Pseudosymmetry in the Structure of Myohemerythrin

(Received for publication, December 13, 1976)

WAYNE A. HENDRICKSON AND KEITH B. WARD*
From the Laboratory for the Structure of Matter, Naval Research Laboratory 6030, Washington, D.C. 20375

The three-dimensional structure of the protein myohemerythrin from retractor muscles of the sipunculan worm Themiste zostericola has been explored for the existence of approximate symmetry operators that locally interrelate portions of the molecule. First, the electron density distribution at 5.5 Å resolution was examined. A local 2-fold axis that transposes the C-D helix pair into the A-B helix pair was found and refined by the method of least squares. The match in electron densities for a pure 2-fold rotation had a correlation coefficient of 0.56. Next, a comprehensive search was made for rotational symmetry in the Patterson function of an isolated molecule. The rotation function based on data to 6 Å spacings showed a major peak, 72% of the self-peak height, that confirmed the result from electron density correlations. In addition, a pattern of lower level peaks revealed approximate point group symmetry as high as D₆. Finally, the amino acid sequence has been inspected for evidence of a repeated structure. The level of amino acid identities between positions in the A-B and C-D helix pairs is 28%. Several factors are discussed to suggest that this homology, although low, is nonetheless significant.

Hemerythrin is an oxygen carrier from the blood of certain invertebrate animals. Like hemoglobin, it too is an iron protein; but unlike hemoglobin, hemerythrin reversibly binds oxygen at binuclear iron centers rather than at heme groups. Hemerythrin most commonly occurs as an octamer with a molecular weight of approximately 108,000 and is contained in erythrocytes of the circulatory system. However, at least some hemerythrin-containing animals also have a monomeric hemerythrin present in their muscles. This protein, myohemerythrin, is homologous with the subunits of octameric hemerythrin.

Both the three-dimensional structure (1, 2) and the amino acid sequence (3) are now known for the myohemerythrin from the sipunculan worm Themiste (syn. Dentrostomum) zostericola. The amino acid sequence of subunits of the octameric hemerythrin from Phascolopsis (syn. Golfingia) gouldii has been known for some time (4, 5). It was evident from an inspection of even the first low resolution electron density map of myohemerythrin that parts of the molecule are related by an approximate local axis of symmetry (1). These parts in turn correspond to apparent internal repeats in both the myohemerythrin and hemerythrin sequences (1, 3). Recently, the structures of two octameric hemerythrins, those from P. gouldii (6) and Themiste dyscrita (7, 8), have also been determined by x-ray crystallographic analysis. The subunits of these hemerythrins have essentially the same tertiary structure as that of myohemerythrin. Thus, intramolecular pseudosymmetry appears to be a general attribute of the "hemerythrin fold."

The purpose of this report is to describe in quantitative terms the nature of this pseudosymmetry in the structure of myohemerythrin. Three lines of evidence are presented. One is from direct examination of the electron density map and of models. A second is from a rotation function analysis. The final one is from the amino acid sequences and iron ligand structure. The pseudosymmetry of myohemerythrin is then discussed in relation to other molecules that exhibit internal symmetry and with regard to evolutionary implications.

ELECTRON DENSITY CORRELATIONS

The most striking features in the electron density map of myohemerythrin at 5.5 Å resolution are four dense rods and an especially dense spheroidal mass that they surround (1). The rods correspond to helical polypeptide segments and the dense spheroid arises from the dimeric iron center. Each helix is 30 to 40 Å long. All four helices are quasi-parallel to one another and in cross-section they form an approximately square array centered around the iron mass.

The pseudosymmetry in myohemerythrin first became apparent after calculation of the angles between the four major helices, A, B, C, and D. The interhelix angles of helix pairs A-B, B-C, and A-C (158, 170, and 25°) are very similar to those of helix pairs C-D, D-A, and D-B (163, 169, and 18°) (1). These data together with a corresponding approximate internal repeat in amino acid sequence suggested that helix pairs A-B and C-D are approximately related by a local symmetry operator.

