A THREE-DIMENSIONAL MODEL OF THE MYOGLOBIN MOLECULE OBTAINED BY X-RAY ANALYSIS

By Drs. J. C. KENDREW, G. BODO, H. M. DINTZIS, R. G. PARRISH and H. WYCKOFF
Medical Research Council Unit for Molecular Biology, Cavendish Laboratory, Cambridge
AND
D. C. PHILLIPS
Davy Faraday Laboratory, The Royal Institution, London

MYOGLOBIN is a typical globular protein, and is found in many animal cells. Like hæmoglobin, it combines reversibly with molecular oxygen; but whereas the role of hæmoglobin is to transport oxygen in the blood stream, that of myoglobin is to store it temporarily within the cells (a function particularly important in diving animals such as whales, seals and penguins, the dark red tissues of which contain large amounts of myoglobin, and which have been our principal sources of the protein). Both molecules include a non-protein moiety, consisting of an iron–porphyrin complex known as the hæm group, and it is this group which actually combines with oxygen; hæmoglobin, with a molecular weight of 67,000, contains four hæm groups, whereas myoglobin has only one. This, together with about 152 amino-acid residues, makes up a molecular weight of 17,000, so that myoglobin is one of the smaller proteins. Its small size was one of the main reasons for our choice of myoglobin as a subject for X-ray analysis.

In describing a protein it is now common to distinguish the primary, secondary and tertiary structures. The *primary structure* is simply the order, or sequence, of the amino-acid residues along the polypeptide chains. This was first determined by Sanger using chemical techniques for the protein insulin[1], and has since been elucidated for a number of peptides and, in part, for one or two other small proteins. The *secondary structure* is the type of folding, coiling or puckering adopted by the polypeptide chain: the α-helix and the pleated sheet are examples. Secondary structure has been assigned in broad outline to a number of fibrous proteins such as silk, keratin and collagen; but we are ignorant of the nature of the secondary structure of any globular protein. True, there is suggestive evidence, though as yet no proof, that α-helices occur in globular proteins, to an extent which is difficult to gauge quantitatively in any particular case. The *tertiary structure* is the way in which the folded or coiled polypeptide chains are disposed to form the protein molecule as a three-dimensional object, in space. The chemical and physical properties of a protein cannot be fully interpreted until all three levels of structure are understood, for these properties depend on the spatial relationships between the amino-acids, and these in turn depend on the tertiary and secondary structures as much as on the primary.

Only X-ray diffraction methods seem capable, even in principle, of unravelling the tertiary and secondary structures. But the great efforts which have been devoted to the study of proteins by X-rays, while achieving successes in clarifying the secondary (though not yet the tertiary) structures of fibrous proteins, have hitherto paid small dividends among the metabolically more important globular, or crystalline, proteins. Progress here has been slow because globular proteins are much more complicated then the organic molecules which are the normal objects of X-ray analysis (not counting hydrogens, myoglobin contains 1,200 atoms, whereas the most complicated molecule the structure of which has been completely determined by X-rays, vitamin B_{12}, contains 93). Until five years ago, no one knew how, in practice, the complete structure of a crystalline protein might be found by X-rays, and it was realized that the methods then in vogue among protein crystallographers could at best give the most sketchy indications about the structure of the molecule. This situation was transformed by the discovery, made by Perutz and his colleagues[2], that heavy atoms could be attached to protein molecules in specific sites and that the resulting complexes gave diffraction patterns sufficiently different from normal to enable a classical method of structure analysis, the so-called 'method of isomorphous replacement', to be used to determine the relative phases of the reflexions. This method can most easily be applied in two dimensions, giving a projection of the contents of the unit cell along one of its axes. Perutz attached a *p*-chloromercuri-benzoate molecule to each of two free sulphydryl groups in hæmoglobin and used the resulting changes in certain of the reflexions to prepare a projection along the *y*-axis of the unit cell[3]. Disappointingly, the projection was largely uninterpretable. This was because the thickness of the molecule along the axis of projection was 63 A. (corresponding to some 40 atomic diameters), so that the various features of the molecule were superposed in inextricable confusion, and even at the increased resolution of 2·7 A. it has proved impossible to disentangle them[4]. It was clear that further progress could only be made if the analysis were extended to three dimensions. As we shall see, this involves the collection of many more observations and the production of three or four different isomorphous replacements of the same unit cell, a requirement which presents great technical difficulties in most proteins.

