

Amino Acid Sequence of Hemerythrin from *Themiste dyscritum**

(Received for publication, January 30, 1978)

Joann S. Loehr,‡ Peter J. Lammers,‡ Bernadine Brimhall,§ and Mark A. Hermodson¶ ||

From the ‡Department of Chemistry, Portland State University, Portland, Oregon 97207, the §Department of Biochemistry, University of Oregon Health Sciences Center, Portland, Oregon 97201, and ¶The Division of Medical Genetics, University of Washington, Seattle, Washington 98195

The amino acid sequence of hemerythrin from the sipunculid worm, *Themiste dyscritum*, was determined by sequenator analyses of the S-pyridylethylated protein and fragments derived by further chemical and enzymatic cleavages. The fragments were obtained by cleavage of the intact protein with hydroxylamine, trypsin digestion of citraconylated intact protein, and subdigestion with *Staphylococcal protease* V8. The COOH-terminal sequence was determined using carboxypeptidases A and B and amino acid analyses. The polypeptide chain was found to contain 113 amino acids. Since heterogeneity was observed at no more than two positions in the amino acid sequence, the native octameric protein appears to be composed of identical subunits. By combining information derived from sequence analyses and x-ray crystallographic studies, it has been possible to identify amino acids responsible for the tertiary and quaternary structure of the protein as well as amino acids serving as iron ligands at the oxygen-binding site.

Hemerythrin is a respiratory protein found in the erythrocytes of certain marine invertebrates. The oxygen-binding site contains 2 iron atoms which, in contrast to hemoglobin, are coordinated only to amino acid residues. Although the physical and chemical properties of the protein have been thoroughly investigated (1), the 3-dimensional structure of the molecule and the nature of the iron complex at the active site are just beginning to be elucidated.

Hemerythrins from several species of sipunculid worms have recently been investigated by x-ray crystallography. The structure of hemerythrin from the coelomic fluid of *Themiste dyscritum* has been obtained at 2.8 Å resolution (2), while the structures of coelomic hemerythrin from *Phascolopsis gouldii* and myohemerythrin from the retractor muscles of *Themiste pyroides* are based on 5.5 Å resolution x-ray data (3). As a result of these studies, two different models have been proposed for the coordination of iron in hemerythrin (2, 3). Spectroscopic comparisons indicate that the active site structures of these hemerythrins must be very similar, making it unlikely that both of the crystallographic interpretations are correct (4). Resolution of this question will require more definitive x-ray data and more accurate amino acid sequence information.

* This research was supported by Grants GM 18865 (J.S.L.), GM 15253 (M.A.H.), and AM 17850 (B.B.) from the United States Public Health Service. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

|| Present address, Department of Biochemistry, Purdue University, West Lafayette, Ind. 47907.

Complete amino acid sequences have been reported for *P. gouldii* hemerythrin (5) and *T. pyroides* myohemerythrin (6) and were utilized in solving the respective crystal structures (2, 3). The interpretation of the 2.8 Å resolution electron density map of *T. dyscritum* hemerythrin (2) was based on the amino acid sequence of *P. gouldii* hemerythrin. The higher resolution study identified the following amino acid residues from *P. gouldii* hemerythrin as iron ligands: His 25, His 54, Gln 58, His 73, His 77, His 101, Asp 106, and Tyr 109. In order to evaluate the proposed active site structure for *T. dyscritum* hemerythrin, it is necessary to determine whether *T. dyscritum* hemerythrin contains these same residues in its amino acid sequence.

The present communication describes the determination of the complete amino acid sequence of hemerythrin from *T. dyscritum*. This protein has been shown to be composed of eight subunits, each having a molecular weight of approximately 13,000 (7). Only one size of subunit was observed using polyacrylamide gel electrophoresis in sodium dodecyl sulfate (7). However, this did not rule out the existence of molecular variants such as have been observed for a number of other coelomic hemerythrins (3). The present work shows that the *T. dyscritum* hemerythrin monomer has a molecular weight of 13,400 and is heterogeneous at one or more positions in its amino acid sequence, indicating the presence of at least two protein components.¹

METHODS

Themiste dyscritum sipunculid worms were provided by the Oregon Institute of Marine Biology in Charleston, Oregon. Hemerythrin was isolated from the coelomic fluid using the method described by Klotz *et al.* (8), with a modified crystallization procedure. The hemerythrin solution from lysed red blood cells was concentrated to greater than 20 mg/ml and dialyzed against solutions of decreasing ionic strengths starting with 0.05 M KCl, 0.01 M Tris-Cl (pH 7.5). The KCl concentration was lowered slowly until crystals appeared. Crystalline hemerythrin was then redissolved in 1.0 M KCl, 0.01 M Tris-Cl (pH 7.5) at a concentration less than 5 mg/ml. The apoprotein was prepared by dialysis against 0.02 M HCl to remove iron, followed by lyophilization. The concentration of apoprotein in acid solution was determined spectrophotometrically using $\epsilon_{280} = 17.7$. This extinction coefficient was obtained by drying the protein to constant weight followed by dissolution in 9% formic acid.

The apoprotein was reduced and S-pyridylethylated with dithiothreitol and 4-vinylpyridine (Sigma, redistilled) according to the

¹ Some of the methods and data are presented as a miniprint supplement immediately following this paper. Unmodified methods, Figs. 3 to 5, and Tables I to III are found on pp. 5730-5731. Miniprint can be easily read with the aid of a standard magnifying glass. Full size photocopies are available from The Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, Md. 20014. Orders should specify the authors, JBC Document Number 78M-149, and the number of copies desired and must be accompanied by a remittance to the order of the Journal in the amount of \$1.50 per set of photocopies.

procedure of Hermodson *et al.* (9). Following the reaction, the solution was acidified to pH 2 with 88% formic acid, and the reagents and salts were removed by dialysis against 9% formic acid. S-Pyridylethylated hemerythrin was used for all of the experiments except the carboxypeptidase digestions.

