# RELATION BETWEEN STRUCTURE AND SEQUENCE OF HÆMOGLOBIN

By Dr. M. F. PERUTZ, F.R.S.
Medical Research Council Laboratory for Molecular Biology,
Hills Road, Cambridge

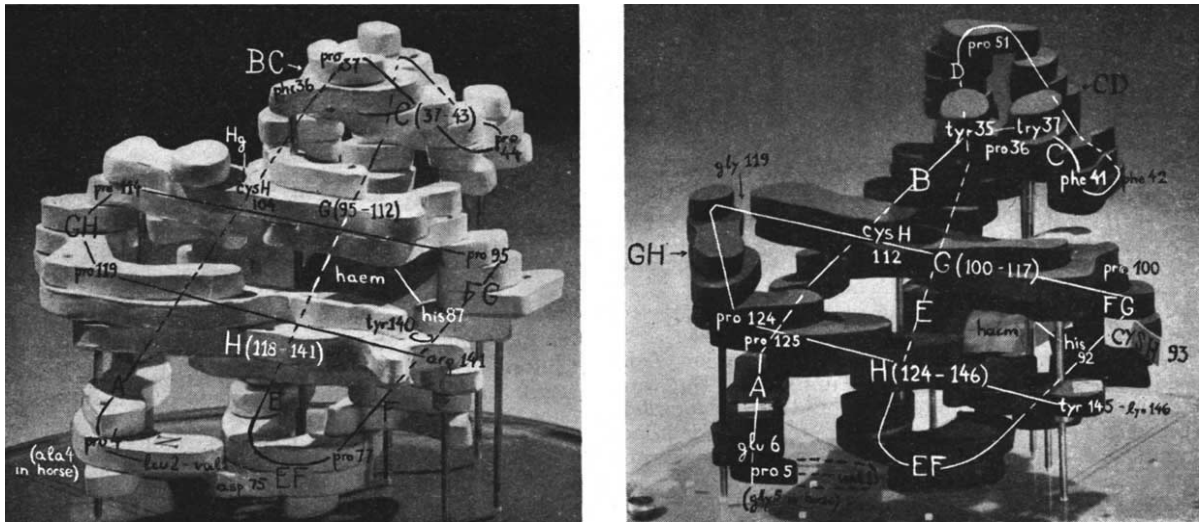
WHEN the tertiary structure of horse hæmoglobin was first published, its amino-acid sequence was still unknown, and the resolution of the X-ray analysis was too low to allow side-chains to be seen[1]. Since then, the amino-acid sequences in the two different chains of human hæmoglobin and in one of the chains of horse hæmoglobin have been elucidated[2–5]. In addition, most of the structure of sperm whale myoglobin is now known in atomic detail[6]. Thanks to the close similarity in tertiary structure between sperm whale myoglobin and horse hæmoglobin the position of almost any residue of a given sequence number can be found by looking up the co-ordinates of the corresponding residue in myoglobin and then transferring those co-ordinates to the model of hæmoglobin. This procedure can be checked, because the positions of certain residues can be found directly in the electron-density maps of hæmoglobin, despite the low resolution, and can then be compared with those inferred by analogy with myoglobin. For example, the cysteines are found by labelling them with mercurials; the indol rings of tryptophan are recognizable as bulges of high electron density protruding from the main chain; by contrast, glycines in non-helical regions sometimes show up by causing a marked constriction in the roughly cylindrical cloud of high electron density which represents the polypeptide chain.

A year ago, Watson and Kendrew published a preliminary comparison of the structure and sequence of hæmoglobin, but at that stage only parts of the chemical sequence were known[7]. In the light of the further knowledge now available, many interesting new features reveal themselves.

In the discussion which follows, the tertiary structures of equine and human hæmoglobin will be assumed to resemble each other closely. This has not yet been proved; but is very probable in view of the similarity in amino-acid sequence of these two hæmoglobins, and in view of the correspondence between the tertiary structures of the different species of myoglobin and hæmoglobin so far examined[9,10].

