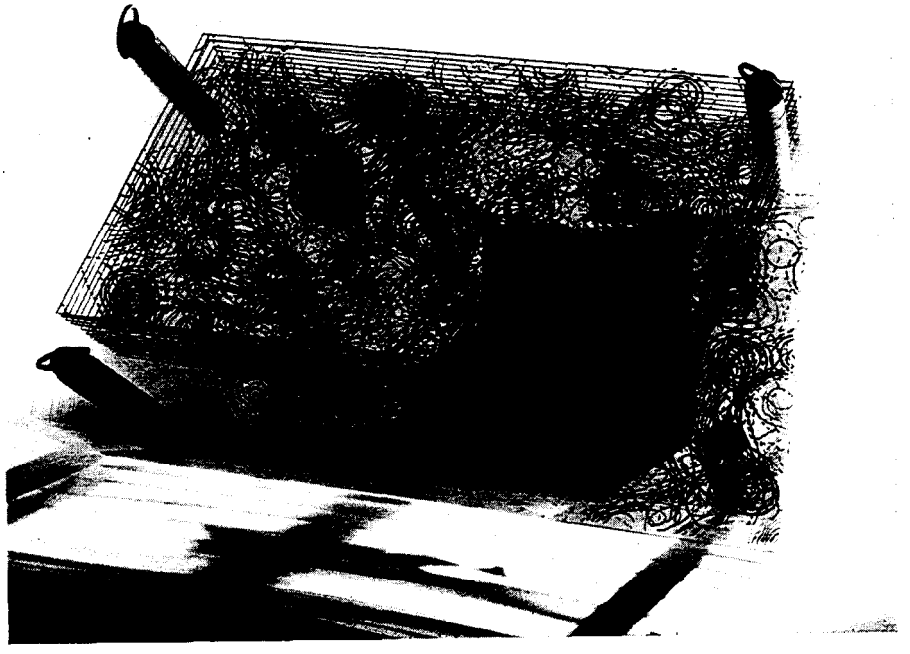

Fig. 2. Fourier synthesis of myoglobin at 6Å resolution.

regarding the nature of the folding of the popypeptide chain, or to see, let alone identify, side chains.

Myoglobin at 2Å *resolution*

To achieve a resolution of 2 Å it was necessary to determine the phases of nearly 10,000 reflexions, and then to compute a Fourier synthesis with the same number of terms. As already indicated, this task represented about the extreme limit of what is practicable by photographic techniques, and the Fourier synthesis itself (excluding preparatory computations of considerable bulk and complexity) required about 12 hours of continuous computation on a very fast machine (EDSAC II). The electron density function was calculated at about 100,00 points in the molecule, and was represented on the three-dimensional contour map shown in Fig. 4. In this photograph we are looking at the density distribution directly along the axis of one of the

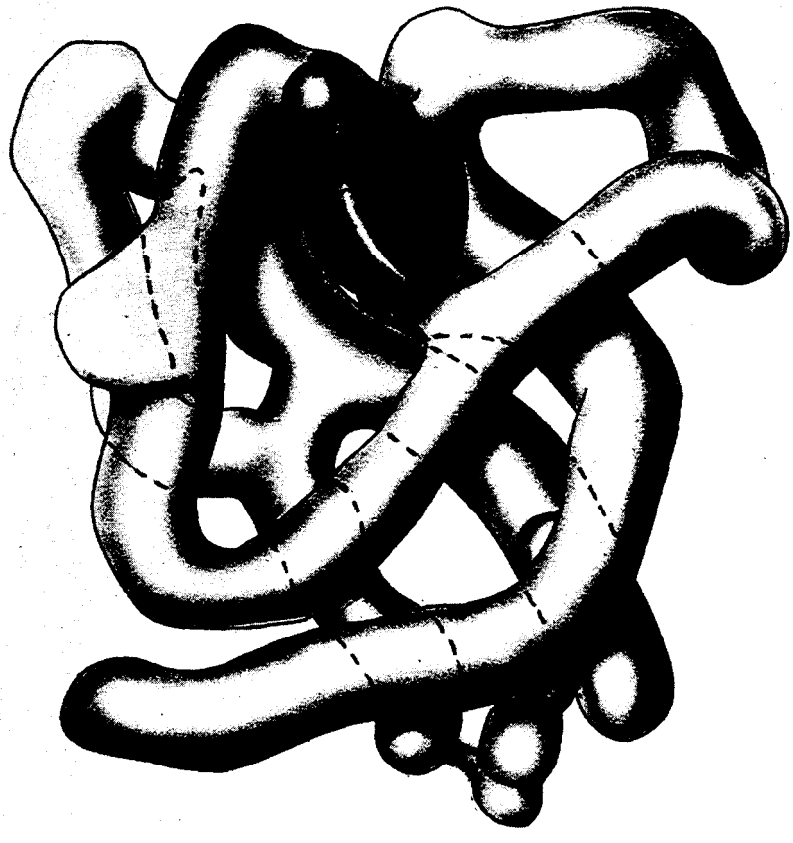

Fig. 3. Model of the myoglobin molecule, derived from the 6Å Fourier synthesis. The haem group is a dark grey disk (*centre top*).

