The Primary Structure of *Golfingia gouldii* Hemerythrin. Interpeptide Overlaps and Sequences from Chymotryptic Peptides*

A. R. Subramanian,† J. W. Holleman,‡ and I. M. Klotz

ABSTRACT: Peptides from chymotryptic digests of hemerythrin were isolated by a rapid method which employed preliminary fractionation by gel filtration and further purification by ion-exchange chromatography where necessary, and by paper electrophoresis and chromatography. They were then characterized chemically. Twenty-five major peptides were obtained. Twelve of these provided overlaps between the tryptic peptides obtained in previous work (Groskopf, W. R., Holleman, J. W., Margoliash, E., and Klotz, I. M. (1966a), *Biochemistry 5*, 3783); the others could be assigned to individual tryptic peptides. Some uncertainties in sequence were resolved by detailed sequence analysis of certain of the chymotryptic peptides. In this way the 113 residues of the hemerythrin monomer were arranged in sequence within four fragments: an N-terminal one containing 29 residues, a C-terminal one containing 46 residues, and two inner fragments of 4 and 34 residues, respectively.

Amino acid substitutions at two points in the chain in samples of hemerythrin from pooled blood were noted.

Hemerythrin is the respiratory protein of a class of marine worms, the sipunculids. It is also found in some brachiopods, annelids, and priapulids. As such it represents one of the three great classes of oxygen-carrying proteins found in nature, the other two being the hemoglobins and the hemocyanins. Study of its structure and function therefore presents considerable comparative biochemical interest. The hemerythrin from the sipunculid species *Golfingia gouldii* has been shown to have a molecular weight of 108,000 arising from the association of eight subunits or monomers of 13,500 molecular weight each (Klotz and Keresztes-Nagy, 1962, 1963). Each of these subunits contains two atoms of iron and can bind one molecule of oxygen. The subunits, which are identical in size, are also apparently identical chemically, except for the presence of some minor differences between individual polypeptide chains (Manwell, 1963; Groskopf *et al.*, 1963, 1966a,b) at certain points in the sequence, which may represent genetically controlled amino acid substitutions. Unlike the hemoglobins, hemerythrin has no porphyrin or other prosthetic group, and therefore the iron atoms must be bound directly, presumably through the side chains of specific amino acid residues. Neither the nature of these iron-coordinating ligands nor their location within the molecule is known. The determination of the amino acid sequence of the protein was undertaken to provide a basis for further experimentation on the questions of mode of iron binding and oxygen fixation and functioning of the protein.

Previous investigations in this laboratory (Groskopf *et al.*, 1966b) have placed all 113 amino acid residues of the hemerythrin subunit within 16 peptides, obtained from tryptic digests of hemerythrin. This paper describes the overlaps between these obtained by the characterization of peptides from chymotryptic digests of the protein.

Experimental Section

Isolation of the Protein and Preparation for Enzymic Digestion. G. gouldii worms were obtained from the Marine Biological Laboratory, Woods Hole, Mass., and hemerythrin was isolated from the erythrocytes of their coelomic fluid as described previously (Groskopf *et al.*, 1966a). A slight modification involved the use of dilute NaCl (0.4%) instead of water at the laking and dialysis steps and in the crystallizing medium to maintain the hemerythrin in the chloride form. Overlay of a small amount of toluene–ether (1:1) was routinely used at the laking stage to facilitate removal and caking of stromata and lipids and to inhibit bacterial action. The protein crystals obtained were kept frozen until use. The concentration of hemerythrin in solutions made from them was determined from the absorption at 280 nm (mμ) using the value $E = 2.58$ cm/g per l. (Keresztes-Nagy, 1962).

The lone cysteine residue of the hemerythrin monomer was protected and the protein was prepared for enzymic digestion by two methods. In the first of these, 50 ml of a 2% solution of hemerythrin in 0.05 M sodium

* From the Biochemistry Division, Department of Chemistry, Northwestern University, Evanston, Illinois 60201. *Received March 13, 1968.* 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-00626) from the National Institute of General Medical Sciences.

† Present address: Department of Bacteriology and Immunology, Harvard Medical School, Boston, Mass. 02115.

