tion, 2nd ed, Academic Press, New York, N.Y.
Tawada, K., & Oosawa, F. (1969a) *J. Mol. Biol. 44,* 309-317.
Tawada, K., & Oosawa, F. (1969b) *Biochim. Biophys. Acta 180,* 199-201.
West, J. J. (1970) *Biochemistry 9,* 3847-3853.
Yagi, K., Mase, R., Sakukibara, I., & Asai, H. (1965) *J. Biol. Chem. 240,* 2448-2454.

Structure of Methemerythrin at 2.8-Å Resolution: Computer Graphics Fit of an Averaged Electron Density Map†

Ronald E. Stenkamp, Larry C. Sieker, Lyle H. Jensen,* and John E. McQueen, Jr.

ABSTRACT: The crystal structure of methemerythrin from *Themiste dyscritum* has been determined at 2.8-Å resolution by single isomorphous replacement techniques combined with anomalous scattering from a K_2HgI_4 derivative. Noncrystallographic symmetry relating the four subunits in the asymmetric unit was used to obtain an average electron density map of the hemerythrin monomer, and a computer graphics system was used to fit a polypeptide model to the electron density. The average map was of sufficient quality to locate most of the amino acid side chains and to confirm the assignment of His-25, His-54, Glu-58, His-73, His-77, His-101, Asp-106, and Tyr-109 as the iron ligands. One of the mercury sites in the heavy atom derivative is located between two Cys-9 residues related by a noncrystallographic twofold axis, although no intersubunit disulfide bond is present in the native structure. The residues responsible for the binding of the subunits to form the octamer are identified.

Low resolution molecular structures for the oxygen-carrying proteins, myohemerythrin (monomer) and hemerythrin (octamer), have been determined for molecules from different species of marine worms (Hendrickson & Ward, 1975; Ward et al., 1975; Stenkamp et al., 1976b). These models show the secondary and tertiary structure of the subunits and the arrangement of the subunits in the octamer, but a detailed view of the proteins sufficient to explain subunit interactions and the mode of binding of small molecules can only be derived from higher resolution electron density maps.

The preliminary results of a 2.8-Å resolution study of hemerythrin from *Themiste dyscritum* have been reported (Stenkamp et al., 1976a) indicating the geometric arrangement of the iron ligands, but polypeptide models were not fit to the electron density because of the effort required and the uncertainties of the locations of amino acid side chains in the map based on a single heavy atom derivative. We have now used the noncrystallographic symmetry relating the four subunits in the asymmetric volume to obtain an improved average electron density map of the hemerythrin monomer and fit a model to this average density with a molecular display system. This paper reports the results of the model fitting and the structure of hemerythrin.

Electron Density Map. Diffraction data for this study were collected on a computer controlled, four-circle diffractometer by the same techniques used for the 5.0-Å resolution investigation (Stenkamp et al., 1976b).

Table I summarizes the crystal and data collection information. Reflections from four crystals, two native and two derivative, were processed to provide a selected set of 9461 unique Friedel pairs out of the ~14 000 to 2.8-Å resolution. The intensities of from 10 to 13 reflections measured at regular time intervals were used to correct the data sets for radiation decay. The data were corrected for absorption by the usual empirical method (Furnas, 1957; North et al., 1968).

After calculating the heavy atom coefficients as suggested by Matthews (1966a), we sought to refine the positions and B values of individual Hg and I atoms so a more definitive identification of the heavy atom complexes binding to the protein could be made. The least-squares refinement of the parameters for the individual atoms was not well behaved, however, presumably because the Hg and I atoms, separated by ~2.7 Å, were not well resolved by the 2.8-Å data set. Residual density in a difference Fourier map calculated with phases from only the Hg atoms suggested that four of the sites correspond to HgI_2 molecules while the other two were either HgI_3^- or HgI_4^{2-} ions. Refinement of the heavy atoms by ΔF maps did not converge rapidly, so we placed single anisotropic Hg atoms at each site and refined by least squares.

We used the refined heavy atom positions to determine phases for the 2.8-Å resolution data set, making use of the Bijvoet differences to resolve the phase ambiguity (Singh & Ramaseshan, 1966; North, 1965; Matthews, 1966b). The overall figure of merit is 0.807. An electron density map was calculated on a grid of 0.87 × 0.87 × 0.81 Å. The quality of the map was sufficient to identify the iron ligands, and to determine the general location of the C^α atoms in each subunit. The four independent subunits in the asymmetric unit allowed the course of the polypeptide backbone to be traced across occasional gaps in the electron density.