The present article describes the application, at low resolution, of the isomorphous replacement method in three dimensions to type *A* crystals of sperm whale myoglobin[5]. The result is a three-dimensional Fourier, or electron-density, map of the unit cell, which for the first time reveals the general nature of the tertiary structure of a protein molecule.

Isomorphous Replacement in Myoglobin

No type of myoglobin has yet been found to contain free sulphydryl groups, so that the method of

attaching heavy atoms used by Perutz for hæmoglobin could not be employed. Eventually, we were able to attach several heavy atoms to the myoglobin molecule at different specific sites by crystallizing it with a variety of heavy ions chosen because they might be expected, on general chemical grounds, to possess affinity for protein side-chains. X-ray, rather than chemical, methods were used to determine whether combination had taken place, and, if so, whether the ligand was situated predominantly at a single site on the surface of the molecule. Among others, the following ligands were found to combine in a way suitable for the present purpose: (i) potassium mercuri-iodide and auri-iodide; (ii) silver nitrate, potassium auri-chloride; (iii) *p*-chloromercuri-benzene sulphonate; (iv) mercury diammine $(Hg(NH_3)^{2+}$, prepared by dissolving mercuric oxide in hot strong ammonium sulphate), *p*-chloro-aniline; (v) *p*-iodo-phenylhydroxylamine. Each group of ligands combined specifically at a particular site, five distinct sites being found in all. The substituted phenylhydroxylamine is a specific reagent for the iron atom of the hæm group[6], and may be assumed to combine with that group; in none of the other ligands have we any certain knowledge of the mechanism of attachment or of the chemical nature of the site involved.

Methods of X-ray Analysis

Type *A* crystals of myoglobin are monoclinic (space group $P2_1$) and contain two protein molecules per unit cell. Only the *h*0*l* reflexions are 'real', that is, can be regarded as having relative phase angles limited to 0 or π, or positive or negative signs, rather than general phases; when introduced into a Fourier synthesis, these reflexions give a projection of the contents of the cell along its *y*-axis. In two dimensions the analysis followed lines[7] similar to that of hæmoglobin. First, the heavy atom was located by carrying out a so-called difference-Patterson synthesis; if all the heavy atoms are located at the same site on every molecule in the crystal, this synthesis will contain only one peak, from the position of which the *x*- and *z*-co-ordinates of the heavy atom can be deduced, and the signs of the *h*0*l* reflexions determined. These signs were cross-checked by repeating the analysis for each separate isomorphous replacement in turn; we are sure of almost all of them to a resolution of 4 A., and of most to 1·9 A. Using the signs, together with the measured amplitudes, we may, finally, compute an electron-density projection of the contents of the unit cell along *y*; but, as in hæmoglobin and for the same reasons, the projection is in most respects uninterpretable (even though here the axis of projection is only 31 A.). On the other hand, knowledge of the signs of the *h*0*l* reflexions to high resolution enabled us to determine the *x*- and *z*-co-ordinates of all the heavy atoms with some precision. This was the starting point for the three-dimensional analysis now to be described.

In three dimensions the procedure is much more lengthy because all the general reflexions *hkl* must be included in the synthesis, and more complicated because these reflexions may have any relative phase angles, not only 0 or π. Furthermore, we need to know all three co-ordinates of the heavy atoms; the two-dimensional analysis gives *x* and *z*, but to find *y* is more difficult, and details of the methods used will be published elsewhere, including among others two proposed by Perutz[8] and one proposed by Bragg[9]. Finally, a formal ambiguity enters into the deduction of general phase angles if only one isomorphous replacement is available; this can be resolved by using several replacements[10], such as are available in the present case. Once the phases of the general reflexions have been determined, one can carry out a three-dimensional Fourier synthesis which will be a representation of the electron density at every point in the unit cell.

