A Comparison of the Heme Binding Pocket in Globins and Cytochrome $b_5^*$

MICHAEL G. ROSSMANN AND PATRICK ARGOS†

From the Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907

SUMMARY

Of the 85 three-dimensionally characterized residues of cytochrome $b_5$, 51 are found to be structurally and topologically equivalent to the globin fold. When these proteins have been superimposed, the heme iron atoms are found to be less than 1.4 Å separated and the heme normals are inclined by less than 9.5°. The proximal histidine of the globins and two adjacent helices are equivalent to the sixth iron ligand and adjacent helices of cytochrome $b_5$. Larger differences in structure are observed on the distal side of the heme, coincident with the most changeable part of the globin structures. The heme itself is rotated by 53° about its normal but such a change is energetically minimal and conservative as the heme side groups are not directly involved in the function of the molecules. The $\beta$-sheet of cytochrome $b_5$ is inserted into a corresponding cavity of the globins forming an additional lining to the heme pocket. The roughly 50 residues missing at the carboxy end of the known cytochrome $b_5$ fragment could correspond in part to the H helix in the globins. While it would seem probable that these similarities represent divergent evolution from a primordial heme-binding protein, the possibility of structural convergence to a functionally satisfactory protein cannot be excluded.

There are now close to 50 known protein structures (1). Some of these belong to families where the three-dimensional structure has been retained and only the specificity has been altered, such as the group of serine proteases exemplified by chymotrypsin (2). In other cases there occur domains within the structure with a given function which are part of a longer polypeptide chain. Examples are the nucleotide-binding proteins (3, 4) and the calcium-binding proteins (5). In still other cases, there exist structural similarities where a functional relationship is not so clear: for example, the superoxide dismutase fold and the immunoglobulin domain.1 In those cases where there is both a strong correlation between amino acid sequences and a common functional role, there can be little doubt that the two different enzymes or protein domains arose from a common precursor. The primordial gene was duplicated permitting each to evolve independently and separately, possibly being subsequently fused with another gene. In those cases where there is but a structural resemblance associated with a less well defined common function, then convergent evolution may be the reason for the similarities. In general if the number of similar characters far exceeds the dissimilarities, it is improbable that each has converged independently, thus suggesting a divergent evolutionary process. Schulz and Schirmer (6) have quantized this argument for the case of analyzing the divergence of structure among nucleotide-binding proteins.

This paper shows that there is a structural relationship between a subunit of hemoglobin and the cytochrome $b_5$ fragment each sharing the function of heme binding. No definite conclusions can, however, be drawn concerning their divergent or convergent evolution.

EXPERIMENTAL PROCEDURES AND RESULTS

The $\beta$ chain of horse oxyhemoglobin (1) was used for the purposes of the comparisons shown in this paper. Coordinates were kindly supplied by Dr. M. F. Perutz (MRC Laboratory of Molecular Biology, Cambridge, England). The structural nomenclature used is that described for myoglobin by Dickerson (8). As the differences in the tertiary structure among the known globins (9–11) are quite small compared to those between the globins and cytochrome $b_5$, the specific use of the horse hemoglobin $\beta$ chain as standard will not significantly affect the results reported here.

Coordinates for the structure of calf liver cytochrome $b_5$ were obtained from Dr. F. S. Mathews (Washington University School of Medicine, St. Louis, Missouri 63110). The amino acid sequence (12, 13) from residue 3 to 87 is shown in Table I together with the proximal histidine of the globins and cytochrome $b_5$, the specific use of the horse hemoglobin $\beta$ chain as standard will not significantly affect the results reported here.

The superposition of the proteins with “Fit 3” resulted in a rather accurate superposition of the heme groups, which had not been included in the refinement. The iron atoms were separated by

* This work was supported by National Science Foundation Grant GB 20596x and National Institutes of Health Grant GM 10704.

† Present address, Department of Physics, Southern Illinois University, Edwardsville, Illinois 62025.