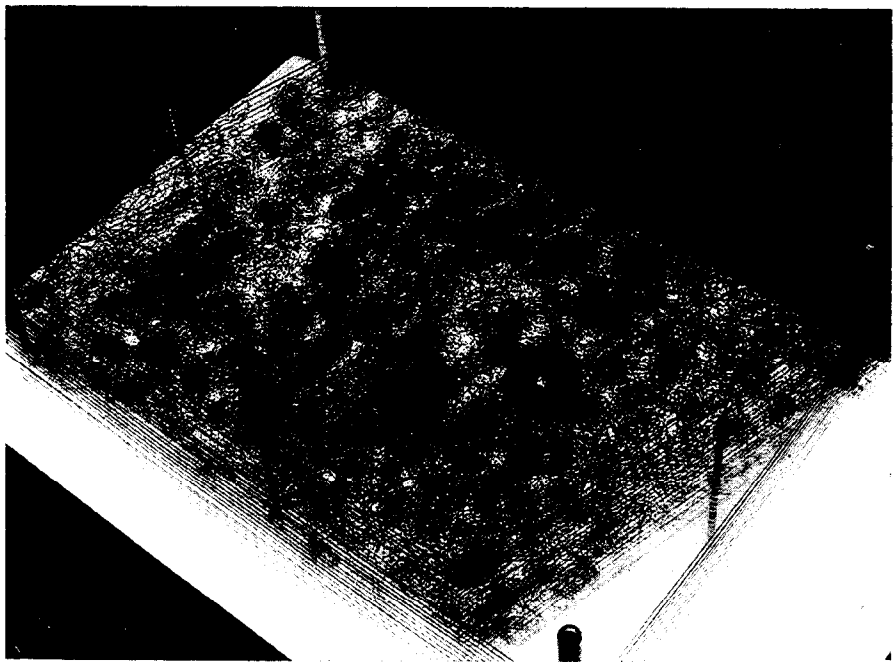

Fig. 4. Fourier synthesis of myoglobin at 2Å resolution, showing a helical segment of polypeptide chain end-on.

straight rod-like sections of polypeptide chain identified at low resolution; it will be seen that the rod has now developed into a straight hollow cylinder. Study of the density distribution on the surface of this (and other) cylinders showed that it fits the arrangement of atoms in the α-helix, postulated by Pauling & Corey in 1951 as the chain configuration in the so-called α-family of fibrous proteins; careful analysis of the density distribution, carried out on the computer, shows that the helical segments are nearly all precisely straight, and that their co-ordinates correspond to those given by Pauling & Corey within the limits of error of the analysis. Furthermore it is possible to see directly the orientation of each side chain relative to the atoms within the helix, and hence, from a knowledge of the absolute configuration of an L-amino acid, to show that all the helices are right-handed.

Another view (Fig. 5) of the contour map shows the haem group edge-on, now appearing as a flat disk with the iron atom at its centre. To our surprise we found that the iron atom lay more than $\frac{1}{4}$ Å out of the plane of the group; it was only later that we heard from Dr. Koenig at Johns Hopkins University that he had observed the same phenomenon in his structure anal-

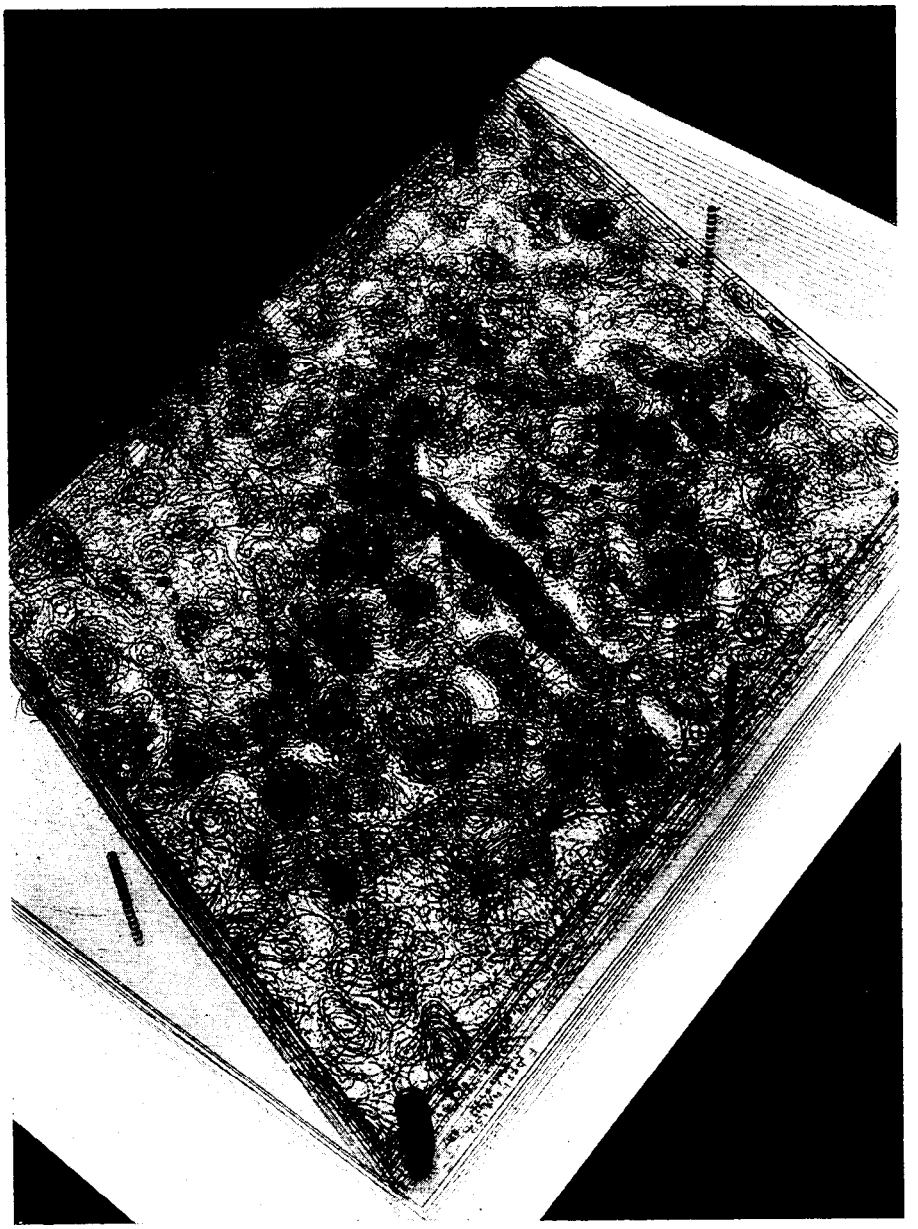

Fig. 5. Fourier synthesis of myoglobin at 2Å resolution, showing haem group edge-on.