The four subunits in the crystallographic asymmetric unit are distributed between the two octamers in the unit cell. Both octamers are located on unique fourfold axes with the crystallographic asymmetric unit composed of four subunits, two from each octamer. The molecule centered on the fourfold axis

† From the Department of Biological Structure and the Department of Biochemistry, University of Washington, Seattle, Washington 98195, and the Department of Biochemistry and Nutrition, University of North Carolina, Chapel Hill, North Carolina 27514. *Received January 11, 1978.* This work was supported by U.S. Public Health Service Grant AM 3288 from the National Institutes of Health and a postdoctoral fellowship awarded to R.E.S. by the American Cancer Society. The GRIP-75 system development has been supported by National Institutes of Health Biotechnology Research Resource Grant No. RR00898, National Science Foundation Grant No. GJ-34697, AEC Contract No. AT-(40-1)-3817, and the IBM Corporation.

0006-2960/78/0417-2499$01.00/0

TABLE I: Crystal and Data Collection Information.

Space group	$P4$
$a = b$	86.62 (2) Å
c	80.89 (2)
Cu Kα radiation,	λ = 1.5418 Å
K_2HgI_4 derivative with anomalous scattering	
9461 reflections to 2.8 Å resolution	
Average figure of merit	0.807

TABLE II: Transformations between Subunits.[a]

$$\begin{pmatrix} -0.385487 & 0.922713 & 0.0 \\ 0.922713 & 0.385487 & 0.0 \\ 0.0 & 0.0 & -1.0 \end{pmatrix}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{IA} + \begin{pmatrix} 0.00 \\ 0.0 \\ 0.11492 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{IB}$$

$$\begin{pmatrix} -0.338598 & -0.940931 & 0.0 \\ 0.940931 & -0.338598 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{pmatrix}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{IA} + \begin{pmatrix} 0.5 \\ 0.5 \\ 0.49104 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{IIA}$$

$$\begin{pmatrix} 0.675148 & -0.737682 & 0.0 \\ -0.737682 & -0.675148 & 0.0 \\ 0.0 & 0.0 & -1.0 \end{pmatrix}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{IA} + \begin{pmatrix} 0.5 \\ 0.5 \\ 0.60596 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_{IIB}$$

[a] Matrices to be applied to subunit IA coordinates (fractional) to generate subunits IB, IIA, and IIB.

(0, 0, z) is designated molecule I, the other, centered on the axis ($\frac{1}{2}$, $\frac{1}{2}$, z) is molecule II. The two subunits making up each octamer are labeled A and B, A having higher z coordinate than B. Subunits A and B are related by noncrystallographic twofold axes. The four subunits in the asymmetric unit are thus identified as IA, IB, IIA, and IIB (Stenkamp et al., 1976a).

Map Averaging. Although the four subunits in the asymmetric unit are not crystallographically equivalent, the hemerythrin subunits appear to be chemically identical, and averaging the electron density for the subunits should improve the map to provide a better view of the monomer. Transformations relating coordinates in the subunits were obtained from 113 C$^\alpha$ markers in each, making use of the fact that, for the octamer to have the assumed D_4 symmetry, the noncrystallographic twofold axes must intersect with and be perpendicular to the crystallographic fourfold axes. Figure 1 shows the location of these twofold axes relative to the crystallographic axes and Table II lists the transformations. Rotating and translating the C$^\alpha$ markers in subunits IB, IIA, and IIB into subunit IA and averaging the four positions resulted in a set of average C$^\alpha$ positions which give an rms distance from the average positions of 1.1 Å, a reasonable value considering the positional errors of each marker.

The electron density map was averaged by use of a PDP 11/70 computer. Because of the limited core available, it was convenient to use direct access input-output routines to randomly access the points in the electron density map. The locations of the density to be averaged were obtained by the transformations in Table II, and an eight-point linear interpolation of the electron density around that location gave the value to be averaged. This process required 12 hours of computer time to average the map of 62 500 grid points. Only 1% of that time was spent in computation, the remainder being involved in positioning the disc heads.

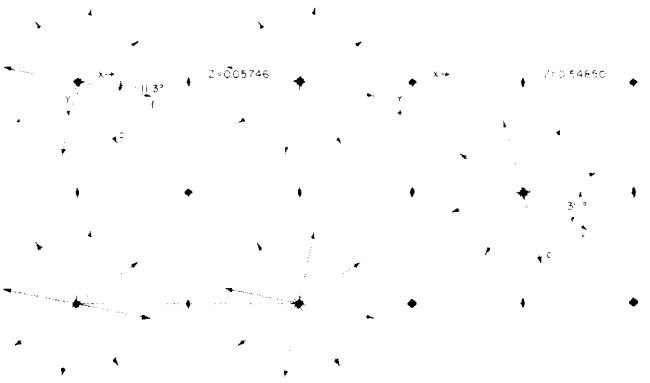

FIGURE 1: Location of noncrystallographic twofold axes relative to crystallographic axes. Axes through corner and faces of octamers designated by c and f, respectively. The errors in the z coordinates of the twofold axes are estimated to be 0.15 Å. Those of the rotations are 0.25°.

