Fisher, J. J., & Jardetzky, O. (1965) *J. Am. Chem. Soc. 87*, 3237.
Franks, F., & Eagland, D. (1975) *CRC Crit. Rev. Biochem. 3*, 165.
Fung, B. M., & Sarney, S. G. (1971) *Biochim. Biophys. Acta 237*, 135.
Gerig, J. T. (1968) *J. Am. Chem. Soc. 90*, 2681.
Handbook of Physical Chemistry (1974) WNT, Warsaw, Poland.
Hayes, M. B., Cohen, J. S., & McNeel, M. L. (1974) *Magn. Reson. Rev. 3*, 1.
Herskovits, T. T., & Jaillet, H. (1969) *Science 163*, 282.
Herskovits, T. T., Gadegbeku, B., & Jaillet, H. (1970) *J. Biol. Chem. 245*, 2588.
Hollis, D. P. (1967) *Biochemistry 6*, 2080.
Jacobson, A. L., & Krueger, P. J. (1975) *Biochim. Biophys. Acta 393*, 274.
Jardetzky, O. (1964) *Adv. Chem. Phys. 7*, 499.
Jardetzky, O., & Wade-Jardetzky, N. G. (1965) *Mol. Pharmacol. 1*, 214.
Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379.
Lubas, B., Soltysik, M., & Wieniewska, T. (1977a) in Abstracts, Sixth International Symposium on Magnetic Resonance, Banff, Canada, May 1977, p 122.
Lubas, B., Witman, B., Wieniewska, T., & Soltysik, M. (1977b) *Rep. IFJ Kraków (Poland)*, 956/P1, 38.
Nemethy, G., & Scheraga, H. A. (1962) *J. Phys. Chem. 66*, 1773.
Nozaki, Y., & Tanford, C. (1965) *J. Biol. Chem. 240*, 3568.
Ohama, H., Suigura, N., & Yagi, K. (1973) *J. Biochem. (Tokyo) 73*, 1123.
Ohama, H., Suigura, N., Tanaka, F., & Yagi, K. (1977) *Biochemistry 16*, 126.
Parodi, R. M., Bianchi, E., & Ciferri, A. (1973) *J. Biol. Chem. 218*, 4047.
Peters, T., Jr. (1975) *Plasma Proteins, 2nd Ed., 1975 1*, 140.
Randall, R. F., Stoddart, R. W., Metcalfe, S. M., & Metcalfe, J. C. (1972) *Biochim. Biophys. Acta 255*, 888.
Roberts, G. C. K., & Jardetzky, O. (1969) *Adv. Protein Chem. 24*, 447.
Robillard, K. A., & Wishnia, A. (1972a) *Biochemistry 11*, 3835.
Robillard, K. A., & Wishnia, A. (1972b) *Biochemistry 11*, 3841.
Russell, A. E. (1973) *Biochem. J. 131*, 335.
Russell, A. E., & Cooper, D. R. (1969) *Biochemistry 8*, 3980.
Schrier, E. E., Ingwall, R. T., & Scheraga, H. A. (1965) *J. Phys. Chem. 69*, 298.
Shrake, A., & Rupley, J. A. (1973) *J. Mol. Biol. 79*, 351.
St. Pierre, T., & Jencks, W. P. (1969) *Arch. Biochem. Biophys. 133*, 99.
Tan, H. K., & Lovrien, R. (1972) *J. Biol. Chem. 247*, 3237.
Tanford, C. (1968) *Adv. Protein Chem. 23*, 121.
Wishnia, A. (1969a) *Biochemistry 8*, 5064.
Wishnia, A. (1969b) *Biochemistry 8*, 5070.
Woodward, C. K., Ellis, L. M., & Rosenberg, A. (1975) *J. Biol. Chem. 250*, 440.
Zia, H., Cox, R. H., & Luzzi, L. A. (1971) *J. Pharm. Sci. 60*, 89.
Zimmerman, J. R., & Brittin, W. E. (1957) *J. Phys. Chem. 61*, 1328.

Structural Comparisons of Heme Binding Proteins†

Patrick Argos* and Michael G. Rossmann

ABSTRACT: Of the 82 three dimensionally characterized residues of cytochrome c_{551}, 49 are found to be structurally and topologically equivalent to the globin fold and 41 are equivalent to the cytochrome b_5 fold, with a respective root mean square separation of 3.5 and 4.9 Å between equivalenced C_α atoms. The common fold represents a central heme binding core, corresponding to the middle exon of certain globin genes. After superposition of the protein folds, the heme irons are found to be separated by 5.4 and 1.6 Å, while their heme normals are inclined by 6° and 32°, respectively. Furthermore, the heme "face", determined by the asymmetric attachment of the vinyl and propionyl side chains, is directed similarly in all three heme proteins. The heme itself is rotated by 72° and 116° about its normal, respectively. The minimum base change per codon for the three pairwise comparisons corresponds to the expected value of random sequence comparisons. While all three heme proteins may have diverged from a common ancestor, their similarity may have arisen from the requirements of heme binding or the utilization of a particularly stable fold. Known structures within commonly accepted divergent families were superimposed in order to discriminate better between convergence and divergence. Minimum base changes per codon, number of deletions and insertions, percentage of equivalenced residues, precision of heme superposition, and root mean square separation of equivalenced C_α atoms were tested as measures of evolutionary relationships.

The three-dimensional structures within three families of heme binding proteins have been well studied: the globins, the cytochrome *c* family, and the mammalian cytochrome b_5 from the endoplasmic reticulum. Rossmann & Argos (1975) have shown a reasonable structural similarity between the globin fold and cytochrome b_5. This paper demonstrates that a comparable structural similarity exists between the cytochrome *c* and globin folds.

The globin family extends further than the oxygen carriers hemoglobin and myoglobin found in higher vertebrates. Proteins with the same function and homologous structure have been found in the annelid worm *Glycera dibranchiata* (Padlan

† From the Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907. *Received April 17, 1979.* The work was supported by a National Institutes of Health grant (No. GM 10704) and a National Science Foundation grant (No. BMS74-23537) to M.G.R. and a National Science Foundation grant (No. PCM77-20287) and an American Cancer Society Faculty Research Award (No. FRA173) to P.A.

0006-2960/79/0418-4951$01.00/0

& Love, 1974), in sea lamprey (Hendrickson et al., 1973), in fruit fly (Weber et al., 1978), and even in the root nodules of leguminous plants (Vainshtein et al., 1975). Amino acid sequences from an even greater array of species have been determined [cf. Dayhoff (1972)], establishing that the basic globin fold must have appeared at least as early as the first eucaryotes, about 1.2 billion years ago.

The tertiary structure and amino acid sequence of eucaryotic cytochromes *c* show that this electron transfer protein in the oxygen respiratory chain has evolved only very slowly in the last billion years [cf. Dickerson & Timkovich (1975)]. The amino acid sequences (Ambler & Bartsch, 1975; Ambler, 1973) and structure determinations (Salemme et al., 1973a; Almassy & Dickerson, 1978; Timkovich & Dickerson, 1976; Weber & Salemme, 1977; Korszun & Salemme, 1977) of a large number of procaryotic cytochrome *c* like molecules have demonstrated that electron transport is mediated by these molecules not only in oxygen respiration but also in sulfide-using photosynthesis, sulfate respiration, cyclic photosynthesis, and water-using photosynthesis (Dickerson et al., 1976). Although these cytochrome *c* like molecules have many structural features in common, there are substantial differences. Their amino acid sequences, varying in length from 82 to 134 residues, cannot always be aligned without reference to their three-dimensional structures (Dickerson, 1979). Thus, the origin of the cytochrome *c* fold must be placed well before the occurrence of photosynthesis, about 2 billion years ago.

The classification associated with cytochrome *b* molecules stems from the spectral properties of these heme proteins and their function as electron transport proteins. The structure of cytochrome b_5 from beef liver endoplasmic reticulum has been determined (Mathews et al., 1971a,b, 1972). It catalyzes the transfer of electrons from an NADH-linked FAD-containing reductase to a nonheme cyanide-sensitive factor (Oshino et al., 1971; Holloway & Katz, 1972; Mathews & Czerwinski, 1976). Its similarity to a heme-linked fragment of yeast cytochrome b_2 has been shown by sequence homology (Guiard et al., 1974) and other physical methods (Keller et al., 1973), thus suggesting that the cytochrome b_5–b_2 family is at least 1.0 billion years old.

Structural similarity is not always easy to perceive due to the frailty of the human mind in remembering three-dimensional topology and recognizing structural arrangements (Rossmann & Argos, 1976). Thus, at first sight, the globin fold and the eucaryotic cytochrome *c* structure appear totally different. However, the presence of many deletions and some insertions significantly alters the structure of *Pseudomonas aeruginosa* cytochrome c_{551} relative to eucaryotic cytochrome *c*. Although these two cytochrome *c* structures resemble each other and exhibit amino acid sequence homology, the small "core" cytochrome c_{551} molecule also has some resemblance to the globin fold. This similarity is masked in eucaryotic cytochrome *c* by the large insertions which dominate the structure.

The recognition of structural similarity must not be taken as sufficient evidence for divergent evolution from a common ancestor. Rather, it may be indicative of convergence to an energetically favorable protein fold suitable for particular functions. Differentiation between these two possibilities must be given in terms of probabilities which are dependent on the extent of common structural and functional characteristics. An examination of such features among the globin, cytochrome *c*, and cytochrome b_5 families is given in the final section of this paper. The lack of a definitive answer on evolutionary origin does not mitigate the importance of recognizing common structural features and their correlation with function.

Experimental Section

The C_α atom coordinates of cytochrome c_{551} were kindly supplied by Professor R. E. Dickerson (Almassy & Dickerson, 1978). All other coordinate sets were taken from the GAPSOM data file (Feldmann, 1975). These included tuna cytochrome *c* (Dickerson et al., 1971; Swanson et al., 1977; Mandel et al., 1977), *Rhodospirillum rubrum* cytochrome c_2 (Salemme et al., 1973a,b), and *Paracoccus denitrificans* cytochrome c_{550} (Timkovich & Dickerson, 1976; Timkovich et al., 1976). The globin family was represented by the horse oxy α and β chains (Ladner et al., 1977), sperm whale myoglobin (Takano, 1977; Watson, 1969), sea lamprey hemoglobin (Hendrickson et al., 1973), and *G. dibranchiata* hemoglobin (Padlan & Love, 1974). The cytochrome b_5 structure was determined by Mathews et al. (1972). The amino acid sequences of the globins and their alignments were taken from the tables of Dayhoff (1972). This included the sequence of the hemoglobin from root nodules of leguminous plants (Ellfolk & Sievers, 1971; Lehtovaara & Ellfolk, 1975). Although the structure of leghemoglobin has been solved (Vainshtein et al., 1975), no atomic coordinates were available for comparisons. The amino acid sequence alignments for the *c* cytochromes are given by Dickerson (1979). Guiard et al. (1974) have compared the sequences of cytochrome b_5 and yeast cytochrome b_2.

The small cytochrome c_{551} molecule was used as the representative cytochrome *c* structure since it has a recognizable resemblance to the globin fold. The hemoglobin β chain served as a model of the globin structure. The specific use of the horse hemoglobin β chain will not significantly affect the results here, as the differences in the tertiary structure of the known globins (Padlan & Love, 1974; Huber et al., 1971; Perutz et al., 1968; Vainshtein et al., 1975) are small compared to those between the globins and cytochrome *c* or *b*.

Structural equivalence was determined by a procedure developed by Rao & Rossmann (1973) and extended by Rossmann & Argos (1975, 1976). This technique alternatively minimizes the sum of the squares of distances between equivalenced atoms of the superimposed molecules and reexamines the set of equivalenced residues. The latter procedure requires a topological relationship between the two folds such that, if residues i and j are equivalenced, then residue $i + n$ can be equivalenced to residue $j + m$ only where n and m are positive integers. Further tests assure spatial and orientational similarities of the peptide bonds. Limits of topological equivalence were set by ambiguities in alignment and not by any maximum distance between C_α positions (Rossmann & Argos, 1976). The algorithm for alignment of molecules was the same for all comparisons discussed in this paper.

Results

The structural relationship which had been found previously between the globin fold and cytochrome b_5 (Rossmann & Argos, 1975) is shown in Figure 1b and is represented by columns 1 and 4 in Table I. The uniqueness of this comparison was established by a universal search of all possible orientations (Rossmann & Argos, 1976). Structural equivalences between cytochrome c_{551} and the globin fold were initially selected by visual inspection of the polypeptide backbones. A number of related, but different, starting equivalences all resulted in the same superposition. The resultant fit is given in columns 1 and 2 of Table I which shows

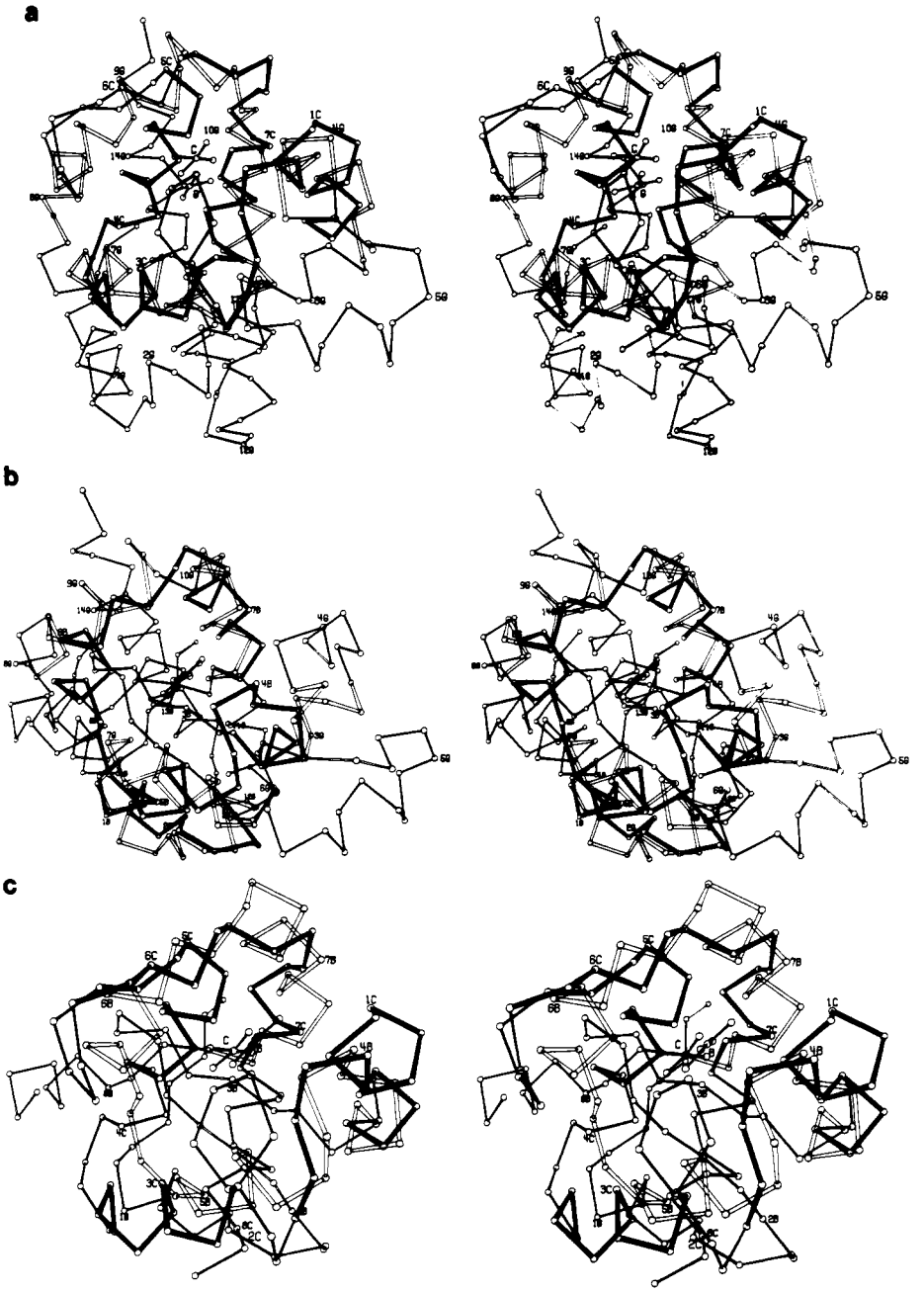


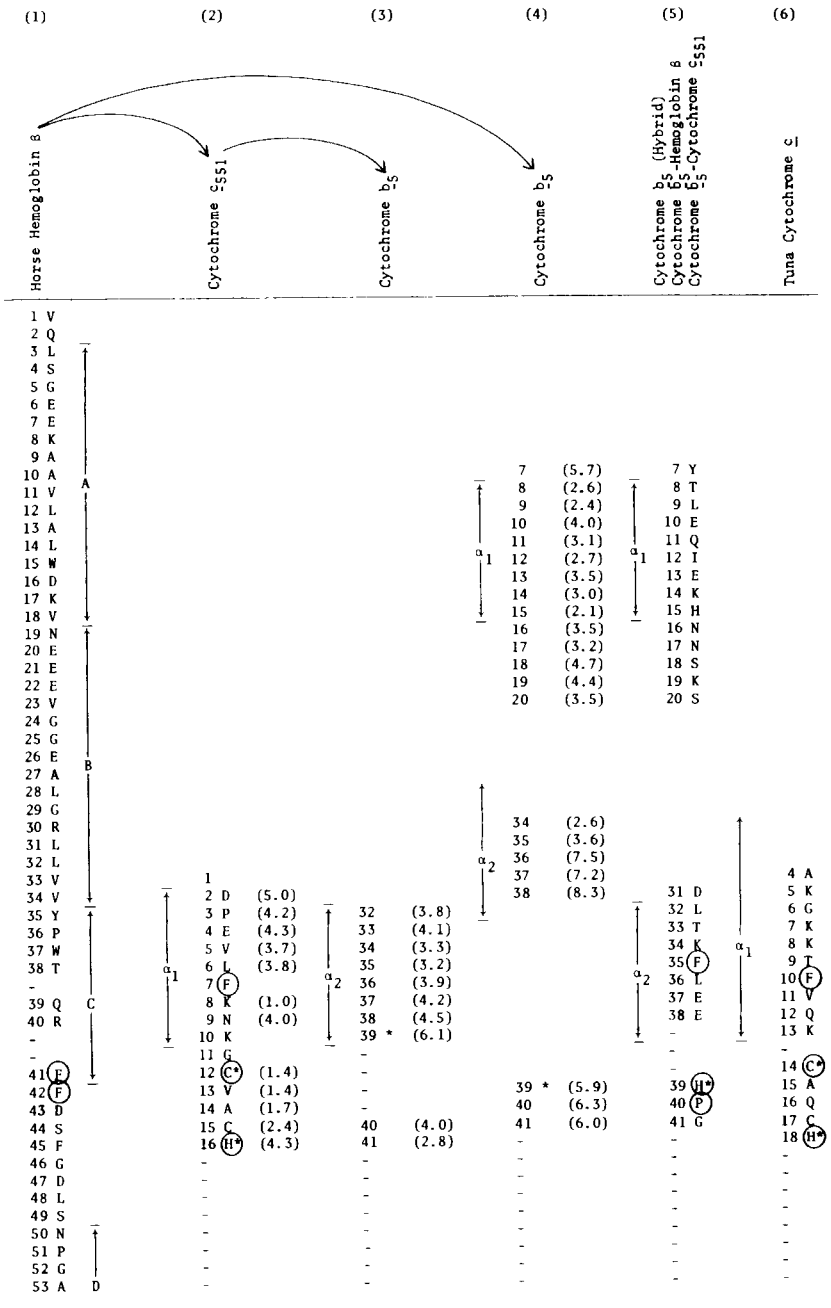
FIGURE 1: Pairwise comparisons of the polypeptide folds for the horse hemoglobin β chain (G), cytochrome c_{551} (C), and cytochrome b_5 (B). Amino acid sequence numbers are given for every tenth residue, for instance, 1G, 2G, 3G, ... for residues 10, 20, 30, ... in the globin fold. Equivalenced parts of the structure are emphasized by thicker virtual bonds. The heme groups are represented diagrammatically by the central iron atom and its four surrounding nitrogen atoms. All comparisons are viewed approximately with the same orientation. (a) Cytochrome c_{551} (black bonds) against globin (open bonds); (b) cytochrome b_5 (black bonds) against globin (open bonds); (c) cytochrome c_{551} (black bonds) against cytochrome b_5 (open bonds).

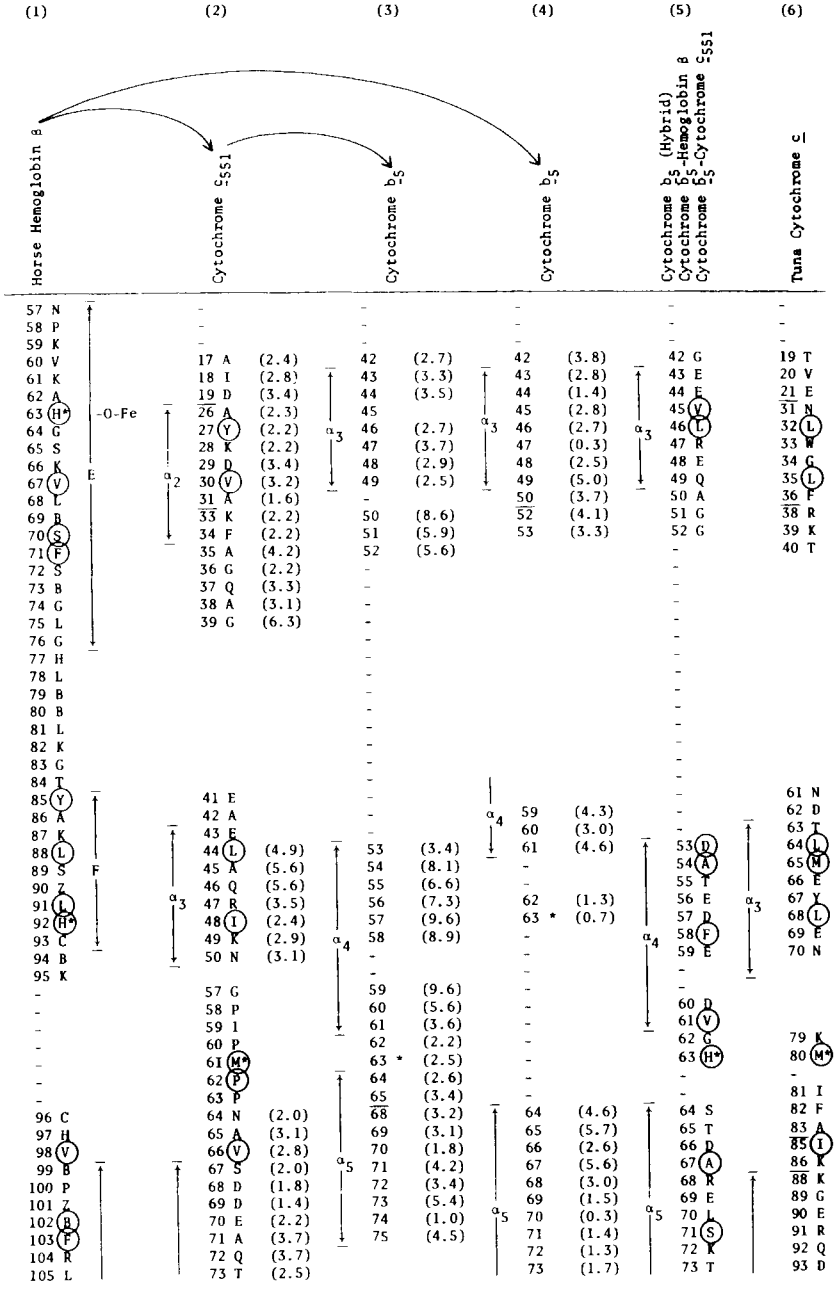
49 equivalenced residues with a root mean square distance between C_α atoms of 3.5 Å. The corresponding structural superposition is shown in Figure 1a, which illustrates the relative closeness of the heme irons (5.4 Å) with the heme normals varying by 6°.

The comparisons of cytochrome b_5 and of cytochrome c_{551} with the globin fold provided an initial alignment for superposition of cytochrome c_{551} and cytochrome b_5. The end point of the refinement is shown in Figure 1c and in columns 2 and 3 of Table I. A total of 41 residues were found to be topologically equivalent with a root mean square deviation of 4.9 Å. The heme iron atoms were separated by 1.6 Å, with an inclination of 32° between their heme normals.

The alignments of cytochrome b_5 with the globin fold and cytochrome c_{551} fold were not entirely consistent; thus, a "hybrid" equivalencing was devised by careful inspection of parts a, b, and c of Figure 1. Since several starting points in equivalencing cytochrome c_{551} and the globin fold converged rapidly to the same final alignment, no alterations were made in their alignments. However, the fit of cytochrome b_5 to either of the other two structures allowed some latitude. Thus, a hybrid cytochrome b_5 alignment was determined and is shown in column 5 of Table I. It is hardly surprising that the root mean square deviations between the "hybrid structure" and globin (4.9 Å) or cytochrome c_{551} (5.0 Å) are slightly larger, yet it provides a consistent "fit" to both the cytochrome c and globin fold. Since eucaryotic cytochrome c and procaryotic cytochrome c_{551} are homologous (Almassy & Dickerson, 1978), the alignment of tuna cytochrome c to the other structures can also be given (column 6 of Table I).

Table I: Alignment of the Horse Hemoglobin β Chain (Column 1) with Cytochrome c_{551} (Column 2) and Cytochrome b_5 (Column 4) Based on Spatial Superposition of C_α Atoms[a]





54 V	.	.	.	.	.	106 (L)		α4	74 (L)	(3.8)	74	(3.8)	74 F	α4 94 (L)
55 M	.	.	.	.	.	107 G			75 (A)	(6.1)	75	(3.4)	75 I	95 (V)
56 G	.	.	.	.	.	108 B	G		76 K	(6.1)				96 A
						109 V			77 W	(5.0)				97 Y
						110 L			78 (V)					98 (L)
						111 A			79 L					99 K
						112 V			80 S					100 S
						113 V			81 Q					101 A
						114 L			82 K					102 T
						115 A								
						116 R								
						117 H								
						118 F								
						119 G								
						120 K								
						121 D								
						122 F								
						123 T								
						124 P								
						.								
						.	H							
						.								
						143 H								
						144 K								
						145 Y								
						146 H								

[a] Distances in angstroms between equivalent atoms are shown in parentheses. Column 3 shows the result of the alignment between cytochrome c_{551} and cytochrome b_5 with distances in angstroms given in parentheses. Column 5 is a "hybrid" fit of cytochrome b_5 with globin and cytochrome c_{551}. Comparison of the eucaryotic cytochrome c structure based on its amino acid sequence alignment with cytochrome c_{551} is shown in column 6. Amino acid types are represented by the one-letter code. Heme liganding residues are asterisked. Residues lining the heme pocket are circled. Results shown here are those produced by the algorithm given by Rossmann & Argos (1976, 1977). Other conditions such that major secondary structural elements cannot have insertions or deletions might improve the present minimum, but in any event the associated secondary structural elements will remain invariant.

Chart I

globin	cytochrome c_{551}	cytochrome b_5
helix A		helix α_1
helix C	helix α_1	helix α_2
helix E	helix α_2	helix α_3
helix F	helix α_3	helix α_4
helix G	helix α_4	helix α_5

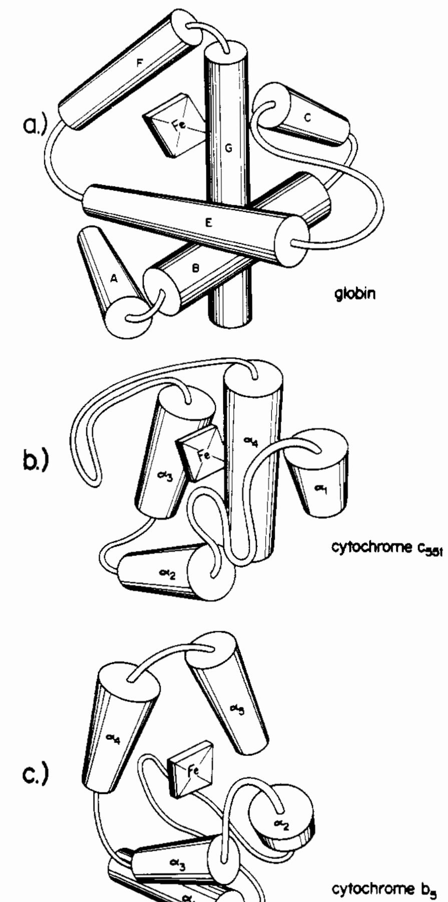


FIGURE 2: Diagrammatic representation of the similar polypeptide chain topology in (a) hemoglobin β chain, (b) cytochrome c_{551}, and (c) cytochrome b_5.

Common Structural Features. The topological alignment of the three heme binding proteins can be roughly summarized as shown in Chart I. This is shown diagrammatically in Figure 2. Table II compares the structural analogies discussed here with other "more established" relationships. The latter standards are the equivalence of (1) the NAD binding domain of lactate and glyceraldehyde-3-phosphate dehydrogenase (Rossmann et al., 1974; Ohlsson et al., 1974) and (2) hen egg white and phage lysozyme (Rossmann & Argos, 1976; Remington & Matthews, 1978). The quality of the comparisons is here measured by the number of topologically equivalenced residues, the percentage of topologically equivalenced residues with respect to the mean number of residues in the molecules, and the root mean square deviation between topological residues. Although the number of equivalenced residues is less for the present comparisons than for the two standards, the total number of residues in these molecules is also less. Thus, the percentage of residues equivalenced compares favorably with the earlier results. The root mean square deviation between residues for the heme protein comparisons is comparable to that of the lysozyme case. Hence, the equivalences reported here appear as significant

Table II: Comparison of Structural Superpositions Reported Here in Relation to Previously Reported Analogies

comparison	mean no. of residues per molecule	no. of topologically equivalent residues	% of topologically equivalent residues	rms[a] deviation of equivalent C_α atoms (Å)
(1) lactate dehydrogenase to glyceraldehyde-3-phosphate dehydrogenase				
(a) whole molecule	319	103	32	3.5
(b) NAD binding domain	147	83	56	2.9
(2) phage to hen egg white lysozyme	147	64	44	3.9
(3) hemoglobin β to cytochrome b_5	114	48	42	4.1
(4) hemoglobin β to cytochrome c_{551}	116	49	43	3.5
(5) cytochrome b_5 to cytochrome c_{551}	84	41	49	4.9

[a] rms, root mean square.

as those given earlier. Furthermore, search functions for dehydrogenase, lysozyme, and cytochrome b_5–globin comparisons are all similar in their peak to background ratios and other measures of significance (Rossmann & Argos, 1976, 1977; Remington & Matthews, 1978).

The heme is roughly bound within the V-shaped cavity formed by representative helices E and F on two sides with the antiparallel helices C and G closing the cavity on the third side. This configuration thus forms an essential core as represented by cytochrome c_{551} (Figure 2). Among the cytochrome c structures only cytochrome c_{551} contains helix α_2 (topologically equivalent to helix E in the globins). Thus, cytochrome c_{551} forms a missing link between the cytochrome c family and the globins. The absence of helix D is not surprising as it is a short helix in the vicinity of the heme pocket present only in the hemoglobin β chain and in myoglobin. Helices A and B do not appear to be involved intimately in the formation of the heme binding pocket. The position of insertions or deletions is between helices both in the structural comparison between different heme binding proteins and in the alignment of sequences within any one family, thus preserving the basic core structure. The four-helical core configuration provides at least 80% of all residues lining the heme pocket. The possibility that a heme binding structural core exists within the globin fold has also been proposed by Blake (1979) on the basis of the presence of a possible intron at the beginning (Tilghman et al., 1978) and an intron around position 110 ± 10 in β (Tilghman et al., 1978; Jeffreys & Flavell, 1977) and γ (Little et al., 1979) chains of hemoglobin. This is the same fragment of the globin chain for which structural equivalence is shown here, suggesting an ancestral exon for all three proteins.

There is another protein fold which also consists primarily of α helices. However, these helices are close packed and almost parallel. Such a fold has been found in the topologically similar structures of tobacco mosaic virus (Champness et al., 1976), hemerythrin (Hendrickson et al., 1975), and cytochrome b_{562} (Mathews et al., 1979a). On the other hand, one domain of Tyr-tRNA synthetase (Irwin et al., 1976) and one of apoferritin (Banyard et al., 1978) show the same helix packing but with different topologies. The possible function (Argos et al., 1977) and topologies (Blow et al., 1977) of such arrangements have been discussed. These variously connected close packed arrangements bear little resemblance in both structure and topology to the loose association of helices in the heme binding proteins discussed here. Yet the globin fold may be one of a few simple ways in which a polypeptide chain can cover a heme group. The similarity of the heme binding structures could, therefore, be an example of convergent evolution representing a preferred packing arrangement of α helices which serves as a heme binding envelope. This arrangement can be considered as a super-secondary structure (Rao & Rossmann, 1973) with a chirality analogous to nucleotide binding to β–α–β structures (Hol et al., 1978). An alternative four-helical topology and heme association is expressed by the hemerythrin-like cytochrome b_{562}.

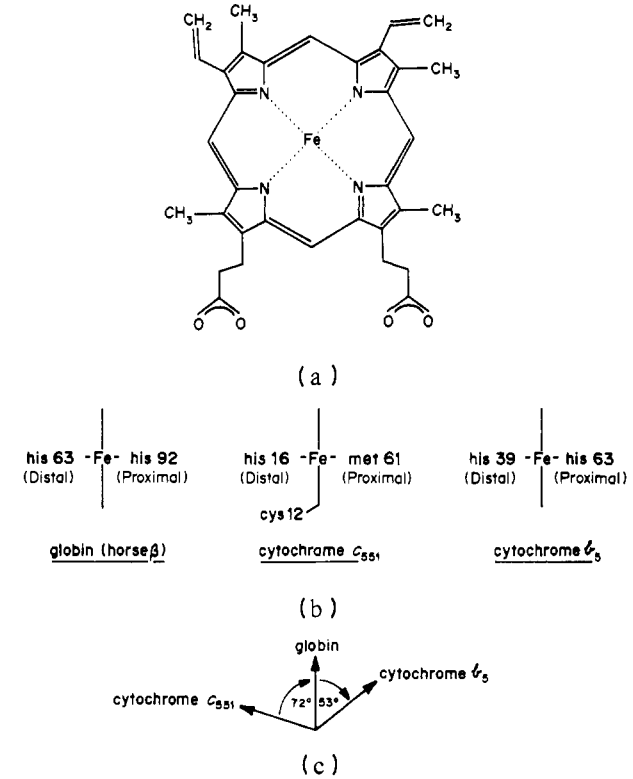


FIGURE 3: (a) View of a heme group with the distal side nearest the observer. For definition of positive heme rotation, the second-named group is rotated clockwise about the heme normal when viewed in this manner. (b) Side view of heme groups showing the distal side on the left and the proximal side on right. (c) Relative rotation of heme groups.

The heme group displays handedness. Its two faces can be distinguished by the asymmetric attachment of its two vinyl and propionyl side chains. As viewed in Figure 3a, the oxygen binding (distal) side in the globins is closest to the observer.[1] This face will thus be referred to as the distal and the other will be referred to as the proximal for all the heme proteins, regardless of their oxygen binding ability. Rotation of the

[1] We define the distal side of the heme face in the globins as that which associates with oxygen. The terms distal and proximal are then "borrowed" for the cytochromes by association of protein structure. These terms would otherwise be meaningless for cytochrome b_5 where the fifth and sixth iron ligands are both histidine coordinated.