RESULTS

The complete amino acid sequence of hemerythrin from *Themiste dyscritum* is presented in Fig. 1. The sequence was determined by sequenator analyses of the intact protein and four peptide fragments and by exopeptidase degradation from the COOH-terminal of the protein. One of the peptide fragments was obtained by chemical cleavage of an Asn-Gly bond. The other three were isolated after enzymatic cleavage at arginyl or glutamyl bonds. The methods used to generate the fragments and the details of the sequence analyses are summarized in Fig. 2 and Table I. The amino acid composition of the intact protein (Table II) agrees very well with the experimentally determined sequence.

NH₂-terminal Sequence—In the sequenator analysis of the intact protein, the first 33 amino acids were identified and were shown to end in the sequence Asn-Gly-Ile-Leu (Table I). Heterogeneity was observed at position 21 in the sequence with valine and isoleucine appearing in approximately equal amounts in the gas chromatograms.

Sequencing of Hydroxylamine Fragment—The presence of an asparaginyl-glycyl bond between residues 30 and 31 which was susceptible to cleavage with NH₂OH provided a useful peptide for continuing the sequence determined with the intact protein. Digestion with NH₂OH resulted in a fragment, designated N-1, which could be separated from unreacted protein by gel filtration (Fig. 3A). Fragment N-1 was further purified by rechromatography (Fig. 3B) and its NH₂-terminal sequence was determined to be Gly-Ile-Leu (Table I). This established that the N-1 fragment began with residue 31. The

high molecular weight of fragment N-1, as evidenced by its elution shortly after intact protein (Fig. 3), indicated that it probably contained residues 31 to 113. Additional sequenator analysis of this peptide identified 22 residues (between 31 and 53) ending in the sequence Arg-Arg-Cys-Thr-Gly-Lys. Residue 36 could not be identified by sequenator analysis, but it was assigned to serine on the basis that the amino acid composition of peptide 16-48 (Table II) showed the presence of 1 serine residue. Residue 36 was the only unidentified position between 16 and 48 and there was no evidence for serine in any of the other residues.

Sequencing of Tryptic Fragments—Cleavage after arginyl residues was effected by trypsin digestion of S-pyridylethyl-hemerythrin containing citraconylated lysyl residues. Gel filtration on Sephadex G-50 superfine yielded three fractions designated T-1, T-2, and T-3 (Fig. 4). These three fragments were easily distinguished from one another and the starting material by their characteristic ultraviolet absorption spectra resulting from different contents of tryptophan, tyrosine, and S-pyridylethylcysteine.

Sequenator analysis of Fraction T-1 began with Cys-Thr-Gly-Lys (Table I), establishing an overlap with the last 4 residues (50 to 53) sequenced in the N-1 fragment. This placed the beginning of the T-1 peptide at residue 50. Extended analysis of the T-1 fragment identified the sequence from residues 50 to 73, ending in Ala-Glu-His. The amino acid composition of fragment T-1 (Table II) agrees well with the proposed sequence for residues 50 to 110 (Fig. 1), supporting the identification of T-1 as peptide 50-110. The composition indicates 1.5 residues of serine which were not detected in the sequenator analysis. One serine residue is very likely located at position 65 (at which no amino acid was identified) and some serine may be present at position 64 (at which less than 1 residue of alanine was detected). The possibility of heterogeneity at position 64 is supported by the low alanine content

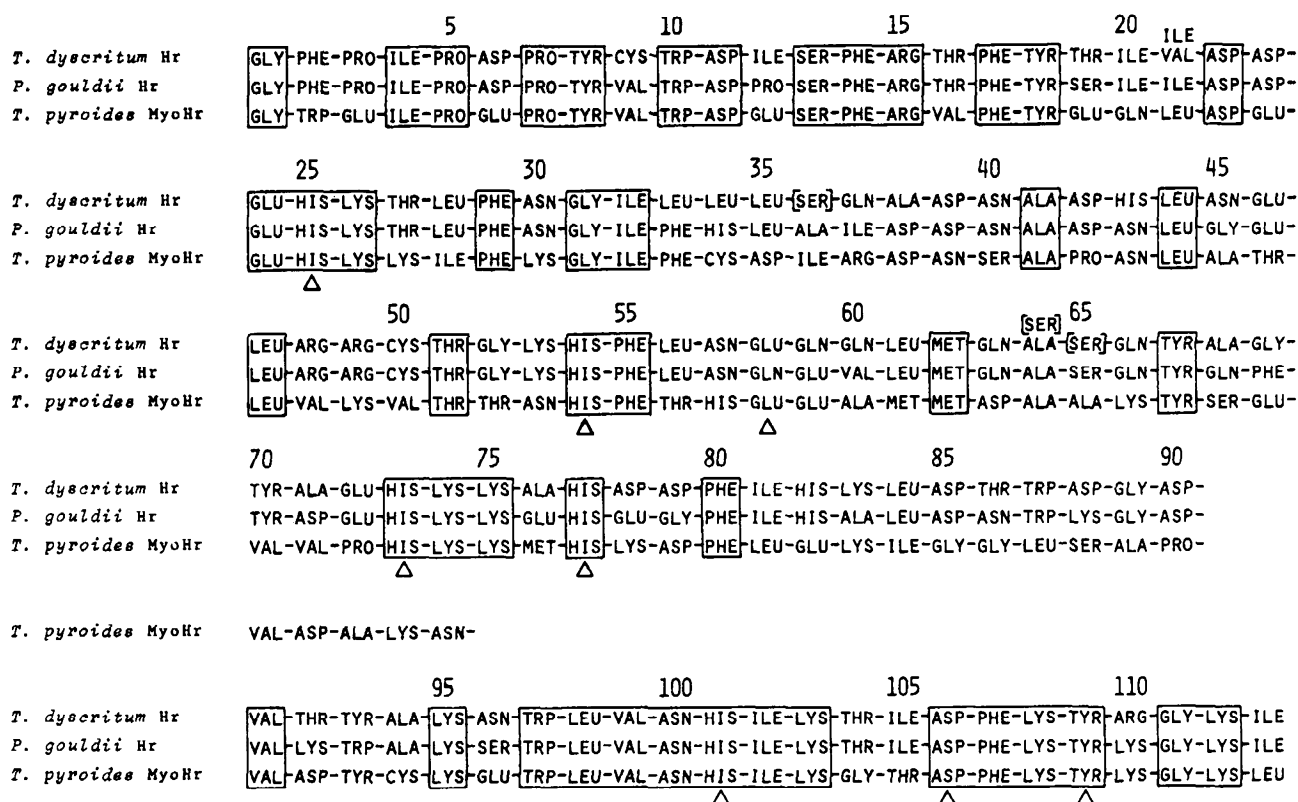


FIG. 1. Amino acid sequence of *T. dyscritum* hemerythrin (Hr). The sequences of *P. gouldii* hemerythrin, major component (5, 10), and *T. pyroides* myohemerythrin (6) are included for comparison. Invariant residues are enclosed by boxes. Residues implicated as iron ligands in crystallographic studies of *T. dyscritum* hemerythrin (2) are indicated by Δ .

Amino Acid Sequence of Hemerythrin

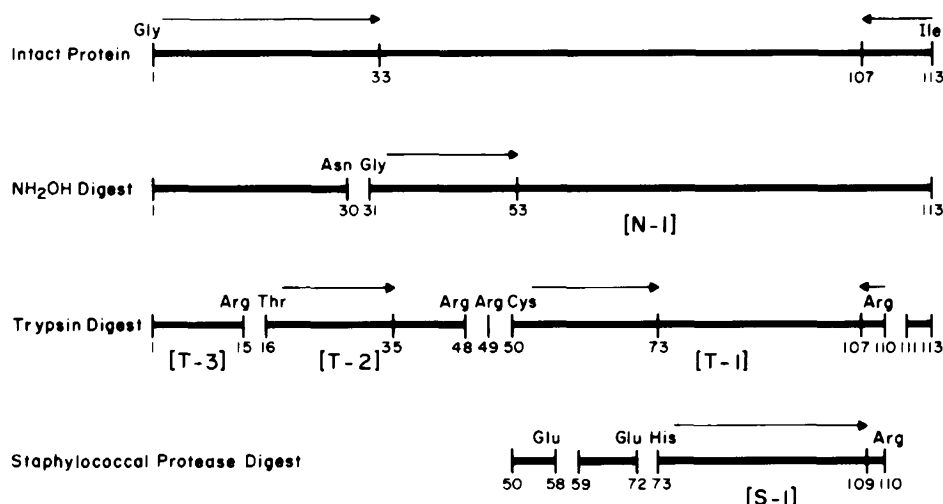


FIG. 2. Diagrammatic representation of intact hemerythrin and peptide fragments utilized for sequence analysis. Arrows indicate regions sequenced by automated Edman degradation (→) and by carboxypeptidases A and B (←). The amino acid residues are numbered according to the proposed sequence (Fig. 1).

of the peptide (Table II) and the fact that considerably greater amounts of alanine were detected for residues 68 and 71 than for 64 (Table I).

Sequenator analysis of fragment T-2 (Table I) revealed it to be the peptide generated by cleavage after arginine 15 and arginine 48. The sequence from 16–35 was identical to that obtained from the sequenator analyses of the intact protein and hydroxylamine fragment, N-1, described above. The amino acid composition of fragment T-2 (Table II) shows excellent agreement with the proposed sequence for residues 16 to 48.

The fragment designated T-3 was not sequenced directly. However, its amino acid composition agreed well with that expected for peptide 1–15 (Table II).

Sequencing of Staphylococcal Protease Fragment—Comparison of the amino acid analysis of fragment T-1 (Table I) with the sequence information indicated that most or all of the glutamic acid occurred prior to residue 73. It was reasoned that treatment of fragment T-1 with staphylococcal protease which cleaves after glutamic acid might generate a long fragment beginning with residue 73 which could be separated from the smaller pieces, 50 to 58 and 59 to 73. Chromatography on Sephadex G-50 superfine revealed several fragments as shown in Fig. 5. Sequenator analysis on the major fragment, S-1, identified 37 residues starting with the sequence His-Lys-Lys and ending in Phe-Lys-Tyr (Table I), placing the beginning of this peptide at histidine 73 and establishing the sequence through residue 109.

Although only one sequenator analysis was performed on residues 70 to 73 in fragment T-1, there is additional evidence for the amino acid assignments and placement of fragments in this region. The generation of fragment S-1 by staphylococcal protease digestion supports the Glu-His assignment for residues 72 and 73, the amino acid composition for fragment T-1 corresponds with the sequence information for fragments T-1 and S-1 (Table II), and the sequence for residues 72 to 75 is identical to that in *P. gouldii* hemerythrin (Fig. 1).

COOH-terminal Sequence—Sequenator analysis of fragment S-1 ended with residues phenylalanine (107), lysine (108), and tyrosine (109). Carboxypeptidase digestion of both intact hemerythrin and tryptic peptide T-1 was utilized in order to establish the sequence of the last 4 residues, 110 to 113. The reaction conditions and results are shown in Table III.

Digestion of the intact protein with carboxypeptidase A released only the terminal residue, isoleucine, and a small amount of the penultimate residue, lysine (Table III). The intact protein was next digested with a mixture of carboxy-

peptidase A and carboxypeptidase B. The rate of release of amino acids (Table III) confirmed the order of the last 2 residues as Lys-Ile and indicated the presence of arginine and glycine between residues 109 and 112. The higher value for lysine than isoleucine in these digests is due to the presence of the additional lysine at 108.

To determine the order of arginine and glycine residues at 110 and 111 we took advantage of the fact that the tryptic fragment T-1 must end in arginine. Hence, digestion of T-1 with carboxypeptidase B would release no glycine if the sequence in the intact protein was Arg-Gly, but would release glycine if the sequence was Gly-Arg. The former case proved correct (Table III), identifying the final COOH-terminal sequence as Tyr-Arg-Gly-Lys-Ile.

DISCUSSION

The strategy for determining the amino acid sequence of hemerythrin was based on the isolation and automated Edman degradation of large peptides (Fig. 2). Although sequencing the NH₂-terminal of the intact protein provided good information through residue 33, proceeding further into the protein gave some difficulties. The tryptic fragment T-2 (residues 16 to 48) could not be sequenced beyond residue 35 due to an extensive yield drop at glutamine 37. A similar problem arose in sequencing tryptic fragment T-1 (residues 50 to 110) because the 4 glutamine residues between position 59 and 66 made it difficult to reach residue 73 for an overlap. Continuing past residue 35 was made possible by the use of hydroxylamine cleavage fragment N-1 (residues 31 to 113). However, this peptide was more difficult to isolate by gel filtration due to the incompleteness of the NH₂OH cleavage and the overlap of this peptide with intact protein. Staphylococcal protease fragment S-1 (residues 73 to 110) provided good yields on the sequenator through its penultimate residue, 109. The amino acid analyses showed excellent agreement with peptide compositions expected from sequencing. A slight discrepancy in proline content was observed for the intact protein. However, this is probably due to an error in the amino acid analysis as no evidence for a fourth proline was obtained in any of the other amino acid or sequence analyses.

The amino acid sequence of *T. dyscritum* hemerythrin is shown in Fig. 1 along with the sequences of *P. gouldii* hemerythrin and *T. pyroides* myohemerythrin. Residues which are invariant in all three proteins are enclosed in boxes. While the degree of sequence identity between coelomic hemerythrins and myohemerythrin is only 42%, the proteins appear to have very similar tertiary structures as judged by x-ray crystallog-

raphy (11, 12). The electron density maps of *T. dyscritum* hemerythrin at 5 Å resolution (12) show peptide backbone structures and conformations which are very close to those observed in *T. pyroides* myohemerythrin at 5.5 Å resolution (11).

Both myohemerythrin and the two coelomic hemerythrins have as their major structural feature four approximately parallel sections of α helix accounting for about 70% of the amino acid residues (3, 12). The first helix begins near residue 20 and the fourth helix ends at the COOH-terminal of the molecule. Although the NH₂-terminal appears to have a less-ordered structure, its amino acid sequence is highly conserved. Thus, the structural analogies in the three proteins appear to be present throughout. The fact that myohemerythrin exists as a monomer in solution whereas the two coelomic hemerythrins are octamers must be due to differences in surface amino acids rather than to differences in protein conformation. The relationship between hemerythrin and myohemerythrin is reminiscent of the relationship between hemoglobin and myoglobin. In the latter case, the proteins have similar tertiary structures containing about 70% α helix, but they are only about 20% homologous in amino acid sequence (13). The muscle protein, myoglobin, is monomeric while the blood protein, hemoglobin, is tetrameric.

The largest number of sequence homologies between hemerythrin and myohemerythrin are observed in the NH₂-terminal region between residues 1 and 18, the COOH-terminal region between residues 95 and 112, and in the regions containing residues implicated as iron ligands (Fig. 1). The high degree of amino acid conservation in the NH₂- and COOH-terminal regions appears to be due to the importance of these two regions in maintaining the tertiary structure of the molecule. Examination of a 2.8 Å resolution, computer-averaged electron density map for *T. dyscritum* hemerythrin (14) indicates that many of the conserved amino acids in these two regions have their nonpolar portions buried inside the molecule. Extending inward from the NH₂-terminal region are phenylalanine 2, tyrosine 8, tryptophan 10, phenylalanine 14, and tyrosine 18. Residues 8 and 14 appear to interact with phenylalanine 29. Residue 2 is close to the carbon chain of lysine 103 and residue 18 points toward the peptide backbone between residues 111 and 113.

The two coelomic hemerythrins show a much higher degree of sequence homology with one another than with myohemerythrin, particularly between residues 19 and 66 (Fig. 1). Analysis of the computer-averaged electron density map for *T. dyscritum* hemerythrin reveals that the two helices in this region of the molecule provide the major points of subunit contact for octamer formation (14). Electrostatic and hydrogen-bonding interactions across one of the 2-fold crystallographic axes involve arginine 15, threonine 19, and aspartic acid 23 of one subunit and aspartic acid 42, arginine 49, and lysine 53, respectively, of an adjacent subunit. Interactions between subunits related by the 4-fold crystallographic axes include arginine 48 with glutamine 66 (peptide oxygen) and lysine 53 (peptide nitrogen) with serine 65. The amino acids at these positions are sufficiently conserved in *P. gouldii* hemerythrin to give rise to the same electrostatic interactions and, thus, the common octameric structure. The same amino acid positions are sufficiently changed in myohemerythrin to eliminate these salt bridges, thus explaining its stability as a monomeric species.

Based on the reported amino acid sequences, seven of the amino acids proposed as iron ligands in *T. dyscritum* hemerythrin are completely conserved in the three proteins (Fig. 1), but there is a lack of agreement for the amino acid at

position 58. In the crystallographic studies of *T. dyscritum* hemerythrin, residue 58 appears to be bridging the 2 iron atoms (2). The glutamic acid at position 58 in *T. dyscritum* hemerythrin fits the role of a bridging ligand better than the glutamine at position 58 in *P. gouldii* hemerythrin. Since spectroscopic studies indicate that the active site structures of the two hemerythrins are conserved (4) and since *T. pyroides* myohemerythrin also contains a glutamic acid at position 58, it seemed possible that residue 58 was mistakenly identified as glutamine in *P. gouldii* hemerythrin. The original determination of glutamine and glutamic acid at positions 58 and 59, respectively, in *P. gouldii* hemerythrin was based on the resistance of residue 59 to digestion by carboxypeptidase A (15). A reinvestigation by automated sequenator analysis indicates that the correct sequence in *P. gouldii* hemerythrin is glutamic acid at 58 and glutamine at 59.² Thus, all eight ligand positions are conserved.

The amino acid sequence of *T. dyscritum* hemerythrin substantiates the crystallographic model for the iron-binding site of that protein in that the amino acids observed at the proposed iron ligand positions fit well with the proposed binding of the 2 iron atoms in a face-sharing biocubane (2). The amino acids identified by sequence analysis correspond to two carboxylates coordinated as bridging ligands, 3 histidine coordinated to one of the iron atoms, and 2 histidines and 1 tyrosine coordinated to the 2nd iron atom at the oxygen-binding site. As further refinements of the crystallographic data become available, comparison with the amino acid sequences should yield additional information about the structures of hemerythrin and myohemerythrin.

Acknowledgments—*Themiste dyscritum* were kindly supplied by Robert C. Terwilliger (Oregon Institute of Marine Biology) and identified as such by Leonard Simpson (Portland State University). Saradell Poddar, Suzanne E. Clarke, and Patricia M. Gormley at Portland State and Walter Mahoney (University of Washington) provided valuable assistance in performing this research. We thank Richard T. Jones and John A. Black (University of Oregon Health Sciences Center) for their help in the design of experiments. We are especially grateful to Lyle H. Jensen, Larry C. Sieker, and Ronald E. Stenkamp (University of Washington) for access to their crystallographic data and for many enlightening discussions. We also thank Thomas M. Loehr and Carol and Jerome Doherty for their involvement in this work.

REFERENCES

- Okamura, M. Y., and Klotz, I. M. (1973) in *Inorganic Biochemistry* (Eichhorn, G. L., ed) Vol. 1, pp. 320-343, Elsevier, Amsterdam
- Stenkamp, R. E., Sieker, L. C., and Jensen, L. H. (1976) *Proc. Natl. Acad. Sci. U. S. A.* **73**, 349-351
- Klotz, I. M., Klippenstein, G. L., and Hendrickson, W. A. (1976) *Science* **192**, 335-344
- Dunn, J. B. R., Addison, A. W., Bruce, R. E., Loehr, J. S., and Loehr, T. M. (1977) *Biochemistry* **16**, 1743-1749
- Klippenstein, G. L., Holleman, J. W., and Klotz, I. M. (1968) *Biochemistry* **7**, 3868-3878
- Klippenstein, G. L., Cote, J. L., and Ludlam, S. E. (1976) *Biochemistry* **15**, 1128-1136
- Loehr, J. S., Meyerhoff, K. N., Sieker, L. C., and Jensen, L. H. (1975) *J. Mol. Biol.* **91**, 521-522
- Klotz, I. M., Klotz, T. A., and Feiss, H. A. (1957) *Arch. Biochem. Biophys.* **68**, 284-299
- Hermanson, M. A., Ericsson, L. H., Neurath, H., and Walsh, K. A. (1973) *Biochemistry* **12**, 3146-3153
- Klippenstein, G. L. (1972) *Biochemistry* **11**, 372-380
- Ward, K. B., Hendrickson, W. A., and Klippenstein, G. L. (1975) *Nature* **257**, 818-821

² P. M. Gormley, J. S. Loehr, and M. A. Hermanson, unpublished results.