‡ Present address: Biology Division, Oak Ridge National Laboratory, Oak Ridge, Tenn. 37830.

borate (pH 9.2), 0.01 M in NaN_3,[1] was treated with 100 mg of cystine (chromatographically pure, Mann Laboratories) at room temperature for 6 hr. The solution was then chilled in ice to precipitate excess dissolved cystine and centrifuged. The supernatant was dialyzed at room temperature against two changes of 6 M guanidinium chloride (pH 7–8), 0.01 M in Na_2EDTA. This operation, in one step, removes the iron and the residual cystine and simultaneously denatures the protein in preparation for the chymotryptic digestion. The salts were removed by exhaustive dialysis against distilled water, and the protein, which had precipitated during the last step, was collected by centrifugation.

In the alternate procedure, 14 mg (75 μmoles) of freshly recrystallized iodacetamide was added to a solution of 203 mg of hemerythrin (15 μmoles of subunit) in 25 ml of 0.1 M Tris-HCl buffer (pH 8.6), 0.01 M in NaN_3. The reaction vessel was wrapped in aluminum foil to protect the reagent from light and the reaction was allowed to proceed for 4 hr at room temperature. The solution was then dialyzed in the cold, first against distilled water, then against 0.015 M HCl to remove iron, and finally against 0.1 M NH_4HCO_3. The protein, which was pinkish-yellow in the beginning, became colorless on dialysis against the hydrochloric acid and precipitated out during the last step. The suspension was used as such for chymotryptic digestion.

Chymotryptic Digestion. Preliminary experiments had shown that the protein was completely solubilized by digestion for 3–4 hr at 38°, that the peptide map remained unchanged for 3–8-hr digestion, and that the map was altered noticeably after longer (24–38 hr) digestion.

Most of the results described in this paper were obtained from a chymotryptic digest prepared from 740 mg (55 μmoles of subunit) of the guanidinium chloride denatured protein. The protein was suspended in 150 ml of 0.1 M NH_4HCO_3 and to the suspension, kept well stirred at 37°, was added 15 mg of bovine α-chymotrypsin (three-times crystallized; Worthington); another 10 mg of the enzyme was added after 2 hr. The digestion was continued for 5 hr, after which the digest was lyophilized. The residue was dissolved in distilled water and relyophilized twice to remove excess ammonium salt, and it was finally dissolved in 10 ml of fresh 0.1 M NH_4HCO_3. This straw yellow solution was centrifuged at high speed to remove suspended particles and fibers.

Gel Filtration. Gel filtration chromatography of the complete digest was carried out in a 180 × 3.8 cm column which was packed with both G-25 and G-50 (fine) grades of Sephadex (Pharmacia). The column was packed first to a height of 40 cm with G-25 and the rest of it was filled with G-50. A 0.1 M solution of NH_4HCO_3 was the medium for both packing and elution. A peristaltic pump was used to obtain constant flow rates; we used flow rates of 30.4 and 38 ml per hr (*i.e.*, 2.7 and 3.4 ml per cm^2 per hr) in the two runs on this column. Five milliliters of the previously described solution was applied for each run. The effluent was monitored by its absorption at 220 or 280 nm, collected in 5-ml fractions.

The dilution of the sample during chromatography was estimated by a calibration run with a 5-ml sample containing hemerythrin octamer, cytochrome *c*, and tyrosine. These three compounds came out of the column at V_{max} of 800, 1020, and 1900 ml, respectively; the widths of their peaks at half-height were 30, 40, and 75 ml. This information was used, along with electrophoretic data, to pool the fractions from this column.

For further purification of peptide fractions that contained the smaller peptides a mixed Sephadex column was used. It was 210 × 2.5 cm in size and was filled with a 1:3 mixture of G-10 and G-15; 0.1 M NH_4HCO_3 was used both for packing and elution. The flow rate with this column was 30 ml/hr and fractions of 2–4 ml were collected.

Ion-Exchange Chromatography. A few of the peptide fractions from the gel filtration were further fractionated on a short column (20 × 1.2 cm) of Dowex 50 (Bio-Rad AG-50W-X2) by means of a stepwise elution procedure. The eluents were 0.1 M ammonium acetate (pH 4.7), 0.1 M ammonium acetate (pH 6.9), and 0.05 M ammonium carbonate. The column was equilibrated with the first of these eluents, and the peptide sample, made acidic with a few drops of 50% acetic acid, was applied and consecutively eluted with 60–100 ml of each eluent. The first solvent desorbs off the column almost all the acidic and neutral peptides, and the third desorbs the basic peptides. The second eluent, 0.1 M ammonium acetate of pH 6.9, was found to selectively desorb certain histidine-containing peptides. The peptides were isolated free of the ammonium salts by repeated lyophilization.