‡ D. C. Richardson, 1974, private communication.
| Residue | F8 L1 | F8 L2 | F8 L3 | F8 L4 | F8 L5 |
|---------|-------|-------|-------|-------|-------|
| 1       |       |       |       |       |       |
| 2       |       |       |       |       |       |
| 3       |       |       |       |       |       |
| 4       |       |       |       |       |       |
| 5       |       |       |       |       |       |
| 6       |       |       |       |       |       |
| 7       |       |       |       |       |       |
| 8       |       |       |       |       |       |
| 9       |       |       |       |       |       |
| 10      |       |       |       |       |       |
| 11      |       |       |       |       |       |
| 12      |       |       |       |       |       |
| 13      |       |       |       |       |       |
| 14      |       |       |       |       |       |
| 15      |       |       |       |       |       |
| 16      |       |       |       |       |       |
| 17      |       |       |       |       |       |
| 18      |       |       |       |       |       |
| 19      |       |       |       |       |       |
| 20      |       |       |       |       |       |
| 21      |       |       |       |       |       |
| 22      |       |       |       |       |       |
| 23      |       |       |       |       |       |

Table I: Equivalenced residues of two heme-binding proteins.

Note: The equivalenced residues are listed by their corresponding positions in the primary structures of F8 L1 and F8 L2 (Xenopus).
The extra pleated sheet ($\beta_3$ and $\beta_4$) occurring in cytochrome $b_6$ can be regarded as an insertion in helix B of hemoglobin $\beta$. In Figs. 1 and 2 are shown the correspondence of the two structures as determined by "$\text{Fit 3}$", which was based upon the protein superposition. Fig. 1 shows the superposition of the polypeptide chains, while Fig. 2 illustrates the relative heme orientations. The environments of the heme groups are compared in Figs. 3 and 4. Fig. 3 again corresponds to "$\text{Fit 5}$" while Fig. 4 relates to "$\text{Fit 6}$" where the proximal histidine $C_{\alpha}$ atoms and, in particular, the heme iron atoms, the refinement converged to a superposition of 48 residues and an iron atom separation of 0.3 Å.

All of the above weighting schemes gave essentially the same results, although there were a few changes in certain of the equivalence of residues of cytochrome $b_6$ and hemoglobin $\beta$. The results can be roughly summarized as:

| Cytochrome $b_6$ | Hemoglobin $\beta$ |
|------------------|---------------------|
| Helix $\alpha_1$ | corresponds to Helix A |
| Helix $\alpha_2$ | corresponds to Helix B |
| Helix $\alpha_3$ | corresponds to Helix E |
| Helix $\alpha_4$ | corresponds to Helix F |
| Helix $\alpha_5$ | corresponds to Helix G |

The above results are based upon an initial superposition obtained by visual inspection. While it would seem improbable that the sequential superposition of roughly 50 residues of the known 85 residues in cytochrome $b_6$ is a chance event, yet a more systematic approach would be desirable. However, since it would not be easy to explore the six variables (three rotational angles and three translational distances) involved in the superposition of two structures, a well chosen line was selected so as to superimpose the molecule onto the other after a rotation of $\kappa$ about this line. For each value of $\kappa$ the $C_{\alpha}$ atoms of the proximal histidines (F8 in hemoglobin $\beta$ and 63 in cytochrome $b_6$) were superimposed, thus determining the three translational parameters. The results for this one-dimensional search, with associated distances between heme iron atoms, are shown in Fig. 5. It will be seen that, when the sequential superposition of residues is at a maximum, then the distance between the heme iron atoms is at a minimum.

**Diagram**

```
    Cytochrome $b_6$
     /   |
     /    |
    /     |
   /      |
  /       |
 /        |
/         |
```

**DISCUSSION**

A comparison of the globins with cytochrome $b_6$ was previously attempted by Ozols and Strittmatter (12) based entirely upon a comparison of amino acid sequences. Although they found some weak homologies, these are totally different from those found here by three-dimensional comparison of structure. In contrast, however, the amino acid sequence of cytochrome $b_{62}$ has been shown to be reasonably homologous to that of sperm whale myoglobin (19) and without an obvious analogy to the cytochrome $b_6$ sequence. Yet Itagaki and Hager (20) have shown that cytochrome $b_{62}$ and cytochrome $b_6$ exhibit similar physical and functional properties thus suggesting the relationship:

The protein comparisons shown in Table I and Figs. 1 and 2 are remarkable not only for the close superposition of the heme groups, but also in conserving the functionally critical parts of the structures. The proximal histidines (cytochrome $b_6$ 63, hemoglobin $\beta$ F8) and the two linked helices ($\alpha_3$ and $\alpha_4$ of cytochrome $b_6$, F and E of hemoglobin $\beta$) are structurally equivalent. Furthermore, the cytochrome $b_6$ helix $\alpha_3$ corresponds to the hemoglobin $\beta$ helix E on the distal side, yet there is no equivalence between the distal histidine of hemoglobin $\beta$ (E7) and the fifth Fe ligand of cytochrome $b_6$ (His 39). The necessary presence of the extra oxygen in the globin destroys structural equivalence on the distal heme side. Furthermore, the CD corner, which corresponds structurally to the position of histidine 39 in cytochrome $b_6$, is the position of maximum change between the $\alpha$ and $\beta$ chains of hemoglobin. The D helix is absent in the $\alpha$ chain of mammalian hemoglobins, and in the single chain glycera hemoglobin (9). This independently suggests that structural change is easily accommodated on the distal side of the heme group.

Although the orientation of the heme groups is maintained to within 9.5°, there is a 53° rotation of the heme groups relative to each other in the 2 molecules. The heme side chains have neither been implicated in the mechanism for hemoglobin (21) nor in cytochrome $b_6$ (22). The only functional necessity is to
keep the propionic acid groups mostly without and the vinyl groups within the heme pocket. Apparently, therefore, the importance of the heme orientation relative to the protein backbone, particularly with respect to the proximal histidine, is in the conservation of the heme binding function making the rotation of the heme less critical. The rotation of the heme will be primarily controlled by the hydrophobic interactions within the heme pocket and hydrogen bonds between the polar propionate groups and the protein main chain or side chains.

One significant difference between cytochrome $b_6$ and myoglobin is the insertion of an anti-parallel $\beta$-pleated sheet (residues 21 to 32) between helices $\alpha_6$ and $\alpha_7$ in cytochrome $b_4$. This sheet is situated in an empty region between helices $E$ and $G$ in the corresponding globin structure forming an extra lining to the heme pocket (Figs. 1 and 3). Corresponding residues lining the heme pocket for cytochrome $b_6$ (22) and in the hemoglobin $\beta$ chain (9) are shown in Table II and mostly shown in Figs. 4 and 5. Of these there are only four pairs of residues, excluding the proximal histidines, which show structural equivalence, yet none show any obvious functional similarity. Apparently the
but are given where a structural equivalent exists. The link peptide must thus be helix cy6 which might on the same surface to expose the negative charge at the COOH terminal residue bound to the endoplasmic reticulum are cleaved all of helix H from cytochrome bs.

Another large difference is the deletion of half of helix G and the membrane-bound domain might correspond in part to helix H of all of the heme-lining residues orients the hemes between liver cytochrome bs and the globins, the established homology between liver cytochrome bs and yeast cytochrome bs (yeast L-(+)-lactate dehydrogenase or EC 1.1.2.3) should also be discussed. The latter is a tetramer with each subunit (MW 58,000) containing a single polypeptide chain associated with one heme and one flavin (FMN) moiety (27). The heme-binding fragment is obtained as a tryptic hydrolysate of the active molecule (28). It has been sequenced and found to bear a reasonable homology to the cytochrome bs fragment (29). Cytochrome bs is a membrane-bound microsomal molecule which catalyzes the transfer of electrons from an NADH-linked FAD containing reductase to a non-heme cyanide-sensitive factor. On the other hand, cytochrome bs is a mitochondrial protein which catalyzes the transfer of electrons from the FMN-containing reductase (which is in fact part of the same polypeptide chain) to cytochrome c. It has been suggested (3) that many nucleotide-binding proteins have similar structures. Thus the cytochrome bs reductase-cytochrome bs system may bear not only a functional but also a structural resemblance to the yeast cytochrome bs molecule. The role of cytochrome bs in the mitochondrial respiratory chain might thus suggest a possible common functional origin for the globin oxygen carriers and the cytochrome bs electron carriers, occurring at least 1.5 X 10^7 years ago (30).