The resulting average map is considerably improved over the initial one as indicated by the more uniform electron density along the chain, although the average density at the C$^\alpha$ markers decreased slightly. The interpretation of the density corresponding to many of the amino acid side chains is readily apparent in the average map, and no breaks appear in the electron density along the polypeptide backbone. Figure 2 shows an example of the electron density in the four subunits which went into forming the average density in the final frame. Although the correspondence among the four subunits appears to be poor, the averaged electron density is clear.

While the method of phase improvement described by Bricogne (1974) would have improved the electron density, the map resulting from this simple averaging should be adequate for the purpose of obtaining an initial model for refinement. By generating a model at this stage and then refining it, rather than after several cycles of averaging and rephasing, the model should be less susceptible to possible errors caused by imposing the noncrystallographic symmetry on the structure.

Graphics. The current trend in protein crystallography is to replace the optical comparator (Richards, 1968) with a computer display system. Thus far most of the practical fitting of electron density maps in such systems has made use of the installation in the Computer Science Department, University of North Carolina, where models have been obtained for tRNA (Sussman & Kim, 1976), superoxide dismutase (Richardson, personal communication), sea snake neurotoxin (Tsernoglou et al., 1977), and flavodoxin (Watenpaugh, personal communication). The techniques and programs developed for these structures were used to fit the average electron density map of the hemerythrin subunit. The density was contoured in three perpendicular planes to produce a cage into which the molecular model was fitted by adjusting the various controls on the graphics console. A three-dimensional view of the density and molecule was obtained by depth cueing and rapid rotation of the picture by manual operation of a toothpick-type joystick.

The only change in the techniques used for this structure was the manner in which the initial model was generated for the display system. Normally, C$^\alpha$ positions are input to a model building program which then calculates the locations of the other atoms in some standard orientation. No knowledge of the presence of secondary structure is used, so the peptides and side chains are not initially oriented to provide the correct hydrogen bonding pattern. After attempting to build the model in this manner, we found that a quicker way to fit the density for hemerythrin was to generate several ideal α helices and fit

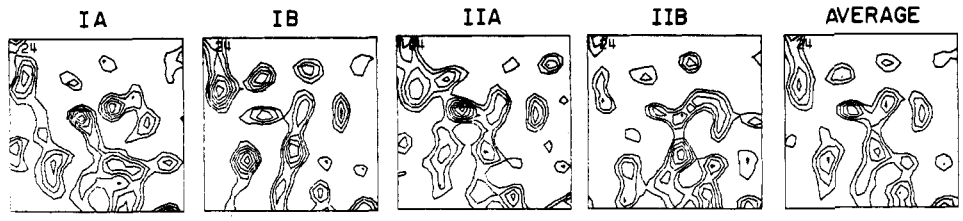

FIGURE 2: Electron density in section at $z = 0.24$, frame 1. Frames 2, 3, and 4 are densities of subunits IB, IIA, and IIB in the same orientation as IA. Frame 5 is the averaged density.