ysis of haemin. We were also able to see that the iron atom was attached to one of the helical segments of polypeptide chain by a group which we were later able to identify as histidine - a striking confirmation of suggestions which had been made as much as thirty years earlier to the effect that histidine was the haem-linked group in haemoglobin and myoglobin.

In our preliminary publication about this Fourier synthesis in 1960, we pointed out that at a resolution of 2Å, neighbouring covalently bonded atoms are not resolved, and gave it as our opinion that systematic identification of side chains would not be possible at this resolution. Events proved that we had been too pessimistic; by studying carefully the shapes of the lumps of density projecting at the proper intervals from the polypeptide

	Leu	Leu	Leu 5
	Phe	Phe	Phe 5
	Lys	Lys	Thr 3, Lys 2
	Ser	Ser	Ser 3, Thr 2
C1	His	His	His 5
	Pro	Pro	Pro 4
	Glu.C	Glu	Glu.C 5
	Thr	Thr	Thr 4
	Leu	Leu	Leu 4
	Glu	Glu	Ser 2, Thr 2
	Lys	Lys	Lys 3
CD1	Phe	Phe	Phe 5
	Asp	Asp	Asp 2
	Arg	Arg	Arg 3
	Phe	Phe	Phe 4
	Lys	Lys	Lys 3
	His	His	His 4
	Leu	Leu	Leu 2, Asp 2
	Lys	Lys	Lys 1
D1	Thr	Thr	Thr 5
	Glu.C	Glu	Glu 4
	Ala	Ala	Ala 5
	Glu.C	Glu	Glu 3
	Met	Met	Met 5
	Lys		Lys 4

Fig. 6. A comparison between chemical and X-ray evidence for part of the amino-acid sequence of myoglobin. *(First column)* tryptic peptides; (secondcolumn) chymotryptic peptides; *(third column)* X-ray evidence. Peptides are enclosed by brackets. In the third column, the figure gives the degree of confidence in the identification (5 = complete confidence, 1 = a guess).

chain we were often able to identify them unambiguously with one of the seventeen different types of side chain known, from the overall composition, to be present in the myoglobin molecule. We were able to seek confirmation and extension of our results from a quite different source. At the time when the myoglobin program was getting seriously under way I discussed with Drs. W. H. Stein and Stanford Moore at the Rockefeller Institute in New York the possibility that some member of their laboratory might undertake a determination of the complete amino acid sequence of myoglobin, using the methods originally employed by Sanger in his studies of insulin, and later developed and extended at the Rockefeller Institute for the analysis of ribonuclease. They were kind enough to arrange that Dr. Allen Edmundson, at that time a graduate student working in their laboratory under the supervision of Dr. C. H. W. Hirs, should undertake this task. By the time our 2 Å synthesis was available, Dr. Edmundson had studied most of the peptides obtained by tryptic digestion of myoglobin, determining their composition and in a few cases the sequence of residues within them. We found that, by laying his peptides along the partial and tentative sequence derived from the X-ray analysis, we were able in many cases to observe correspondences which confirmed both our identifications and his analysis, and to clear up ambiguities and confusions in each (Fig. 6). All in all it was possible to identify about two-thirds of all the residues in the molecule with some assurance, though some certain pairs of residues of similar shape were difficult to distinguish. We were able to summarize the results of the analysis up to this stage in the form of a model (Fig. 7) which showed the positions in space of the helical polypeptide chain segments, of the haem group, and of most of the side chains; it included, less precisely, the positions of the atoms in most of the non-helical regions and in many of the remaining side chains.

Myoglobin at 1.4Å *resolution*

During the past two years we have been concerned with improving the resolution of the electron density map by including virtually all the observable reflexions in the pattern, about 25,000 in number and extending to spacing of 1.4 A; we now plot the electron density at half a million points in the molecule. It has already been pointed out that this extension of the analysis was made possible by the availability of automatic data-collecting equipment