Some ultraviolet-absorbing material can be leached out of the Dowex resin, and therefore it was necessary to wash the column just prior to use.

Criteria of Purity of the Peptides. The purity of the isolated peptides was checked by electrophoresis chromatography (peptide mapping) and by amino acid analysis. Electrophoresis was carried out on Whatman No. 3MM paper at 2000 V (40 V/cm) for 90–150 min using a water-cooled, horizontal strip apparatus (Savant Instruments). Pyridine–acetic acid–water (100:4:900) buffer (pH 6.4) was used routinely. Occasionally two other buffers, pyridine–acetic acid–water (25:78:900) (pH 3.5) and formic acid–acetic acid–water (25:78:900) (pH 1.9), were also used. Chromatography was routinely done for 16 hr using 1-butanol–acetic acid–water (200:30:75) or occasionally pyridine–1-butanol–acetic acid–water (60:10:12:48). For revealing the peptides and specific amino acid residues on the paper standard techniques were used (Groskopf *et al.*, 1966b).

Isolation of Peptides from Peptide Maps. Peptides from the above fractionations were then subjected to preparative peptide mapping. Samples of 20–100 μl containing up to 5 μmoles of the individual peptides in some cases were applied as a single spot on Whatman No. 3MM paper and development was effected as described for the analytical peptide mapping. The peptides were revealed nondestructively initially by the use of a 0.1%

[1] Complexing the iron with ligands such as N_3^- enhances the reactivity of the sulfhydryl group (Keresztes-Nagy and Klotz, 1965).

solution of 2,4-dinitrofluorobenzene in 1-butanol–acetone (7:3). The paper was dipped in this solution and was hung at room temperature. In about 4–8 hr, most of the peptides revealed themselves as yellow spots. A dilute solution of ninhydrin, 0.01% in 1-butanol–acetone (7:3), was also used for the same purpose. This solution (100 ml) was mixed with 10 ml of acetic acid just prior to use. The sheets were dipped in this solution and the color was allowed to develop at room temperature overnight. Virtually all the peptides that were revealed by the usual ninhydrin reagent (0.2%) were also revealed by this dilute solution provided they were present in quantities of 10–50 nmoles.

For extraction of the peptides the periphery of each of the colored spots was marked with a pencil and the disk was cut out using a draftsman's blade. Each disk was cut into small squares which were placed in a screw-capped culture tube and the peptide was extracted with mechanical shaking for 15 min at room temperature with three 2-ml portions of 1% acetic acid. The extract was filtered through glass wool and lyophilized.

Characterization of the Peptides. The amino acid compositions of the peptides were determined with a Beckman–Spinco Model 120C analyzer equipped with a digital integrator (Infotronics, Model CRS-12AB), after hydrolysis at 110° for 24 hr with 5.7 M HCl (triply distilled) in a sealed, evacuated tube. Tryptophan residues were assigned on the basis of Ehrlich reaction of the peptides on the peptide maps. It was assumed from chymotryptic specificity that each of the Ehrlich-positive peptides contained only a single tryptophan residue.

Digestions with leucine aminopeptidase and carboxypeptidase A (Worthington) for determination of terminal residues were carried out as indicated previously (Groskopf *et al.*, 1966b). For digestions with pepsin (Worthington), subtilisin (Sigma), Aspergillus peptidase (Sigma), and Pronase (Calbiochem), 0.5–1.0 μmole of the peptide was taken in 1.0 ml of 0.02 M HCl (for pepsin) or 0.02 M NH_4HCO_3 (for the others) and 10 μl of a freshly made enzyme solution (1 mg/ml) was added. Digestion was allowed to proceed at 30° for various lengths of time and the solution was then lyophilized. The residue was dissolved in 0.1–0.2 ml of distilled water and subjected to peptide mapping. The individual peptides were isolated, when needed, from the preparative peptide maps.