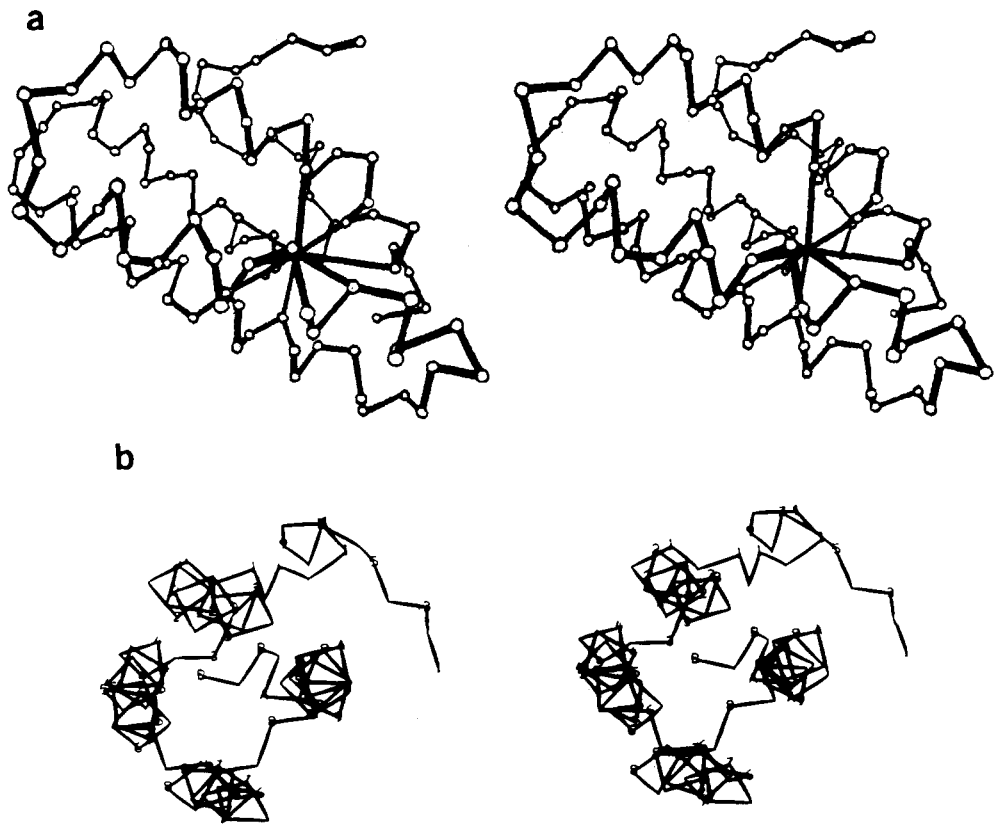

FIGURE 3: Stereo views of C^{α} skeleton of hemerythrin monomer. (a) View approximately normal to helix axes. Ligand residues designated by lines connected to center of Fe complex. Computer generated drawing produced by program written by Dr. C. M. Anderson, Yale University. (b) View approximately parallel to helix axes. Drawing produced by GRIP computer display system, University of North Carolina.

these to the helical regions in the map. The rest of the model was adjusted into the electron density using the joystick and rotational controls.

The amino acid sequence of hemerythrin from *T. dyscritum* had not been completed when the graphics work was undertaken, so the sequence of the *Phascolopsis gouldii* was used for the undetermined residues (Subramanian et al., 1968; Klippenstein et al., 1968). Most of the side chains fit into density reasonably well. One branch of density connects a neighboring subunit through density lying on the noncrystallographic twofold axis and could thus be ruled out as main chain density. The other branch corresponds more closely to the structure of the hemerythrin subunit obtained from previous studies and leads to a more compact subunit.

α Carbon Atoms. The C^{α} coordinates for subunit IA are listed in Table III. Figure 3 is a stereoview of the polypeptide backbone and the connections between the protein and the iron atoms. As in myohemerythrin, the overall dimensions of the subunits are ~20 × ~25 × ~40 Å.

Figure 4 is a contoured plot of the distances between the C^{α} atoms in one subunit (Nishikawa et al., 1972). The four nearly parallel helices give rise to the symmetrical pattern of bands of short distances parallel and perpendicular to the diagonal of the plot. The nearly square cross section of the subunit perpendicular to the helical axes shown in Figure 3b explains the shorter distances between the antiparallel helices.

As shown in Figure 5, the helical regions (A, B, C, and D) are roughly from residues 21 to 37, 41 to 64, 69 to 86, and 90 to 103.

Subsequent to fitting the model on the computer display, it was edited to reflect the chemical sequence which had been completed (Loehr et al., 1978).

Fe–Fe Ligands. On the basis of chemical and low resolution crystallographic studies, Tyr-67 and His-73 have been proposed as iron ligands (Klotz et al., 1976), but the model derived from our 2.8-Å resolution unaveraged map (Stenkamp et al., 1976a) differs in having His-73 and His-77 in the corresponding positions. In addition, we find electron density leading from Glu-58[1] and Asp-106 to the Fe atoms, implicating them as Fe ligands.

[1] Residue 58 in *T. dyscritum* is Glu (Loehr et al., 1978) instead of Gln as reported for *Ph. gouldii*.

TABLE III: Fractional Coordinates of C^{α} Atoms for Model Fit to the Averaged Electron Density Map.