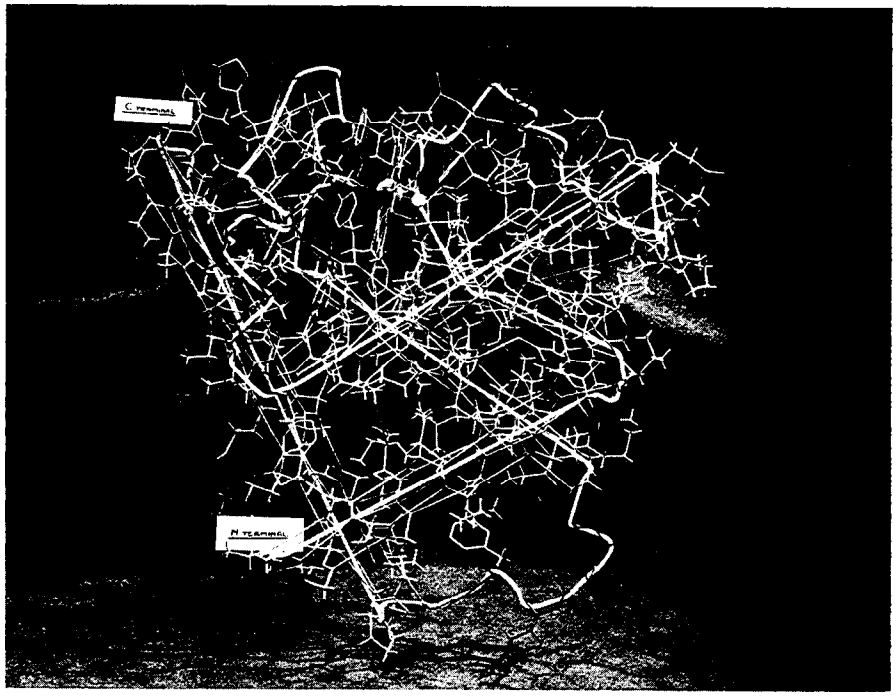

Fig. 7. Model of the myoglobin molecule, derived from the 2Å Fourier synthesis. The *white cord* follows the course of the polypeptide chain; the iron atom is indicated by a *grey sphere,* and its associated water molecule by a *white sphere.*

using proportional counters, and of still larger computers such as the IBM 7090. Even so the task would have been a very formidable one if we had continued to use the method of isomorphous replacement, involving the collection of data from a number of different isomorphous derivatives. Instead we have reverted to a more conventional method, that of successive refinement, and have abandoned the use of heavy-atom derivatives. From a study of the 2 Å Fourier synthesis we were able to assign spatial co-ordinates to about three quarters of the atoms in the molecule. Owing to the limited resolving power of this synthesis, the accuracy with which atoms could be located was a good deal less than is desirable, but this imprecision was compensated for by their number, a good deal higher in proportion to the size of the structure than is generally necessary for the success of the refinement method. This method consists in calculating the phases of all the reflexions from the co-ordinates of the atoms which have already been located; a Fourier synthesis is then computed using *observed* amplitudes and *calculated*

phases. This synthesis necessarily shows all the atoms which have been used for calculating phases, but should reveal "ghosts" of additional ones, with reduced density; it also indicates any minor errors in the positions of the atoms previously located, and if their positions are not found to coincide exactly with those assumed. One is now in a position to embark on the next cycle of refinement, using the previous set of atoms with corrected co-ordinates together with additional atoms located after the first cycle. After a few such cycles the successive Fourier syntheses should converge to a precise representation of the whole structure. We have so far carried out two cycles of refinement, including 825 atoms in the first, and 925 atoms in the second (myoglobin contains in all 1,260 atoms excluding hydrogen; in addition there are some 400 atoms of liquid and salt solution, a proportion of which are bound to fixed sites on the surface of the molecule). One or two further cycles of refinement will probably be necessary, but in the meantime the 1.4 Å Fourier synthesis based on the second cycle is very much better resolved than the 2 Å synthesis. In many cases neighbouring covalently bonded atoms are just resolved, the background between groups of atoms is much cleaner than before and, finally, many of the disturbances found in the region of the heavy atom sites in the 2 Å synthesis have disappeared. Figs. 8 *(i-iii)* will give some impression of this synthesis.

Meanwhile Dr. Edmundson has greatly advanced his study of the amino acid sequence of myoglobin; in particular he has characterized a large number of chymotryptic peptides in addition to the tryptic ones previously mentioned. Taking the results of the X-ray and chemical studies together, the situation today is that some 120 amino acid residues are known with almost complete certainty, and many of the remaining 30 with fair probability. There is little doubt that the residual ambiguities will shortly be resolved, and that the positions of all the atoms in the structure will be known with reasonable accuracy, with the exception of a few long side chains (such as lysine) which are apparently flexible and do not occupy defined positions in the crystal.

The general nature of the structure

What is the nature of the molecule which has emerged with progressively increasing clarity from successive Fourier syntheses? Some 118 out of the total of 151 amino acid residues make up 8 segments of right-handed α-helix,