Dilute acid hydrolysis for selective cleavage at aspartyl residues, essentially according to the procedure of Tsung and Fraenkel-Conrat (1965), was carried out on 0.5–1.0 μmole of the peptides using 2.0 ml of 0.02 M HCl in sealed, evacuated tubes for 2–16 hr at 110°. An aliquot of the hydrolysate was used on the amino acid analyzer for estimating the free aspartic acid and the rest was lyophilized. The residue, dissolved in 0.1–0.2 ml of distilled water, was used for preparative peptide mapping.

The amide content of one of the peptides was determined by hydrolyzing it with 1.0 M HCl at 100° for 4 hr and estimating the released NH_3 on the amino acid analyzer. Difficulty was encountered from the ubiquitous ammonia contamination of peptide preparations. This was overcome by the following procedure. The peptide

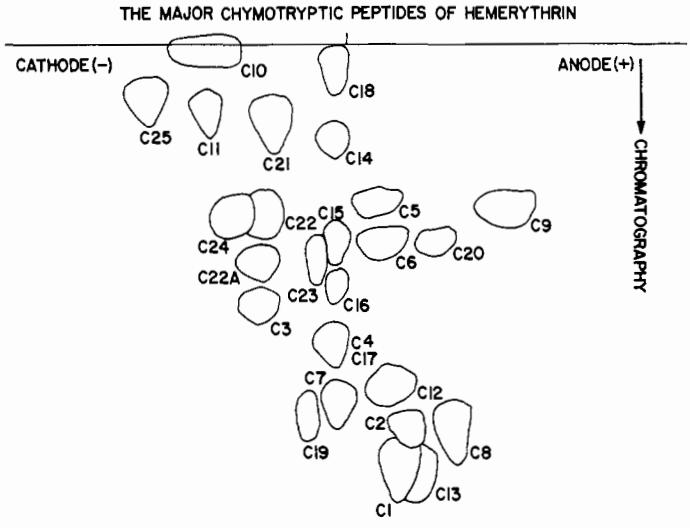

FIGURE 1: Schematic drawing of a peptide map of the major chymotryptic peptides of hemerythrin of *G. gouldii*. Electrophoresis in pyridine–acetate buffer (pH 6.4) 90 min, 40 V/cm); chromatography in 1-butanol–acetic acid–water (200:30:75, v/v) 16 hr. Staining reactions: Arg, C3, C10; His, C11, C18, C19, C5, C6, C9, C23; sulfur C10, C13; Tyr, C24, C15, C4/C17, C1; Trp, C21, C22, C22A, C2, C20; rust color, C8; and no special color, C25, C14, C16, C7, C12. The single spot designated C4/C17 is free tyrosine, derived from two loci in the chain.

solution (5 μl) containing 0.2 μmole of the peptide was pipetted into a culture tube, and was made alkaline with 10 μl of 0.1 M NaOH. The tube was kept overnight over concentrated H_2SO_4 in a vacuum desiccator to remove the free ammonia from the peptide. The residue was dissolved in 0.2 ml of 1.0 M HCl (freshly made from triply distilled 5.7 M HCl and distilled water), tightly closed in a vial with a Teflon-lined cap, and kept in an oven at 100° for 4 hr. The solution was diluted with 0.8 ml of distilled water and immediately applied on the amino acid analyzer. Determination of the amide content of the intact protein was made by the Conway (1962) procedure.

Edman degradations were carried out by a procedure adapted from that of the Protein Sequenator (Edman and Begg, 1967). The peptide solution (5 or 10 μl) containing 0.1–0.5 μmole of a peptide was placed in a culture tube and dried overnight in a vacuum desiccator over KOH and H_2SO_4. The residue was dissolved in a dioxane–*N*-ethylmorpholine–acetic acid–water buffer (50:6:0.15:44), and the tube was flushed with N_2. After addition of 20 μl of PITC[2] the tube was tightly closed with a Teflon-lined cap and immersed in a water bath at 40° for 1 hr. Excess PITC was extracted out with five 2-ml portions of benzene and the residual aqueous layer was freeze dried. The residue was further desiccated while on the lyophilizer by heating it at 60° for 15 min by means of a heating tape. After cooling, the residue was dissolved in 100 μl of trifluoroacetic acid, and the tube was flushed with N_2, capped, and placed in a

[2] Abbreviations used that are not listed in *Biochemistry 5*, 1445 (1966), are: PITC, phenyl isothiocyanate; PTH, phenylthiohydantoin derivative.