	x/a	y/b	z/c		x/a	y/b	z/c
C-1A	0.0931	0.4213	0.2355	C-58A	0.0331	0.1691	0.1855
C-2A	0.0743	0.4306	0.1994	C-59A	0.0367	0.1490	0.2246
C-3A	0.0460	0.4524	0.1776	C-60A	0.0429	0.1140	0.1970
C-4A	0.0097	0.4352	0.1653	C-61A	0.0731	0.1354	0.1703
C-5A	−0.0079	0.4413	0.1237	C-62A	0.0988	0.1417	0.2050
C-6A	−0.0488	0.4542	0.1113	C-63A	0.0987	0.1029	0.2208
C-7A	−0.0577	0.4262	0.0769	C-64A	0.1128	0.0760	0.1889
C-8A	−0.0522	0.3998	0.1124	C-65A	0.1514	0.0945	0.1971
C-9A	−0.0334	0.3705	0.0823	C-66A	0.1418	0.1013	0.2437
C-10A	0.0095	0.3526	0.0552	C-67A	0.1468	0.1431	0.2512
C-11A	0.0457	0.3771	0.0611	C-68A	0.1593	0.1521	0.2922
C-12A	0.0813	0.3527	0.0494	C-69A	0.1360	0.1903	0.2985
C-13A	0.0999	0.3609	0.0877	C-70A	0.1015	0.1684	0.2829
C-14A	0.0778	0.3280	0.1073	C-71A	0.0823	0.1754	0.3201
C-15A	0.0882	0.2971	0.0835	C-72A	0.0845	0.2168	0.3350
C-16A	0.1113	0.2613	0.1100	C-73A	0.0646	0.2178	0.2928
C-17A	0.1202	0.2553	0.0736	C-74A	0.0340	0.1865	0.2967
C-18A	0.1042	0.2164	0.0778	C-75A	0.0214	0.2083	0.3307
C-19A	0.0642	0.2209	0.0527	C-76A	0.0162	0.2467	0.3131
C-20A	0.0320	0.1942	0.0668	C-77A	−0.0099	0.2331	0.2800
C-21A	0.0388	0.2094	0.1111	C-78A	−0.0318	0.2082	0.3071
C-22A	0.0401	0.2486	0.0953	C-79A	−0.0367	0.2434	0.3325
C-23A	0.0053	0.2411	0.0627	C-80A	−0.0504	0.2668	0.2942
C-24A	−0.0170	0.2312	0.1029	C-81A	−0.0786	0.2345	0.2835
C-25A	−0.0067	0.2695	0.1231	C-82A	−0.0938	0.2425	0.3263
C-26A	−0.0260	0.2895	0.0897	C-83A	−0.0972	0.2878	0.3270
C-27A	−0.0623	0.2670	0.0992	C-84A	−0.1218	0.2754	0.2857
C-28A	−0.0602	0.2736	0.1454	C-85A	−0.1428	0.2470	0.3047
C-29A	−0.0590	0.3185	0.1345	C-86A	−0.1536	0.2728	0.3377
C-30A	−0.0886	0.3224	0.1005	C-87A	−0.1562	0.3072	0.3067
C-31A	−0.1138	0.3104	0.1361	C-88A	−0.1628	0.3509	0.3133
C-32A	−0.1035	0.3435	0.1650	C-89A	−0.1466	0.3687	0.2725
C-33A	−0.1146	0.3733	0.1329	C-90A	−0.1169	0.3989	0.2762
C-34A	−0.1493	0.3480	0.1287	C-91A	−0.1089	0.3874	0.2344
C-35A	−0.1604	0.3565	0.1757	C-92A	−0.0757	0.4158	0.2458
C-36A	−0.1451	0.3972	0.1736	C-93A	−0.0595	0.3895	0.2762
C-37A	−0.1692	0.4021	0.1378	C-94A	−0.0596	0.3564	0.2450
C-38A	−0.2032	0.3895	0.1622	C-95A	−0.0474	0.3762	0.2092
C-39A	−0.2021	0.3588	0.1972	C-96A	−0.0146	0.3861	0.2339
C-40A	−0.2293	0.3346	0.1901	C-97A	−0.0081	0.3477	0.2508
C-41A	−0.2301	0.2984	0.1996	C-98A	−0.0003	0.3222	0.2127
C-42A	−0.2202	0.2768	0.1670	C-99A	0.0220	0.3526	0.1914
C-43A	−0.1840	0.2994	0.1623	C-100A	0.0489	0.3558	0.2279
C-44A	−0.1720	0.2859	0.2047	C-101A	0.0461	0.3107	0.2292
C-45A	−0.1727	0.2444	0.1922	C-102A	0.0571	0.3039	0.1848
C-46A	−0.1521	0.2549	0.1522	C-103A	0.0906	0.3336	0.1830
C-47A	−0.1210	0.2668	0.1825	C-104A	0.1074	0.3215	0.2275
C-48A	−0.1189	0.2349	0.2150	C-105A	0.0970	0.2808	0.2416
C-49A	−0.1197	0.2052	0.1794	C-106A	0.0917	0.2639	0.2014
C-50A	−0.0874	0.2202	0.1540	C-107A	0.1271	0.2717	0.1811
C-51A	−0.0664	0.2278	0.1934	C-108A	0.1483	0.2583	0.2193
C-52A	−0.0806	0.1915	0.2041	C-109A	0.1400	0.2187	0.2016
C-53A	−0.0579	0.1667	0.1753	C-110A	0.1569	0.2361	0.1638
C-54A	−0.0236	0.1917	0.1734	C-111A	0.1818	0.2080	0.1480
C-55A	−0.0133	0.1886	0.2172	C-112A	0.1844	0.1744	0.1766
C-56A	−0.0147	0.1467	0.2095	C-113A	0.1421	0.1627	0.1622
C-57A	0.0009	0.1460	0.1674				

Two aspects of the fitting of the averaged map confirm this assignment of the Fe ligands. First, the density in the loops between the helices was readily fitted with the appropriate number of amino acids. Second, density was found corresponding to the large side chains, particularly Tyr-67, Phe-80, Trp-87, and Trp-97, as well as many others. These observations provide compelling evidence that the iron ligands in this molecule are His-25, His-54, Glu-58,[1] His-73, His-77, His-101, Asp-106, and Tyr-109 (Stenkamp et al., 1976a).