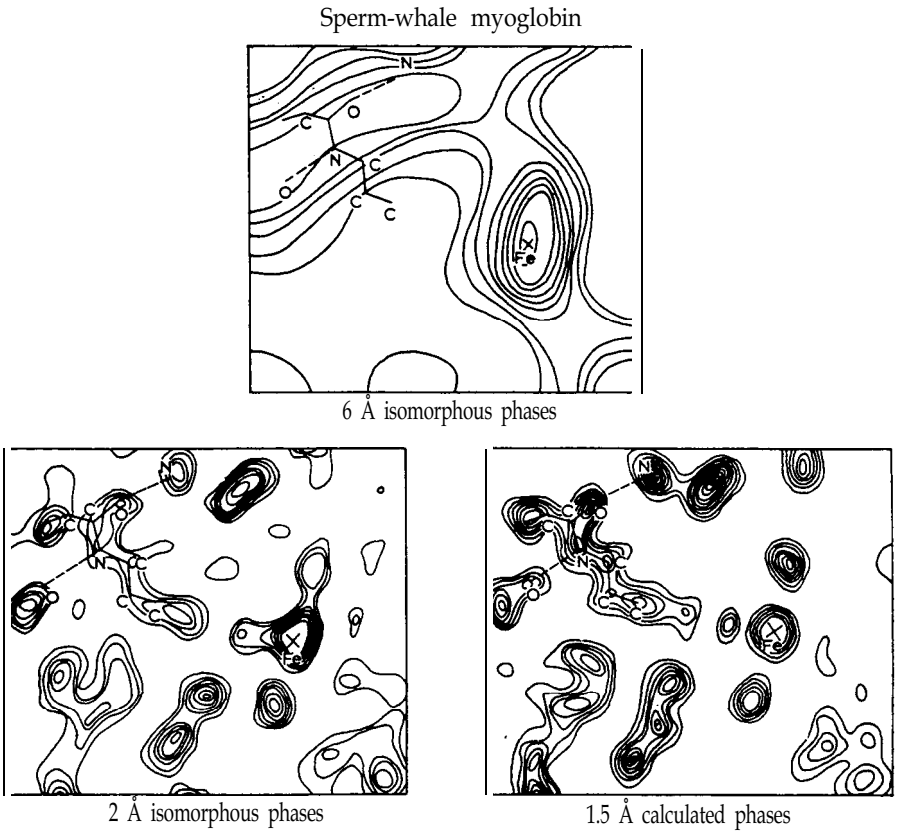

Fig. 8. Comparison between the same section through the myoglobin molecule, *(i)* at 6Å resolution, *(ii)* at 28 resolution, *(iii)* at 1.4 Å resolution. *(Top left)* longitudinal section through a helix; *(right centre)* haem group edge-on. The atoms marked are part of the distal histidine (see text); note that several neighbouring atoms are resolved at 1.4Å

of lengths ranging from 7 to 24 residues. These segments are joined by 2 sharp corners (containing no non-helical residues) and 5 non-helical segments (of 1 to 8 residues) ; there is also a non-helical tail of 5 residues at the carboxyl end of the chain. The whole is folded in a complex and unsymmetrical manner to form a flattened, roughly triangular prism with dimensions about 45 x 35 x 25 Å. The whole structure is extremely compact; there is no water inside the molecule, with the probable exception of a very small number (less than 5) of single water molecules presumably trapped at the time the molecule was folded up; there are no channels through it, and the volume of internal empty space is small. The haem group is disposed almost

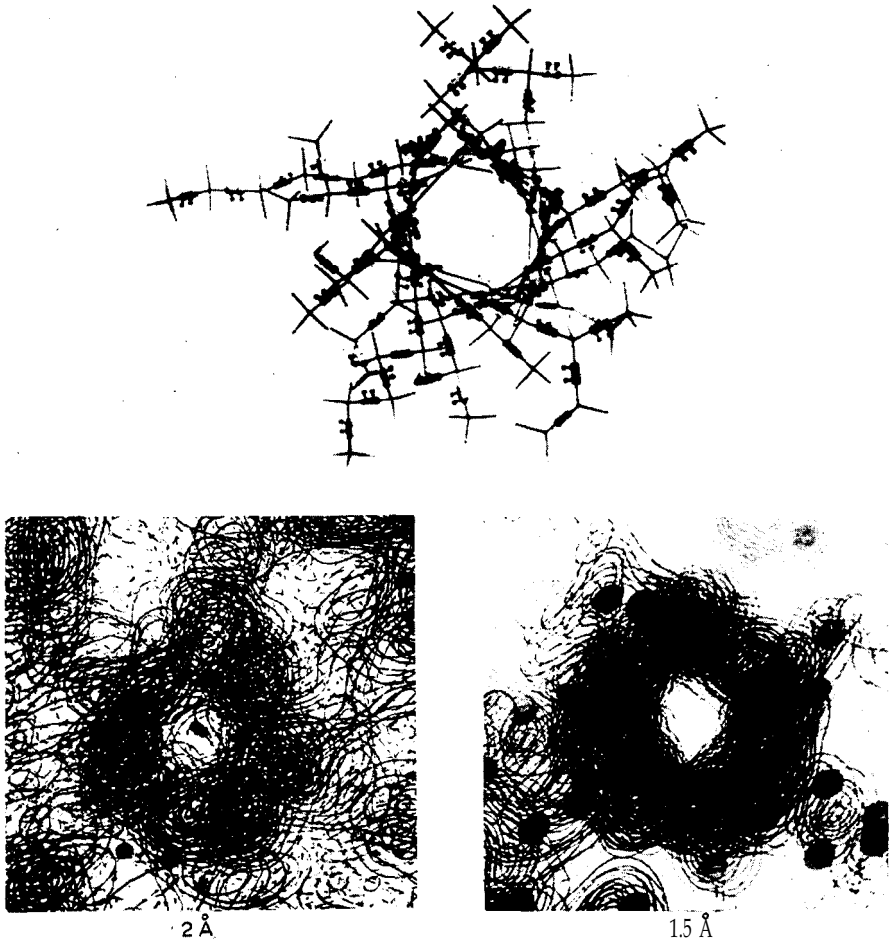

Fig. 9. End-on view of a helix at 2Å and 1.4Å resolution; *(above)* a model helix for comparison.

normally to the surface of the molecule, one of its edges (that containing the polar propionic acid groups) being at the surface and the rest buried deeply within.

Turning now to the side chains, it is found that almost all those containing polar groups are on the surface. Thus with very few exceptions all the lysine, arginine, glutamic, aspartic, histidine, serine, threonine, tyrosine, and tryptophan residues have their polar groups on the outside (the rare exceptions

appear to have some special function within the molecule, e.g. the haem-linked histidine). The interior of the molecule, on the other hand, is almost entirely made up of non-polar residues, generally close-packed and in Van der Waals' contact with their neighbours.