Before such a programme is embarked upon, however, the resolution to be aimed at must be decided. The number of reflexions needed, and hence the amount of labour, is proportional to the cube of the resolution. To resolve individual atoms it would be necessary to include at least all terms of the series with spacings greater than 1·5 A.—some 20,000 in all; and it is to be remembered that the intensities of all the reflexions would have to be measured for *each* isomorphous derivative. Besides this, introduction of a heavy group may cause slight distortion of the crystal lattice; as the resolution is increased, this distortion has an increasingly serious effect on the accuracy of phase determination. In the present stage of the analysis the most urgent objective was an electron-density map detailed enough to show the general layout of the molecule—in other words, its tertiary structure. If the α-helix, or something like it, forms the basis of the structure, we need only work to a resolution sufficient to show up a helical chain as a rod of high electron density. For this purpose we require only reflexions with spacings greater than about 6 A.; in all there are some 400 of these, of which about 100 are *h*0*l*'s already investigated in the two-dimensional study. The Fourier synthesis described here is computed from these 400 reflexions only, and is in consequence blurred; besides this, it is distorted by an unknown amount of experimental error, believed to be small but at the moment difficult to estimate. Thus while the general features of the synthesis are undoubtedly correct, there may be some spurious detail which will require correction at a later stage.

The Three-dimensional Fourier Synthesis

The synthesis was computed in 70 min. on the EDSAC Mark I electronic computer at Cambridge (as a check, parts of the computation were repeated on DEUCE at the National Physical Laboratory). It is in the form of sixteen sections perpendicular to *y* and spaced nearly 2 A. apart; these must be piled on top of one another to represent the electron density throughout the cell, containing two myoglobin molecules together with associated mother liquor (which amounts to nearly half the whole). Unfortunately, the synthesis cannot be so represented within the two-dimensional pages of a journal; furthermore, if the sections are displayed side by side, they give no useful idea of the structure they represent. The examples reproduced in Fig. 1 illustrate some of the more striking features.

A first glance at the synthesis shows that it contains a number of prominent rods of high electron density; these usually run fairly straight for distances of 20, 30 or 40 A., though there are some curved ones as well. Their cross-section is generally nearly circular, their diameter about 5 A., and they tend to lie at distances from their neighbours of 8–10 A. (axis to axis). In some instances two segments of rod are joined by fairly sharp corners. Fig. 1*a*

