Bessler, W., Shafer, J. A., and Goldstein, I. J. (1974), *J. Biol Chem. 249,* 2819–2822.
Brewer, C. F., Sternlicht, H., Marcus, D. M., and Grollman, A. P. (1973a), *Proc. Natl. Acad. Sci. U.S.A. 70* 1007–1011.
Brewer, C. F., Sternlicht, H., Marcus, D. M., and Grollman, A. P. (1973b), *Biochemistry 12,* 4448–4457.
Cunningham, B. A., Wang, J. L., Waxdal, M. J., and Edelman, G. M. (1975), *J. Biol. Chem. 250,* 1503–1512.
Dickerson, R. E., Weinzierl, J. E., and Palmer, R. A. (1968), *Acta Crystallogr., Sect B 24,* 997–1003.
Doyle, R. J., and Roholt, O. A. (1968), *Life Sci. 7,* 841–846.
Edelman, G. M., Cunningham, B. A., Reeke, G. N., Jr., Becker, J. W., Waxdal, M. J., and Wang, J. L. (1972), *Proc. Natl. Acad. Sci. U.S.A. 69,* 2580–2584.
Goldstein, I. J., Hollerman, C. E., and Smith, E. E. (1965), *Biochemistry 4,* 876–883.
Goldstein, I. J., Reichert, C. M., Misaki, A., and Gorin, P. A. J. (1973), *Biochim. Biophys. Acta 317* 500–504.
Hardman, K. D. (1973), *Adv. Exp. Med. Biol. 40,* 103–123.
Hardman, K. D., and Ainsworth, C. F. (1972), *Biochemistry 11,* 4910–4919.
Hardman, K. D., and Ainsworth, C. F. (1973), *Biochemistry 12,* 4442–4448.
Hardman, K. D., Wood, M. K., Schiffer, M., Edmundson, A. B., and Ainsworth, C. F. (1971a), *Proc. Natl. Acad. Sci. U.S.A. 68* 1393–1397.
Hardman, K. D., Wood, M. K., Schiffer, M., Edmundson, A. B., and Ainsworth, C. F. (1971b), *Cold Spring Harbor Symp. Quant. Biol. 36,* 271–276.
Hassing, G. S., and Goldstein, I. J. (1972), *Biochim. Biophys. Acta 271,* 388–399.
Hassing, G. S., Goldstein, I. J., and Marini, M. (1971), *Biochim. Biophys. Acta 243,* 90–97.
Johnson, C. K. (1965), *Oak Ridge Tech. Manual,* 3794.
Karle, J. M., and Karle, I. L. (1972), *J. Am. Chem. Soc. 94,* 9182–9189.
Lis, H., and Sharon, N. (1973), *Annu, Rev. Biochem. 42* 541–574.
Lonsdale, K., Ed. (1952), International Tables for X-ray Crystallography, Vol. I, Birmingham, England, Kynoch Press.
Matthews, B. W. (1968), *J. Mol. Biol. 33,* 491–497.
Matthews, B. W. (1974), *J. Mol. Biol. 84,* 512–526.
Poretz, R. D., and Goldstein, I. J. (1970), *Biochemistry 9,* 2890–2896.
Reeke, G. N., Jr., Becker, J. W., and Edelman, G. M. (1975), *J. Biol. Chem. 250,* 1525–1547.
So, L. L., and Goldstein, I. J. (1969), *Carbohyd. Res. 10,* 231–244.
Sumner, J. B., Gralén, N. and Eriksson-Quensel, I. B. (1938), *J. Biol. Chem. 125,* 45–48.
Villafranca, J. J., and Viola, R. E. (1974), *Arch. Biochem. Biophys. 160,* 465–468.
Wyckoff, H. W., Doscher, M., Tsernoglou, D., Inagami, T., Johnson, L. N., Hardman, K. D., Allewell, N. M., Kelley, D. M., and Richards, F. M. (1967), *J. Mol. Biol. 27,* 563–578.

The Primary Structure of Myohemerythrin†

Gerald L. Klippenstein,* Joseph L. Cote, and Suzanne E. Ludlam

ABSTRACT: The complete amino acid sequence of muscle hemerythrin (myohemerythrin) from the sipunculid *Themiste* (syn. *Dendrostomum*) *pyroides* has been determined by analysis of tryptic, chymotryptic, and cyanogen bromide peptides. The primary structure of myohemerythrin differs substantially from that of coelomic hemerythrins of *Phascolopsis* (syn. *Golfingia*) *gouldii* and *Themiste pyroides,* the amino acid sequence of the muscle protein being only 46 and 45% homologous with the respective coelomic hemerythrins. The most extensive regions of homology between muscle and coelomic proteins occur near the terminii. These and other shorter regions of homology are interpreted in terms of the essential iron ligand residues of the active center.

A number of chemical, spectroscopic, and magnetic techniques have been used in the characterization of the oxygen-binding site of the non-heme iron protein hemerythrin. Chemical modification experiments directed toward identification of the amino acid side chains which provide ligands to the two iron atoms have indicated that lysyl (Fan and York, 1969, 1972), tryptophanyl (York and Fan, 1971), and carboxyl (Klippenstein, 1972a) groups are not iron linked while perhaps four histidyl (Fan and York, 1969) and two or three tyrosyl (Rill and Klotz, 1970, 1971; York and Fan, 1971; Fan and York, 1972) residues may be iron ligands. To confirm and extend these results, analysis and comparison of the amino acid sequences of hemerythrins from a variety of sources are being done. A limitation in this approach has been the small number of amino acid interchanges between hemerythrins studied, e.g., five among variants of *Phascolopsis* (syn. *Golfingia*) *gouldii* coelomic hemerythrin (Klippenstein, 1972b) and four between *P. gouldii* and *Themiste* (syn. *Dendrostomum*) *pyroides* coelomic hemerythrins (Ferrell and Kitto, 1971).

The isolation and characterization of myohemerythrin

† From the Department of Biochemistry, Spaulding Life Science Building, University of New Hampshire, Durham, New Hampshire 03824. *Received July 24, 1975.* Supported in part by a grant from the National Science Foundation (GB 35610) and a grant from the Central University Research Fund at the University of New Hampshire. Published with the approval of the Director of the New Hampshire Agricultural Experiment Station as Scientific Contribution No. 788.

from *Themiste pyroides* (Klippenstein et al., 1972), a monomeric protein which differs substantially in primary structure from coelomic hemerythrins, presented the opportunity to identify a number of variable amino acid residues and to see the emergence of invariant residues, some of these presumably being part of the oxygen-binding site. This paper describes the determination of the complete amino acid sequence of myohemerythrin and some conclusions which can be drawn about the iron-linked groups in the active center of the protein.