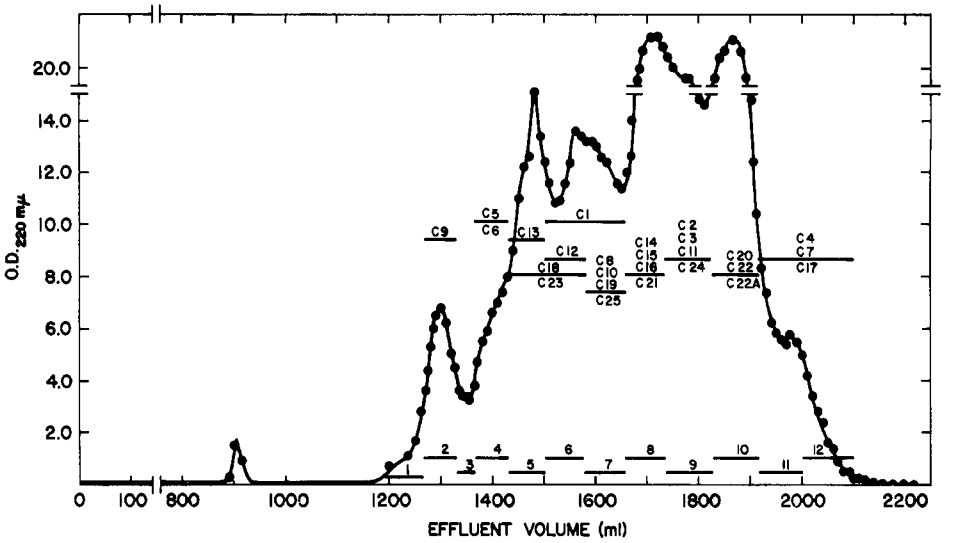

FIGURE 2: Gel filtration pattern of the chymotryptic peptides of *G. gouldii* hemerythrin. The effluent fractions were pooled as indicated. OD_{220} values are for 10-mm path lengths, calculated from measurements with 1- and 2-mm cells. Top bars show from which fraction or fractions peptides were isolated.

40° bath for 15 min for the cleavage reaction to proceed. The trifluoroacetic acid was driven off with a stream of N_2. The thiazolinone of the amino acid, which is formed under these conditions, was extracted with two 1-ml portions of benzene followed by two 1-ml portions of ethyl acetate. The extract was kept in a freezer until several were accumulated for the subsequent analysis. The residual peptide was freed of benzene by a stream of N_2 and was dried in a vacuum desiccator over KOH and H_2SO_4 for 1 hr. It was then used for determination of the next residue by repetition of the cycle. The complete sequence of operations, which is carried out in a single culture tube, takes about 4 hr.

The thiazolinones were converted into the corresponding PTH's, from which the original amino acids were regenerated for identification and analysis. After removal of benzene the sample was dissolved in 0.2 ml of 1.0 M HCl and was immersed for 10 min in a bath at 80°. The thiohydantoin was extracted with three 1-ml portions of ethyl acetate. The solvent was evaporated and the residue was hydrolyzed in a sealed, evacuated tube with 0.1 M NaOH for 12 hr at 120° according to the procedure of Van Orden and Carpenter (1964). The regenerated amino acid was analyzed on the amino acid analyzer.

Benzene, trifluoroacetic acid, and PITC were purified as described by Edman and Begg (1967). Dioxane was purified according to the procedure of Fieser (1957). Acetic acid and ethyl acetate were redistilled reagent grade materials and *N*-ethylmorpholine was triply distilled.

Results

The Peptide Map. The electrophoresis chromatography map of the major chymotryptic peptides of *G. gouldii* hemerythrin is shown in Figure 1. The peptides are designated C1 (for chymotrypsin), C2, etc., according to their position in the final sequence as determined in the succeeding paper (Klippenstein *et al.*, 1968). All of the peptides could be separated well from one another except for two of the basic peptides, C22 and C24, and three of the acidic ones, C1, C2, and C13, which gave complex spots. The peptides in these spots were distinguished by specific color reactions. Furthermore, they came out in different fractions in the Sephadex chromatography. Peptide C2 was not stained by ninhydrin; it was detected by the NaOCl–starch–KI stain and by the Ehrlich reaction. Peptide C8 appeared rust brown with ninhydrin. It strikingly resembled, in color as well as in location, a peptide isolated from the tryptic digest. This peptide, which was produced by chymotryptic contamination of trypsin, had an amino-terminal β-aspartyl residue that was responsible for the distinctive ninhydrin color (John and Young, 1954).