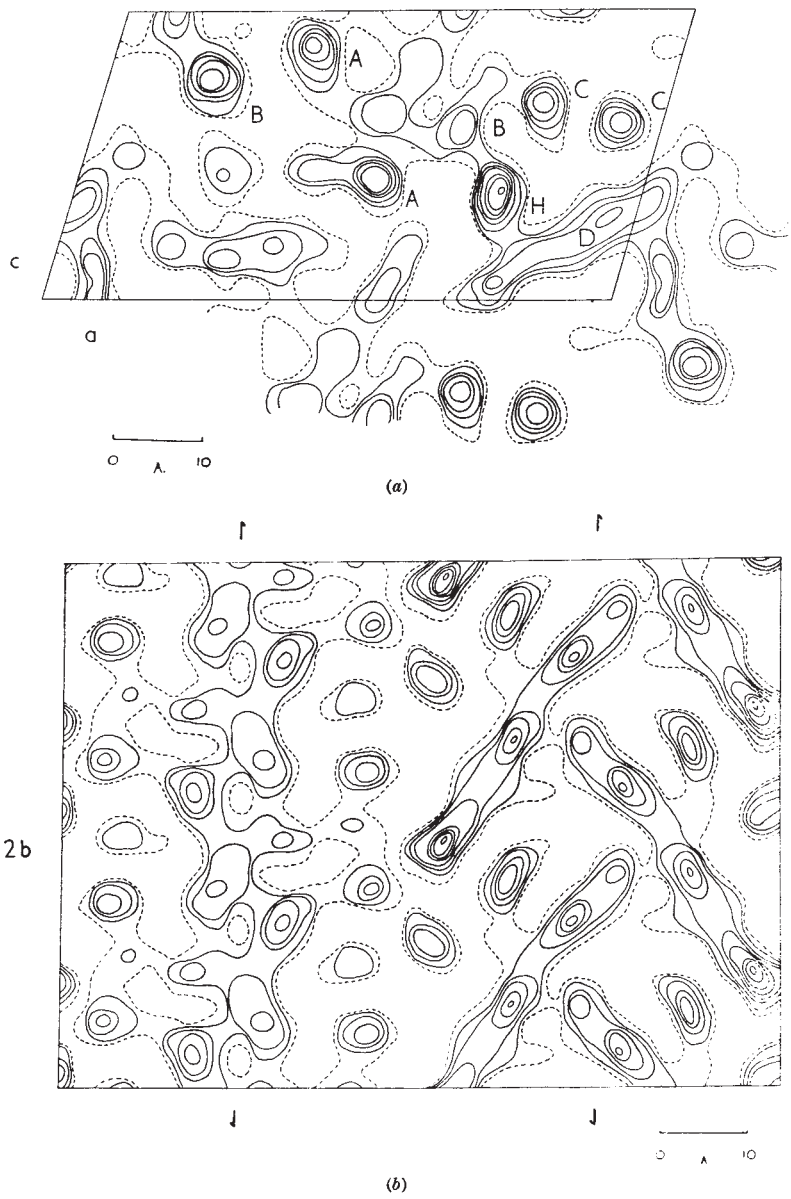

Fig. 1. (*a*) Section of three-dimensional Fourier synthesis of type *A* myoglobin at $y = -1/8$ *b*. *A*–*D*, polypeptide chains; *H*, hæm group. (*b*) Section parallel to [20$\bar{1}$] at $x = 0$, showing polypeptide chain *A* (on the right)

shows several rods—three of them (*A*, *B* and *C*) cross the plane of the section almost at right angles, while one (*D*) lies nearly in that plane. *D* is part of a nearly straight segment of chain about 40 A. long, of which some 20 A. is visible in this section. It seems virtually certain that these rods of high density are the polypeptide chains themselves—indeed, it is hard to think of any other features of the structure which they could possibly be. Their circular cross-section is what would be expected if the configuration were helical, and the electron density along their axes is of the right order for a helical arrangement such as the α-helix. The various rods in the structure are intertwined in a very complex manner, the nature of which we shall describe later.

Another prominent feature is a single disk-shaped region of high electron density which reaches a peak value greater than at any other point in the cell. A section through this disk is shown at *H* in Fig. 1*a*. We identify this feature as the hæm group itself, for the following reasons: (i) the hæm group is a flat disk of about the same size; (ii) its centre is occupied by an iron atom and therefore has a higher electron density than any other point in the whole molecule; (iii) a difference-Fourier projection of the *p*-iodophenylhydroxylamine derivative shows that, at least in *y*-projection, the position of the iodine atom is near that of our group; this is what we should expect, since this reagent specifically combines with the hæm group; (iv) the orientation of the disk corresponds, as closely as the limited resolution of the map allows one to determine it, with the orientation of the hæm group deduced from measurements of electron spin resonance[5,11].

We cannot understand the structure of the molecules in the crystal unless we can decide where one ends and its neighbours begin. In a protein crystal the interstices are occupied by mother liquor, in this case strong ammonium sulphate, the electron density of which is nearly equal to the average for the whole cell. Hence it is to be expected that in the intermolecular regions the electron density will be near average (the density of coiled polypeptide chains is much above average, and that of side-chains well below). It should also be fairly uniform; these regions should not be crossed by major features such as polypeptide chains. Using these criteria, it is possible to outline the whole molecule with minor uncertainties. It was gratifying to find that the result agreed very well, in projection, with a salt-water difference-Fourier projection made as part of the two-dimensional programme (for the principles involved, see ref. 12). Moreover, the dimensions of the molecule agreed closely with those deduced from packing considerations in various types of unit cell.

The Myoglobin Molecule

We are now in a position to study the tertiary structure of a single myoglobin molecule separated from its neighbours. Fig. 2 illustrates various views of a three-dimensional model constructed to show the regions of high electron density in the isolated molecule. Several points must be noticed. First, the model shows only the general distribution of dense regions. The core of a helical polypeptide chain would be such a region; but if the chain were pulled out, into a β-configuration, for example, its mean density would drop to near the average for the cell and the chain would fade out at this resolution.