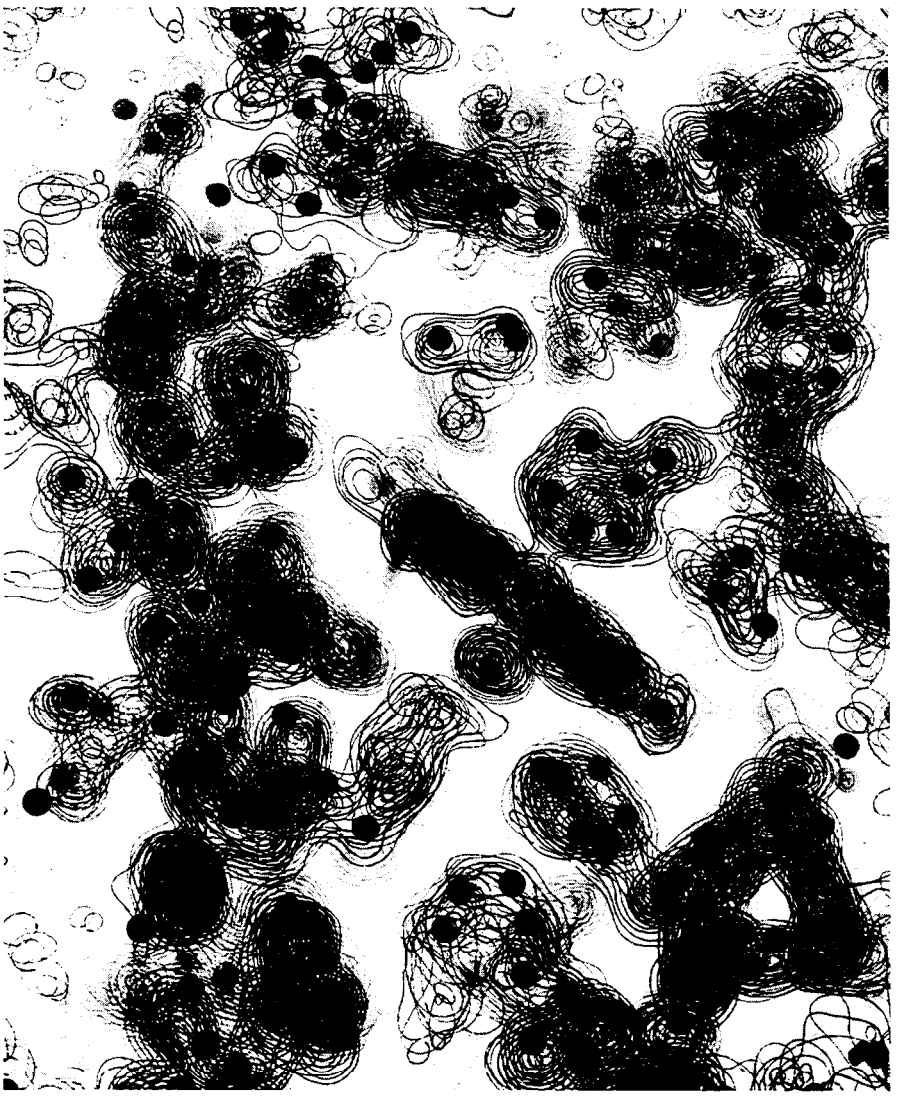

Fig.10. Part of the 1.4Å Fourier synthesis. *(Centre)* the haem group (edge-on), showing haem-linked and distal histidines, and water molecule attached to iron atom. *(Top right)* a helix end-on. (*Bottom*) a helix seen longitudinally, together with several side chains.