12. Stenkamp, R. E., Sieker, L. C., Jensen, L. H., and Loehr, J. S. (1976) *J. Mol. Biol.* **100**, 23-34
13. Dayhoff, M. O. (1969) *Atlas of Protein Sequence and Structure*, Vol. 4, pp. 17-24, National Biomedical Research Foundation, Bethesda, Md.
14. Stenkamp, R. E., Sieker, L. C., and Jensen, L. H. (1978) *Biochemistry* **17**, 2499-2505
15. Subramanian, A. R., Holleman, J. W., and Klotz, I. M. (1968) *Biochemistry* **11**, 3859-3867

SUPPLEMENTAL MATERIAL TO
AMINO ACID SEQUENCE OF HEMERYTHRIN
FROM *Thomiste dyscritum*

by

Joann S. Loehr, Peter J. Lambers, Bernie Brimhall, and Mark A. Hermodson

METHODS

Preparation and Isolation of Peptides. Cleavage of asparaginyl-glycyl bonds was accomplished by treatment of S-pyridylethylated (PE) hemerythrin with NH_4OH (15). Solutions containing 5 mg/ml of protein in 2 M NH_4OH , 6 M guanidine-HCl (mp 182-184°) were incubated at 45° with vigorous stirring and maintenance of the pH at 9.0 by dropwise addition of 4.5 M LiOH. After four hours the mixture was acidified to pH 2.0 with 88% formic acid and dialyzed versus 9% formic acid to desalt the peptides. The lyophilized peptides were separated by gel filtration on Sephadex G-75 Superfine (10-40 μ) in 9% formic acid.

Cleavage after arginyl bonds was effected with trypsin after blocking the lysyl residues with citraconyl groups (16). Lyophilized, PE-hemerythrin was dissolved in 6 M guanidine-HCl at 20-30 mg/ml. Citraconic anhydride was added with vigorous stirring in 50 μ l portions at 20 minute intervals up to a 50-fold molar excess of citraconic anhydride to lysine. The pH was maintained between 8 and 9 with 3N NaOH during the course of the reaction. The acylated protein was desalted by dialysis against 0.2 M NH_4HCO_3 (pH 8.5) followed by dialysis against 0.05 M NH_4HCO_3 (pH 8.5). The modified protein was treated with trypsin (1 mg of Worthington TPK-trypsin/50 mg of substrate) at pH 8.5 for 2 hours at 37° followed by acidification with 88% formic acid to pH 1.5 at the same temperature to remove the citraconyl groups. A three hour incubation was allowed to ensure complete hydrolysis. The products were lyophilized and separated by gel filtration on Sephadex G-50 Superfine equilibrated in 9% formic acid.

Cleavage after glutamyl bonds in peptide T3 was accomplished by treatment with Staphylococcal protease (17). Solutions containing 1 mg/ml of peptide with 0.05 M ammonium acetate (pH 4.0) were incubated at 37° for 20 hours with 1 mg enzyme per 25 mg peptide. After lyophilization of the reaction mixture, the products were separated by gel filtration on Sephadex G-50 Superfine in 9% formic acid.

Sequence Analyses. Edman degradations were performed with a Beckman Sequencer, Model 890, by the method of Edman and Begg (18) as modified by Hermodson et al. (19, 20). The sequenator reagents were of "sequen" grade (Pierce Chemical Co.). Sequenator products were identified by gas chromatography, thin layer chromatography, or spot tests, as described previously (19). Each reported sequence is the result of at least two sequenator analyses with the exception of residues 70-73, which were only analyzed once.

Digestions with bovine carboxypeptidase A and porcine carboxypeptidase B (Worthington, treated with phenylmethylsulfonyl fluoride) were carried out according to Ambler (21). Hemerythrin was dissolved at a concentration of 0.1 mM in 0.3% sodium lauryl sulfate, 0.2 M N-ethylmorpholine acetate (pH 8.5). Carboxypeptidases A and B were added in the ratios of 0.01 mg and 0.04 mg per 0.1 μ mole protein, respectively. Reactions were performed at 23° and ended by the addition of 88% formic acid to pH 1.5 between 10 minutes and 2.5 hours after the start of the reaction. Products were separated by gel filtration on Sephadex G-25 in 1% formic acid and the resulting amino acids were analyzed as described below.

Amino Acid Analyses. Amino acid compositions were determined using a Beckman amino acid analyzer, Model 120C. Amino acid and peptide samples were S-pyridylethylated and subjected to at least two acid hydrolyses (between 24 and 96 hours digestion) prior to analysis. Cysteine was quantitated as the S-pyridylethylcysteine derivative (22).

Tryptophan residues in S-pyridylethylated protein and peptides were determined by amino acid analysis after alkaline hydrolysis at 135° for 48 hours (23). The tryptophan composition of the apoprotein was also verified spectrophotometrically by the method of Benzene and Schmid (24).