Tests were then performed to determine the kind of pseudosymmetry operator that is present and to quantify the degree of approximation with which it holds. Three locations in the A-B helix pair in the Fourier map and a set of corresponding points from the C-D pair were selected and used to find a trial transformation relating the two substructures. Initial parameters for the transformation

\[x' = R(\theta_1, \theta_2, \theta_3) \cdot x + t(x_1, t_y, t_z)\]

takes the C-D pair into the A-B pair were found by an exhaustive survey seeking to minimize the discrepancy be-
between superposed test points. \( R \) is the rotation matrix, here in terms of Eulerian angles in the convention used by Rossmann and Blow (9), and \( t \) is the translational vector with components along the crystal axes. A point \( x \) is transformed into \( x' \) by these operations.

The initial parameters for the transformation were refined by a least squares procedure that minimized the sum of squared residuals,

\[
E = \sum_{\text{test pairs}} (\rho(x') - \rho(x))^2
\]

between the electron density, \( \rho(x') \), at grid points in the A-B pair and the corresponding density, \( \rho(x) \), at transformation related points in the C-D pair. The electron density distribution used in these refinements was that of an individual molecule isolated from the five-derivative, solvent constraint refined map described earlier (1). This map was sampled on a grid of \( a/26, b/48, \) and \( c/24 \) which gives sampling intervals of approximately 1.7 \( \AA \). Those grid points in the volume containing the A-B helix pair and its associated half were included in the summation only if the transformation-related point also fell within the bounds of the isolated molecule. A 64-point cubic spline interpolation was made at each transformation-related point in order to evaluate the density and the positional partial derivatives that are needed for the normal equations of the least squares minimization.

Three different refinements were performed. In the first minimization, the orientation angles and translations were allowed to vary in an unrestrained manner. This results in a general screw axis as the permitted symmetry operator. The orientation angles used in this and the next refinement case were the quasi-orthogonal Eulerian angles of Lattman (10) \((\theta_x = \theta_1 + \theta_2, \theta_y = \theta_1 - \theta_2)\). In the second case, the symmetry operator was restricted to be a 2-fold screw axis by constraining \( \theta_1 = 180^\circ \). Since \( \cos(\theta/2) = \cos(\theta_1/2) \cos(\theta_2/2) \) (9), fixing \( \theta \) at \( 180^\circ \) has the desired effect of constraining the actual rotation, \( \chi \), to be \( 180^\circ \). In the final refinement, the symmetry operator was constrained to be a strict 2-fold rotation axis. This was accomplished by using spherical polar angles, \( \phi \) and \( \psi \), in the Rossmann and Blow convention (9) as orientational parameters while holding the rotational throw angle, \( \chi \), at \( 180^\circ \) and introducing as an added parameter a Lagrange multiplier for the constraining equation

\[
t_s = \sin \xi \cos \phi \ tau + \cos \xi \ tau - \sin \xi \sin \phi \ t_x = 0.
\]

This fixes the symmetry operator as a true diad by maintaining a precisely 2-fold rotation with a zero translational component, \( t_x \), along the rotation axis.

Results of these three refinements of the pseudosymmetry operator are given in Table I. The operator may be described as a local pseudodiad. Although errors on the parameters from the unrestrained refinement show that the departures from a strict 2-fold rotation are significant, it is also clear that the local operator does not differ radically from a true diad. In particular, the translational components of the screw operators are insignificant. Moreover, the differences among the three refinements in the root mean square discrepancy in density, \( \langle (\Delta \rho)^2 \rangle^{1/2} \), between the pseudosymmetrically related helix pairs are hardly significant in light of the standard deviation in density of 0.07 e/\( \AA^2 \) for the map as a whole (1).