Gel filtration of the total digest gave the elution pattern shown in Figure 2. The first small peak is presumed to be the chymotrypsin used for digestion. The peptides were then eluted out in a continuous band, with several peaks, in about 900 ml of the eluate. There was no clear separation of these peaks, but an effective distribution of the peptides on the basis of size did take place.

The eluate was pooled into 12 fractions as indicated in Figure 2. The fractions were lyophilized to remove water and ammonium bicarbonate. The cuts were made on the basis of (1) electrophoresis of an aliquot from every fifth tube, and (2) the $V_{1/2\,h}$ (volume between sides of the peak at half-height) of pure substances at the various elution volumes as determined during calibration. We arbitrarily chose about 1.5 × $V_{1/2\,h}$ as the volume of a fraction.

Amino Acid Composition of the Peptides. The amino acid compositions of the isolated peptides are given in Table I. That C4/C17 and C7 were free tyrosine and phenylalanine, respectively, was confirmed by placing the fractions which contained them directly, without hydrolysis, on the amino acid analyzer. Peptides C5 and C6 had the same composition except for the lack of a

phenylalanine in C5; they come from the same segment of the molecule and reflect incomplete cleavage of a Leu–Phe bond. Peptides C12 and C13 were derived similarly from partial cleavage of a Leu–Met bond. The recovery of methionine in C13 was low, indicating that this residue is badly destroyed during peptide isolation. This is to be expected since the peptides, spread out on paper, are exposed to air for long periods in our isolation procedure.

Analysis of the peptide designated C16 was unsatisfactory due to contamination of the spot containing it with minor peptides. Glutamic acid (glutamine) and phenylalanine were, however, predominant in material extracted from this spot and hydrolyzed. That the glutamic acid was present as the amide was shown by the lack of movement in electrophoresis. The position of the spot is as would be expected for the peptide Gln-Phe. This sequence is known to exist as the first two residues of the tryptic peptide T57elN (T9 in present nomenclature).

The tetrapeptides C22 and C22A have the same composition except for a serine-alanine substitution. The molar ratio of amounts of C22 to C22A isolated was 6:4. The neutral peptide C18, on the other hand, has two amino acids, glycine and threonine, each present as less than one residue, but the sum of which does add up to one residue. The molar ratio of glycine to threonine was 6:4. We consider that the spot C18 actually consists of two peptides, identical except for the fact that one contains glycine and the other threonine. These peptides arise from polymorphism of the hemerythrin subunits, as will be discussed later.

The peptides in Table I account for the complete amino acid composition of hemerythrin. The molar ratios of the variant peptides satisfactorily account for the nonintegral values for serine and alanine and glycine and threonine reported previously (Groskopf *et al.*, 1966a).

The peptide mapping isolation procedure gave us peptides which were, with the exceptions mentioned above, virtually pure. The amino acid compositions of the peptides indicate that only a small fraction, perhaps no more than 5%, was used up in color development with the dilute ninhydrin. We had little difficulty from impurities eluted from the filter paper; blank spots showed only minute amounts of a few amino acids, in particular glycine and alanine, which were present to the extent of 2–10 nmoles. In contrast the peptide content of the spots was generally 100–300 nmoles.

The Minor Chymotryptic Peptides. Besides the major peptides several others present in small amounts are listed with their amino acid compositions in Table II. Most of these are derived from nonspecific cleavage by chymotrypsin; one of them is from incomplete cleavage of a Phe–Arg bond. All of them, however, are contained in the major peptides described above.

Isolation of a Few Peptides in Large Amounts. Among the tryptic peptides of hemerythrin are four, the complete sequence of which had not been worked out. These segments are contained in the chymotryptic peptides C5 (or C6), C9, C11, C12, and C20 and therefore these were isolated in large amounts.

Peptide C12 was contained in fraction 6. It was puri-

TABLE I: The Major Chymotryptic Peptides of *G. gouldii* Hemerythrin.