These side chains fit the density near the iron atoms remarkably well, but the map is not of sufficient resolution to give a clear indication of the orientation of all the side chains. The side chains of all five histidine residues, however, can be oriented to place the N^{ϵ} atoms toward the Fe atoms. Furthermore, the hydroxyl O atom of Tyr-109 can be placed in density close to one of the Fe ligands. The current map indicates that both Glu-58 and Asp-106 have a carboxyl O atom in a bridging position, but the possibility of both O atoms on each carboxyl bridging the iron pair cannot be ruled out.

Heavy Atom Sites. The six Hg atom sites found in the heavy atom derivative are distributed in corresponding sites among the four subunits in the asymmetric unit, each subunit having

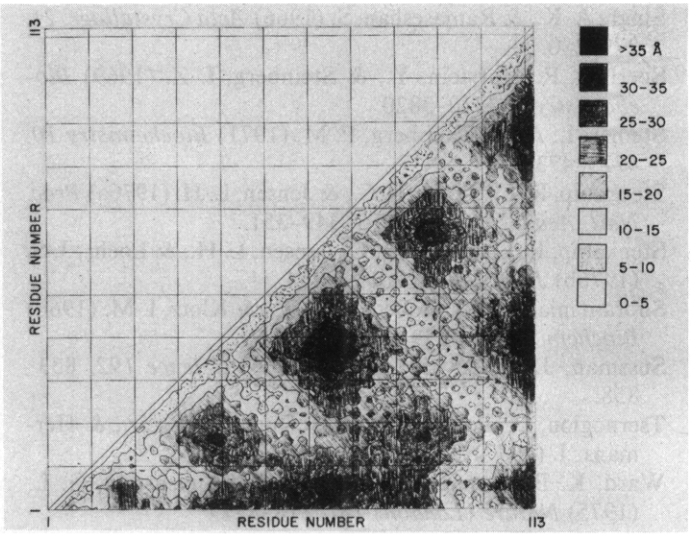

FIGURE 4: Plot of distances between C^{α} atoms of hemerythrin subunit.

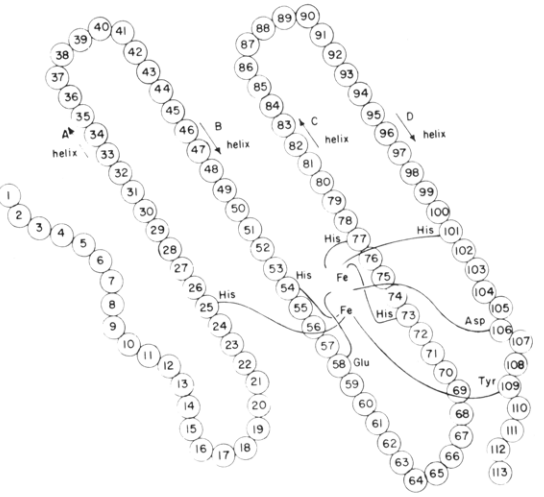

FIGURE 5: Schematic drawing of subunit indicating helical regions and Fe ligands.

an isolated binding site and sharing one with a neighboring subunit. The isolated site is near Cys-50 in the middle of the B helix inside the central cavity of the molecule. In the *Ph. gouldii* molecule, this is the only Cys in the sequence, and binding of mercurials such as *p*-hydroxymercuribenzoic acid causes the octamer to dissociate; however, it appears that, in *T. dyscritum* hemerythrin, Cys-50 can bind small ions without greatly disrupting the subunit interactions. The chemical binding experiments of Clarke & Loehr (personal communication) provide additional evidence that such is the case and that crystal packing forces are not responsible for retaining the octamer upon binding of small mercurials.

The other type of heavy atom site is located between two subunits on the noncrystallographic twofold axes near the N terminal loops of the monomers. In *Ph. gouldii,* residue 9 is Val while in *T. dyscritum* it is Cys. The heavy atom is found to bind between the two Cys-9 residues related by the noncrystallographic symmetry axes, presumably in an arrangement similar to the S–Hg–S bonds formed in immunoglobulins (Steiner & Blumberg, 1971; Schiffer et al., 1973), ribonuclease (Sperling et al., 1969), and papain (Arnon & Shapira, 1969). What prevents the disulfide bridge from forming in the native molecule is not clear, since the two cysteine residues are reasonably close to one another.