## Chemical Identification of Chains in the Crystallographic Model

The two different chains of horse hæmoglobin have the *N*-terminal sequences valyl-leucyl and valyl-glutaminyl, and are similar in composition to the valyl-leucyl and valyl-histidyl-leucyl chains of human hæmoglobin[4,5], which have become known respectively as the α- and β-chains. Smith and Perutz have shown the valyl-glutaminyl chain to correspond to the black chain of the crystallographic model[8], from which it would follow that the black chain also corresponds to the β-chain of human




*a* *b*

Fig. 1. *a*, A model of the cloud of high electron density representing the α-chain of horse hæmoglobin. The letters *A*, *C*, *E*, *F*, *G* and *H* refer to various regions of chain identified as right-handed α-helices in myoglobin. The numbers in brackets give the sequential residue numbers associated with them. *EF*, *FG*, *GH* refer to corners or non-helical regions. Except for proline $4_\alpha$ all the residues marked in this figure occupy the same sequential positions in human and equine hæmoglobin. Note the proximity of aspartic acid 75 to valine 1, the *N*-terminus

*b*, The β-chain. The residues marked on the model are taken from the sequence of human hæmoglobin. Cysteine $112_\beta$ is known to be replaced by another residue in equine hæmoglobin; there may be some other differences between the two hæmoglobins which affect this figure and which are not yet known. Note the position of glutamic acid 6, the residue which is replaced by valine in sickle-cell hæmoglobin; also the bulges marking the positions of tyrosine 35, tryptophan 37 and of phenylalanines 41 and 42, and the gap which marks the position of glycine 119 in the non-helical region *GH*. The hæm group lies at the back of the models

Table 1. ABBREVIATED AMINO-ACID SEQUENCES OF VARIOUS HAEMOGLOBIN CHAINS, ILLUSTRATING POINTS MENTIONED IN THE TEXT
For the α-chain of horse hæmoglobin only those residues are shown which differ from the corresponding ones in human hæmoglobin. In the β-chain of horse hæmoglobin only the *N*-terminal sequence is shown which differs from human hæmoglobin in three places. The vertical lines indicate tryptic peptides the composition of which is either the same as that of the corresponding peptide of human hæmoglobin or differs from it by only one amino-acid residue. In the γ-chain only those residues are shown which differ from the corresponding ones of the β-chain or (bracketed) where only the composition, but not the sequence of the peptide is known

| Human α | | Horse α | Human β | | Horse β | Human γ | Remarks |
|---|---|---|---|---|---|---|---|
| No. | Residue | | No. | Residue | | | |
| | | | 1 | val | | gly | |
| 1 | val | | 2 | his | gluN | | These residues form a hook at the *N*-terminal end which has no counter-part in myoglobin |
| 2 | leu | | 3 | leu | | phe | |
| 3 | ser | | 4 | thr | ser | | |
| 4 | pro | ala | 5 | pro | gly | glu | |
| 5 | ala | | 6 | glu | | | glu 6β is replaced by valine in hæmoglobin S |
| 14 | try | | 15 | try | | | Tryptophan responsible for photodissociation effect |
| 15 | gly | | 16 | gly | | | |
| 36 | phe | | 35 | tyr | | | |
| 37 | pro | | 36 | pro | | | Proline in *CD* corner |
| 38 | thr | | 37 | try | | | Tryptophan replaced by lysine in 'fast' horse hæmoglobin |
| 44 | pro | | 43 | glu | | asp | |
| 45 | his | | 44 | ser | | | |
| 46 | phe | | 45 | phe | | | |
| | | | 46 | gly | | | α-Chain shortened by one residue |
| 47 | asp | | 47 | asp | | | |
| 48 | leu | | 48 | leu | | | |
| 49 | ser | | 49 | ser | | | |
| 50 | his | | 50 | thr | | ser | |
| | | | 51 | pro | | ala | Proline at *N*-terminal end of helix *D* in β-chain |
| | | | 52 | asp | | ser | |
| | | | 53 | ala | | | α-Chain shortened by five residues |
| | | | 54 | val | | ileu | |
| | | | 55 | met | | | |
| 51 | gly | | 56 | gly | | | |
| 52 | ser | | 57 | asp | | | |
| 53 | ala | | 58 | pro | | | Proline in corner *DE* in β-chain |
| 54 | gluN | | 59 | lys | | | |
| 55 | val | | 60 | val | | | |
| 56 | lys | | 61 | lys | | | |
| 57 | gly | ala | 62 | ala | | | |
| 58 | *his* | | 63 | *his* | | | Histidine on distal side of hæm |
| 59 | gly | | 64 | gly | | | Glycine essential for short contact between helices *E* and *B* |
| 60 | lys | | 65 | lys | | | |
| 61 | lys | | 66 | lys | | | Three lysines in neighbourhood of distal histidine forming Braunitzer's 'basic centre' |
| 76 | met | leu | 81 | leu | | | |
| 77 | pro | | 82 | lys | | | |
| 78 | aspN | gly | 83 | gly | | | Proline in non-helical region *EF* of α-chain |
| 86 | leu | | 91 | leu | | | |
| 87 | *his* | | 92 | *his* | | [his ?] | Hæm linked histidine |
| 88 | ala | | 93 | cysH | | [cysH?] | Reactive cysteine in β-chain |
| 89 | his | | 94 | asp | | [asp ?] | |
| 95 | pro | | 100 | pro | | | Proline in corner *FG* |
| 104 | cysH | | 109 | val | | [val?] | Unreactive cysteine in α-chain |
| 107 | val | ser | 112 | cysH | not cysH | thr | Unreactive cysteine in human β-chain |
| 114 | pro | | 119 | gly | | | Proline in non-helical region *GH* of α-chain |
| 115 | ala | asp | 120 | lys | | | |
| 116 | glu | asp | 121 | glu | | | |
| 117 | phe | | 122 | phe | | | |
| 118 | thr | | 123 | thr | | | |
| 119 | pro | | 124 | pro | | | Prolines at *C*-terminal end of helix *H* |
| 120 | ala | | 125 | pro | | glu | |
| 139 | lys | | 144 | lys | | arg | |
| 140 | tyr | | 145 | tyr | | | Tyrosine forming hydrogen bond with main chain carbonyl in *FG* corner |
| 141 | arg | | 146 | his | | | |

hæmoglobin. We therefore have the following relations:

| | *N-Terminal sequences* | | *Crystallographic model* |
|---|---|---|---|
| | *Horse* | *Human* | |
| α-Chain | valyl-leucyl | valyl-leucyl | white chain |
| β-Chain | valyl-glutaminyl | valyl-histidyl-leucyl | black chain |

Table 1 lists those parts of the sequences of human and equine hæmoglobin which are discussed in this article, while Fig. 1 shows the tertiary structures of the α- and β-chains with the positions of certain residues superimposed. The sequences of the α- and β-chains are similar, but to bring them into register all along their length appropriate gaps have to be left in certain positions. Thus a gap of one, and another of five, residues have to be assumed between residues Nos. 46 and 51 in the α-chain in order to keep its sequence in register with that of the β-chain. These gaps evidently correspond to a shortening of the α-chain which should be noticeable in the model. By looking up the positions of the corresponding residues in myoglobin the shortening can be traced to the non-helical region *CD* and the helical region *D* which lie in the loop at the top of the models in Fig. 1. It will be noted that in the white chain this loop is cut short compared with the black chain, showing that the shortening of the α-sequence has its corollary in the electron density maps.

## Co-ordination of Hæm Groups

The hæm-linked residue in myoglobin has been identified as histidine. The α- and β-sequences each show a histidine in the corresponding position ($87_\alpha$ and $92_\beta$), thus confirming that histidine is the hæm-linked residue in hæmoglobin, as had long been predicted from physico-chemical data[12–14]. Watson and Kendrew have pointed out that a second histidine occurs on the distal side of the hæm group in all

three chains (62 in myoglobin; $58_\alpha$ and $63_\beta$); this histidine is not in contact with the iron atom, but is within hydrogen-bonding distance of the hæm-linked water molecule, and is the residue which is replaced by tyrosine in two of the forms of hæmoglobin M (ref. 15). This may be the second histidine in close proximity to the iron atom, but not directly linked to it, the presence of which has long been surmised[11–13,16,17]. It is most satisfactory to find these expectations confirmed by crystallographic analysis.

## Position of Cysteines

All vertebrate hæmoglobins appear to contain at least one pair of cysteines the sulphydryl groups of which are reactive in the native protein. Their positions can be determined by X-ray analysis, using mercurials as a label. From the X-ray results, combined with model building, it was possible to show that in horse hæmoglobin the reactive cysteine is the residue immediately following the hæm-linked histidine in the β-chain, and a cysteine in this position has since been found by chemical sequence work in human hæmoglobin ($93_\beta$). The histidine and cysteine are part of an α-helix, and the side-chain of cysteine points in the direction away from the iron atom.

Horse hæmoglobin contains one other pair of cysteines which is unreactive in the native protein[18,19], but can be labelled with mercurials under appropriate conditions. This is attached to the helical regions *G* of the α-chains, and lies tucked away in the interior of the molecule, sandwiched between neighbouring sub-units. The position deduced by labelling with mercury corresponds exactly to that inferred for residue No. $104_\alpha$, the sequential position of the corresponding cysteine in human hæmoglobin, by analogy with the myoglobin structure.

Adult human hæmoglobin, but not fœtal human or horse hæmoglobin, contains a third pair of cysteines ($112_\beta$), the position of which in the model is similar to that of $104_\alpha$ in that it is part of the helical region *G* and lies in the interior of the molecule. This pair is also unreactive in the native protein[19,20]. Neither of the unreactive pairs has any contact with the hæm group, nor are they in contact with each other or with residues of any chain other than their own. Thus their function is obscure. Their lack of reactivity in the native protein can be attributed to their situation in the interior of the molecule, though covering of the sulphydryl groups by immediate neighbours cannot be excluded at this stage.

## Tryptophan

As has already been mentioned, the indol groups of tryptophan are recognizable as marked bulges of high electron density protruding from the main polypeptide chain, even at a resolution of only 5·5 Å. Tryptophan $14_\alpha$ and $15_\beta$ appear in the exact positions occupied by tryptophan 12 in myoglobin and are part of the helical region *A*. The indol ring lies in the interior of the molecule, sandwiched between helices *A* and *E*; it is almost parallel to the plane of the porphyrin ring, but separated from the latter by helix *E*, and lies at a distance of 14 Å. from the iron atom. Since this is the only tryptophan which all chains of hæmoglobin and myoglobin so far analysed have in common, it is probably the residue responsible for the photodissociation observed when carbon monoxyhæmoglobin is illuminated at the ultra-violet absorption band of tryptophan. In this process a quantum must be transferred from the indol ring, across helix *E*, to the porphyrin ring. However, the photodissociation of carbon monoxide is unlikely to be of biological significance, and the common occurrence of tryptophan in this position suggests that it may actually play a part in the oxygen-combining activity in a manner which we cannot yet guess. The β-chain of human hæmoglobin contains another tryptophan in position 38. This is also present in the slow component of horse hæmoglobin, but in the fast component it is replaced by lysine[21]. Myoglobin contains a second tryptophan in position 5 which is absent in hæmoglobin. Evidently these tryptophans are not essential.

## Prolines and Corner-turning

The two chains of human hæmoglobin each contain seven prolines which all lie in corners or non-helical regions of the chain, as had been suggested by Watson and Kendrew[7]. However, there are only three positions (the corner *CD*, the *N*-terminal end of helix *G* and the corner *GH*) where prolines occur both in myoglobin and in the α- and β-chains. The fourth proline of myoglobin (98) occurs in a position where the hæmoglobin chains have none, and curiously enough, the prolines which occur in each of the chains of hæmoglobin do not all occur at the same corners. For example, in the neighbourhood of the short helical region *D* the α-chain has no proline and the β-chain has two (51 and 58). On the other hand, the α-chain has a proline near the *C*-terminal end of the region *C* (44) where the β-chain has none. Admittedly, the regions *C*, *CD* and *D* are those where the two chains also differ most markedly in structure, but there is also proline $77_\alpha$ which occurs in the non-helical region *EF* and has no counterpart in the β-chain. In the γ-chain of human fœtal hæmoglobin, which replaces the β-chain in the adult form, the prolines in positions 5 and 51 are replaced by glutamic acid and alanine respectively[22]. The two chains of horse hæmoglobin contain only six prolines each, and differ from the human chains by the absence of prolines $4_\alpha$ and $5_\beta$ at the *N*-terminal end of helix *A*; these are replaced respectively by alanine and glycine. These results suggest that the corner-turning mechanism which depends on the presence of proline is readily replaced by others, and is therefore not of great evolutionary importance.