Similarly, side-chains should, in general, scarcely show up, so that the polypeptide rods in the model must be imagined as clothed in an invisible integument of side-chains, so thick that neighbouring chains in reality touch. Third, features other than polypeptide chains may be responsible for some of the regions of high density; patches of adsorbed salt, for example. Fourth, the surface chosen to demarcate a molecule cannot be traced everywhere with certainty, so it is possible that the molecule shown contains parts of its neighbours, and correspondingly lacks part of its own substance.

Making due allowance for these difficulties, we may note the main features. It is known[13] that myoglobin has only one terminal amino-group: it is simplest to suppose that it consists of a single polypeptide chain. This chain is folded to form a flat disk of dimensions about 43 A. × 35 A. × 23 A. Within the disk chains pursue a complicated course, turning through large angles and generally behaving so irregularly that it is difficult to describe the arrangement in simple terms; but we note the strong tendency for neighbouring chains to lie 8–10 A. apart in spite of the irregularity. One might loosely say that the molecule consists of two layers of chains, the predominant directions of which are nearly at right angles in the two layers. If we attempt to trace a single continuous chain throughout the model, we soon run into difficulties and ambiguities, because we must follow it around corners, and it is precisely at corners that the chain must lose the tightly packed configuration which alone makes it visible at this resolution (an α-helix, for example, cannot turn corners without its helical configuration being disrupted). Also, there are several apparent bridges between neighbouring chains, perhaps due to the apposition of bulky side-chains. The model is certainly compatible with a single continuous chain, but there are at least two alternative ways of tracing it through the molecule, and it will not be possible to ascertain which (if either) is correct until the resolution has been improved. Of the secondary structure we can see virtually nothing directly at this stage. Owing to the corners, the chain cannot be in helical configuration throughout; in fact, the total length of chain in the model is 300 A., whereas an α-helix of 152 residues would be only 228 A. long. The 300 A. might correspond, for example, to 70 per cent α-helix and 30 per cent fully extended chain, but of course intermediate configurations are probably present, too. The hæm group is held in the structure by links to at least four neighbouring chains; nevertheless, one side of it is readily accessible from the environment to oxygen and to larger reagents such as *p*-iodo-phenylhydroxylamine (in the difference-Fourier projection of this complex, referred to above, the position of the iodine atom indicates that the ligand is attached to the outside of the group). Clearly, however, the model cannot at present be correlated in detail with what we know of the chemistry of myoglobin; this must await further refinement.