Finally it is necessary to consider divergent as opposed to convergent evolution of the heme-binding protein represented by cytochrome bs and the hemoglobin bs chain. On the basis of the criteria described in the “Appendix,” 51 of the 85 (60%) structurally known residues of cytochrome bs are equivalent to those of the hemoglobin bs chain. Similar comparisons of the nucleotide binding domains of lactate dehydrogenase and glyceraldehyde-3-phosphate dehydrogenase show 92 of 144 (64%) residues in lactate dehydrogenase to be structurally equivalent to glyceraldehyde-3-phosphate dehydrogenase. Thus the percentage of structurally equivalent residues is essentially the same, although the complexity of the globin structure is significantly smaller. Divergent evolution of the nucleotide binding domains has been proposed (3, 4, 21). However, no divergent evolution must be established by the presence of the number of similar independent characters (e.g. amino acids) in relation to the number of total

joint effect of all of the heme-lining residues orients the hemes within their respective pockets.

| Residues lining heme pocket | Cytochrome bs | Hemoglobin bs |
|-----------------------------|---------------|---------------|
| Leu 23 (a1)                 |               |               |
| Leu 25 (a2)                 |               |               |
| Leu 32 (a3)                 |               |               |
| Phe 35 (a4)                 | [Ala 27 (B9)] | Phe 41 (C7)   |
| His 30 (a5) (heme ligand)   | [Asp 43 (C12)]| Phe 42 (CD1)  |
| Pro 40                      |               | [Gly 46 (E8)] |
| Val 45 (a6)                 | His 63 (E7) (distal) | [Val 67 (E11)] |
| Val 16 (a7)                 | [Glu 49 (a1)] | [Gly 52 (a2)] |
| [Gly 50 (a3)]               |               | Ser 70 (E14)  |
| [Asp 53 (a4)]               | Phe 71 (E15)  |               |
| Ala 54 (a5)                 | Phe 85 (E1)   |               |
| [Asp 57 (a6)]               | Leu 88 (E4)   |               |
| Phe 38 (a7)                 | [Glu 90 (E6)] |               |
| Val 61 (a8)                 | His 63 (E8) (proximal) | [His 92 (F8)] |
| [Thr 65 (a9)]               | Lys 95 (FG2)  | Val 98 (FG5)  |
| Ala 67 (a10)                |               |               |
| [Leu 70 (a11)]              | Asn 102 (G4)  |               |
| Ser 71 (a12)                | Phe 103 (G5)  |               |
| [Phe 74 (a13)]              | Leu 106 (G8)  |               |
|                             | Val 134 (H18) |               |
|                             | Leu 141 (H18) |               |

While considering a possible evolutionary relationship between liver cytochrome bs and the globins, the established homology between liver cytochrome bs and yeast cytochrome bs (yeast L-(+)-lactate dehydrogenase or EC 1.1.2.3) should also be discussed. The latter is a tetramer with each subunit (MW 58,000) containing a single polypeptide chain associated with one heme and one flavin (FMN) moiety (27). The heme-binding fragment is obtained as a tryptic hydrolysate of the active molecule (28). It has been sequenced and found to bear a reasonable homology to the cytochrome bs fragment (29). Cytochrome bs is a membrane-bound microsomal molecule which catalyzes the transfer of electrons from an NADH-linked FAD containing reductase to a non-heme cyanide-sensitive factor. On the other hand, cytochrome bs is a mitochondrial protein which catalyzes the transfer of electrons from the FMN-containing reductase (which is in fact part of the same polypeptide chain) to cytochrome c. It has been suggested (3) that many nucleotide-binding proteins have similar structures. Thus the cytochrome bs reductase-cytochrome bs system may bear not only a functional but also a structural resemblance to the yeast cytochrome bs molecule. The role of cytochrome bs in the mitochondrial respiratory chain might thus suggest a possible common functional origin for the globin oxygen carriers and the cytochrome bs electron carriers, occurring at least 1.5 X 10^7 years ago (30).

Finally it is necessary to consider divergent as opposed to convergent evolution of the heme-binding protein represented by cytochrome bs and the hemoglobin bs chain. On the basis of the criteria described in the “Appendix,” 51 of the 85 (60%) structurally known residues of cytochrome bs are equivalent to those of the hemoglobin bs chain. Similar comparisons of the nucleotide binding domains of lactate dehydrogenase and glyceraldehyde-3-phosphate dehydrogenase show 92 of 144 (64%) residues in lactate dehydrogenase to be structurally equivalent to glyceraldehyde-3-phosphate dehydrogenase. Thus the percentage of structurally equivalent residues is essentially the same, although the complexity of the globin structure is significantly smaller. Divergent evolution of the nucleotide binding domains has been proposed (3, 4, 21). However, no divergent evolution must be established by the presence of the number of similar independent characters (e.g. amino acids) in relation to the number of total...
TABLE III
Analysis of minimum base changes per codon (MBC/C)
for Fit 5 of Table I

| By separation, d, of Cn atoms | d (Å) | 0.2-2.0 | 2.0-4.0 | 4.0-9.0 |
|-------------------------------|-------|---------|---------|---------|
| MBC/C                        |       | 1.21    | 1.25    | 1.75    |
| Number                       |       | 14      | 20      | 12      |