The measures of goodness-of-fit for the refinements also show that the match between A-B and C-D helix pairs is imperfect. The observed values of 0.16 to 0.17 e/\( \AA^2 \) for the root mean square discrepancy in density are significantly greater than the value of 0.10 e/\( \AA^2 \) expected just from errors in the electron density function. Values of the correlation coefficient between densities,

\[
C = \frac{\sum \rho_{\text{A-B}} - \rho_{\text{C-D}}} {\sqrt{\sum (\rho_{\text{A-B}} - \rho_{\text{C-D}})^2}} \frac{\sum (\rho_{\text{C-D}} - \rho_{\text{A-B}})^2} {\sum (\rho_{\text{C-D}} - \rho_{\text{A-B}})^2}
\]

also make manifest the approximate nature of the local symmetry. These range from 0.60 for a general screw to 0.56 for a 2-fold rotation. These values are appreciably lower than those from similar comparisons between independent maps of essentially identical molecules. Both in the case of \( \alpha \)-versus \( \gamma \)-chymotrypsin (11) and in the case of monoclinic versus orthorhombic Glycera hemoglobin,\(^2\) the correlation coefficient was found to be about 0.75. The density projections in Fig. 1 and the backbone models in Fig. 2 show pictorially the inexact, yet striking, internal symmetry in myohemerythrin.

\( ^2 \) W. A. Hendrickson, unpublished results.
Fig. 2. Stereodrawings of α carbon backbones of myohemerythrin in pseudosymmetrically related views. Stereopair A is of the molecule as viewed looking into the c axis of the crystal with b across to the right. Stereopair B is of the molecule after transformation by parameters of the general screw axis reported in Table I. These drawings are based on the initial atomic model for the polypeptide backbone (2). It is now known that this model is inaccurate in detail but that it is faithful to the general features of the polypeptide folding (12).

The pseudodiad does not relate the whole of the molecule onto itself, but it does apply to about two-thirds of the molecular volume. A total of 1598 grid points for the A-B helix pair were used in the refinements as compared with the 4829 grid points in the entire isolated molecule. The molecular portions included in these refinements may actually have underestimated the extent of pseudosymmetric relatedness. As can be seen in Fig. 2, the transformation that relates the C-D helix pair to the A-B pair places the E helical stub into approximate correspondence with the B-C corner as well.

ROTATION FUNCTION COMPARISON

A separate attempt at locating and quantifying the pseudosymmetry in the structure of myohemerythrin was made by searching for rotational symmetry in the Patterson function of the molecule. The Patterson function, which can be derived directly from the intensities of diffraction from a crystal, corresponds to a weighted superposition of the interatomic vectors of a structure, all referred to the origin. Thus, rotational symmetry in the Patterson function is indicative of a corresponding symmetry in the structure. The rotation function provides a sensitive means for detecting such symmetry (9). This approach has the advantage of providing a comprehensive survey of any rotational symmetry that may be present without depending on prior conjecture about the possible symmetry. However, information about translational components of the symmetry is not directly accessible in the Patterson.

The rotation function in this study was computed from the Fourier transform of an isolated molecule rather than from the diffraction pattern of the crystal. This was done in order to eliminate the complications of the crystallographic symmetry that would be represented in the Patterson function by the interatomic vectors between molecules of the real crystal. The electron density distribution of an isolated molecule at 5.5 Å resolution, shown in Fig. 1 of Ref. 1, was "crystallized" in space group P1 in a cell of a = 60 Å, b = 88 Å, and c = 56 Å. The axial directions are the same as in the real P2₁₂₁ crystal. Cell dimensions were chosen to be just sufficient to preclude overlap of the Patterson vectors from crystallographically related molecules, namely twice the molecular dimensions. Structure factors for this hypothetical crystal were calculated by Fourier inversion of the electron density function.

The rotation function was computed with Crowther’s program for implementing his fast rotation function algorithm (13). The F² values for all reflections of spacing between 12 Å and 6 Å were included in the calculation for a total of 2503 data. A cutoff sphere of radius 35 Å was used in the expansion of the Patterson function into spherical harmonic coefficients. The rotation function comparing the Patterson function of myohemerythrin with itself was computed on a 5° grid in the Eulerian system (α, β, γ). These angles are related to the convention of Kossmann and Blow (9) by θ₁ = α + π, θ₂ = β, and θ₃ = γ - π. Angular ranges of 0 ≤ α ≤ 2π, 0 ≤ β ≤ π, and 0 ≤ γ ≤ 2π were required in order to conveniently encompass the asymmetric unit. However, since the rotational self-comparison of a Patterson function has special added symmetry that gives mirror planes at α + γ = ±π, this region is actually 2-fold redundant.