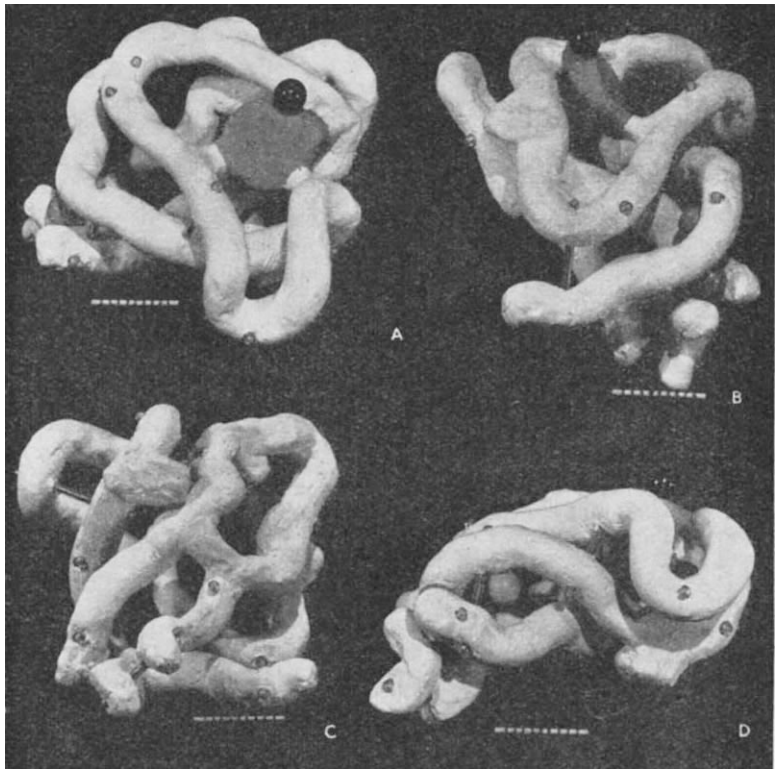

Fig. 2. Photographs of a model of the myoglobin molecule. Polypeptide chains are white; the grey disk is the hæm group. The three spheres show positions at which heavy atoms were attached to the molecule (black: Hg of *p*-chloro-mercuri-benzene-sulphonate; dark grey: Hg of mercury diammine; light grey: Au of auri-chloride). The marks on the scale are 1 A. apart

Perhaps the most remarkable features of the molecule are its complexity and its lack of symmetry. The arrangement seems to be almost totally lacking in the kind of regularities which one instinctively anticipates, and it is more complicated than has been predicated by any theory of protein structure. Though the detailed principles of construction do not yet emerge, we may hope that they will do so at a later stage of the analysis. We are at present engaged in extending the resolution to 3 A., which should show us something of the secondary structure; we anticipate that still further extensions will later be possible—eventually, perhaps, to the point of revealing even the primary structure.

Full details of this work will be published elsewhere. We wish to record our debt to Miss Mary Pinkerton for assistance of all kinds; to the Mathematical Laboratory, University of Cambridge, for computing facilities on the EDSAC; to Dr. J. S. Rollett and the National Physical Laboratory for similar facilities on the DEUCE; to Mrs. Joan Blows and Miss Ann Mansfield for assistance in computing; for fellowships to the U.S. Public Health Service (H. W.), the Merck Fellowship Board (R. G. P.), the U.S. National Science Foundation (R. G. P. and H. M. D.), and the Rockefeller Foundation (H. M. D.); and to Sir Lawrence Bragg for his interest and encouragement. Finally, we wish to

express our profound gratitude to the Rockefeller Foundation, which has actively supported this research from its earliest beginnings.

[1] Sanger, F., and Tuppy, H., *Biochem. J.*, **49**, 481 (1951). Sanger, F., and Thompson, E. O. P., *ibid.*, **53**, 353, 366 (1953).
[2] Green, D. W., Ingram, V. M., and Perutz, M. F., *Proc. Roy. Soc.*, A, **225**, 287 (1954).
[3] Bragg, W. L., and Perutz, M. F., *Proc. Roy. Soc.*, A, **225**, 315 (1954).
[4] Dintzis, H. M., Cullis, A. F., and Perutz, M. F. (in the press).
[5] Kendrew, J. C., and Parrish, R. G., *Proc. Roy. Soc.*, A, **238**, 305 (1956).
[6] Jung, F., *Naturwiss.*, **28**, 264 (1940). Keilin, D., and Hartree, E. F., *Nature*, **151**, 390 (1943).
[7] Bluhm, M. M., Bodo, G., Dintzis, H. M., and Kendrew, J. C. (in the press).
[8] Perutz, M. F., *Acta Cryst.*, **9**, 867 (1956).
[9] Bragg, W. L. (in the press).
[10] Bokhoven, C., Schoone, J. C., and Bijvoet, J. M., *Acta Cryst.*, **4**, 275 (1951).
[11] Ingram, D. J. E., and Kendrew, J. C., *Nature*, **178**, 905 (1956).
[12] Bragg, W. L., and Perutz, M. F., *Acta Cryst.*, **5**, 277 (1952).
[13] Schmid, K., *Helv. Chim. Acta*, **32**, 105 (1949). Ingram, V. M. (unpublished work).