TABLE I

Sequenator Analyses of Hemerythrin and Hemerythrin Fragments^a

Position	Hemerythrin (700 nmol)		N-1 (400 nmol)		T-1 (550 nmol)		T-2 (1000 nmol)		S-1 (850 nmol)	
	PTH Deriv- ative	yield (nmol)	PTH Deriv- ative	yield (nmol)	PTH Deriv- ative	yield (nmol)	PTH Deriv- ative	yield (nmol)	PTH Deriv- ative	yield (nmol)
1	Gly		Gly	335	Cys ^b		Thr		His ^c	
2	Phe	715	Ile	240	Thr		Phe	595	Lys	
3	Pro	760	Leu	265	Gly	370	Tyr	745	Lys	
4	Ile	505	Leu	345	Lys		Thr		Ala	
5	Pro	650	Leu	360	His ^c		Ile	550	His ^c	
6	Asp	410	X		Phe	310	Val, Ile	340, 165	Asp	525
7	Pro	530	Gln		Leu	275	Asp	295	Asp	525
8	Tyr ^b	580	Ala	170	Asn		Asp	300	Phe	785
9	Cys ^b		Asp	125	Glu		Glu	345	Ile	345
10	Trp		Asn		Gln		His ^c		His ^c	
11	Asp	315	Ala	195	Gln		Lys		Lys	
12	Ile	275	Asp ^c	80	Leu	160	Thr		Leu	380
13	Ser		His ^c		Met	140	Leu	525	Asp	245
14	Phe	285	Leu	130	Gln		Phe	345	Thr	
15	Arg ^c		Asn		Ala	45	Asn		Trp	
16	Thr		Glu		X		Gly	180	Asp	110
17	Phe	285	Leu ^c	120	Gln		Ile	355	Gly	355
18	Tyr	260	Arg ^c		Tyr	75	Leu	340	Asp	230
19	Thr		Ala	85	Leu		Leu	430	Val	270
20	Ile	215	Cys ^b		Gly	85	Leu	325	Thr	
21	Val, Ile	160, 165	Thr		Tyr	50			Tyr	220
22	Asp	120	Gly	85	Ala	60			Lys	255
23	Asp	260	Lys		Glu	20			Ala	
24	Glu				His ^c				Asn	
25	His ^c								Trp	
26	Lys								Leu	195
27	Thr								Val	180
28	Leu	210							Asn	
29	Phe	150							His ^c	
30	Asn								Ile	105
31	Gly	165							Lys	
32	Ile	155							Thr	
33	Leu								Ile	85
34									Asp	
35									Phe	55
36									Lys	
37									Tyr	35

Residues 1-33 31-53 50-73 16-35 73-109

^aPhenylthiohydantoin (PTH) amino acids were identified and quantitated by gas chromatography (19) unless otherwise indicated. Repetitive stepwise yields were 94-97%. Degradations were terminated when the peak intensity of the amino acid being detected was less than 1.5 times that of background peaks or less than 0.7 times that of carryover peaks. At any given position, all amino acids which were unambiguously identified are reported, with the exception of amino acids carried over from previous steps.

^bPyridylethylcysteine derivative identified by thin layer chromatography (19).

^cIdentified by spot tests (19).

REFERENCES

- Bornstein, P. (1970) *Biochemistry* **9**, 2408-2421
- Atassi, M.Z., and Habeeb, A.F.S.A. (1972) *Methods Enzymol.* **25**, 546-553
- Houmar, J., and Drapeau, G.R. (1972) *Proc. Nat. Acad. Sci. U.S.A.* **69**, 3506-3509
- Edman, P., and Begg, G. (1967) *Eur. J. Biochem.* **1**, 80-90
- Hermodson, M.A., Ericsson, L.H., Tien, K., Neurath, H., and Walsh, K.A. (1972) *Biochemistry* **11**, 4493-4502
- Hogg, R.W., and Hermodson, M.A. (1977) *J. Biol. Chem.* **252**, 5135-5141
- Ambler, P. (1972) *Methods Enzymol.* **25**, 142-154
- Cavins, J.F., and Friedman, M. (1970) *Anal. Biochem.* **35**, 489-493
- Hugli, T.E., and Moore, S. (1972) *J. Biol. Chem.* **247**, 2828-2834
- Bencze, W.L., and Schmid, K. (1957) *Anal. Chem.* **29**, 1193-1196

TABLE II

Amino acid composition of hemerythrin and fragments generated by trypsin digestion of the citraconylated protein^a

Amino Acid	Hemerythrin ^b	T-1 ^c	T-2 ^d	T-3 ^d
Lys	9.3 (9)	7.0 (7)	1.0 (1)	0.4 (0)
His	6.8 (7)	4.7 (5)	1.7 (2)	0.0 (0)
Cys ^e	1.7 (2)	0.9 (1)	0.0 (0)	0.7 (1)
Arg	3.9 (4)	1.2 (2)	1.0 (1)	0.9 (1)
Asp+Asn	17.5 (18)	8.7 (9)	7.1 (7)	2.1 (2)
Thr	7.2 (7)	3.7 (4)	2.5 (3)	0.1 (0)
Ser	2.9 (3)	1.5 (1)	1.2 (1)	1.0 (1)
Glu+Gln	8.4 (9)	6.7 (6)	3.2 (3)	0.2 (0)
Pro	4.2 (3)	0.0 (0)	0.0 (0)	2.8 (3)
Gly	6.1 (6)	3.2 (3)	1.4 (1)	1.3 (1)
Ala	6.2 (6.5)	3.9 (4.5)	1.8 (2)	0.2 (0)
Val	2.9 (2.5)	1.9 (2)	0.8 (0.5)	0.1 (0)
Met	1.3 (1)	1.1 (1)	0.0 (0)	0.0 (0)
Ile	8.9 (8.5)	2.8 (3)	2.2 (2.5)	1.9 (2)
Leu	10.3 (10)	3.9 (4)	5.8 (6)	0.2 (0)
Tyr	6.1 (6)	4.2 (4)	1.2 (1)	1.0 (1)
Phe	7.1 (7)	3.2 (3)	1.9 (2)	1.8 (2)
Trp ^f	2.8 (3)	1.9 (2)	0.0 (0)	0.5 (1)
Total	114	61	33	15
Residue #'s	1-113	50-110	16-48	1-15

^aValues listed are relative molar quantities for each polypeptide. Numbers in parentheses are those expected from amino acid sequence of polypeptide. Amino acid analyses were corrected for decomposition of serine and threonine during hydrolysis and for incomplete hydrolysis of isoleucine and valine.

^bAverage of 6 analyses of samples hydrolyzed 22, 46, or 70 hr.

^cAverage of 4 analyses of samples hydrolyzed 24 or 96 hr.

^dAverage of 2 analyses of samples hydrolyzed 24 and 96 hr.

^eDetermined as S-pyridylethylcysteine.