The resulting rotation map was normalized relative to the standard deviation from the mean of the entire map. That is, the normalized value R¹ is related to the raw rotation function value R according to

\[ R¹ = (R - \bar{R})/\Delta \]

where

\[ \Delta² = \sum w R² - \bar{R}²/\sum w \]

\[ \bar{R} = \sum w R/\sum w \]

The weight, w, given to each value in computing the mean, \( \bar{R} \), and the standard deviation from the mean, \( \Delta \), was in proportion to the volume of rotation space that is associated with each grid point. This is \( w = \frac{1}{2} \sin β \) for an evenly sampled Eulerian grid (10). The sections at β = 0 and β = π were treated specially and given weights equal to the ratio of the number actually taken, i.e. w = 1/72. Weighting made an
appreciable difference in the normalization parameters. After weighted normalization, the maximal and minimal values of the map are 8.8 and −2.0 whereas unit weighting would have given 7.4 and −1.6 for these values.

The grid of rotation values was searched for all peaks of $R'$ greater than 2.0. Peaks of that magnitude were found at a total of 19 unique orientations. The location and peak values for each of these maxima were then evaluated more precisely by an interpolation procedure. Next, the Eulerian orientation angles for the maxima were converted to the spherical polar system so as to make the symmetry more apparent. The orientations for the most significant peaks, those with maximal values of 4.0 or greater, are listed in Table II. The next highest peak has a value of 3.3. All but three of the other 10 peaks of $R'$ greater than 2.0 are at $\chi = 180^\circ$ and rise from shoulders or ridges associated with the principal maxima at $\chi = 180^\circ$. One of the exceptions is at $\chi = 158^\circ$ and at an orientation within 8" of that of $\delta_2$. The other two are maxima of 2.9 and 3.0 that occur at $\chi = 56^\circ$ and $\chi = 119^\circ$. Both of these are for each of these maxima were then evaluated more precisely by an interpolation procedure. Next, the Eulerian orientation angles that are shown in stereographic projection in Fig. 3 were converted to the spherical polar system so as to make the symmetry more apparent. The pseudodiad $\gamma$, that corresponds to this symmetry operator is highly significant at 72% of the identity peak value. However, a number of other unexpectedly significant peaks also occur in the rotation map. Reference to the sketch of myohemerythrin shown in Fig. 4 is helpful in understanding the correspondence of these additional symmetry operators to elements of the molecular structure. The peak designated $\zeta$ corresponds to an approximate 4-fold axis that is nearly coincident with the principal pseudodiad at $\alpha$. It transposes each helix successively into the next. The other highly significant peaks are all pseudodiads approximately at right angles to the diad at $\alpha$. These peaks are also orthogonal to one another in pairs: $\beta$ with $\gamma$, $\delta_1$ with $\epsilon_1$, and $\delta_2$ with $\epsilon_2$. The $\delta_1$, $\delta_2$, and $\epsilon_1$, set of peaks is at 45º with respect to the $\beta$ and $\gamma$ pair, but their satellites, $\delta_3$, $\delta_4$, and $\epsilon_2$, deviate somewhat from this special orientation and are also farther from being mutually perpendicular to the $\alpha$ and $\zeta$ orientation. The pseudodiad at $\beta$ takes helix A into D and B into C while that at $\gamma$ takes A into B and D into C. Pseudodiads $\delta_1$ and $\delta_2$, rotate A and C into themselves and B into D whereas $\epsilon_1$ and $\epsilon_2$ rotate B and D into themselves and A into C.