Intra- and Intersubunit Interactions. In general, the tertiary structure of the subunit in hemerythrin appears to be maintained by the Fe ligands and hydrophobic interactions between the helices. A few hydrogen bonds may also contribute to the packing of the helices, but a detailed picture of the interactions will await the results of crystallographic refinement.

The hydrogen bonding and other Coulombic interactions contributing to the quaternary structure are the most easily identifiable. In accord with the interpretation of Ward et al. (1975), no hydrogen bonds can be identified between subunits related by the twofold axes through the center of the faces of the square donut-shaped octamer. However, interactions involving intermediary water molecules cannot be ruled out.

The strongest interactions appear between subunits related by the other twofold axes, the ones through the corners of the octamer. Due to a large mass of density between Cys-9 and the twofold related Cys-9 from an adjacent subunit, presumably caused by phasing errors associated with the derivative site located there, it is not possible to determine whether the two side chains are bonded but there is no chemical evidence for an intersubunit disulfide bridge (Loehr, personal communication). There are several places where side chains are sufficiently close to interact and hold the subunits together. For example, the carboxyl group of Asp-23 is near the hydroxyl of Thr-27 and the guanidinium group of Arg-49. Arg-49 is also close to Thr-19, and a Coulombic attraction between Arg-15 and Asp-42 is possible. The terminal oxygen of Asn-30 can be hydrogen bonded to the backbone amide of Trp-10 and/or N^{ζ} of Lys-26.

Slightly fewer interactions occur between fourfold related subunits. The side chain of Arg-48 is hydrogen bonded to the peptide carbonyl of Gln-66, and the hydroxyl of Ser-65 is bonded to the main chain amide of Lys-53. The side chains of Asn-45 and Lys-112 are close together, and Arg-49 appears to interact with the C terminus, possibly through a water molecule.

The model described here is currently being refined crystallographically with 2.5-Å resolution data to provide a more reliable basis for assessing the subunit interactions and determining the detailed arrangement of the Fe ligands. We intend to use the refined model in ligand binding studies to determine how the subunits bind small molecules such as N_3^-, SCN^-, ClO_4^-, H_2O, and O_2.

Acknowledgments

We thank Professors F. M. Richards and T. A. Steitz of the Department of Molecular Biophysics and Biochemistry, Yale University, for the use of their PDP 11-70 computer for averaging the electron density map. Some of our results were derived with the aid of the GRIP-75 molecular graphics system built by the University of North Carolina Department of Computer Science with the help of many users. Major system contributions were made by E. G. Britton, J. S. Lipscomb, and M. E. Pique, under the direction of W. V. Wright; F. P. Brooks, Jr., is Principal Investigator.

References

Arnon, R., & Shapira, E. (1969) *J. Biol. Chem. 244,* 1033–1038.

Bricogne, G. (1974) *Acta Crystallogr., Sect. A 30,* 395–405.

Furnas, T. C. (1957) *Single Crystal Orienter Instruction Manual,* General Electric Co., Milwaukee, Wis.

Hendrickson, W. A., & Ward, K. B. (1975) *Biochem. Biophys. Res. Commun. 66,* 1349–1356.

Hendrickson, W. A., Klippenstein, G. L., & Ward, K. B. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72,* 2160–2164.
Klippenstein, G. L., Holleman, J. W., & Klotz, I. M. (1968) *Biochemistry 7,* 3868–3878.
Klotz, I. M., Klippenstein, G. L., & Hendrickson, W. A. (1976) *Science 192,* 335–344.
Loehr, J. S., Lammers, P. J., Brimhall, B., & Hermondson, M. A. (1978) *J. Biol. Chem.* (in press).
Matthews, B. W. (1966a) *Acta Crystallogr. 20,* 230–239.
Matthews, B. W. (1966b) *Acta Crystallogr. 20,* 82–86.
Nishikawa, K., Ooi, T., Isogai, Y., & Saito, N. (1972) *J. Phys. Soc. Jpn. 32,* 1331–1337.
North, A. C. T. (1965) *Acta Crystallogr. 18,* 212–216.
North, A. C. T., Phillips, D. C., & Mathews, F. S. (1968) *Acta Crystallogr., Sect. A 24,* 351–359.
Richards, F. M. (1968) *J. Mol. Biol. 37,* 225–230.
Schiffer, M., Girling, R. L., Ely, K. R., & Edmundson, A. B. (1973) *Biochemistry 12,* 4620–4631.
Singh, A. K., & Ramaseshan, S. (1966) *Acta Crystallogr. 21,* 279–280.
Sperling, R., Burstein, Y., & Steinberg, I. Z. (1969) *Biochemistry 8,* 3810–3820.
Steiner, L. A., & Blumberg, P. M. (1971) *Biochemistry 10,* 4725–4739.
Stenkamp, R. E., Sieker, L. C., & Jensen, L. H. (1976a) *Proc. Natl. Acad. Sci. U.S.A. 73,* 349–351.
Stenkamp, R. E., Sieker, L. C., Jensen, L. H., & Loehr, J. S. (1976b) *J. Mol. Biol. 100,* 23–34.
Subramanian, A. R., Holleman, J. W., & Klotz, I. M. (1968) *Biochemistry 7,* 3859–3867.
Sussman, J. L., & Kim, S.-H. (1976) *Science 192,* 853–858.
Tsernoglou, D., Petsko, G. A., McQueen, J. E., Jr., & Hermans, J. (1977) *Science 197,* 1378–1381.
Ward, K. B., Hendrickson, W. A., & Klippenstein, G. L. (1975) *Nature (London) 257,* 818–821.