RECENT CLIMATIC CHANGES

By Dr. E. B. KRAUS
Snowy Mountains Hydro-Electric Authority, Cooma North, New South Wales, Australia

CLIMATE is nowhere invariant. The existence of a never-ending change is most apparent along the shifting boundaries of the various climatic provinces. During the past few hundred years the boundaries of the Alpine and Sub-Arctic regions have experienced some striking changes in glaciation and considerable variations in the extent and the seasonal duration of sea ice. It has now been established from the analysis of rainfall and stream-flow records that fluctuations of corresponding or even larger amplitude occurred in the sub-tropics at the boundary of the arid zone. During the same time the climates of the Antarctic ice cap, the central Sahara or the Amazon rain forest may not have altered a great deal.

Changes in a time series can be illustrated to advantage by graphs of cumulative percentual deviations from a mean. Graphs of this type accentuate changes in a time series, while running averages or orthodox statistical analyses tend to smooth them out. They demonstrate, at a glance, the existence of a trend and the time when it changes.

In the following diagrams (Figs. 1–6), the ordinate :

$$y(n) = 100 \sum_{l=1881}^{n} \left(\frac{r_l}{\bar{r}} - 1 \right)$$

where n represents the running calendar year, r_l the rainfall or stream discharge for the year l, and $\bar{r}$ the mean for the years 1881–1940. It is easily seen that y must be zero at the beginning of 1881 and the end of 1940. The graph of y rises for $r_l > \bar{r}$ and vice versa. It will be concave upwards during periods when r is increasing with time and convex when it is decreasing. The percentual deviation from $\bar{r}$ of the mean for any other period, say the years n to $(n + m)$, is given simply by $[y\,(n + m) - y(m)]/m$.

The accompanying six diagrams have been brought together to demonstrate how fluctuations of the same type occur at about the same time all around the globe. Fig. 1 illustrates the changes which occurred in the rainfall regime of semi-arid, north-western New South Wales (Australian rainfall district No. 48). It can be deduced from the curve that the mean rainfall for the years 1879–97 was 26 per cent above the 1881–1940 average. It was 11 per cent below that average during the period 1898–1946. In absolute terms, the mean annual rainfall from the beginning of records to 1897 was 16·90 in ; during the following forty-nine years it was 12·02 in. During the past ten years conditions were again more similar to those of the nineteenth century.

Fig. 2 deals with the east coast of Queensland (Australian rainfall district No. 40). The rainfall there is more seasonal but also more than three times as heavy as in north-western New South Wales. The onset of the dry period occurred two years later. It was interrupted by wetter conditions after about 1923, though a second severe drought occurred in the late '30's and early 1940's.

Conditions on the east coast of North America are illustrated by Fig. 3, which gives the mean rainfall for Charleston and Cape Hatteras. In an earlier paper[4] it had been shown that fluctuations there were paralleled by the records of other stations along the American east coast, and that the records suggest the persistence of relatively wet conditions all through the nineteenth and the later eighteenth centuries.

Fig. 4 shows the changes which occurred in the space-average of a large number of rainfall stations used by the Indian Meteorological Service to evaluate and forecast monsoon rainfall over the Indian peninsula. The change in the rainfall regime there was comparatively small and the vertical scale of the diagram has therefore been doubled. The same pattern of change with somewhat larger amplitude occurred in the more arid districts of north-west India and West Pakistan. All over this region the dry period which began at the end of the past century was ameliorated or terminated by about 1910.

Fig. 5 treats the discharge of the Nile at Aswan in a similar way. The difference in the mean discharge before and after 1898 amounts, in this case, to more than 30 per cent or 27×10^9 m.^{3}/year.

The changes which occurred in parts of South Africa are illustrated by Fig. 6, which represents the South African rainfall district No. 16*A* in the Central Cape Province. Unfortunately, data after 1946 were not available in Australia for the two African cases.

Arakawa[1] and I have shown that changes similar to those demonstrated here have occurred in maritime parts of East Asia and in the West Indies, Hawaii and other tropical stations. They can also be found in Rhodesian rainfall records. Together, this material would seem to prove the existence of parallel and rather abrupt climatic fluctuations over a large part of the Earth.

The amplitude of the change in rainfall regime was largest at the fringes of the arid zone. In the extreme case of Aden, the rainfall record shows a difference of 84 per cent between the means for the thirty years after 1894 and the preceding thirteen years of avail-