^fDetermined by alkaline hydrolysis with quantitation relative to histidine content.

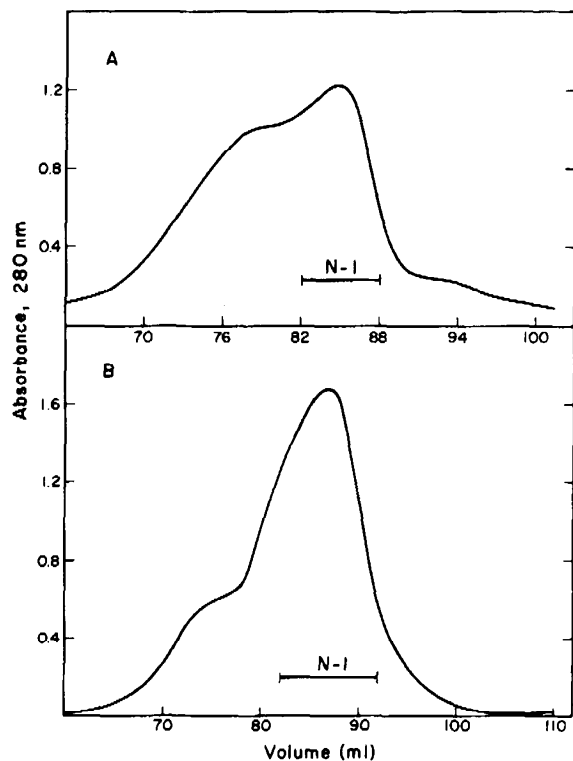


Figure 3. Gel filtration of NH_4OH digestion products of hemerythrin on Sephadex G-75 Superfine. The column (1.5 x 87 cm) was eluted with 9% formic acid at a flow rate of 3 ml/hr. Fractions were pooled as indicated by horizontal lines. Shoulder to left of fragment N-1 corresponds to elution volume of intact protein. A. Sample contained 23 mg of NH_4OH -treated hemerythrin. B. Sample contained 11 mg of peptide N-1 pooled from two columns similar to A.

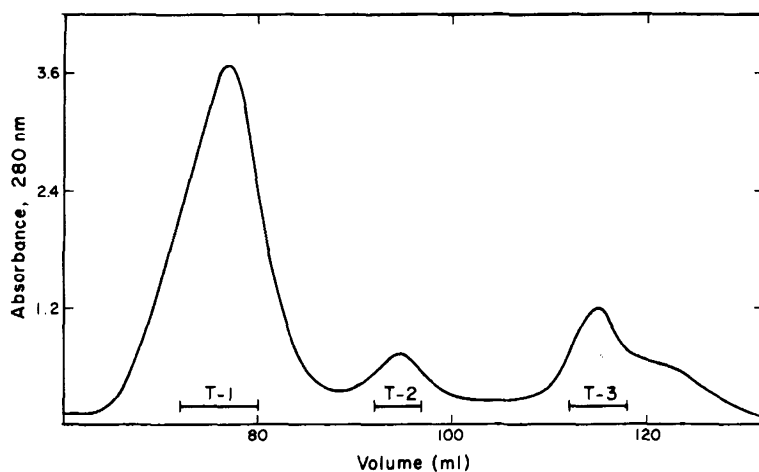


Figure 4. Gel filtration of trypsin digestion products of citraconylated hemerythrin on Sephadex G-50 Superfine. A 50 mg sample was applied to the column (1.5 x 86 cm) and eluted with 9% formic acid at a flow rate of 6 ml/hr. Fractions were pooled as indicated by horizontal lines.

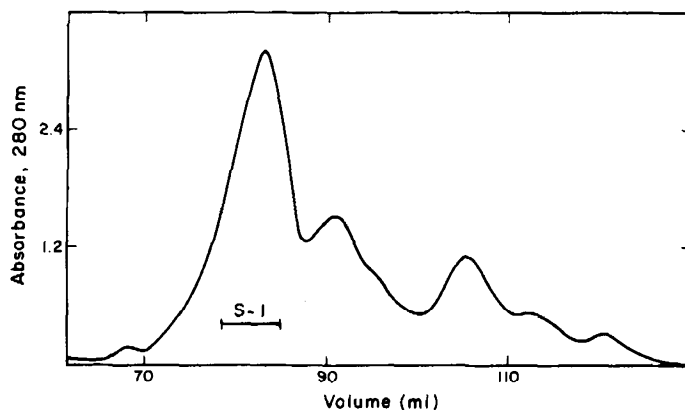


Figure 5. Gel filtration of Staphylococcal protease digestion products of hemerythrin fragment T-1 (34 mg) on Sephadex G-50 Superfine. Conditions as described in Figure 3.

Table III

Digestion of Hemerythrin and Fragment T-1 with Carboxypeptidases A and B

Protein Sample	Carboxypeptidase Treatment	Amino Acids Liberated ^a					
		Phe	Tyr	Arg	Gly	Lys	Ile
ApoHr (1-113)	A (150 min)	-	-	-	0.2	0.1	1.0
	A+B (10 min)	-	0.2	0.2	0.2	1.0	1.1
	A+B (150 min)	0.4	0.9	0.8	0.8	1.4	0.9
T-1 (50-110)	B (20 min)	0.4	1.6	1.2	0.1	0.5	0.1

Proposed Sequence (107-113): Phe-Lys-Tyr-Arg-Gly-Lys-Ile

^aData have been normalized to the amount of isoleucine produced (70 nmol) upon prolonged hydrolysis of apohemerythrin (100 nmol).