Primary Structure of Histone H2B from Trout (*Salmo trutta*) Testes†

Arend Kootstra and George S. Bailey*

ABSTRACT: Large scale purification of histone H2B from trout (*Salmo trutta*) testes was achieved by a simple procedure consisting of extraction enrichment and chromatography on Bio-Gel P10 and Sephadex G-100, without the use of ion exchange chromatography. The complete amino acid sequence of histone H2B was deduced from automated sequence analysis and compositional analysis of intact H2B, cyanogen bromide, and *N*-bromosuccinimide fragments, thermolysin peptide TH 1, and maleylated and demaleylated tryptic fragments. Comparison of the primary structure of histone H2B from calf and trout revealed eight changes out of 125 residues. The calculated rate of change for H2B among vertebrates is 0.16 residue/100 residues per 10^7 years. All changes except one have occurred in the amino-terminal 37 residues of H2B, which is in striking contrast to the highly conserved sequence of this region in histones H2A, H3, and H4. An even more striking series of insertions, deletions, and substitutions is observed on comparison of the vertebrate sequences with the amino-terminal regions of H2B from two genera of sea urchins. In addition, H2B appears to have evolved among echinoderms at a rate approximately tenfold higher than among vertebrates. These results suggest that it is the core regions of histones which have the most critically conserved sequences, in line with the proposed role of these regions in forming the histone–histone contacts necessary for nucleosome formation. By contrast, the variability of the amino-terminal region of H2B suggests a more species-specific interaction with chromatin components, possibly in addition to, or other than, nucleosomal DNA.

Histones are well known to be intimately associated on an approximately equal weight basis with the DNA of all eukaryotes (for a recent review, see Felsenfeld, 1978). The structural basis for this interaction appears to be a stoichiometric association of two each of histones H2A, H2B, H3, and H4 per unit core length of DNA, the core particle containing approximately 140 base pairs of DNA (Kornberg, 1974; Olins & Olins, 1974; Van Holde et al., 1974; Shaw et al., 1976; Compton et al., 1976; Lohr et al., 1977). These core particles, or nucleosomes, are separated by a variable and less well-defined length of nuclease-accessible (Hewish & Burgoyne, 1973) DNA of 20–60 base pairs (Noel, 1976; Morris, 1976; Lohr et al., 1977; Thomas & Furber, 1976) which also appears to be the sight of DNA interaction with histone H1 (Kornberg & Thomas, 1974; Varshavsky et al., 1976). However, the precise details of histone:histone:DNA interaction, as well as the possible role(s) and relationships of the nucleosomal structure to such events as DNA replication (Weintraub, 1976; Seale, 1976), gene activation, and transcription (Piper et al., 1976; Reeves & Jones, 1976; Garel & Axel, 1976) and histone modification are at present not clearly established.

One useful approach toward an understanding of some of the above problems is the delineation through comparative sequence studies of crucial structural features among each class of histones. Indeed the range of evolutionary variability established among histones (Delange & Smith, 1971; Bailey & Dixon, 1973), and in particular the pair-wise conservatism of H3 and H4 formed part of the basis for the original nucleosome model (Kornberg, 1974). We wish to present here the complete primary sequence of histone H2B from trout. The results, when compared with calf H2B, show that H2B is less conservative than H3 or H4, and evolves among vertebrates at a rate identical with H2A. This would further strengthen the suggestion that H2A and H2B are complementary in their interaction

† From the Department of Biochemistry, University of Otago, Dunedin, New Zealand. *Received October 6, 1977; revised manuscript received March 15, 1978.* This research was supported by grants from the Medical Research Council of New Zealand and the University Grants Committee.

0006-2960/78/0417-2504$01.00/0