Thus, the rotational symmetry in myohemerythrin forms an intricate and quite systematic pattern. The three highest rotation function peaks, $\alpha$, $\beta$, and $\gamma$, correspond to diads at mutually orthogonal positions and indicate a quasi-D$_4$ (222) point symmetry in the molecule. These symmetry elements taken together with the approximate 4-fold axis at $\zeta$, coincident with the 2-fold at $\alpha$, and the appropriately placed additional diads at $\delta_1$ and $\epsilon_1$, reveal a D$_4$ (422) pseudosymmetry in the structure. The D$_4$ symmetry is, of course, quite inexact; it involves rotation function peaks as low as 45% of the identity peak. However, the positions of these symmetry elements do not deviate by more than 2º from their ideal orientations and the stereographic projections shown in Fig. 3. The drawing was simplified from a computer-generated picture of the $\alpha$ carbon backbone; thus it is faithful to the actual geometry of the structure. The NH$_2$ terminus is designated by N and the four principal helices are identified by A, B, C, and D. The dimeric iron center is represented by the spheroid. A spike has been tacked into this spheroid to show the common orientation of the principal diad, $\alpha$, and the tetrad $\zeta$. The positions of other symmetry elements from the rotation function are shown in reference to the projection of the mean of the two iron positions as the origin. They are identified in correspondence with the labels of Table II. Drawing by Diane Ward.

### Table II

<table>
<thead>
<tr>
<th>Peak</th>
<th>$\phi$</th>
<th>$\psi$</th>
<th>$\chi$</th>
<th>$R'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity</td>
<td>0</td>
<td>8.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>153.1</td>
<td>6.2</td>
<td>180.0</td>
<td>6.3</td>
</tr>
<tr>
<td>$\beta$</td>
<td>−15.5</td>
<td>83.2</td>
<td>180.0</td>
<td>5.4</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>−100.5</td>
<td>64.2</td>
<td>190.0</td>
<td>5.2</td>
</tr>
<tr>
<td>$\delta_1$</td>
<td>31.4</td>
<td>85.4</td>
<td>190.0</td>
<td>4.2</td>
</tr>
<tr>
<td>$\delta_2$</td>
<td>19.1</td>
<td>93.4</td>
<td>180.0</td>
<td>4.0</td>
</tr>
<tr>
<td>$\epsilon_1$</td>
<td>−59.1</td>
<td>86.1</td>
<td>180.0</td>
<td>4.2</td>
</tr>
<tr>
<td>$\epsilon_2$</td>
<td>−70.5</td>
<td>74.5</td>
<td>180.0</td>
<td>4.1</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>166.4</td>
<td>6.4</td>
<td>90.7</td>
<td>4.0</td>
</tr>
</tbody>
</table>

![Fig. 3](left): Stereographic projection of the rotation function density at $\chi = 90^\circ$ (A) and at $\chi = 180^\circ$ (B). These are zenithal equiarea projections that have been evaluated by interpolation from the Eulerian grid of the actual rotation map. The projection is oriented with a across, c down, and b out of the page. Angle $\phi$ runs counterclockwise from a zero coincident with a. Lines of longitude are marked at 45º intervals. Angle $\psi$ runs out from the pole and lines of latitude are marked at 30º intervals. The Greek letter designations on peaks correspond to those in Table II.

![Fig. 4](right): Schematic drawing of the myohemerythrin structure. The molecule is oriented to correspond exactly to the view in the stereographic projections shown in Fig. 3. The drawing was simplified from a computer-generated picture of the $\alpha$ carbon backbone; thus it is faithful to the actual geometry of the structure. The NH$_2$ terminus is designated by N and the four principal helices are identified by A, B, C, and D. The dimeric iron center is represented by the spheroid. A spike has been tacked into this spheroid to show the common orientation of the principal diad, $\alpha$, and the tetrad $\zeta$. The positions of other symmetry elements from the rotation function are shown in reference to the projection of the mean of the two iron positions as the origin. They are identified in correspondence with the labels of Table II. Drawing by Diane Ward.
throw of the pseudotetrad peak is within 1° of ideal. It should be noted that the values of precisely 180° do not necessarily reflect exact 2-fold rotations. These peaks all occur on the mirror planes at \( \alpha + \gamma = \pm \pi \) and could very well arise from the unresolved superposition of peaks somewhat off precisely 2-fold rotations.

One aspect of the rotation map that is not entirely clear is the origin of the ridge of 2-fold rotational density that occurs between \( \alpha \) and \( \gamma \) in the plane perpendicular to the peak at \( \beta \). In part, this ridge must be due to the pairwise rotational displacement of helices that is seen in projection down \( \alpha \) as shown in Fig. 3 of Ref. 1. However, the significant density at all rotations about the \( \beta \) peak axis is another manifestation of the ridge and it is difficult to see the reason for the 3-fold and 6-fold rotation peaks at this orientation.

**SEQUENCE HOMOLOGY**

The pseudosymmetry that is apparent in the tertiary structure of myohemerythrin leads one to anticipate an accompanying expression of symmetry in the primary structure. Indeed, approximate repeats that correspond to major portions of the A-B and C-D helix pairs have been noticed in the amino acid sequences of both hemerythrin and myohemerythrin (1, 3). In hemerythrin a core sequence containing residues 23 to 54 is repeated at residues 71 to 101 with identities at 8 of 32 positions. In myohemerythrin a similar core segment, 23 to 54, is repeated at 73 to 106 with identities at 11 of 30 positions if one ignores a 4-residue insertion. An alignment of the sequences of *Phascolopsis gouldii* coelomic hemerythrin (4) and of *The- miste zostericola* myohemerythrin (3) that makes these internal repetitions readily evident is shown in Fig. 5. This alignment quite naturally places into register the pairs of histidine residues that have been identified as iron ligands.

The extent of homology for various comparisons between hemerythrin sequences are tabulated in Table III. The measure of homology for the presumed repeat in sequence that derives from a description in terms of core segments that begin and end with identities would be unduly high. In an attempt to avert this bias, the realm of comparison has been chosen on a structural basis. It corresponds to the minimal extent of helical structure as identified in Table I of Hendrickson and Ward (2). This places the limits at the start of helix C and the end of helix D.

The level of homology between the repeated segments in the hemerythrin and myohemerythrin sequences is rather low at 21 and 28%, respectively. Moreover, Klippenstein et al. (3) report statistical tests indicating that although the fraction of identities is somewhat greater than expected by chance, many of the nonidentical residues are actually mutationally quite dissimilar. Comparisons of segments 30 residues long gave an alignment score 3.7 \( \sigma \) greater than that for randomized sequences when using a unitary matrix, but the score was only 0.4 \( \sigma \) when computed with a mutation data matrix (18). Thus the relatedness of these internally repeated sequences seems to be distant at best.

However, several factors combine to suggest that these apparent sequence repeats are significant. First, added significance derives from the occurrence of the repeat in both known sequences; indeed the joint homology for the two structures is 38%. Secondly, just as for the low homology repeats in the carp muscle calcium-binding protein (19), the repeated segments of hemerythrin also correspond precisely with structural and functional elements related by operations of pseudosymmetry. Thirdly, the level of internal sequence homology in hemerythrin is greater than that between several globin sequences where relatedness is quite clearly established. Fourthly, a low level of homology is to be expected on the basis of a plausible

**Table III**

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Length</th>
<th>Insertions or deletions</th>
<th>Sequence homology</th>
<th>Structural dissimilarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hemerythrin vs myohemerythrin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>114</td>
<td>1</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>21-58 vs 21-58</td>
<td>38</td>
<td>0</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>69-105 vs 69-110</td>
<td>38</td>
<td>1</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Hemerythrin vs myohemerythrin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-58 vs 69-105</td>
<td>38</td>
<td>1</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Myohemerythrin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21-58 vs 69-110</td>
<td>39</td>
<td>1</td>
<td>28</td>
<td>1.6</td>
</tr>
<tr>
<td>Carp calcium-binding protein</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39-69 vs 78-108</td>
<td>31</td>
<td>0</td>
<td>23</td>
<td>2.0</td>
</tr>
<tr>
<td>Bacterial ferredoxin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-25 vs 27</td>
<td>26</td>
<td>1</td>
<td>46</td>
<td>1.4</td>
</tr>
<tr>
<td>Lactate dehydrogenase</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22-28, 35-66 &amp; 77-83 vs 92-95, 121-152 &amp; 199-164</td>
<td>41</td>
<td>2</td>
<td>11</td>
<td>2.1</td>
</tr>
</tbody>
</table>

\( a \) Each insertion or deletion of residues has been given one position in a length of sequence.