Materials and Methods

Purification of Myohemerythrin. Myohemerythrin was prepared from the retractor muscles of 100 *Themiste pyroides* (Pacific Biomarine Supply Co., Venice, Calif.) by a modification of the procedure of Klippenstein et al. (1972). All steps were carried out at 0–4 °C. The tissue was homogenized in 0.1 M Tris-chloride buffer (pH 8.0) and centrifuged at 750*g* for 20 min and the supernatant made 0.1 M in sodium azide and 5×10^{-4} M in dithioerythritol. After standing overnight the solution was made 70% and then 80% saturated in ammonium sulfate by addition of the solid and the precipitates obtained upon centrifugation were discarded. The supernatant was then made 100% saturated in ammonium sulfate and the precipitate collected by centrifugation. The precipitate was dissolved in 0.1 M sodium azide, which was 2.5×10^{-4} M in dithioerythritol and dialyzed against the same solution. The azidometmyohemerythrin solution was concentrated by pressure ultrafiltration and applied to a 2.5 × 90 cm column of Sephadex G-75, equilibrated with 0.1 M NaCl which was 5×10^{-4} M in dithioerythritol. The tubes containing the myohemerythrin, as detected by its absorption at 445 nm, were combined.

S-Carboxyamidomethylmyohemerythrin. The two cysteine residues in myohemerythrin were blocked by the procedure of Crestfield et al. (1963) which was modified only in that iodoacetamide was substituted for iodoacetic acid and that exhaustive dialysis against cold 0.1 M acetic acid replaced the gel filtration step. The *S*-carboxyamidomethylated protein was subsequently dialyzed against distilled water and then made 0.1 M in ammonium bicarbonate by addition of the solid, and pH 8.0 by addition of ammonium hydroxide.

Enzymatic Digestions. All enzymes used were obtained from Worthington Biochemical Corp., Freehold, N.J. Digestions of *S*-carboxyamidomethylmyohemerythrin with α-chymotrypsin or TPCK[1]-treated trypsin were carried out at 37 °C by adding 1% enzyme by weight with respect to protein at zero time and another 1% at 3 h. Digestion was terminated at 6 h by lyophilization.

Carboxypeptidase A and B digestions were performed as described by Groskopf et al. (1966) using in both cases 0.05 M sodium borate buffers at pH 8.5. Leucine aminopeptidase digestions were done as described by Margoliash and Smith (1962). Amino acids released in exopeptidase digests were determined on the amino acid analyzer.

Cyanogen Bromide Cleavage. To 1.8 μmol of *S*-carboxyamidomethylmyohemerythrin in 1.5 ml of 70% formic acid was added a 600-fold molar excess (0.114 g) of CNBr in an additional 1.5 ml of 70% formic acid. The mixture was allowed to react at room temperature for 24 h after which the sample was lyophilized.

Fractionation of Peptides. Tryptic and chymotryptic digests were fractionated on a 1.5 × 90 cm column of Sephadex G-15, equilibrated in 0.1 M ammonium bicarbonate (pH 8.0), and eluted at a flow rate of 7 ml/h; 1.5-ml fractions were collected, the tubes monitored at 254 or 280 nm, and the peak fractions combined. Cyanogen bromide peptides were fractionated on a G-50 Sephadex column of the same dimensions and run under the conditions described above, and the effluent was monitored at 220 nm. Some of the fractions obtained from Sephadex columns were further fractionated on columns of Dowex 50-X 2 (1.2 × 65 cm) by the procedure described by Schroeder (1967). The convex gradient in pyridine acetate concentration and pH was generated by a two-chambered device, the mixing chamber containing 500 ml of pH 3.1 pyridine acetate buffer (0.2 M in pyridine) and the reservoir containing 1 l. of pH 5.0 pyridine acetate (2 M in pyridine); 2.5-ml fractions were collected and monitored by the alkaline hydrolysis–ninhydrin procedure of Hirs (1967).

One of the Sephadex fractions, T-I, was purified by chromatography on a 1 × 30 cm column of DEAE-Sephadex A-25 eluted with a 500-ml linear gradient of 0.05–0.5 M ammonium bicarbonate (pH 8.0) at a flow rate of 15 ml/h; 2-ml fractions were collected and their absorbance at 254 nm was measured.

Purification of peptides by preparative paper electrophoresis, chromatography, or peptide mapping was carried out as described previously (Klippenstein, 1972b). Peptides were eluted from paper using 1% v/v aqueous pyridine.

Sequence Analysis. Amino acid analysis was done as described previously (Klippenstein, 1972b) on a Beckman 120C amino acid analyzer. Peptides were hydrolyzed for 24 h at 110 °C, whereas intact myohemerythrin was hydrolyzed for 24, 48, and 72 h and values for threonine and serine were extrapolated to zero time. Sequencing was done using the dansyl-Edman technique (Gray, 1967) with identification of dansyl-amino acids by chromatography on thin-layer polyamide sheets (Cheng-Chin, obtained from Gallard-Schlesinger, Carle Place, N.Y.) (Hartley, 1970) or by direct identification of Pth-amino acids on thin-layer silica gel plates, using solvent system V of Jeppsson and Sjöquist (1967) to monitor the progress of the Edman procedure of Peterson et al. (1972).

Nomenclature. Peptides are designated T for tryptic digest, C for chymotryptic digest, and CNBr for cyanogen bromide digest; this designation is followed by the numbers, in parentheses, of the amino-terminal and carboxyl-terminal residues in the total amino acid sequence of myohemerythrin.

Results

Elution patterns from gel filtration of the tryptic and chymotryptic digests of *S*-carboxyamidomethylmyohemerythrin are shown in Figure 1. The amino acid compositions of the purified tryptic and chymotryptic peptides are shown in Tables I and II. The elution pattern for fractionation of the cyanogen bromide peptides on Sephadex G-50 is shown in Figure 2. Of these peptides, only CNBr(77–118) required further purification which was done by rechromatography on the same Sephadex column. The amino acid compositions of the cyanogen bromide peptides are shown in Table III. Included in Table III is the amino acid compo-

[1] Abbreviations used are: TPCK-, L-1-tosylamido-2-phenylethyl chloromethyl ketone; Dns, 5-dimethylaminonaphthalene-1-sulfonyl-; Pth, phenylthiohydantoin; Cam-, carboxyamidomethyl-; Cm-, carboxymethyl-.

Table I: Tryptic Peptides of *T. pyroides* Myohemerythrin[a] (Residues/mole of Peptide).

	T(1–15)	T(16–27)	T(28–30)	T(31–37)	T(38–49)	T(50–66)	T(67–75)	T(76–83)	T(79–83)	T(84–94)	T(95–100)	T(101–108)	T(109–113)	T(114–115)	T(116–118)
Lysine		1.96 (2)	1.05 (1)		0.95 (1)	1.02 (1)	2.02 (2)	2.04 (2)	1.02 (1)	0.98 (1)	1.03 (1)	1.01 (1)	0.98 (1)	1.02 (1)	0.98 (1)
Histidine		0.95 (1)				1.85 (2)	0.91 (1)	0.95 (1)				0.94 (1)			
Arginine	0.97 (1)			1.00 (1)											
Aspartic acid	1.00 (1)	1.02 (1)		1.04 (1)	3.02 (3)	2.10 (2)		0.96 (1)	0.99 (1)	1.03 (1)	1.95 (2)	1.02 (1)	1.03 (1)		
Threonine[b]					0.95 (1)	2.82 (3)							1.00 (1)		
Serine[b]	0.85 (1)				0.87 (1)		1.00 (1)			0.87 (1)					
Glutamic acid	3.04 (3)	3.98 (4)				2.14 (2)	1.17 (1)	1.22 (1)	0.96 (1)			1.03 (1)			
Proline	2.39 (2)				1.04 (1)		1.02 (1)			1.03 (1)					
Glycine	1.00 (1)			1.02 (1)						1.96 (2)			1.00 (1)		0.99 (1)
Alanine					1.97 (2)	2.93 (3)				2.00 (2)					
Cm-cysteine[c]				0.80 (1)							0.69 (1)				
Valine	1.05 (1)	1.03 (1)			1.04 (1)	1.08 (1)	1.60 (2)			1.03 (1)	1.05 (1)	1.00 (1)			
Methionine[d]						1.85 (2)		0.80 (1)							
Isoleucine	1.03 (1)		1.00 (1)	1.99 (2)						0.97 (1)		0.96 (1)			
Leucine		1.03 (1)			1.96 (2)			1.01 (1)	0.99 (1)	0.99 (1)		1.03 (1)			1.03 (1)
Tyrosine	0.96 (1)	1.00 (1)					0.88 (1)				0.96 (1)			0.98 (1)	
Phenylalanine	0.98 (1)	0.91 (1)	0.94 (1)	0.95 (1)		0.97 (1)		0.90 (1)	1.02 (1)				0.98 (1)		
Tryptophan	(2)											(1)			
Total residues	15	12	3	7	12	17	9	8	5	11	6	8	5	2	3
% yield of purified peptide	30	21	45	50	64	19	15	6	51	69	46	33	48	82	53

[a] Residue ratios in parentheses represent number of amino acids found in the sequence of the peptides. [b] Uncorrected for hydrolytic losses. [c] Includes 1/2-cystine. [d] Includes methionine sulfoxides.

Table II: Chymotryptic Peptides of *T. pyroides* Myohemerythrin[a] (Residues/mole of Peptide).

	C(1–8)	C(9–14)	C(15–17)	C(18–29)	C(30–33)	C(34–43)	C(44–47)	C(48–55)	C(63–67)	C(68–73)	C(74–76)	C(77–80)	C(81–87)	C(88–98)	C(99–102)	C(103–112)	C(113–114)	C(115–118)
Lysine				2.05 (2)	0.90 (1)			0.99 (1)	1.16 (1)		1.97 (2)	0.96 (1)	1.03 (1)	1.03 (1)	0.90 (1)	0.97 (1)	1.07 (1)	2.34 (2)
Histidine				1.00 (1)				1.02 (1)		0.92 (1)		0.82 (1)				0.88 (1)		
Arginine			0.90 (1)			0.95 (1)												
Aspartic acid		1.07 (1)		1.14 (1)		4.00 (4)		0.99 (1)	1.16 (1)			1.12 (1)		3.12 (3)		2.03 (2)		
Threonine[b]							0.94 (1)	1.94 (2)								0.98 (1)		
Serine[b]		0.87 (1)				1.03 (1)				0.90 (1)				0.93 (1)				
Glutamic acid	2.11 (2)	1.12 (1)		3.98 (4)						1.06 (1)			1.21 (1)		1.16 (1)			
Proline	1.77 (2)					1.03 (1)				1.12 (1)				1.01 (1)				
Glycine	1.18 (1)				1.14 (1)								2.10 (2)			1.09 (1)		0.93 (1)
Alanine						1.10 (1)	1.06 (1)		1.89 (2)					2.02 (2)				

	T-I	T-II	T-III	T-IV	T-V	C-I	C-II	C-III	C-IV	C-V	CNBr-I	CNBr-II	CNBr-III	CNBr-IV	CNBr-V
Cm-cysteine[c]			0.95 (1)						0.89 (1)						
Valine	0.99 (1)	1.12 (1)		2.03 (2)	1.57 (2)				2.00 (2)	1.04 (1)					
Methionine[d]						1.03 (1)									
Isoleucine	0.96 (1)		1.03 (1)	0.99 (1)	0.92 (1)						0.90 (1)			0.92 (1)	
Leucine			0.95 (1)			2.02 (2)					1.78 (2)			0.96 (1)	0.74 (1)
Tyrosine	0.97 (1)		0.83 (1)				0.74 (1)					0.91 (1)			0.94 (1)
Phenylalanine		0.95 (1)	0.98 (1)	0.97 (1)	0.98 (1)			1.09 (1)						1.05 (1)	
Tryptophan	(1)	(1)					1.10 (1)			(1)					
Total residues	8	6	3	12	4	10	4	8	5	6	3	11	4	7	11	4	10	2	28	4
% yield of purified peptide	10	25	23	14	31	30	34	15	32	23	11	8	29	17	1	16	70			

[a–d] See footnotes to Table I.

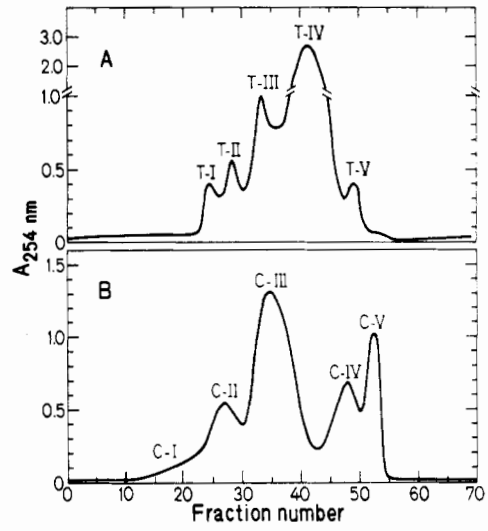

FIGURE 1: Elution pattern of peptides of myohemerythrin from Sephadex G-15 columns (1.5 × 90 cm). (A) Tryptic peptides; (B) chymotryptic peptides.

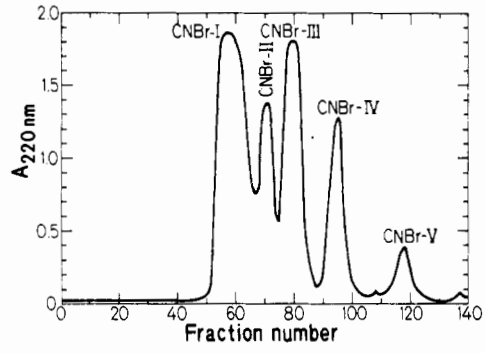

FIGURE 2: Elution pattern of cyanogen bromide peptides of myohemerythrin from a Sephadex G-50 column (1.5 × 90 cm).

sition of myohemerythrin, calculated on the basis of 118 residues per molecule.

Tryptic Peptides. Fractions T-II through T-IV from Sephadex G-15 were subjected to chromatography on columns of Dowex 50-X2. Fraction T-I was further purified on a column of DEAE-Sephadex and T-V was purified by paper electrophoresis. All 15 tryptic peptides described below were sufficiently pure for sequencing following these purification steps.

Peptide T(1-15). Eleven steps of the dansyl-Edman procedure gave the sequence Gly-X-Glx-Ile-Pro-Glx-Pro-Tyr-Val-Y-Asx. Residues X and Y were assumed to be Trp because of the destruction during hydrolysis of their dansyl derivatives. This was confirmed for residue X by the identification of Dns-Trp at this position after chymotryptic hydrolysis of the dansyl derivative of the N-1 peptide.

Carboxypeptidase B digestion for 1 h liberated arginine and phenylalanine in equimolar quantities, indicating, by invoking tryptic specificity, a C-terminal sequence of Phe-Arg. Chymotryptic digestion of peptide T(1–15) yielded two major peptides, one of composition Glu(2.2), Pro(2.1), Gly(0.78), Ile(1.1), Tyr(0.92), Trp(1) and the second of composition Arg(0.90), Asp(1.1), Ser(0.95), Glu(1.0), Val(0.98), Phe(1.1), Trp(1). The partial sequence of T(1–15) is, therefore, Gly-Trp-Glx-Ile-Pro-Glx-Pro-Tyr-Val-Trp-Asx-(Glx, Ser)-Phe-Arg. The remaining sequence and amide assignments were made on corresponding chymo-

Table III: Cyanogen Bromide Fragments of *T. pyroides* Myohemerythrin[a] (Residues/mole).

	CNBr(1–61)	CNBr(62)	CNBr(63–76)	CNBr(77–118)	Myohemerythrin
Lysine	4.2 (4)		3.0 (3)	7.6 (8)	15.2 (15)
Histidine	3.0 (3)		1.0 (1)	1.9 (2)	5.9 (6)
Arginine	1.9 (2)				2.2 (2)
Homoserine + homoserine lactone	0.86 (1)	1.0 (1)	0.79 (1)		
Aspartic acid	7.1 (7)		1.1 (1)	6.0 (6)	13.8 (14)
Threonine[b]	3.6 (4)			1.0 (1)	5.0 (5)
Serine[b]	2.3 (2)		1.0 (1)	1.2 (1)	4.1 (4)
Glutamic acid	9.0 (9)		1.2 (1)	2.2 (2)	12.1 (12)
Proline	3.0 (3)		0.87 (1)	1.1 (1)	5.3 (5)
Glycine	2.4 (2)			4.2 (4)	6.2 (6)
Alanine	3.2 (3)		2.1 (2)	2.1 (2)	7.0 (7)
Cm-cysteine[c]	0.82 (1)			0.74 (1)	(2)
Valine	4.0 (4)		1.3 (2)	3.1 (3)	8.7 (9)
Methionine[d]	0.34 (0)				2.8 (3)
Isoleucine	3.6 (4)			1.9 (2)	5.7 (6)
Leucine	3.1 (3)			4.0 (4)	7.0 (7)
Tyrosine	2.0 (2)		0.92 (1)	1.8 (2)	4.9 (5)
Phenylalanine	4.7 (5)			2.1 (2)	7.0 (7)
Tryptophan	(2)			(1)	(3)
Total residues	61	1	14	42	118
% yield	27	10	23	23	

[a] Residue ratios in parentheses represent integral numbers of residues in sequence of peptide or protein. [b] Uncorrected for hydrolytic losses in peptides, corrected in myohemerythrin. [c,d] See footnotes to Table I.

tryptic peptides C(1–8) and C(9–14) and cyanogen bromide peptide CNBr(1–61) described below.

Peptide T(16–27). Nine dansyl-Edman determinations and 10 steps with direct identification of Pth-amino acids gave the sequence Val-Phe-Tyr-Glu-Gln-Leu-Asp-Glu-Glu-His. By difference, the remaining two positions are occupied by -Lys-Lys.

Peptide T(28–30). The dansyl-Edman procedure gave the sequence Ile-Phe-Lys for this peptide.

Peptide T(31–37). Six dansyl-Edman determinations resulted in the sequence Gly-Ile-Phe-Cys-Asx-Ile. Since this arginine-containing peptide was neutral in electrophoresis at pH 6.4, the Asx must be aspartic acid. Carboxypeptidase B liberated equimolar quantities of arginine and isoleucine after 1 h of digestion. This peptide therefore has the sequence Gly-Ile-Phe-Cys-Asp-Ile-Arg.

Peptide T(38–49). Three steps of the Edman with direct Pth identification gave the N-terminal sequence Asp-Asn-Ser. Chymotryptic digestion of the intact peptide produced, as a major product, a fragment containing all but one valyl and one lysyl residue. By tryptic specificity, these residues must have been removed from the C-terminus of the original peptide. Carboxypeptidase A digestion of the chymotryptic fragment indicated a C-terminal leucine. The partial sequence of T(38–49) is thus Asp-Asn-Ser-(Ala_2, Pro, Asn, Leu, Thr)-Leu-Val-Lys. The amide assignment of the Asn in parentheses was made on the basis of electrophoretic mobility (neutral) of the peptide.

Peptide T(50–66). This peptide was obtained in nearly pure form from the G-15 Sephadex fractionation of the tryptic digest and was further purified on a column of DEAE-Sephadex. The same peptide was also obtained from a Dowex 50 separation of the adjacent Sephadex fraction. Eight dansyl-Edman determinations gave the N-terminal sequence Val-Thr-Thr-Asx-His-Phe-Thr-His. A chymotryptic digest of T(50–66) was fractionated on Sephadex G-15 and, by paper electrophoresis and chromatography, peptides of the following compositions were obtained: C(50–55), His(0.85), Asp(1.2), Thr(2.0), Val(0.97), Phe(0.93); C(50–62), His(1.7), Asp(1.2), Thr(2.9), Glu(2.4), Ala(1.3), Val(0.89), Met(1.6), Phe(0.99); C(63–66), Lys(1.3), Asp(0.85), Ala(1.8). Peptide C(50–55) corresponds to the N-terminal six residues of T(50–66), the sequence of which is shown above. The peptide was slightly basic in electrophoresis, indicating that residue 4 is asparagine. Peptide C(50–62) contains all but four of the amino acids of the original peptide. Carboxypeptidase A digestion for 24 h liberated 1.9 residues of methionine and 0.66 residue of alanine, indicating a C-terminal sequence Ala-Met-Met. The sequence of C(50–62) is therefore Val-Thr-Thr-Asn-His-Phe-Thr-His-Glx-Glx-Ala-Met-Met. This peptide is acidic and therefore the two Glx residues must both be glutamic acid. Peptide C(63–66) was found to have an N-terminal Asx by dansylation. The peptide is neutral and therefore Asx is aspartic acid. On the basis of tryptic specificity, the sequence of this tetrapeptide is Asp-Ala-Ala-Lys.

The sequence of T(50–66) is therefore Val-Thr-Thr-Asn-His-Phe-Thr-His-Glu-Glu-Ala-Met-Met-Asp-Ala-Ala-Lys. Confirmation of this sequence was obtained by CNBr cleavage of T(50–66). Reaction with 400 mol of CNBr/mol of peptide yielded three fractions when chromatographed on Sephadex G-15. The compositions of these fragments were: CNBr(50–61), His(1.9), Asp(1.1), Thr(2.7), Glu(2.3), Ala(1.0), Val(1.1), Phe(0.95), homoserine lactone(0.42); CNBr(62), homoserine(1); CNBr(63–66), Lys(1.0), Asp(0.91), Ala(2.1). The presence of free homoserine is a confirmation of the Met-Met sequence. Carboxypeptidase A digestion of CNBr(50–61) for 24 h released 0.93 residue of homoserine and 0.32 residue of alanine. These results are consistent with the sequence given above.

T(67–75). Six dansyl-Edman steps gave the sequence Tyr-Ser-Glx-Val-Val-Pro. The Val-Val linkage was incompletely hydrolyzed after dansylation of the N-3 peptide and therefore Dns-Val-Val was observed on thin-layer polyamide chromatography (Hartley, 1970). The N-4 peptide yielded Dns-Val.

Carboxypeptidase B digestion for 1 h yielded 1.9 residues of lysine. The Glx in the third position from the N-terminus was shown to be glutamic acid by direct identification of

Pth-Glu. The sequence of T(67–75) is therefore Tyr-Ser-Glu-Val-Val-Pro-His-Lys-Lys.

T(76–83). The amino terminal residue of this peptide was found to be methionine by dansylation. Further dansyl-Edman sequencing gave ambiguous results. Carboxypeptidase B liberated 0.95 residue of lysine in 4 h of digestion suggesting that the second lysine in the peptide is not penultimate. The remainder of the sequence of this peptide was obtained with information from T(79–83) below.

T(79–83). This peptide had the composition of T(76–83) except that one less lysine, no methionine, and no histidine were present. Four dansyl-Edman determinations gave the sequence Asx-Phe-Leu-Glx and carboxypeptidase B digestion (1 h) released lysine. The peptide is acidic and therefore neither Asx nor Glx is amidated. The sequence of this peptide, Asp-Phe-Leu-Glu-Lys, suggests the sequence Met-His-Lys-Asp-Phe-Leu-Glu-Lys for peptide T(76–83) above. Additional confirmation of this sequence was obtained from peptides CNBr(77–118) and C(77–80) described below.

T(84–94). The N-terminal sequence Ile-Gly-Gly-Leu-Ser-Ala-Pro-Val was determined by the dansyl-Edman method. Hydrolysis of the peptide with dilute acid followed by purification by preparative peptide mapping yielded three major fragments: A(84–91), Ser(0.94), Pro(1.2), Gly(1.7), Ala(1.0), Val(1.1), Ile(0.80), Leu(1.2); A(92), Asp(1); A(93–94), Lys(0.96), Ala(1.0). The compositions of these fragments place the Asx between valine and Ala-Lys. The parent peptide was electrophoretically neutral, indicating that aspartic acid rather than asparagine occurs at this position. The complete sequence of T(84–94) is Ile-Gly-Gly-Leu-Ser-Ala-Pro-Val-Asp-Ala-Lys.

T(95–100). This peptide was sequenced by direct identification of the first four Pth-amino acids giving the sequence Asn-Val-Asp-Tyr. Carboxypeptidase A or B released only lysine. The sequence is therefore Asn-Val-Asp-Tyr-Cys-Lys.

T(101–108). Six Edman steps with Pth identification and seven dansyl-Edman determinations gave the sequence Glu-Trp-Leu-Val-Asn-His-Ile-Lys.

T(109–113). Four dansyl-Edman determinations on this electrophoretically neutral peptide resulted in the sequence Gly-Thr-Asp-Phe-Lys.

T(114–115). This dipeptide gave an N-terminal tyrosine upon dansylation, demonstrating the sequence Tyr-Lys.

T(116–118). Three dansyl-Edman determinations gave the sequence Gly-Lys-Leu for this tripeptide. The lack of a C-terminal lysine suggested that this peptide is the C-terminal peptide of the protein.

Chymotryptic Peptides. Fractions C-III, C-IV, and C-V from chromatography on Sephadex G-15 were further fractionated on Dowex 50-X2 and several of the resulting Dowex fractions were then subjected to preparative paper electrophoresis or preparative peptide mapping. Fractions C-I and C-II contained large peptides which were undoubtedly products of incomplete chymotryptic cleavage. None of these large peptides was successfully purified. Several small chymotryptic peptides were detected but not successfully purified from this chymotryptic digest. These peptides (C(15–17), C(63–67), C(74–76), and C(113–114)) were obtained from a separate chymotryptic digest using 1 μmol of carboxyamidomethylmyohemerythrin, and were purified by gel filtration and preparative peptide mapping.

C(1–8). Four Edman steps with direct identification of Pth derivatives gave the sequence Gly-Trp-Glu-Ile. Carboxypeptidase A digestion for 4 h liberated only tyrosine. This peptide thus constitutes the N-terminal portion of T(1–15) and by comparison with this tryptic peptide, has the sequence Gly-Trp-Glu-Ile-Pro-Glx-Pro-Tyr. The amide assignment for Glx was made upon sequencing cyanogen bromide peptide CNBr(1–61).

C(9–14). Four Edman steps with direct Pth identification gave the sequence Val-Trp-Asp-Glu. Carboxypeptidase A digestion for 4 h liberated only phenylalanine. The sequence of the peptide is therefore Val-Trp-Asp-Glu-Ser-Phe.

C(15–17). Dansylation of this peptide identified arginine as the N-terminus. Invoking chymotryptic specificity, the sequence of this tripeptide is Arg-Val-Phe. This chymotryptic peptide overlaps T(1–15) and T(16–27).

C(18–29). Twelve dansyl-Edman determinations gave the sequence Tyr-Glx-Glx-Leu-Asx-Glx-Glx-His-Lys-Lys-Ile-Phe. Carboxypeptidase A digestion for 10 min, 40 min, and 4 h confirmed the C-terminus as Lys-Ile-Phe. This peptide provides an overlap between T(16–27) and T(28–30).

C(30–33). Four dansyl-Edman identifications gave the sequence Lys-Gly-Ile-Phe and carboxypeptidase A digestion confirmed the Ile-Phe sequence. This peptide overlaps T(28–30) and T(31–37) but this assignment is ambiguous since, like T(28–30), peptides T(109–113) and T(114–115) have penultimate chymotrypsin-sensitive residues and C-terminal lysines. However, these latter two possibilities were eliminated by the inclusion of T(109–113) and T(114–115) in overlapping peptides C(103–112) and CNBr(77–118), respectively.

C(34–43). Nine dansyl-Edman determinations gave the sequence Cys-Asx-Ile-Arg-Asx-Asx-Ser-Ala-Pro. The C-terminal residue is Asx by difference, in fact identified as asparagine on the basis of the electrophoretic mobility of peptide T(38–49). This peptide overlaps T(31–37) and T(38–49).

C(44–47). Four dansyl-Edman analyses gave the sequence Leu-Ala-Thr-Leu. Carboxypeptidase A digests for 1 and 4 h confirmed the C-terminal Thr-Leu. This peptide is contained within T(38–49) and completes the sequence of T(38–49).

C(48–55). This peptide was completely sequenced as Val-Lys-Val-Thr-Thr-Asx-His-Phe by the dansyl-Edman procedure. Carboxypeptidase A liberated only phenylalanine after 4 h of digestion. This peptide is the overlap between T(38–49) and T(50–66).

C(63–67). The composition of this peptide identified it as an overlap between T(50–66) and either T(65–75) or T(114–115). The overlap was established unambiguously by peptide CNBr(63–76) as involving T(67–75).

C(68–73). This hexapeptide was entirely sequenced by the dansyl-Edman technique as Ser-Glx-Val-Val-Pro-His. This sequence is contained within T(67–75).

C(74–76). This tripeptide (Lys_2,Met) was purified by preparative peptide mapping but remained contaminated with the peptide C(115–118) Lys-Gly-Lys-Leu described below. Since it did not appear feasible to separate these peptides, they were sequenced simultaneously by the dansyl-Edman procedure. The amino acid composition and yield of the tripeptide (Table II) were based on the quantity of methionine in the fraction. Dansylation of the intact peptide mixture yielded only Dns-Lys. While the N-1 peptide gave Dns-Lys plus Dns-Gly, the N-2 peptide gave Dns-Lys and Dns-Met. These results suggest C(74–76) has the sequence Lys-Lys-Met. This provides an overlap between peptides T(67–75) and T(76–83). A confirmation of this overlap was obtained using CNBr(63–76) described below.

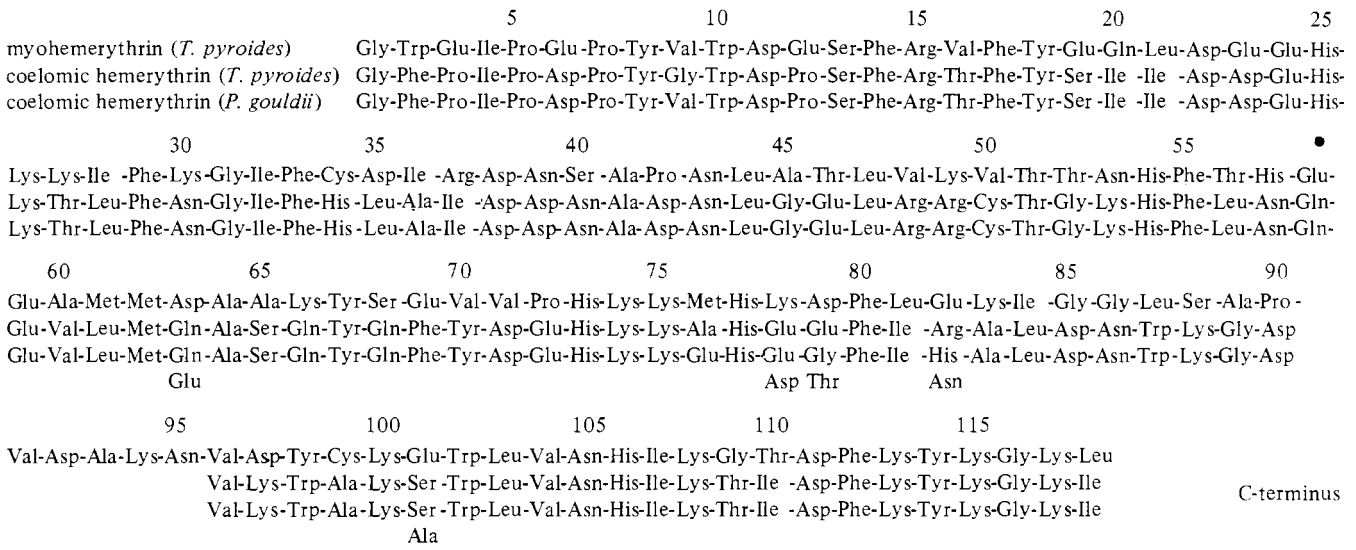

FIGURE 3: Amino acid sequence of *T. pyroides* myohemerythrin. The sequence of *P. gouldii* coelomic hemerythrins (Klippenstein, 1972b) and the proposed sequence of *T. pyroides* coelomic hemerythrin (Ferrell and Kitto, 1971) are included for comparison.

C(77–80). Four dansyl-Edman determinations gave His-Lys-Asx-Phe and carboxypeptidase A digestion confirmed the C-terminal phenylalanine. The Asx was determined to be aspartic acid by electrophoretic mobility. This peptide is contained within T(76–83).

C(81–87). The sequence Leu-Glx-Lys-Ile-Gly-Gly-Leu was obtained with dansyl-Edman steps and leucine confirmed as the C-terminus by carboxypeptidase A digestion. Glx is glutamic acid in this electrophoretically neutral peptide. This peptide is the overlap between T(76–83) and T(84–94).

C(88–98). The complete sequence of this peptide by the dansyl-Edman method was Ser-Ala-Pro-Val-Asx-Ala-Lys-Asx-Val-Asx-Tyr. The C-terminal tyrosine was confirmed by carboxypeptidase A digestion. This peptide overlaps tryptic peptides T(84–94) and T(95–100).

C(99–102). Four dansyl-Edman steps gave the sequence X-Lys-Glx-Trp. By difference, the N-terminal residue was identified as Cam-cysteine. This was confirmed by a dansyl-Edman degradation on a second preparation of peptide, giving the sequence Cam-Cys-Lys. Glx was found to be glutamic acid by electrophoretic mobility (neutral at pH 6.4). This peptide, Cam-Cys-Lys-Glu-Trp, overlaps peptide T(95–100) and T(101–108).

C(103–112). Carboxypeptidase A digestion for 4 h liberated only phenylalanine from this peptide while dansyl-Edman sequencing gave Leu-Val-Asx-His-Ile-Lys. The entire peptide, Leu-Val-Asx-His-Ile-Lys-(Gly, Thr, Asp)-Phe, is an overlap between T(101–108) and T(109–113).

C(113–114). This dipeptide was shown to be Lys-Tyr by dansylation. It provides an overlap between T(109–113) and T(114–115).

C(115–118). The sequence Lys-Gly-Lys-Leu was obtained by use of the dansyl-Edman method. Results of a carboxypeptidase A time study were consistent with this sequence. This peptide is the overlap between T(114–115) and T(116–118).

Cyanogen Bromide Peptides. Fragmentation with CNBr followed by fractionation on Sephadex G-50 yielded three peptides and free homoserine or homoserine lactone. A fourth fraction was obtained (CNBr-II) which was a mixture of fragments resulting from incomplete cleavage at one of the methionine residues.

CNBr(1–61). The largest fragment (CNBr-I) contained about a third of a residue of methionine in addition to 0.86 residue of homoserine + homoserine lactone. This result suggests the occurrence of the Met-Met sequence found in T(50–66) at the C-terminus of this fragment. The Edman degradation with direct Pth identification gave the sequence Gly-Trp-Glu-Ile-Pro-Glu at the N-terminus, the sequence contained in peptides T(1–15) and C(1–8).

CNBr(62). Free homoserine and its lactone were obtained from the CNBr fragmentation of the protein (fraction CNBr-V), this resulting from the Met-Met sequence at residues 61–62.

CNBr(63–76). Dansyl-Edman sequencing of CNBr-IV gave Asx-Ala-Ala-Lys-Tyr-Ser-Glx-Val-Val-Pro(His, Lys, Hse), an overlap between tryptic peptides T(50–66), T(67–75) and, assuming homoserine + homoserine lactone to be C-terminal, T(76–83).

CNBr(77–118). Direct identification of Pth derivatives was used with CNBr-III to obtain the N-terminal sequence His-Lys-Asp-Phe-Leu-Glu-Lys-Ile-Gly-Gly. This peptide contains no homoserine or homoserine lactone and therefore is the C-terminal CNBr fragment.

The overall order of CNBr fragments was established by overlap of CNBr(1–61), CNBr(62), and CNBr(63–76) by T(50–66); the overlap of CNBr(63–76) and CNBr(77–118) by T(76–83) and C(74–76). CNBr(77–118) was placed at the C-terminus by the absence of homoserine or its lactone.

The complete amino acid sequence of myohemerythrin is shown in Figure 3.

Discussion

The primary structure of *T. pyroides* myohemerythrin differs from the coelomic hemerythrins of *T. pyroides* (Ferrell and Kitto, 1971) and *P. gouldii* (Klippenstein et al., 1968) at 60 and 59 residue positions, respectively (Figure 3). In addition, myohemerythrin has a chain length of 118 rather than 113 residues. The alignment of homologous sequences suggests that the five extra residues in myohemerythrin are located between residues 90 and 91 of the coelomic hemerythrins, although this assignment is by no means certain. Homology between myohemerythrin and *P. gouldii* coelomic hemerythrin is higher near the C-terminus (79% homologous, residues 100–118), and N-terminus

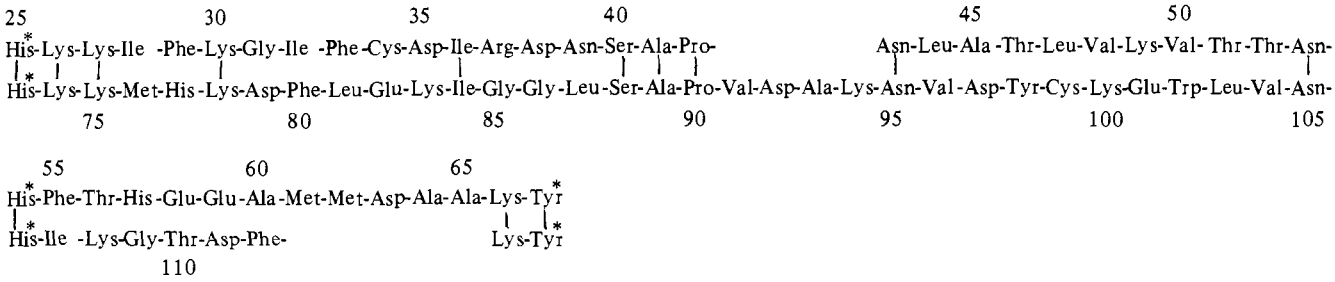

FIGURE 4: Alignment of residues 25-67 with residues 73-114 in myohemerythrin. Proposed iron ligand residues are designated by asterisks. Homologous residues are connected by lines.

(65% homologous, residues 1-26) and lower for the remainder of the chain (30%).

The contrast between the extent of homology between these two coelomic hemerythrins from different species (96%) and myohemerythrin and coelomic hemerythrin of the same species (45%) is striking. As in the case of the hemoglobins and myoglobins, the location and specialized functions of the respective blood and muscle oxygen carriers are reflected in specialized primary structures. However, it is likely that the similarity between *T. pyroides* and *P. gouldii* coelomic hemerythrins is not typical since another coelomic hemerythrin, that of the sipunculid *Phascolosoma agassizii,* is substantially different from these hemerythrins in both quaternary structure (Liberatore et al., 1974) and primary structure (Liberatore, 1974).

The identification of iron-linked amino acid side chains in the hemerythrins is one objective of this work. While tyrosine residues 8, 114, and 67 have been implicated as iron ligands (York and Fan, 1971; Rill and Klotz, 1971) four histidine residues postulated to be iron-linked (Fan and York, 1969) remain unidentified. Histidines which occur at positions 34 and 82 in *P. gouldii* hemerythrin are replaced by cysteine and glutamic acid, respectively, in myohemerythrin and thus cannot be iron ligands. Histidine-82 has been excluded as an iron ligand on the basis of its absence in the primary structures of *P. gouldii* hemerythrin B (Klippenstein, 1972b) and *T. pyroides* coelomic hemerythrin (Ferrell and Kitto, 1971). The cysteine at residue 34 cannot be linked because it is possible to block both cysteines in myohemerythrin with *p*-hydroxymercuribenzoate without apparent iron loss (Klippenstein et al., 1972). Thus, histidines-25, -54, -73, -77, and -106 are found in all hemerythrins examined to date.

Tyrosines-8, -18, -67, and -114 are found in all hemerythrins examined whereas tyrosine-70 is replaced by valine in myohemerythrin. Tyrosine-18 or -70 may be involved in intersubunit contacts in octameric hemerythrins (Rill and Klotz, 1971). Chemical modification results have consistently excluded tyrosine-18 as an iron ligand (Rill and Klotz, 1971; York and Fan, 1971), but its invariance suggests some important functional role.

While methionine-62 is found in all hemerythrins sequenced and could be considered a potential iron ligand, chemical modification experiments (Morrissey, 1971) have revealed that the single methionine in *P. gouldii* hemerythrin can be modified with iodoacetamide without loss of iron. Methionine-62 must therefore be preserved in hemerythrins for a role other than as an iron-binding group.

Histidines-25, -54, -73, -77, and -106, and tyrosines-8, -67, and -114 remain as possible iron ligands. X-ray crystallography of myohemerythrin at 5.5 Å resolution, interpreted with the aid of the primary structure of the molecule, has revealed the course of the polypeptide chain and some features of the active center (Hendrickson et al., 1975). Tyrosine-8 and histidine-77 are some distance from the iron site while tyrosines-67 and -114 and histidines-25, -54, -73, and -106 have electron density connections with the dimeric iron center.

Invariant histidine-77 and tyrosines-8 and -18, while not linked to iron in azidometmyohemerythrin, may function in oxygen binding. Two of these residues could serve as the potentially basic groups postulated by Garbett et al. (1971) to be coordinated to iron in deoxyhemerythrins while the third might act to stabilize the bridging hydroperoxo- group in oxyhemerythrins in a manner which results in the observation of two iron environments in this form of the protein (York and Bearden, 1970; Garbett et al., 1971).

The primary structure of myohemerythrin contains an apparent sequence repeat within the chain which is also manifest in an apparent pseudosymmetry in the tertiary structure of the protein (Hendrickson et al., 1975). The segment containing residues 25-54 is 32% homologous with the segment 73-106 if a four residue gap is left after residue 42 to accommodate extra residues following position 90. These segments can also be extended to residues 67 and 114, respectively, and retain 32% homology if a five-residue gap is left after residue 112 (Figure 4). This alignment places in register suspected iron ligands histidine-25 and histidine-73, histidine-54 and histidine-106, and tyrosine-67 and tyrosine-114. It also places in register the identical sequences, Ser-Ala-Pro, at residues 40-42 and 88-90 corresponding to the approximate positions of corners between helical segments A and B and C and D, respectively (Hendrickson et al., 1975). A similar alignment of large segments is also possible in *P. gouldii* coelomic hemerythrin, but this protein lacks the Ser-Ala-Pro repeat.

Computerized tests for internal duplication in myohemerythrin using unitary matrix and mutation data matrix approaches (Dayhoff, 1972) showed that sections of the molecule have an unusually large number of identities. Comparison of fragments 30 residues long resulted in an alignment score of 3.7 using the unitary matrix. However, the mutation data matrix gave an alignment score of only 0.44, indicating that many residues in these 30-residue fragments are quite dissimilar. It seems likely that since many of the homologous residues in the segments described above have essential functional roles in iron binding and formation of tertiary structure, they may have been conserved, while other, less essential residues changed drastically as a result of mutational events. Thus intragenic duplication, though less apparent than in the analogous case of bacterial ferredoxins (Tanaka et al., 1966), would seem a plausible explanation for the similarities in primary and tertiary structures of these segments. As in the case of the bacterial ferredoxins,

the sequence repeat is best visualized as a specific arrangement of iron ligands which has undergone duplication. It is also interesting, though possibly fortuitous, that this alignment of myohemerythrin segments places the histidine iron ligands 29 residues apart, exactly the spacing of proximal and distal histidines in myoglobin.

Differences in structure between myohemerythrin and coelomic hemerythrins should contribute to understanding of other properties of the hemerythrins as well. For example, myohemerythrin has Val-Lys-Val replacing Arg-Arg-Cys at residues 48–50, demonstrating that Cys-50 is not essential to the oxygen-binding function of this molecule. The Arg-Arg sequence has been considered a possible binding site for ClO_4^- (Garbett et al., 1971; DePhillips, 1971) and Cl^- (Rao and Keresztes-Nagy, 1973). It would be interesting, therefore, to know whether myohemerythrin binds these anions.

References

Crestfield, A. M., Moore, S., and Stein, W. H. (1963), *J. Biol. Chem. 238,* 622.

Dayhoff, M. O. (1972), Atlas of Protein Sequence and Structure, Vol. 5, Washington, D.C., National Biomedical Research Foundation, p 101.

DePhillips, H. A. (1971), *Arch. Biochem. Biophys. 144,* 122.

Fan, C. C., and York, J. L. (1969), *Biochem. Biophys. Res. Commun. 36,* 365.

Fan, C. C., and York, J. L. (1972), *Biochem. Biophys. Res. Commun. 47,* 472.

Ferrell, R. E., and Kitto, G. B. (1971), *Biochemistry 10,* 2923.

Garbett, K., Johnson, C. E., Klotz, I. M., Okamura, M. Y., and Williams, R. J. P. (1971), *Arch. Biochem. Biophys. 142,* 574.

Gray, W. R. (1967), *Methods Enzymol. 11,* 469.

Groskopf, W. R., Holleman, J. W., Margoliash, E., and Klotz, I. M. (1966), *Biochemistry 5,* 3783.

Hartley, B. S. (1970), *Biochem. J. 119,* 805.

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

Hirs, C. H. W. (1967), *Methods Enzymol. 11,* 325.

Jeppsson, J., and Sjöquist, J. (1967), *Anal. Biochem. 18,* 264.

Klippenstein, G. L. (1972a), *Biochem. Biophys. Res. Commun. 49,* 1474.

Klippenstein, G. L. (1972b), *Biochemistry 11,* 372.

Klippenstein, G. L., Holleman, J. W., and Klotz, I. M. (1968), *Biochemistry 7,* 3868.

Klippenstein, G. L., Van Riper, D. A., and Oosterom, E. A. (1972), *J. Biol. Chem. 247,* 5959.

Liberatore, F. A. (1974), Ph.D. Dissertation, University of New Hampshire, Durham, N.H.

Liberatore, F. A., Truby, M. F., and Klippenstein, G. L. (1974), *Arch. Biochem. Biophys. 160,* 223.

Margoliash, E., and Smith, E. L. (1962), *J. Biol. Chem. 237,* 2151.

Morrissey, J. A. (1971), Ph.D. Dissertation, University of New Hampshire, Durham, N.H.

Peterson, J. D., Nehrlich, S., Oyer, P. E., and Steiner, D. F. (1972), *J. Biol. Chem. 247,* 4866.

Rao, A. L., and Keresztes-Nagy, S. (1973), *Biochim. Biophys. Acta 313,* 249.

Rill, R. L., and Klotz, I. M. (1970), *Arch. Biochem. Biophys. 136,* 507.

Rill, R. L., and Klotz, I. M. (1971), *Arch. Biochem. Biophys. 147,* 226.

Schroeder, W. A. (1967), *Methods Enzymol. 11,* 351.

Tanaka, M., Tadayoshi, N., Benson, A., Mower, H., and Yasunobu, K. T. (1966), *Biochemistry 5,* 1666.

York, J. L., and Bearden, A. J. (1970), *Biochemistry 9,* 4549.

York, J. L., and Fan, C. C. (1971), *Biochemistry 10,* 1659.