The correspondence between structure and sequence found in this work provides an independent check for the interpretation of the X-ray results. If the meaning of the electron density maps of either hæmoglobin or myoglobin had been misinterpreted, the correspondence between the observed and predicted positions of cysteine and other residues could not have been found.

A full account of this work, giving a complete description of the methods and results of the X-ray analysis, has been published by Cullis, Muirhead, North, Perutz and Rossmann[23].


[1] Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., and North, A. C. T., *Nature*, **185**, 416 (1960).

[2] Braunitzer, G., Gehring-Müller, R., Hilschmann, N., Hilse, K., Hobom, G., Rudloff, V., and Wittmann-Liebold, B., *Hoppe-Seyl. Z.*, **325**, 283 (1961).

[3] Konigsberg, W., Guidotti, G., and Hill, R. J., *J. Biol. Chem.*, **236**, PC55 (1961).

[4] Braunitzer, G., and Matsuda, G., *Hoppe-Seyl. Z.*, **324**, 91 (1961).

[5] Smith, D. B. (private communication).

[6] Kendrew, J. C., Watson, H. C., Strandberg, B. E., Dickerson, R. E., Phillips, D. C., and Shore, V. C., *Nature*, **190**, 668 (1961).

[7] Watson, H. C., and Kendrew, J. C., *Nature*, **190**, 663 (1961).
[8] Smith, D. B., and Perutz, M. F., *Nature*, **188**, 406 (1960).
[9] Scouloudi, H., *Proc. Roy. Soc.*, A, **258**, 181 (1960).
[10] North, A. C. T., and Green, D. W. (unpublished results).
[11] Conant, J. B., *Harvey Lectures*, **28**, 159 (1932).
[12] Wyman, J., *J. Biol. Chem.*, **127**, 581 (1939).
[13] Coryell, C. D., and Pauling, L., *J. Biol. Chem.*, **132**, 769 (1940).
[14] Keilin, J., *Nature*, **187**, 365 (1960).
[15] Gerald, P. S., and Efron, M. L., *Proc. U.S. Nat. Acad. Sci.*, **47**, 1758 (1961).
[16] Haurowitz, F., *J. Biol. Chem.*, **193**, 443 (1951).
[17] Keilin, D., and Hartree, E. F., *Nature*, **170**, 161 (1952).
[18] Ingram, V. M., *Biochem. J.*, **59**, 653 (1955).
[19] Allison, A. C., and Cecil, R., *Biochem. J.*, **69**, 27 (1958).
[20] Cecil, R., and Snow, N. S., *Biochem. J.*, **82**, 247, 255 (1962).
[21] Braunitzer, G. (private communication).
[22] Schroeder, W. A., Shelton, J. R., Shelton, J. B., and Cormick, J., *Proc. U.S. Nat. Acad. Sci.*, **48**, 284 (1962).
[23] Cullis, A. F., Muirhead, H., North, A. C. T., Perutz, M. F., and Rossmann, M. G., *Proc. Roy. Soc.*, A, **265**, 15, 187 (1961-2).

# OBITUARIES

## Dr. Ernst Schwarz

SYSTEMATIC mammalogy suffered a severe loss with the death on September 23, at Bethesda, Maryland, of Dr. Ernst Schwarz. Although in recent years Schwarz had retired from active work in systematic mammalogy to devote his energies to public health problems his name will long be remembered for his earlier fundamental contributions to the diagnosis of a very wide range of mammalian groups from marsupials to primates.

Schwarz was born at Frankfurt am Main on December 1, 1889. His academic studies (1907–12) were pursued at the Universities of Berlin, London and Munich, and he graduated Ph.D. at the last mentioned in 1912. During 1912–15 he was in charge of the mammal collections at the Senckenbergische Naturforschende Gesellschaft in Frankfurt, thereafter (1915–18) transferring his interests to the clinical laboratory of the University Hospital in the same city. His interest in medical science was continued after the First World War when, in 1919, he became research assistant to the Paul Ehrlich Institute for Experimental Therapy, where his interests were directed towards cancer research. In 1925 he returned to mammalogy, spending the next four years in research at the Zoological Museum of the University of Berlin. During 1929–33 he was custodian of the Natural History Museum of the University of Greifswald, but fled from Germany to London on the advent of the Nazi régime. He spent the next four years attached unofficially to the mammal room of the British Museum, whence many of his best-known papers emanated. In 1937 he transferred to the U.S. National Museum in Washington, where he remained for ten years, with numerous intermediate assignments such as a guest lectureship at the U.S. Navy Medical School at Bethesda (1942–43) and later (1947–53) to the Graduate Medical School, U.S. Naval Hospital, Bethesda. During 1948 he was organizing, planning and participating in the U.S. Navy African Expedition. In 1950 he went on a mission to Venezuela studying bubonic plague, becoming a member of the World Health Organization Commission on this disease. His last years, 1956–59, were spent as a guest research worker at the Armed Forces Research Institute of Pathology at Washington, where he was attached to the Section of Neuropathology.

Schwarz was decorated with the Medal of King Leopold II of Belgium and with the Distinguished Service Medal of the United States.

As a systematic mammalogist, Schwarz is considered a lumper; in fact, he was perhaps overzealous in this direction, as witnessed especially by his treatment of the *Colobus* monkeys—a fact he himself admitted after the publication of his studies. His very last publication dealt with pygmy marmosets, where here too he oversimplified the problems. The position is tersely assessed by Booth, who declared that "Schwarz's descriptions are meticulously careful, but his conception of the species is, in the case of the Primates, often too broad".

Of some sixty-eight scientific papers published by Schwarz, about two-thirds were devoted to the systematics of mammals, living and extinct. The remainder deal with problems of applied medicine dating from his early interest in carcinogenesis and some few to general problems of distribution, genetics and evolution.

Deep sympathy goes to his widow, Henrietta E. Schwarz, whom he married during his period in London, and who was a constant help to him in his work.

W. C. OSMAN HILL

## Dr. N. K. Sen

DR. N. K. SEN, who died on March 21, was born on February 26, 1917, the eldest son of late Jitendra Prosad Sen, subordinate judge, who was the son of the well-known Dewan Bahadhur Sarada Prosad Sen of Dacca. Dr. Sen graduated with honours in botany from the Presidency College, Calcutta, in 1936, and M.Sc. in 1939.

During 1939–41, he was lecturer in botany, Jagannath College, and during 1942–43 in the Dacca Intermediate College and the Rajsahi College (now in East Pakistan). He joined the Botany Department of Presidency College, Calcutta, as a lecturer in botany in 1943 and stayed there until 1948. For two years (1948–50) he was a graduate student, University of California. In 1951 he was awarded the Ph.D. degree in genetics by that University. In 1951 he was also awarded D.Phil. (Science) degree of the University of Calcutta.

In 1952, Sen joined the Agricultural Engineering Department (Applied Botany Section) of the Indian Institute of Technology as assistant professor of applied botany. He published more than fifty scientific papers on genetics, cytogenetics and breeding of pulses.

He started the first course in India on applied botany. He was the editor of *Bulletin of the Botanical Society of Bengal* (1961–62), and first editor of *Bijnan Parichaya*, a scientific Bengali monthly, published from Dacca and founded by Prof. S. N. Bose, Prof. P. N. Mazumdar, Prof. S. G. Mukherjee, the late S. K. Sen, etc., under the inspiration of the late poet Rabindranath Tagore.

R. M. DATTA