\( b \) The value cited for structural dissimilarity is the root mean square deviation of the \( \alpha \) carbon positions in one set from those of the other set after a rigid-body transformation determined by the method of least squares. The \( \alpha \) carbon coordinates and sequences used in these comparisons were taken from Hendrickson and Ward (2), Krebsinger and Nockolds (15), Adman et al. (16), and Adams et al. (17) for myohemerythrin, the calcium-binding protein, ferredoxin, and lactate dehydrogenase, respectively. In the case of myohemerythrin, positions 40 to 42 are compared with 92 to 94 for structural similarity but with 86 to 90 for sequence homology.
Pseudosymmetry in Hemerythrin

3017

evolutionary model. If the repeated segments are assumed to
derive from a duplicated common sequence and if coelomic and
muscle hemerythrin are assumed to be objects of divergent
evolution, then the duplication event that produced a proto-
hemerythrin must have preceded the split into the two tissue
lines. Thus, the repeated segments will have had a longer
mutational history and therefore less residual similarity can
be expected between duplicated segments than exists between
any two hemerythrins. As is shown in Table III, the homology
between hemerythrin and myohemerythrin is only about 39%
in the repeated segments. This sets an upper bound for the
repeat homology.

The low level of homology makes it very difficult to estimate
the true extent of the putative duplicated sequence. The core
sequence repeat can be extended in two different ways to
include additional identities between conserved residues. One
possibility would bring tyrosines 18 and 67 into alignment by
introducing a single residue deletion somewhere between 67
and 71 (1). A second possible extension would align tyrosine 67
with tyrosine 109 or 114 at the expense of a 5-residue gap (3).
This proposal has the advantage of relating tyrosines that may
both be iron ligands. However, it is the first proposed exten-
sion that best matches the structural relatedness. This can be
seen in Fig. 2. The pseudodiad brings residues 59 to 66 from
the B-C corner into close proximity with residues of the E stub,
106 to 113 or 111 to 118 as the case may be. Residue 67 then
falls into a position quite near residue 18 at the start of the A
helix. The transformed position of residue 67 is also near
tyrosine 109 (114) as would be expected from the second pro-
posal, but neighboring positions then are not aligned. The
alignment given by the first proposed extension, shown in Fig.
5, also very naturally suggests plausible bounds for the dupli-
cated regions.

DISCUSSION

The evidence clearly establishes the existence of approxi-
mate symmetry within the myohemerythrin molecule. More-
over, this internal symmetry is a simplifying feature of the
hemerythrin fold in general since the structure of the proto-
mer of octameric hemerythrin has been shown to be very
similar to the myohemerythrin structure (6, 7). The symmetri-
cal nature of this structure can be described as a hierarchy of
symmetries of decreasing fidelity. In the strictest of terms the
molecule is completely asymmetric, point group C1. However,
pairs of helices are closely related to one another by quite a
good pseudodiad, and the realm of applicability of this symme-
try operator can be extended, albeit at a lower level of approxi-
mation, to include over 80% of the molecule. Thus, the mole-
cule has approximate C3 symmetry if the first 17 residues are
ignored. At right angles to this principal pseudodiad there are
additional, less exact diads that impart a quasi-D3 symmetry
to the structure. Finally, at a still lower level of exactness
other symmetry elements combine with these to give an ultimate
symmetry that is very approximately D1. These dihedral
aspects of the symmetry, D2 and D3, are in a sense only
adventitious since helices are transposed without regard to
polarity by these operations.