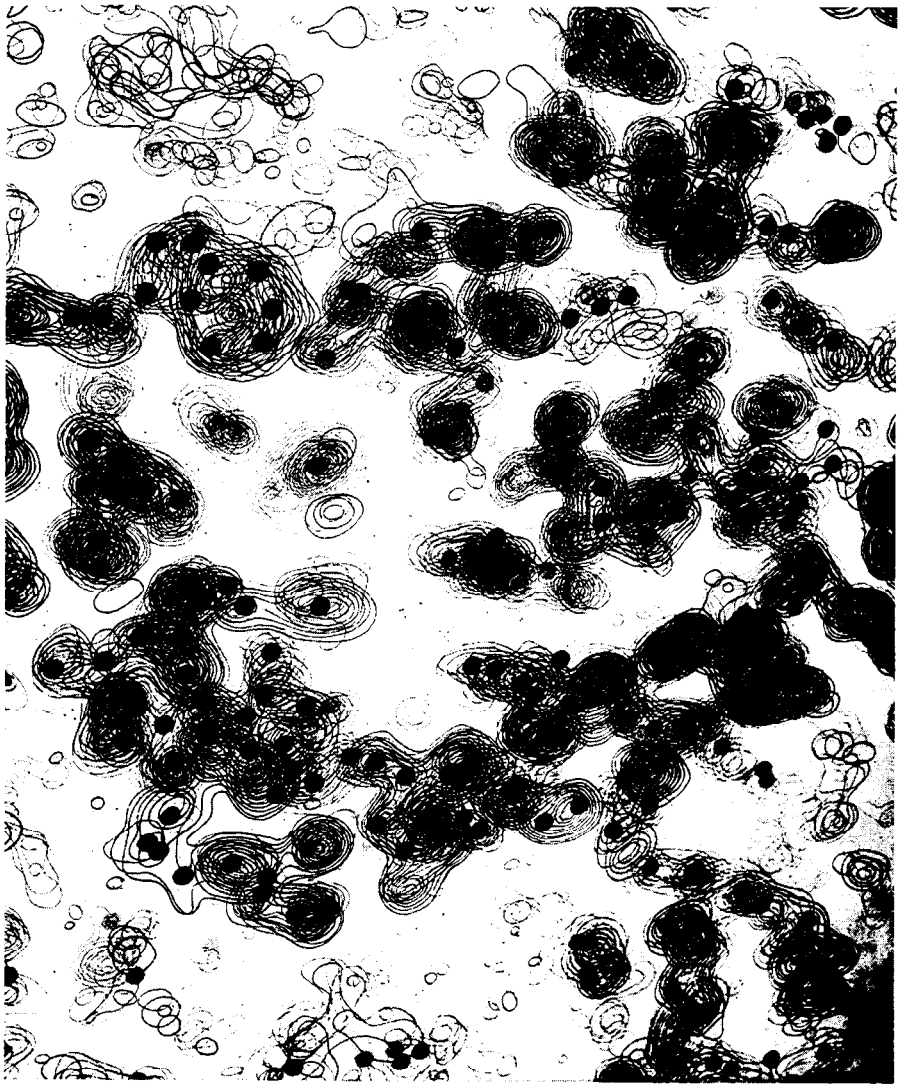

Fig. 11. Part of the 1.4Å Fourier synthesis. *(Left centre)* a tryptophan residue; *(to the left)* a liquid region between two molecules.

We may ask what forces are responsible for maintaining the integrity of the whole structure. The number of contacts between neighbouring groups in the molecule is very large, and to analyse these it has been necessary to use a large computer to calculate all the interatomic distances and to determine which of these lie within the limits corresponding to each type of bonding.

These results have not yet been studied in detail, but it is clear that by far the most important contribution comes from the Van der Waals forces between non-polar residues which make up the bulk of the interior of the molecule. It is true that there are a number of charge interactions and hydrogen bonds between neighbouring polar residues on the surface of the molecule, but one gains the strong impression that many, or even most, of these are, so to say, incidental - a polar group on the surface is quite content to bond with a water molecule or ion in the ambient solution, and only links up with a neighbouring side chain if it can do so without departing too far from its normal extended configuration. This statement should perhaps be qualified by remarking that one observes a number of polar interactions of side chains such as glutamic acid, aspartic acid, serine, and threonine with free amino groups on the last turn of helical segments; and it may be that these have some significance in determining the point at which a helix is broken and gives way to an irregular segment of chain. If so, these special interactions will be important in a wider context, as determinants of the three-dimensional structure of proteins, and might be of service in predicting the nature of the structure from a knowledge of the amino acid sequence.

The interactions of the haem group deserve special consideration. It is these which are responsible for the characteristic function of myoglobin, since an isolated haem group does not exhibit the phenomenon of reversible oxygenation. At the present time we can merely enumerate the haem group interactions; it is a task for the future to explain reversible oxygenation in terms of them. As already mentioned, the fifth co-ordination position of the iron atom is occupied by a ring nitrogen atom of a histidine residue, the so-called haem-linked histidine. On the other (distal) side of the iron atom, occupying its sixth co-ordination position, is a water molecule, as would be expected in ferrimyoglobin, the form of myoglobin used for X-ray analysis; beyond the water molecule, in a position suitable for hydrogen-bond formation, is a second histidine residue. It is noteworthy that the same arrangement of two histidines also exists in haemoglobin. For the rest the environment of the haem group is almost entirely non-polar; it is held in place by a large number of Van der Waals interactions. In haemoglobin it seems that the environment of the haem group is closely analogous, and for both proteins it is clear that a rich field of knowledge awaits exploration, for we may hope that the very extensive studies of the oxygenation reaction made during the past half-century may now be interpreted in precise structural terms.

Some implications

The oxygenation reaction of myoglobin and haemoglobin may be held to be interesting and important enough in its own right to justify the choice of these two proteins for study. In fact, as was indicated at the beginning of this lecture, the choice was originally made on different grounds, such as availability, ease of crystallization, and molecular weight. There are very many proteins which have specific functions as important, or more important; every enzyme - and many hundreds of these have been characterized - has its own specific function vital to some particular process in cell function. A number of enzymes are being studied by X-ray methods in laboratories all over the world, and in several cases the analysis is on the brink of success; a knowledge of the detailed structure of each of these will give insight into some essential biological process, by resolving the molecular architecture of the active site and permitting the same kind of interpretation of function in molecular terms as we may soon anticipate in the haem proteins. From this point of view there is no forseeable limit to the number of proteins whose structure is worth analysing, since each will have its own unique function which demands explanation in structural terms.

From another angle we may rather enquire what features are common to all proteins, and study the structure of myoglobin in its context as a typical member of this vast class of substances. Probably more experimental work has been done on proteins than on any other kind of compounds, and a huge corpus of knowledge has been built up by the organic chemist and the physical chemist. Many generalizations have been observed, but always they have been limited in scope by the fact that they could not be based on a precise molecular model. The emergence of such a model even for a single protein, such as myoglobin, makes it possible to test and to add precision to the chemist's generalizations. Already sperm whale myoglobin is being studied by biochemists in a number of laboratories with this end in view; to give only a few examples, it is being examined from the standpoint of optical rotatory power and helix content, of titration behaviour, of metal binding, chemical modification of side chains, of hydrodynamic characteristics. Such studies, and others like them, will serve to deepen our understanding of the ways in which proteins behave and of the reasons why they are uniquely capable of occupying so central a position in living organisms.

The geneticists now believe - though the point is not yet rigorously proved - that the hereditary material determines only the amino acid sequence of a

protein, not its three-dimensional structure. That is to say, the polypeptide chain, once synthesized, should be capable of folding itself up without being provided with additional information; this capacity has, in fact, recently been demonstrated by Anfinsen *in vitro* for one protein, namely ribonuclease. If the postulate is true it follows that one should be able to predict the three-dimensional structure of a protein from a knowledge of its amino acid sequence alone. Indeed, in the very long run, it should only be necessary to determine the amino acid sequence of a protein, and its three-dimensional structure could then be predicted; in my view this day will not come soon, but when it does come the X-ray crystallographers can go out of business, perhaps with a certain sense of relief, and it will also be possible to discuss the structures of many important proteins which cannot be crystallized and therefore lie outside the crystallographer's purview.

We have taken a preliminary look at the structure of myoglobin from this point of view and have to confess that the difficulties are formidable. The structure is highly irregular; the seven "corner" regions between helical segments are all different, so that generalization is impossible; the interactions between side chains are numerous and of many different types, and one cannot easily see which are crucial in determining the structure. The complexity of myoglobin is very great, yet it is probably simpler than most proteins, not only by virtue of its low molecular weight, but also in respect of its high helix content, probably much higher than that of most others. As things stand we cannot even hazard a guess as to why the helix content of myoglobin is so high, let alone see how to predict its structure in detail.

Much help with these problems may come from a comparison of myoglobin with the sub-units of haemoglobin, which Perutz has shown to resemble it very closely in spite of notable differences in amino acid sequence. By laying alongside one another the sequences of myoglobin and of the α- and β-chains of haemoglobin, and making certain plausible assumptions to explain the (fairly small) differences between their lengths, it is possible to observe homologies - points at which the same ammo acid appears in a corresponding position in all three chains. The number of these homologies is surprisingly small, but presumably it is these which are responsible for the crucial interactions which determine that all three chains have the same three-dimensional arrangement (though some of the homologies may be accidents of evolutionary development). Study of homology will soon be extended-by examinationof other species - human myoglobin, human, horse, rabbit and human foetal haemoglobin - and of the aberrant haemoglobins

whose "mistakes" in amino acid sequence have been shown in recent years to be associated with so many hereditary diseases of the blood.

Perutz and I, with our collaborators, have already spent some time looking at these homologies, and a number of interesting facts have come to light. Yet, even in this narrow field, our studies are in their infancy; and in any case I suspect that only generalization of limited scope can be made from myoglobin and haemoglobin alone. The detailed structures of a few other proteins should soon become known, but it will be clear from many of the topics I have touched upon that we have pressing need to know the structures of very many others, for proteins are unique in combining great diversity of function and complexity of structure with a relative simplicity and uniformity of chemical composition. In determining the structures of only two proteins we have reached, not an end, but a beginning; we have merely sighted the shore of a vast continent, waiting to be explored.

The work described in this lecture has been done by many hands, and a list of those who have contributed to it, formally or informally, would be long. They come from many countries and many disciplines, and their contributions, decisive in sum, cannot be assessed in detail, and are of such varying magnitudes that any list must be invidious and incomplete. I nevertheless wish to record the following names of colleagues whose ideas and whose collaboration have been particularly important and sometimes essential.

J. M. Bennet, C. Blake, Joan Blows, M. M. Bluhm, G. Bodo, Sir Lawrence Bragg, C.-I. Bräden, D. A. G. Broad, C. L. Coulter, Ann Cullis, D. Davies, R. E. Dickerson, H. M. Dintzis, A. B. Edmundson, R. G. Hart, Ann Hartley, W. Hoppe, V. M. Ingram, L. H. Jensen, J. Kraut, R. G. Parrish, P. Pauling, M. F. Perutz, D. C. Phillips, Mary Pinkerton, Eva Rowlands, Helen Scouloudi, Violet Shore, B. Strandberg, I. F. Trotter, H. C. Watson, Joyce Wheeler, Ann Woodbridge, H. W. Wyckoff and the staff of the Mathematical Laboratory, Cambridge.

J. C. Kendrew and R. G. Parrish, The crystal structure of myoglobin. III. Sperm-whale myoglobin, *Proc. Roy. Soc. London, A238 (1956) 305.*

M. M. Bluhm, c. Bodo, H. M. Dintzis, and J. C. Kendrew, The crystal structure of myoglobin. IV. A Fourier projection of sperm-whale myoglobin by the method of isomorphous replacement, *Proc. Roy. Soc. London,* A246 *(1958) 369.*

J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. W. Wyckoff, and D. C. Phihips, A three-dimensional model of the myoglobin molecule obtained by X-ray analysis, *Nature,* 181 (1958) 666.
G. Bodo, H. M. Dintzis, J. C. Kendrew, and H. W. Wyckoff, The crystal structure of myoglobin. V. A low resolution three-dimensional Fourier synthesis of sperm-whale myoglobin crystals, *Proc. Roy. Soc. London,* A253 (1959) 70.
J. C. Kendrew, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, and V. C. Shore, Structure of myoglobin: a three-dimensional Fourier synthesis at 2Å resolution, *Nature,* 185 (1960) 422.
J. C. Kendrew, H. C. Watson, B. E. Strandberg, R. E. Dickerson, D. C. Phillips, and V. C. Shore, The amino-acid sequence of myoglobin : a partial determination by X-ray methods, and its correlation with chemical data, *Nature,* 190 (1961) *666.*
H. C. Watson and J. C. Kendrew, Comparison between the amino-acid sequences of sperm-whale myoglobin and of human haemoglobin, *Nature,* 190(1961) 670.
J. C. Kendrew, Side-chain interactions in myoglobin, *Brookkaven Symposia in Biology,* 15(1962) 216.