Designation	Amino Acid Composition
C1	Gly, 1.0 (1); Phe, 1.2 (1); Pro, 2.8 (3); Ile, 1.0 (1); Asp, 1.3 (1); Tyr, 1.2 (1)
C2	Val, 0.8 (1); Asp, 1.1 (1); Trp (1);[a] Pro, 0.9 (1); Ser, 0.9 (1); Phe, 1.1 (1)
C3	Arg, 1.0 (1); Thr, 1.0 (1); Phe, 1.0 (1)
C4	Tyr (1)[b]
C5	Ser, 0.9 (1); Ile, 1.4 (2);[c] Asp, 2.2 (2); Glu, 1.1 (1); His, 0.9 (1); Lys, 0.9 (1); Thr, 1.0 (1); Leu, 1.0 (1)
C6	Ser, 0.9 (1); Ile, 1.4 (2);[c] Asp, 2.2 (2); Glu, 1.1 (1); His, 1.0 (1); Lys, 1.0 (1); Thr, 1.0 (1); Leu, 1.0 (1); Phe, 0.9 (1)
C7	Phe (1)[b]
C8	Asp, 0.8 (1);[d] Gly, 1.1 (1); Ile, 1.2 (1); Phe, 1.0 (1)
C9	His, 1.0 (1); Leu, 3.2 (3); Ala, 2.1 (2); Ile, 1.0 (1); Asp, 4.8 (5);[e] Gly, 1.0 (1); Glu, 1.0 (1)
C10	Arg, 1.8 (2); CMCys,[f] 0.7 (1); Thr, 1.2 (1)
C11	Gly, 0.8 (1); Lys, 1.2 (1); His, 1.0 (1); Phe, 1.0 (1)
C12	Leu, 1.8 (2); Asp, 1.1 (1);[d] Glu 2.1 (2);[g] Val, 1.0 (1)
C13	Leu, 1.9 (2); Asp, 1.1 (1);[d] Glu, 2.2 (2);[g] Val, 0.8 (1); Met, 0.2 (1)[h]
C14	Glu, 0.9 (1);[d] Ala, 1.2 (1); Ser, 0.5 (1)
C15	Glu, 1.1 (1);[d] Tyr, 0.9 (1)
C16	Glu (1);[d] Phe (1)[i]
C17	Tyr (1)[b]
C18	Asp, 1.2 (1); Glu, 2.9 (3); His, 1.9 (2); Lys, 2.0 (2); Gly, 0.6;[j] Thr, 0.4;[j] Phe, 1.0 (1)
C19	Ile, 0.7 (1); His, 1.0 (1); Ala, 1.2 (1); Leu, 1.1 (1)
C20	Asp (2);[k] Trp (1)[a]
C21	Lys, 2.0 (2); Gly, 1.1 (1); Asp, 1.1 (1); Val, 0.9 (1); Trp (1)[a]
C22	Ala, 1.0 (1); Lys, 1.1 (1); Ser, 0.9 (1); Trp (1)[a]
C22A	Ala, 2.0 (2); Lys, 1.0 (1); Trp (1)[a]
C23	Leu, 0.7 (1); Val, 1.0 (1); Asp, 2.2 (2);[g] His, 1.0 (1); Ile, 1.9 (2); Lys, 1.0 (1); Thr, 1.0 (1); Phe, 1.0 (1)
C24	Lys, 1.0 (1); Tyr, 1.0 (1)
C25	Lys, 2.0 (2); Gly, 1.0 (1); Ile, 1.0 (1)

[a] Assumed from positive Ehrlich reaction. [b] Confirmed by analysis of an unhydrolyzed sample. [c] This peptide contains an Ile-Ile sequence, only 68% of which is hydrolyzed in 24 hr. [d] Amidated. [e] Of which two are amidated. [f] Carboxymethylcysteine. [g] Of which one is amidated. [h] Methionine is extensively destroyed during the isolation. [i] Peptide contaminated with minor peptides. See text. [j] Glycine and threonine add up to one residue. [k] Aspartic acid recovered on hydrolysis. Presence of two residues, of which one is amidated, deduced from electrophoretic mobility.