Symmetry within the tertiary structure of a protein is not
unique to hemerythrin. A number of other proteins are
known to fold in a series of domains that reflect repeats in
amino acid sequence. Immunoglobulins (20) and molluscan
hemocyanins (21) are notable examples. However, the symme-
try of the hemerythrins is distinct from that in these proteins.
In the hemerythrin case, the symmetry is within a domain so
that symmetrically related parts are in intimate contact. Other
instances of intradomain symmetry have been charac-
terized in the ccmp muscle calcium-binding (20) and in bacte-
rial ferredoxin (16). In both of these cases, as in hemerythrin,
approximate 2-fold axes relate metal-binding sites. Another
case of intradomain symmetry occurs in the pseudodiad
relationship between the two mononucleotide binding units of the
NAD-binding domain in lactate dehydrogenase (22).

It is of interest to compare the degree of sequence and
structural similarity between the symmetrically related por-
tions of these 4 molecules. To provide a common measure for
structural similarity, a carbon positions have been compared
in each case. The root mean square deviation between equiva-
 lent a carbon positions was calculated after transposing one
set into the other with a least squares refined rigid body
transformation. Results of these comparisons are given in
Table III along with the corresponding homologies in se-
quence. Bacterial ferredoxin has a match of a carbon positions
that is very good as might be expected from its obvious dupli-
cation in sequence. On the other hand, the marginal level of
sequence homology in the calcium-binding protein is reflected
in an appreciably lower agreement between a carbon posi-
tions. Nonetheless, this match is quite good and is, in fact,
much better than was reported earlier (23). The symmetry
within the nucleotide-binding domain of lactate dehydrogen-
ase is, in contrast, considerably less exact. Even at the ex-
 pense of introducing additional gaps at corners (45 to 46 and 56
to 57), the structural dissimilarity can only be reduced to 2.3
Å. The pseudosymmetric relatedness in myohemerythrin com-
pares favorably with that in these other proteins. Despite
substantially less sequence homology than occurs in ferre-
doxin, the structural match in myohemerythrin is nearly as
close as for that protein. Interestingly, the transformation
based only on a carbon positions also brings the iron atoms of
myohemerythrin to within 1.1 Å of superposition.

The evolutionary mechanism for the origin of pseudosym-
mmetry in myohemerythrin and these other proteins is an un-
resolved issue. Two very different alternatives can be consid-
ered. It may be that the symmetry observed in hemerythrins is
merely a natural consequence of a particularly stable supersec-
ondary folding (22, 24). The helix topology of the hemerythrin
fold also occurs in the tobacco mosaic virus protein (25) and is
very likely present in the purple membrane protein (26) as
well. Perhaps hemerythrin evolved from some other protein
that possessed such helix topology and acquired the capacity
for reversible oxygenation through a suitable combination of
mutations to produce ligands for a dimeric iron center. Thus
the requirements for symmetry in function and folding will
have imposed on the sequence the approximate repeats that
are now evident. This is the argument for convergent evolu-

Another possible evolutionary route is that of gene duplica-
tion and fusion. At some distant time, part of a gene coding for
a polypeptide sequence symbolized as N'A'B' could have been
duplicated and the copy then fused to the original to give a
sequence N'A'B'A'B'. This protohemerythrin would then
have been subject to a history of perfecting as well as indiffer-
ent mutations to yield the contemporary sequence NABCD.
In this, the view of divergent evolution, the approximate repeats
seen in current sequences are vestiges of an earlier complete
identity. The proposition that hemerythrin has evolved by
way of gene doubling is somewhat strengthened by the near
identity in the lengths of repeated segments. It is attractive
to speculate that the original A'B' helix pair was itself a chelate
for iron so that upon duplication, a dimeric iron center could arise quite naturally from the juxtaposition of two preformed metal-binding sites. However, the process could not have been so simple since the coordinating ligand for at least one, and perhaps both, of the iron atoms are furnished jointly from the two repeat segments (1, 8). This situation is like that in bacterial ferredoxin where the coordination of iron/sulfur clusters also bridges the symmetry related halves of the molecule (16).

REFERENCES
Pseudosymmetry in the structure of myohemerythrin.
W A Hendrickson and K B Ward


Access the most updated version of this article at http://www.jbc.org/content/252/9/3012

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/252/9/3012.full.html#ref-list-1