By probability $P$

|                        |       | 1.0-0.7 | 0.7-0.4 | 0.4-0.0 |
|------------------------|-------|---------|---------|---------|
| MBC/C                  |       | 1.10    | 1.26    | 1.65    |
| Number                 |       | 10      | 19      | 17      |

Characters, the comparison with nucleotide-binding proteins cannot alone establish divergence. Similarity of function can also be taken as another character set (6). In this respect the remarkable similarity in the binding of the prosthetic group to cytochrome $b_5$ and the globins is comparable to the similarity of NAD conformation when bound to various dehydrogenases (4). The constant orientation and conformation is maintained in both cases in the face of vastly different amino acid residues lining the binding pockets. Rossmann et al. (3) discuss the use of evaluating minimum base changes in the genetic code and show that the nucleotide binding domains in known dehydrogenases differ by between 1.1 to 1.2 minimum base changes per codon. Table I shows slightly larger values of between 1.24 and 1.37 minimum base changes per codon for the heme-binding proteins considered here. However, Table III shows that the greater the certainty of equivalencing residues the lower is the minimum base change per codon, approaching the value found among the nucleotide binding domains of dehydrogenases. No good estimates exist, however, on how much the random value ($-1.45$) might be lowered by the convergence of two protein structures to a similar fold from different ancestors, as a consequence of requiring certain amino acid types as structural requirements in the folding process.  

In summary, the similarity of protein structure of cytochrome $b_5$ and the globins together with the striking similarity of the heme position and orientation show the presence of a divergent or convergent evolutionary process. The proportion of structurally equivalent residues and the relatively low minimum base changes per codon approach the situation found in the NAD binding domains of dehydrogenases. Nevertheless, as the structure of the globins is somewhat simpler, the case for divergence, while substantial, cannot be considered proven.

APPENDIX: SYSTEMATIC COMPARISON OF TWO MOLECULAR STRUCTURES

The procedure, described here, is a modification of the method of Rao and Rossmann (18). The modified procedure is described in some detail (as opposed to the outline given earlier) due to frequent requests for this information. The first step is to devise an initial rotation matrix $[C]$ and a translation vector $d$ which orients the 2nd molecule (identified by the subscripts $2$) similarly to the 1st molecule (identified by the subscripts $1$). Thus the new position $z_2'$ of $(x_2'y_2'z_2')$ of a point $z_2$ in the 2nd molecule is given by

$$
    z_2' = [C]z_2 + d
$$

A number of positions are presumed to be roughly equivalent. These are used to evaluate linearly the nine elements of $[C]$ and three of $d$ by minimizing the condition

$$
    S = \sum (C_{ij} - C'_{ij})^2
$$

where the summation is over at least four equivalent positions. This gives rise to the three sets of four simultaneous linear equations given by the coefficients (where the summations are over $n$ equivalenced points):

Equation Set 1  2  3

$$
    \left( \begin{array}{ccc}
    \Sigma x_2^2 & \Sigma x_2y_2 & \Sigma x_2z_2 \\
    \Sigma x_2y_2 & \Sigma y_2^2 & \Sigma y_2z_2 \\
    \Sigma x_2z_2 & \Sigma y_2z_2 & \Sigma z_2^2 \\
    \end{array} \right) \left( \begin{array}{ccc}
    c_{11} & c_{12} & c_{13} \\
    c_{12} & c_{22} & c_{23} \\
    c_{13} & c_{23} & c_{33} \\
    \end{array} \right) = \left( \begin{array}{ccc}
    \Sigma x_1^2 & \Sigma x_1y_1 & \Sigma x_1z_1 \\
    \Sigma x_1y_1 & \Sigma y_1^2 & \Sigma y_1z_1 \\
    \Sigma x_1z_1 & \Sigma y_1z_1 & \Sigma z_1^2 \\
    \end{array} \right)
$$

The next step is to find the three Eulerian angles $(\theta_1, \theta_2, \theta_3)$ which best fit the nine linearly determined rotation matrix elements. Using the Eulerian rotation matrix given by Rossmann and Blow (32) it can be shown that:

$$
    \theta_1 = \tan^{-1} \frac{c_{31}}{c_{32}}
$$

$$
    \theta_2 = \cos^{-1} c_{33}
$$

and

$$
    \theta_3 = \tan^{-1} \frac{c_{13}}{c_{23}}
$$

where $\theta_3$ is defined to be in the range $0 \leq \theta_3 \leq \pi$. Whether $\theta_1$ and $\theta_2$ are greater or less than $\pi$ can be determined by inspection of the matrix elements. However, a special situation occurs when $\cos \theta_2 = 1$, then

$$
    \theta_1 + \theta_3 = \cos^{-1} c_{11}
$$

and

$$
    \theta_1 - \theta_3 = \cos^{-1} c_{12}
$$

These expressions are used to find initial values of the Eulerian angles. It is, however, necessary to minimize the expression

$$
    S_2 = \sum \sum (c_{ij}(\text{obs}) - c_{ij}(\text{calc}))^2
$$

with respect to $\theta_1$, $\theta_2$, and $\theta_3$.
where $c_{ij}(\text{obs})$ and $c_{ij}(\text{calc})$ refer to their "observed" and "calculated" values. The observed values refer to the linear determination above while the calculated values depend upon the evaluation of trigonometric expressions (32) with the current Eulerian angles. The three normal equations ($i = 1$ to 3) can be derived from the nine observational equations and shifts $\Delta \theta_i$ ($j = 1$ to 3) may be computed. The $i$th normal equation is of the form shown in Equation 1. This procedure is applied iteratively until values of $\Delta \theta_i$ are less than a preset value (e.g. 0.01°). Expressions for evaluation of the differentials $\partial c_{ij}(\text{calc})/\partial \theta_i$ are easily obtained from the expressions for $c_{ij}(\text{calc})$.

Reasonable values of the three Eulerian angles and three translational components are now available, but it is still necessary to refine these directly in order to minimize the sum of the square of the distances between equivalenced atoms. It is necessary to minimize

$$s_1 = \sum_{N} \omega \left[ (x_1 - x'_1)^2 + (y_1 - y'_1)^2 + (z_1 - z'_1)^2 \right]$$

The weighting factor, $\omega$, may be taken as unity in every case, arbitrarily increased on certain equivalenced atoms to assure their good superposition, or set equal to the probability, $P$, with which the given pair of atoms are equivalenced. $N$ represents the number of equivalenced atoms.

Successive cycles of nonlinear least squares are then used to refine the shifts in the six parameters $\xi_i$ ($i = 1$ to 6). The $i$th normal equation has the form shown in Equation 2.

For example, the derivative $\partial x_2'/\partial \theta_1$, given that

$$x'_2 = c_{11}x_2 + c_{12}y_2 + c_{13}z_2 + d_1$$

becomes

$$\frac{\partial x_2'}{\partial \theta_1} = \frac{\partial c_{11}}{\partial \theta_1} x_2 + \frac{\partial c_{12}}{\partial \theta_1} y_2 + \frac{\partial c_{13}}{\partial \theta_1} z_2 + \frac{\partial d_1}{\partial \theta_1}$$

When the best rotation angles and translation components have been determined with respect to the presumed equivalent atoms (or alternatively a rotation matrix and translation vector has been otherwise given), it is then critical to determine whether other atoms or residues can be equivalenced and whether the previous set of equivalences was the most reasonable. After a revised set of equivalences has been obtained, the above nonlinear least squares procedure is again applied in order to minimize, as before, the sum of the squares of the distances, $s_3$, between equivalenced atoms. This will result in modified Eulerian angles and translational components which will modify the set of structurally equivalenced residues. The procedure is repeated until no further changes in equivalences is obtained.

In practice it was found that sometimes as many as 25 cycles of equivalence determinations were needed in comparing structures as different as the hemoglobin $\beta$ and cytochrome $b_5$ chains before convergence had been reached. However, the basic set of equivalences was invariably found after 1 or 2 cycles. The remaining cycles only added or dropped just a few equivalences. Furthermore, the same doubtful equivalences were invariably the cause of the many small changes before the above criterion of convergence had been satisfied. An appreciation of these changes can be obtained by inspecting Table I where only slightly different results are obtained by altering the weighting system.

The possible equivalence of 2 given residues in the molecules being compared was expressed as a probability, $P$, dependent upon three essentially independent estimates. These were:

1. The distance, $d_{ij}$, between $C_\alpha$ atoms $i$ and $j$ in the 1st and 2nd molecule. This gives rise to the probability

$$P_1 = \exp \left( -\frac{d_{ij}^2}{2E_1^2} \right)$$

which estimates the probability of spatial superposition.

2. The scatter, $S_m$, given by the root mean square deviation from the mean of the distance $d_{i-1,j+1}$, $d_{i,j}$, $d_{i+1,j+1}$. When this scatter is low the polypeptide main chain in the 2 molecules must be oriented similarly. Thus

$$P_2 = \exp \left( -\frac{S_m^2}{2E_2^2} \right)$$

3. The scatter, $S_e$, given by the root mean square deviation from the mean of the distances between corresponding atoms in

$$P_3 = \exp \left( -\frac{S_e^2}{2E_3^2} \right)$$

which also estimates the probability of similar orientation of residues.

$P_3$ mostly applies to the main chain atoms and $C_\beta$. The joint probability will then be given by

$$P = P_1 P_2 P_3$$

For the comparisons given in this paper $P_3$ was set to unity in all
cases. Values of $E_1$, $E_2$, $E_3$ were redetermined for each cycle as the root mean square value of $d_{ij}$, $S_M$, and $S_R$ for the currently equivalenced residues, respectively. Convergence of the $E$ values was good. For the hemoglobin $\beta$ versus cytochrome $b_5$ comparisons it was found that $E_1 \approx 4.2$ Å and $E_2 \approx 0.7$ Å.

The probabilities for the $i$th residue of the 1st molecule interacting with the $i$, $j$, ..., $j$th residue of the 2nd molecule were then sorted into descending order. Equivalenced residues were, tentatively, assumed to be those with the largest probability relating the $i$th and $j$th residues. Such assignments, however, take no account of the topological and genetic (in the case of divergent evolution) requirements of similar folds. That is, the equivalencing of residues must be progressive along each polypeptide chain. More precisely, if

$$R_{1i} \equiv R_{2j}$$

represents equivalence of residues where the subscripts refer to the molecule and residue numbers, then

$$R_{1i+n} \equiv R_{2j+m}$$

only if both $n \geq 1$ and $m \geq 1$. Clearly the previous tentative assignments, based upon maximum probabilities alone, do not necessarily conform to the above "progression rule." Some equivalences must thus be removed, or others of lower probabilities added, before a satisfactory equivalenced set has been found.

"Runs" of equivalences are now identified from the tentative assignments. A "run" is defined as a series of the form:

$$R_{1i} \equiv R_{2j} \equiv R_{1i+1} \equiv R_{2j+1} \equiv R_{1i+2} \equiv R_{2j+2} \equiv \cdots \equiv R_{1i+n} \equiv R_{2j+n}$$

For each run the total

$$T = \sum_{i=1}^{i=n} p_i$$

is evaluated. Adjacent runs can then be compared in pairs. Both runs are accepted if they obey the progression rule; but if that is not the case, that run with the smaller total, $T$, is rejected. Emphasis is thus given to the longer and more similar folds of the protein. This process of rejection is continued until every run, and hence every residue, obeys the progression rule. Finally the runs are extended at either end using the largest available probabilities among the largest five probabilities consistent with the progression rule. The extensions are themselves terminated when no further acceptable probabilities can be found or when the downward extension of one run meets the upward extension of another. The purpose of the run extensions is to include those parts of the protein fold where there may be amino acid homology in the absence of precise structural equivalence, particularly at bends in the polypeptide chain or at deletions or insertions.

Acknowledgment—We are grateful to Sharon Wilder for technical assistance in the preparation of this manuscript.
A comparison of the heme binding pocket in globins and cytochrome b5.
M G Rossmann and P Argos

J. Biol. Chem. 1975, 250:7525-7532.

Access the most updated version of this article at http://www.jbc.org/content/250/18/7525

Alerts:
  • When this article is cited
  • When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/250/18/7525.full.html#ref-list-1