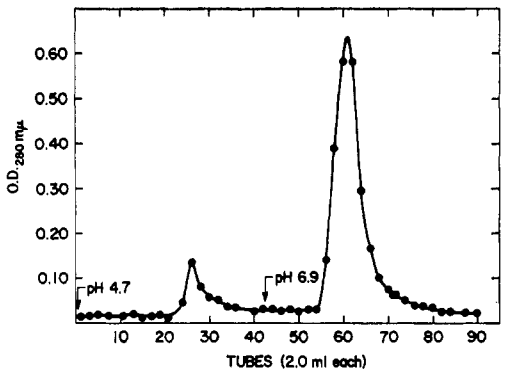

FIGURE 3: Ion-exchange purification of the peptide C9. A 20 × 1.2 cm column of Dowex 50 (Bio-Rad AG-50W-X2) was used and the elution was carried out stepwise using 0.1 M ammonium acetate solutions adjusted to the indicated pH. The second peak was pure C9.

fied by preparative paper chromatography. Peptides C5, C11, and C20 were in more complicated mixtures, and were obtained, in amounts of 0.5–1 μmole, from preparative peptide maps. Peptide C9, which was contained in fraction 2, was purified on a short Dowex 50 column (Figure 3) in 88% yield. The purified peptide contained no impurities that could be detected by peptide mapping or by amino acid analysis.

Sequence of C5. The partial sequence of this peptide was deduced from the known sequences of the tryptic peptides of hemerythrin as Ser-Ile-Ile-Asp-(Asp, Glu)-His-Lys-Thr-Leu. Selective cleavage of this peptide at the aspartyl residues by means of dilute acid should give one of two possible histidine-containing peptides: His-Lys-Thr-Leu or Glu-His-Lys-Thr-Leu. A peptide map of the dilute acid (0.02 M HCl) hydrolysate of C5 (110°, 7 hr) revealed only one major histidine-containing peptide. This peptide was isolated and it gave the following composition: Glu, 1.1 (1); His, 0.8 (1); Lys, 1.1 (1); Thr, 1.0 (1); Leu, 1.1 (1). Hence the sequence of C5 is: Ser-Ile-Ile-Asp-Asp-Glu-His-Lys-Thr-Leu.

Sequence of C9. Comparison of the composition of this peptide with the tryptic peptides gives the following incomplete sequence: His-Leu-Ala-Ile-Asp-(Asp_4, Ala, Leu, Gly)-Glu-Leu. Results of degradation of the

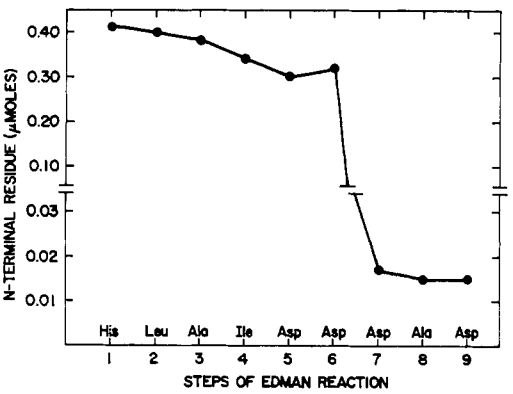

FIGURE 4: Yield of N-terminal amino acids on sequential Edman degradation of 0.4 μmole of peptide C9. Ile (step 4) is the sum of isoleucine and alloisoleucine produced by the alkaline hydrolysis of PTH-isoleucine.

TABLE II: Minor Chymotryptic Peptides of *G. gouldii* Hemerythrin.

Charge at pH 6.4	Part of Major Peptide(s)	Amino Acid Composition
0	C23	Ile, 1.7 (2); Lys, 0.9 (1); Thr, 1.0 (1); Asp, 1.3 (1); Phe, 1.0 (1)
0	C20	Asp, (1);[a] Trp (1)[b]
0	C2; C3	Val, 1.0 (1); Asp, 1.0 (1); Trp, (1);[b] Pro, 1.2 (1); Ser, 1.0 (1); Phe, 1.9 (2); Arg, 1.0 (1); Thr, 0.8 (1)
0	C22	Ser (1); Trp (1)[b]
0	C22A	Ala (1); Trp (1)[b]
−	C2	Val, 0.9 (1); Asp, 1.1 (1); Trp (1);[b] Pro, 1.1 (1); Ser, 1.0 (1)
+	C22; C22A	Ala, 1.0 (1); Lys, 1.0 (1)

[a] Amidated. [b] