Amino Acid Composition and Terminal Sequences of *Golfingia gouldii* Hemerythrin*

W. R. Groskopf,† J. W. Holleman, E. Margoliash, and I. M. Klotz

ABSTRACT: The amino acid composition of the hemerythrin from the sipunculid worm *Golfingia gouldii* is, in residues per molecule of protein subunit of 13,500 molecular weight: $Lys_{11.0}$, $His_{6.7}$, $Arg_{3.2}$, $Asp_{16.9}$, $Thr_{4.6}$, $Ser_{3.6}$, $Glu_{9.9}$, Pro_{4r2}, $Gly_{6.4}$, $Ala_{5.5}$, $CySH_{1.1}$, $Val_{3.9}$, $Met_{1.1}$, $Ile_{8.7}$, $Leu_{7.7}$, $Tyr_{5.0}$, $Phe_{8.8}$, and $Trp_{4.0}$. The iron content is 0.83%.

The amino acid sequence at the amino-terminal end was shown to be Gly-Phe-Pro, while that at the carboxyl-terminal end is (Phe,Lys,Tyr,Lys,Gly)-Lys-Ile. The hemerythrin subunits appear to consist of single polypeptide chains approximately 112.3 residues long, containing only amino acids and two atoms of iron.

Hemerythrin is the oxygen-carrying pigment of sipunculid worms, some brachiopods, and a few annelids. The native protein from the sipunculid *Golfingia gouldii* has a molecular weight of 107,000 and can be dissociated into eight subunits of equal size, each subunit containing two iron atoms capable of combining with one oxygen molecule (Klotz and Keresztes-Nagy, 1963; Keresztes-Nagy and Klotz, 1963, 1965).

Pooled preparations of hemerythrin from *G. gouldii* do not appear to be entirely homogeneous. It has been shown by starch gel electrophoresis (Manwell, 1963) and by paper electrophoresis chromatography (Manwell, 1963; Groskopf *et al.*, 1963) that at least two classes of the octameric protein exist in different individual worms in this species. Paper electrophoresis chromatography reveals that the differences are apparently restricted to no more than two or three peptides (Groskopf *et al.*, 1963), while the vast preponderance of peptides from preparations from individual worms are identical. Amino acid sequence studies using pooled preparations of the protein should therefore be feasible and should reveal a unique structure except for a few positions. Since the proportions of the different classes of *G. gouldii* hemerythrin are not far from being equal, it might even be possible to determine the complete amino acid sequences of the variants directly from pooled preparations. There seems to be no point in preparing material separately from individual worms, since each class may well consist of different proportions of various monomeric units. This paper is the first of a series describing the amino acid sequence of *G. gouldii* hemerythrin.

Experimental Section

Preparation and Composition of G. gouldii Hemerythrin. The worms were obtained from the Marine Biological Laboratory, Woods Hole, Mass. Oxyhemerythrin was prepared according to Klotz *et al.* (1957) and stored as a crystalline suspension in 20% ethanol and under nitrogen at 6°. The yield from 100 worms was generally about 1.5–2 g. The oxyhemerythrin became slowly converted to the deoxygenated ferric form unless oxygen was rigorously excluded.

Iron-free hemerythrin was prepared by dissolving about 15 mg of either the oxy or ferric forms in 0.5 ml of 0.2 N HCl. After the solution became colorless, the protein was precipitated with eight volumes of acetone. The protein was suspended in acid, reprecipitated twice more, washed three times with acetone by centrifugation, and dried.

Iron was determined by the *o*-phenanthroline procedure of Fortune and Mellon (1938a,b). Nitrogen analyses by the Dumas method were carried out by Miss H. Beck (Northwestern University) and tests for reducing sugars were done by Mr. R. Katzman in Dr. Edwin H. Eylar's laboratory at the University of Southern California.

The amino acid composition was determined on iron-free as well as native hemerythrin. The protein was hydrolyzed in triply glass-distilled 6 N HCl under nitrogen at 108° for 12–110 hr. The amino acids were estimated according to Spackman *et al.* (1958), using the Beckman Model 120B amino acid analyzer. Cysteine and methionine were determined as cysteic acid and methionine sulfone, respectively (Moore, 1963), norleucine being used as an internal standard (Walsh and Brown, 1962). Tryptophan was determined spec-

* From the Biochemistry Division, Department of Chemistry, Northwestern University, Evanston, Illinois (W. R. G., J. W. H., I. M. K.), and the Research Division Abbott Laboratories, North Chicago, Illinois (E. M.). *Received August 29, 1966.* This investigation was supported in part by a grant (HE-08299) from the National Heart Institute, U. S. Public Health Service. It was also assisted by support made available by a U. S. Public Health Service Training Grant (5T1-GM-626) from the National Institute of General Medical Services.

† Present address: Michael Reese Research Foundation Chicago, Ill.

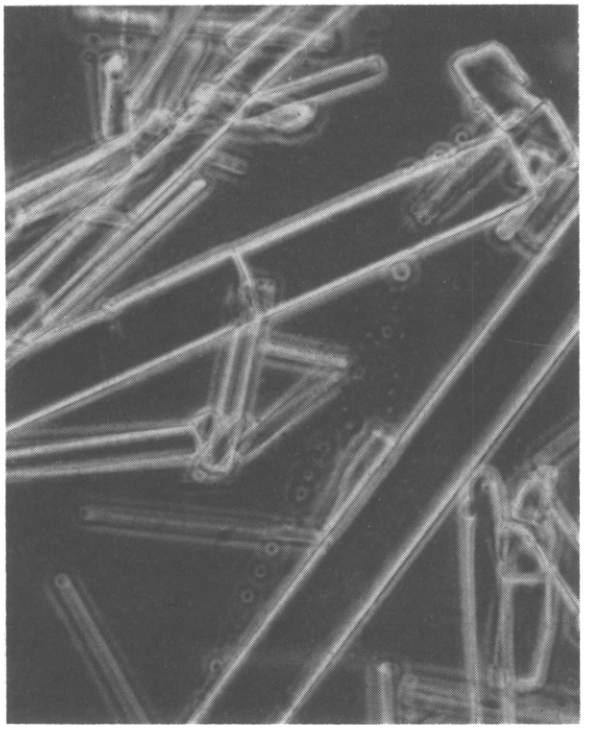

FIGURE 1: Cyrstals of oxyhemerythrin from *G. gouldii*, crystallized from 20% ethanol containing 0.4% NaCl (phase-contrast photograph; magnification, 400×).

trophotometrically (Beaven and Holiday, 1952), a correction being made for the absorption of the peptide bonds.

Terminal Sequences. The sequence at the carboxyl-terminal end of the protein was determined by digestion with DFP[1]-treated carboxypeptidase A (Worthington Biochemical Corp.). Iron-free hemerythrin (1 μmole of subunit) was dissolved in 0.5 ml of 0.01 N NaOH and the solution was added dropwise to 4.5 ml of 0.05 N sodium borate buffer, pH 8.5. The solution was slightly cloudy. Carboxypeptidase A (0.05 μmole) and an internal analytical standard (norleucine, 2.5 μmoles) were added. The mixture was incubated at room temperature. Aliquots were removed at various times for paper chromatography and amino acid analysis. Similar digestions were also performed with native hemerythrin and with hemerythrin heated at 65° in the pH 8.5 borate buffer until a cloudy white suspension was obtained.

The amino-terminal residue of the polypeptide chain or chains was identified by dinitrophenylation. Pooled, crystallized oxyhemerythrin was dissolved in 4 M guanidine hydrochloride and treated with FDNB[1] (Phillips, 1958) at pH 9.0 (NaOH) for 2 hr at 40°. Condensation was also effected in the presence of ethanol (Fraenkel-Conrat and Singer, 1956; Biserte *et al.*, 1960), and with guanidine hydrochloride plus ethanol, to promote coupling with histidine if present as end group. A further portion of the protein was treated by the method of Bettelheim (1955) for possible N-terminal cysteine, which involves oxidation by performic acid, condensation with FDNB, and separation of DNP-cysteic acid from other water-soluble derivatives and free amino acids by ion-exchange chromatography. The ether, ethyl acetate, ethyl acetate–*sec*-butyl alcohol, and residual water phases obtained where appropriate on treatment of hydrolysates from the above condensations were subjected to paper chromatography in the phthalate buffer–*t*-amyl alcohol system of Blackburn and Lowther (1951) and in the phosphate buffer system of Levy (1954).

Edman degradations (Edman, 1950, 1953) were carried out on iron-free hemerythrin. Trimethylamine (5%) was used as the initial solvent in which the PTC[1] protein was formed. Cyclization was effected in trifluoracetic acid (Konigsberg and Hill, 1962). The released PTH[1] derivatives were identified by paper chromatography with Sjöquist's solvent A (Sjöquist, 1953), and located by examination in ultraviolet light or by development with the iodine–azide reagent (Chargaff *et al.*, 1948). In addition, the free amino acid was regenerated from the PTH derivatives by hydrolysis with 0.1 N NaOH under vacuum at 108° for 15 hr, and identified by amino acid analysis (Van Orden and Carpenter, 1964).

Results

Figure 1 shows a photograph of crystals of *G. gouldii* oxyhemerythrin obtained by dialysis of a 0.4% NaCl solution against a water solution 20% in ethanol and 0.4% in NaCl.

Composition of G. gouldii Hemerythrin. The iron content of oxyhemerythrin was found to be 0.83%, in agreement with previous determinations (Klotz *et al.*, 1957). The native protein had a nitrogen content of 15.94 ± 0.11%, which decreased to 14.95 ± 0.02% on removal of iron. This decrease is of a magnitude exactly equal to that expected for formation of the protein hydrochloride estimated from the amino acid composition. Tests for reducing sugars performed in Dr. Eylar's laboratory and tests for reducing sugars, amino sugars, and sialic acid in this laboratory all proved negative (Keresztes-Nagy, 1962).

The results of the amino acid analyses of iron-free hemerythrin are given in Table I. The values listed were obtained by making the appropriate extrapolations from the values at various hydrolysis times (Hirs *et al.*, 1954; Mahowald *et al.*, 1962). Essentially identical results were obtained following hydrolysis of native hemerythrin. It is interesting to note that with the iron-containing protein, threonine was only slightly degraded, even after 72 hr of hydrolysis (5%), while serine was destroyed to the usual extent (65%). In contrast, with the iron-free protein after the same hydrolysis time, the destructions were 37% for threonine

[1] Abbreviations: DFP, diisopropylphosphorofluoridate; FDNB, 1-fluoro-2,4-dinitrobenzene; PTC, phenylthiocarbamyl derivative; PTH, phenylthiohydantoin derivative.

TABLE I: Amino Acid Composition of Iron-Free *G. gouldii* Hemerythrin.

	Residues (g)/100 g of Protein	Residues/ Subunit (13,500 g)	Residues Accounted for by Tryptic Peptides[a]
Lysine	9.959	11.0	11
Histidine	6.596	6.7	7
Arginine	3.467	3.2	3
Aspartic acid	13.811	16.9	17
Threonine	3.286	4.6	5
Serine	2.186	3.6	4
Glutamic acid	9.077	9.9	10
Proline	2.875	4.2	4
Glycine	2.619	6.4	6
Alanine	2.786	5.5	5
Half-cystine[b]	0.816	1.1	1
Valine	2.716	3.9	4
Methionine[b]	0.971	1.1	1
Isoleucine	7.038	8.7	9
Leucine	6.201	7.7	8
Tyrosine	5.793	5.0	5
Phenylalanine	9.257	8.8	9
Tryptophan[c]	5.512	4.0	4
H_2O	0.133		
Moisture	3.810		
Ash	0.125		
Total	99.034	112.3	113

[a] Groskopf *et al.* (1966). [b] Determined as cysteic acid and methionine sulfone (Moore, 1963). [c] Determined spectrophotometrically (Beaven and Holiday, 1952).

and 74% for serine. The total recovery of 99.03% given in Table I for amino acid residues and inorganic constituents indicates that within experimental error hemerythrin contains only amino acid residues and iron.

Amino Acid Sequences of the Amino- and Carboxyl-Terminal Segments. Figure 2 shows the release of amino acids from iron-free hemerythrin by carboxypeptidase A at periods of digestion up to 24 hr. Isoleucine is released first and approaches one residue per molecule of protein subunit, followed by lysine, the amount of which continues to increase as digestion proceeds. Glycine, tyrosine, and phenylalanine are the only other residues released in significant amounts, and they reproducibly appeared in that order. It is notable that the amount of lysine was always equal to or slightly greater than the sum of the amounts of glycine, tyrosine, and phenylalanine. Since lysine is known to be released by carboxypeptidase A far more slowly than any of the other residues observed, even at pH 8.5 (Olson and Kuby, 1964), and the amount of lysine calculated on the assumption of

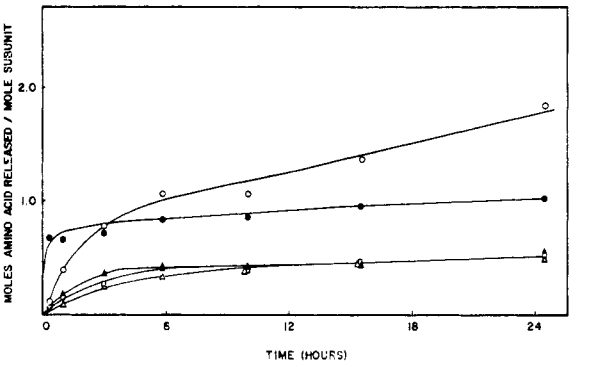

FIGURE 2: Time course of reaction of hemerythrin with carboxypeptidase A at pH 8.5. See text for experimental details. ○ = Lys. ● = Ile. ▲ = Gly. □ = Tyr. △ = Phe.

100% yield of glycine, tyrosine, or phenylalanine approaches three, it appears that there must be a total of three lysyl residues in the carboxyl-terminal sequence. This conclusion was confirmed by the structure of the carboxyl-terminal sequence of hemerythrin detailed in the following paper (Groskopf *et al.*, 1966; see peptides T-50ch2, T-59elB2, and T-45ch2elN).

It is notable that in the carboxyl-terminal sequence Thr-Ile-Asp-Phe-Lys-Tyr-Lys-Gly-Lys-Ile, an aspartyl residue, which is known to be a particularly poor substrate for carboxypeptidase A (Neurath, 1960), precedes the amino acids released by the enzyme. This circumstance probably explains why only traces of other amino acids were detected. No amino acids were released by the action of carboxypeptidase A on native, nondenatured hemerythrin, even after incubation for 8 hr.

Dinitrophenylation showed glycine to be the only amino-terminal residue in the protein. DNP-glycine was recovered in 15–25% yield following hydrolysis of the dinitrophenyl (DNP) protein in 6 N HCl in sealed, evacuated tubes at 108° for 10 hr. Such a yield is reasonable in view of the extreme lability of DNP-glycine under such conditions (Porter, 1951), and, indeed, 18 hr of hydrolysis resulted in the complete destruction of the amino-terminal DNP-glycine. On shorter times of hydrolysis, ε-DNP-lysine peptides were observed, identified as such by reaction with ninhydrin. Attempts to detect bis-DNP-histidine and amino-terminal cysteine as DNP-cysteic acid were consistently negative.

Three steps of the Edman degradation on iron-free hemerythrin demonstrated that the amino-terminal sequence is Gly-Phe-Pro, confirming and extending the dinitrophenylation results above. No other PTH-amino acids were detected, and the amount of free amino acid recovered from the alkaline hydrolysates of the separated PTH derivatives was approximately 20% at each step. This amino-terminal sequence fits the structure of peptide TT-5ch4 recovered from the tryptic digest of the protein (Groskopf *et al.*, 1966).

Discussion

The amino acid composition of the subunit of *G. gouldii* hemerythrin is remarkable in that it contains single residues of methionine and cysteine, and only three residues of arginine, making it amenable to cleavage into large fragments by specific chemical and enzymic procedures. Taking into account the known number of amidated residues, as estimated from the structures of the tryptic peptides (Groskopf *et al.*, 1966), there is a slight preponderance of acidic over basic residues. The hydrophobic residues isoleucine, leucine, valine, phenylalanine, tryptophan, tyrosine, and methionine account for approximately 35% of the total residues of the polypeptide chain. Normalization of the amino acid analysis on the basis of a subunit molecular weight of 13,500 results in a calculated chain length of 112.3 residues for each subunit. This compares to within a fraction of a residue to the sum of the compositions of all the tryptic peptides described in the following paper (Groskopf *et al.*, 1966; see Table I) which accounts for 113 residues.

The presence of glycine at the amino terminus and isoleucine at the carboxyl terminus confirms the results previously obtained by Manwell (1963). The demonstration of unique amino- and carboxyl-terminal sequences indicates that any differences which may exist in the primary structures of the subunits of different classes of hemerythrin must be located elsewhere than in the terminal portions of the subunit molecules. The failure of carboxypeptidase A to release any amino acids from native hemerythrin in the met form is probably due to a conformation of the native protein which prevents access of the enzyme to the carboxyl terminus of the subunits.

References

Beaven, G. H., and Holiday, E. R. (1952), *Advan. Protein Chem. 7*, 319.

Bettelheim, F. R. (1955), *J. Biol. Chem. 212*, 235.

Biserte, G., Holleman, J. W., Holleman-Dehove, J., and Sautiere, P. (1960), *Chromatog. Rev. 2*, 59.

Blackburn, S., and Lowther, A. G. (1951), *Biochem. J. 48*, 126.

Chargaff, E., Levine, C., and Green, C. (1948), *J. Biol. Chem. 175*, 67.

Edman, P. (1950), *Acta Chem. Scand. 4*, 283.

Edman, P. (1953), *Acta Chem. Scand. 7*, 700.

Fortune, W. B., and Mellon, M. G. (1938a), *Ind. Eng. Chem., Anal. Ed. 10*, 13.

Fortune, W. B., and Mellon, M. G. (1938b), *Ind. Eng. Chem., Anal. Ed. 10*, 60.

Fraenkel-Conrat, H., and Singer, B. (1956), *Arch. Biochem. Biophys. 60*, 64.

Groskopf, W. R., Holleman, J. W., Klotz, I. M., Keresztes-Nagy, S., and Margoliash, E. (1963), *Science 141*, 166.

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

Hirs, C. H. W., Stein, W. H., and Moore, S. (1954), *J. Biol. Chem. 211*, 941.

Keresztes-Nagy, S. (1962), Ph.D. Dissertation, Northwestern University, Evanston, Ill.

Keresztes-Nagy, S., and Klotz, I. M. (1963), *Biochemistry 2*, 923.

Keresztes-Nagy, S., and Klotz, I. M. (1965), *Biochemistry 4*, 919.

Klotz, I. M., and Keresztes-Nagy, S. (1963), *Biochemistry 2*, 445.

Klotz, I. M., Klotz, T. A., and Fiess, H. A. (1957), *Arch. Biochem. Biophys. 68*, 284.

Konigsberg, W., and Hill, R. J. (1962), *J. Biol. Chem. 237*, 2547.

Levy, A. L. (1954), *Nature 174*, 126.

Mahowald, T. A., Noltman, E. A., and Kuby, S. A. (1962), *J. Biol. Chem. 237*, 1138.

Manwell, C. (1963), *Science 139*, 755.

Moore, S. (1963), *J. Biol. Chem. 238*, 235.

Neurath, H. (1960), *Enzymes 4*, 11.

Olson, O. E., and Kuby, S. A. (1964), *J. Biol. Chem. 239*, 460.

Phillips, D. M. P. (1958), *Biochem. J. 68*, 35.

Porter, R. R. (1951), *Methods Med. Res. 3*, 256.

Sjöquist, J. (1953), *Acta Chem. Scand. 7*, 447.

Spackman, D. H., Stein, W. H., and Moore, S. (1958), *Anal. Chem. 30*, 1190.

Van Orden, H. O., and Carpenter, F. H. (1964), *Biochem. Biophys. Res. Commun. 14*, 399.

Walsh, K. A., and Brown, J. R. (1962), *Biochim. Biophys. Acta 58*, 596.