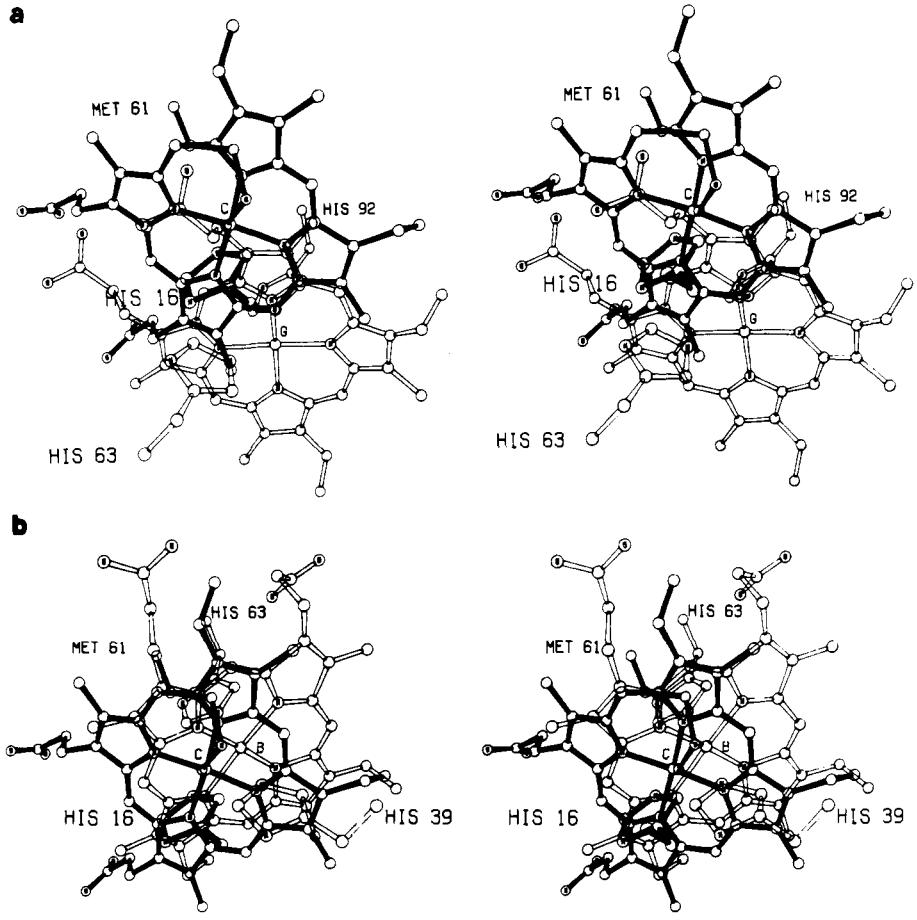


FIGURE 4: Superposition of the heme groups as determined by the best topological equivalencing of the polypeptide backbones. Distances between iron atoms are less than 5.5 Å in all cases, and the heme normals incline to each other by less than 32°. All hemes face the same way. Part a shows the superposition of cytochrome c_{551} (black bonds) on globin (open bonds) while part b shows the superposition of cytochrome c_{551} (black bonds) on cytochrome b_5 (open bonds). A similar comparison of cytochrome b_5 and globin was previously published (Rossmann & Argos, 1975).

heme face about the heme normal, when viewed from the distal side, will be taken as positive when the second-named molecule is rotated clockwise with respect to the first. The heme group is liganded directly or indirectly to the protein at the fifth and sixth iron coordinate position for all three heme binding proteins (Figure 3b). However, the heme is also covalently bound to the vinyl group via a cysteine residue in cytochrome *c*.

With these definitions, it is now easy to recognize that the heme group faces the same way in the globins and cytochrome c_{551} (Figure 4a). Although Figure 4b illustrates the same handedness of cytochrome b_5 and cytochrome c_{551}, recent results (F. S. Mathews, private communication) indicate that the vinyl groups had been wrongly assigned. Thus, the handedness of cytochrome b_5 is opposite to that in the globins and cytochrome c_{551}. Furthermore, erythrocruorin (Steigemann & Weber, 1979) shows the opposite handedness to mammalian globin folds. The Fe–Fe distance is less than 2 Å in the comparisons with cytochrome b_5 although it is 5.4 Å in the cytochrome c_{551}–globin superposition. The heme normals incline to each other by no more than 10° in the comparisons with the horse hemoglobin β chain although the inclination is 32° in relating cytochrome b_5 with cytochrome c_{551}. Nevertheless, the orientation and position of the hemes are reasonably similar in all three heme binding proteins. However, the heme rotation differs substantially with an angle greater than 90° between the hemes of cytochrome c_{551} and cytochrome b_5 (Figure 3c). Presumably, the rotation of the heme with respect to the binding pocket is not relevant to the function of electron transfer. Indeed, exposure of different heme edges to the solvent might facilitate transfer between various electron carriers (Salemme, 1976; Salemme et al., 1973b).

The distal heme ligands His-16 in cytochrome c_{551} (closely associated with the covalently bound Cys-12) and His-39 in cytochrome b_5 have their C_α atoms equivalenced. While the distal liganding group of globin (His-31 of helix E) is spatially close to the other distal coordinating residues, its C_α atom does not equivalence to the corresponding secondary structural component. This contrasts to the proximal side of the heme where the liganding groups as well as their corresponding C_α atoms superimpose approximately. Thus, the positions of the iron coordinating ligands within each polypeptide chain are essentially similar except on the distal side of the heme in the globins where provision must be made for the binding of oxygen.

Conservation of the hydrophobic character of residues lining the heme binding pocket can be observed. However, there is no consistent spatial equivalencing of substituted residues. Table I shows residues associated with the heme pocket for the globins (Padlan & Love, 1974), cytochrome *c* (Almassy & Dickerson, 1978), and cytochrome b_5 (Mathews et al., 1979b). When the side chains of residues lining the heme binding pocket superimpose in space, they only rarely correspond to the C_α atom alignment given in Table I.

Origin of the Heme Binding Fold. When structures bear a resemblance as close as those of the globins, then it is generally assumed that divergent evolution has occurred from

Table III: Measurements of Structural Differences within and between Families of Heme Binding Proteins[a]

criteria	G–G	C–C	G–C	G–B	C–B
(1) mean no. of equivalent residues	134	91	49	54	43
(2) % of nonequivalent residues	7	16	57	53	42
(3) mean heme Fe–Fe separation after alignment of the polypeptide chains (Å)	1.0	0.9	5.3	4.2	3.3
(4) mean no. of insertions (deletions) scored wherever at least one residue has been inserted (deleted)	6	5	10	9	7
(5) mean minimum base change per codon	1.08	1.02	1.46	1.51	1.56
(6) mean rms[b] separation between equivalent atoms (Å)	1.9	2.0	4.4	4.9	5.3
(7) mean angular separation between heme normals after alignment of polypeptide chains (deg)	10	7	9	13	22
(8) mean rotation of the heme in the second compound relative to that in the first when viewed from the distal face (deg)	$\overline{10}$	$\overline{13}$	$\overline{69}$	62	114

[a] G–G implies globin–globin; similarly, C refers to the cytochrome *c* family and B refers to the cytochrome *b* family. [b] rms, root mean square.

a common ancestor. This hypothesis has been amply confirmed from a wealth of other data relating to the evolution of species (Romero-Herrera et al., 1973). The situation is not quite as clear for the evolution of the cytochrome *c* family (Dickerson & Timkovich, 1975). Its common origin occurred sufficiently long ago to make reliable independent data scarce. This is particularly salient when dealing with the evolution of procaryotes where lateral gene transfer is quite feasible.

When there are significant changes between structures, such as those discussed here, the possibility of convergence toward a stable, functionally useful form must also be entertained. Distinction between divergence and convergence must always be phrased in terms of probabilities unless an actual historical record becomes available. Where there are many common features and few differences, divergence is likely.

A variety of properties have been tabulated in Table III in order to obtain some perspective on the evolutionary origin of the heme binding structures discussed here. The globin family, the cytochrome *c* family, and the cytochrome *b* family were each taken as examples of likely cases of divergence. The globin family was represented by the horse hemoglobin α and β chains, sperm whale myoglobin, sea lamprey hemoglobin, *G. dibranchiata* hemoglobin, and leghemoglobin. Only sequence comparisons were possible for leghemoglobin as no coordinates have yet been published. The cytochrome *c* family was represented by eucaryotic tuna cytochrome *c* and procaryotic cytochromes c_2, c_{550}, and c_{551}. The cytochrome *b* family included cytochrome b_5 and yeast cytochrome b_2. Although the structure of cytochrome b_2 is in progress (Mathews & Lederer, 1976), no results are yet available; thus, only sequence comparisons were possible within the cytochrome *b* family. Cytochrome b_5 is also represented by the "hybrid" structure for the purpose of alignment with cytochrome c_{551} and the horse hemoglobin β chain. Comparisons were extended across the boundaries of each family in order to discern the change of various characteristics within and between families.

The criteria listed in Table IV (but given only as means between or within families) include the number of equivalenced residues, percentage of nonequivalenced residues, heme Fe–Fe separation after alignment of polypeptide backbones, number of insertions and deletions, minimum base change per codon, root mean square separation between equivalenced atoms, angular separation of heme normals, and rotation of the heme groups relative to each other. Although only mean values are given here, it was observed that these characters display a continuous trend within and between families. Significant changes of structure or chemical character within divergent families do occur, although these are generally less than those which occur between families. The increased differences

Table IV: Ranking of Various Criteria as Measures of Structural Similarity[a]

criteria	G–G	C–C	G–C	G–B	C–B
(1) no. of equivalent residues	1	2	4	3	5
(2) % of nonequivalent residues	1	2	5	4	3
(3) Fe–Fe distances	2	1	5	4	3
(4) no. of deletions	2	1	5	4	3
(5) minimum base change per codon	2	1	3	4	5
(6) rms[b] C_α separation	1	2	3	4	5
(7) inclination of heme normals	3	1	2	4	5

[a] G–G implies globin–globin; similarly, C refers to the cytochrome *c* family and B refers to the cytochrome *b* family. Relative heme rotations, listed in Table III, were not considered as functionally important and, hence, are probably a poor criterion of structural relatedness. [b] rms, root mean square.

observed between families are, however, not dramatically greater.

Cytochrome c_{551} and leghemoglobin showed the greatest differences within their respective families. However, Dickerson and co-workers proposed that cytochrome c_2 diverged from the other cytochromes *c* discussed here at an earlier time (Dickerson & Timkovich, 1975; Almassy & Dickerson, 1978). Provided the rate of change of characters is roughly constant, such criteria as were examined in Table III suggest that cytochrome c_{551} branched off from the other cytochromes longer ago.

The conservation of the mode of NAD binding to the dehydrogenases is highly conserved, suggesting that divergence preserves function faithfully (Rossmann et al., 1974). A similar example can be found for Ca binding protein domains (Kretsinger & Nockolds, 1973). Hence, the mode of heme binding was also examined in Table III. The difference within and between families is quite marked. Although the Fe–Fe distance may be small in any one comparison between families, the heme normals or heme rotations will usually be substantially different. Thus, while the mode of heme binding between families is similar, it is not as consistently preserved as NAD binding to dehydrogenases.

The various measures listed in Table III have been ranked in Table IV on a scale of 1–5, where 1 represents the greatest similarity for a particular type of comparison between or within families. It is clear that the relationship within families is greatest. However, the inconsistent ranking between families obtained with different criteria might suggest (1) poor selection of criteria, (2) divergence from a common ancestor early in biological time such that differentiation between various

sequential possibilities is obscured, or (3) unequal rates of convergence of different characters toward a functional heme binding protein.

Conclusions

The similarity of the three well-studied heme binding structures and the different cytochrome b_{562} topology enhances the interest in the determination of other heme binding proteins. The structures of yeast cytochrome *c* peroxidase (Poulos et al., 1978) and catalase (Eventoff et al., 1976) can be expected within a short while. Their determination may provide a sufficient collection of structures to better distinguish between divergence from a common ancestor and preference for stable and functionally useful folds. It may then also be possible to discern essential functional and structural features of these folds.

Acknowledgments

We are most grateful to Dr. R. E. Dickerson for supplying us with the *Pseudomonas* cytochrome c_{551} C_α atom coordinates. The authors are indebted to Dr. Rosemarie Swanson for pointing out the nature of the handedness of the heme group. We are most thankful to Sharon Wilder for her careful work in the preparation of this manuscript.

References

Almassy, R. J., & Dickerson, R. E. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2674–2678.

Ambler, R. P. (1973) *Biochem. J. 135*, 751–758.

Ambler, R. P., & Bartsch, R. G. (1975) *Nature (London) 253*, 285–288.

Argos, P., Rossmann, M. G., & Johnson, J. E. (1977) *Biochem. Biophys. Res. Commun. 75*, 83–86.

Banyard, S. H., Stammers, D. K., & Harrison, P. M. (1978) *Nature (London) 271*, 282–284.

Blake, C. C. F. (1979) *Nature (London) 277*, 598.

Blow, D. M., Irwin, M. J., & Nyborg, J. (1977) *Biochem. Biophys. Res. Commun. 76*, 728–734.

Champness, J. N., Bloomer, A. C., Bricogne, G., Butler, P. J. G., & Klug, A. (1976) *Nature (London) 259*, 20–24.

Dayhoff, M. O. (1972) *The Atlas of Protein Sequence and Structure*, Vol. 5 and subsequent supplements, National Biomedical Research Foundation, Silver Spring, MD.

Dickerson, R. E. (1979) in *Biomolecular Structure, Conformation, Function and Evolution* (Srinivasan, R., Ed.) Pergamon Press, New York (in press).

Dickerson, R. E., & Timkovich, R. (1975) *Enzymes, 3rd Ed. 11*, 397–547.

Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A., & Margoliash, E. (1971) *J. Biol. Chem. 246*, 1511–1535.

Dickerson, R. E., Timkovich, R., & Almassy, R. J. (1976) *J. Mol. Biol. 100*, 473–491.

Ellfolk, N., & Sievers, G. (1971) *Acta Chem. Scand. 25*, 3532–3534.

Eventoff, W., Tanaka, N., & Rossmann, M. G. (1976) *J. Mol. Biol. 103*, 799–801.

Feldmann, R. J. (1975) *Global Atlas of Protein Structure on Microfiche*, Division of Computer Research and Technology, National Institutes of Health, Bethesda, MD 20014.

Guiard, B., Groudinsky, O., & Lederer, F. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 2539–2543.

Hendrickson, W. A., Love, W. E., & Karle, J. (1973) *J. Mol. Biol. 74*, 331–361.

Hendrickson, W. A., Klippenstein, G. L., & Ward, K. B. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2160–2164.

Hol, W. G. J., van Duijnen, P. T., & Berendsen, H. J. C. (1978) *Nature (London) 273*, 443–446.

Holloway, P. W., & Katz, J. T. (1972) *Biochemistry 11*, 3689–3696.

Huber, R., Epp, O., Steigemann, W., & Formanek, H. (1971) *Eur. J. Biochem. 19*, 42–50.

Irwin, M. J., Nyborg, J., Reid, B. R., & Blow, D. M. (1976) *J. Mol. Biol. 105*, 577–586.

Jeffreys, A. J., & Flavell, R. A. (1977) *Cell 12*, 1097–1108.

Keller, R., Groudinsky, O., & Wüthrich, K. (1973) *Biochim. Biophys. Acta 328*, 233–238.

Korszun, Z. R., & Salemme, F. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5244–5247.

Kretsinger, R. H., & Nockolds, C. E. (1973) *J. Biol. Chem. 248*, 3313–3326.

Ladner, R. C., Heidner, E. J., & Perutz, M. F. (1977) *J. Mol. Biol. 114*, 385–414.

Lehtovaara, P., & Ellfolk, N. (1975) *Eur. J. Biochem. 54*, 577–584.

Little, P. F. R., Flavell, R. A., Kooter, J. M., Annison, G., & Williamson, R. (1979) *Nature (London) 278*, 227–231.

Mandel, N., Mandel, G., Trus, B. L., Rosenberg, J., Carlson, G., & Dickerson, R. E. (1977) *J. Biol. Chem. 252*, 4619–4636.

Mathews, F. S., & Czerwinski, E. W. (1976) *Enzymes Biol. Membr. 4*, 143–197.

Mathews, F. S., & Lederer, F. (1976) *J. Mol. Biol. 102*, 853–857.

Mathews, F. S., Levine, M., & Argos, P. (1971a) *Nature (London), New Biol. 233*, 15–16.

Mathews, F. S., Argos, P., & Levine, M. (1971b) *Cold Spring Harbor Symp. Quant. Biol. 36*, 387–395.

Mathews, F. S., Levine, M., & Argos, P. (1972) *J. Mol. Biol. 64*, 449–464.

Mathews, F. S., Bethge, P. H., & Czerwinski, E. W. (1979a) *J. Biol. Chem. 254*, 1699–1706.

Mathews, F. S., Czerwinski, E. W., & Argos, P. (1979b) *Porphyrins 7, Part B*, 107–147.

Ohlsson, I., Nordström, B., & Brändén, C. I. (1974) *J. Mol. Biol. 89*, 339–354.

Oshino, N., Imai, Y., & Sato, R. (1971) *J. Biochem. (Tokyo) 69*, 155–167.

Padlan, E. A., & Love, W. E. (1974) *J. Biol. Chem. 249*, 4067–4078.

Perutz, M. F., Muirhead, H., Cox, J. M., & Goaman, L. C. G. (1968) *Nature (London) 219*, 131–139.

Poulos, T. L., Freer, S. T., Alden, R. A., Xuong, N., Edwards, S. L., Hamlin, R. C., & Kraut, J. (1978) *J. Biol. Chem. 253*, 3730–3735.

Rao, S. T., & Rossmann, M. G. (1973) *J. Mol. Biol. 76*, 241–256.

Remington, S. J., & Matthews, B. W. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 2180–2184.

Romero-Herrera, A. E., Lehmann, H., Joysey, K. A., & Friday, A. E. (1973) *Nature (London) 246*, 389–395.

Rossmann, M. G., & Argos, P. (1975) *J. Biol. Chem. 250*, 7525–7532.

Rossmann, M. G., & Argos, P. (1976) *J. Mol. Biol. 105*, 75–95.

Rossmann, M. G., & Argos, P. (1977) *J. Mol. Biol. 109*, 99–129.

Rossmann, M. G., Moras, D., & Olsen, K. W. (1974) *Nature (London) 250*, 194–199.

Salemme, F. R. (1976) *J. Mol. Biol. 102*, 563–568.
Salemme, F. R., Freer, S. T., Xuong, Ng. H., Alden, R. A., & Kraut, J. (1973a) *J. Biol. Chem. 248*, 3910–3921.
Salemme, F. R., Kraut, J., & Kamen, M. D. (1973b) *J. Biol. Chem. 248*, 7701–7716.
Steigemann, W., & Weber, E. (1979) *J. Mol. Biol. 127*, 309–338.
Swanson, R., Trus, B. L., Mandel, N., Mandel, G., Kallai, O. B., & Dickerson, R. E. (1977) *J. Biol. Chem. 252*, 759–775.
Takano, T. (1977) *J. Mol. Biol. 110*, 537–568.
Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sullivan, M., Maizel, J. V., & Leder, P. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 725–729.
Timkovich, R., & Dickerson, R. E. (1976) *J. Biol. Chem. 251*, 4033–4046.
Timkovich, R., Dickerson, R. E., & Margoliash, E. (1976) *J. Biol. Chem. 251*, 2197–2206.
Vainshtein, B. K., Harutyunyan, E. H., Kuranova, I. P., Borisov, V. V., Sosfenov, N. I., Pavlovsky, A. G., Grebenko, A. I., & Konareva, N. V. (1975) *Nature (London) 254*, 163–164.
Watson, H. C. (1969) *Prog. Stereochem. 4*, 299–333.
Weber, E., Steigemann, W., Jones, T. A., & Huber, R. (1978) *J. Mol. Biol. 120*, 327–336.
Weber, P., & Salemme, F. R. (1977) *J. Mol. Biol. 117*, 815–820.

A Correlation of the Visible and Soret Spectra of Dioxygen- and Carbon Monoxide–Heme Complexes and Five-Coordinate Heme Complexes with the Spectra of Oxy-, Carboxy-, and Deoxyhemoglobins†

Chi-Ming Wang‡ and William S. Brinigar*

ABSTRACT: A systematic investigation was carried out concerning relationships between visible and Soret spectra of heme complexes and the nature of the axial ligands. Dioxygen (O_2), carbon monoxide (CO), and five-coordinate complexes were prepared from proto-, meso-, and 2,4-dimethyldeuteroheme dimethyl esters in *N*,*N*-dimethylformamide (DMF) solution. A variety of axial ligands was employed, including imidazoles, pyridines, aliphatic amines, and very weak bases such as DMF and acetone. Variations in the ligands included differences in basicity and differences in substituents which sterically hindered coordination to heme iron. In the five-coordinate complexes, a shift to shorter wavelength in both the visible and Soret peaks accompanied a change to a more hindered axial ligand. Difference spectra obtained by taking the spectrum of a heme complex with a more hindered ligand minus that of a less hindered ligand approximated the T minus R state difference spectrum reported by Perutz et al. [Perutz, M. F., Ladner, J. E., Simon, S. R., & Ho, C. (1974) *Biochemistry 13*, 2163] for deoxyhemoglobin (deoxy-Hb) Kempsey ± inositol hexaphosphate (IHP). In the CO complexes, a decrease in the basicity of the ligand or an increase in the steric hindrance of the axial ligand also resulted in a blue-shifted λ_{max} accompanied by an increase in the ratio of intensities of the long-wavelength visible peak (α) to that of the shorter wavelength visible peak (β). The latter parameter is termed the α/β ratio. Spectra of the O_2 complexes were observed only at temperatures below −40 °C because of limited stability at higher temperatures. In O_2 complexes, shifts in λ_{max} and changes in the α/β ratio with unhindered ligands showed much the same pattern as with the CO complexes, but hindered ligands such as 1,2-dimethylimidazole and 1,2,4,5-tetramethylimidazole gave red-shifted visible λ_{max} and low α/β ratios compared to the unhindered O_2–heme–1-*n*-butylimidazole complex. This observation is interpreted as being due to the greater ease with which the Fe is distorted from the porphyrin plane in O_2 complexes than in CO complexes. Comparison of α/β ratios of model O_2– and CO–protoheme complexes with those of hemoglobins and myoglobins formed the basis for the suggestion that the proximal histidine is restrained by the protein, producing a relatively weak axial Fe–N interaction, both in liganded R-state hemoglobins and in common myoglobins. T minus R state difference spectra reported by Perutz et al. [Perutz, M. F., Kilmartin, J. V., Nagai, K., Szabo, A., & Simon, S. R. (1976) *Biochemistry 15*, 378] for O_2- and CO-Hb Kansas ± IHP were approximated by model protoheme complexes of a stronger minus a weaker axial ligand. These results are interpreted in terms of T-state steric conflict between coordinated O_2 or CO and amino acid side chains on the distal side of the heme. Transition to T state would result in the heme complex being forced toward the proximal imidazole, thereby strengthening the axial Fe–N bonding interaction and introducing strain in both the heme complex and the protein.

The visible spectra of oxyhemoglobins and oxymyoglobins are characterized by two peaks, α (~575 nm) and β (~540 nm), of which the α peak usually has slightly the greater intensity. The spectra of the CO complexes are similar, but the α peak is rarely observed to have a higher absorbance than the β peak. Deoxyhemoglobins have a single absorption band in the visible spectrum with a long-wavelength shoulder. Difference spectra have been reported recently for O_2-, CO-, and deoxy-Hb's[1] under a variety of conditions: for example,

† From the Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122. *Received June 4, 1979*. This work is taken from the dissertation of C.-M.W. submitted to the Graduate School of Temple University in partial fulfillment of the requirements for the Ph.D. degree.
‡ Present address: Department of Biochemistry, China Medical College, Taichung, Taiwan 400, Republic of China.

[1] Abbreviations used: Hb, hemoglobin; Mb, myoglobin; IHP, inositol hexaphosphate; TPP, tetraphenylporphyrin; DMF; *N*,*N*-dimethylformamide; BI, 1-*n*-butylimidazole; CMI, 5-chloro-1-methylimidazole; DMI, 1,2-dimethylimidazole; MEI, 1-methyl-2-ethylimidazole; TMI, 1,2,4,5-tetramethylimidazole; Fe–N_ϵ, bond between heme iron and $N_{\epsilon 2}$ of F8 His.

0006-2960/79/0418-4960$01.00/0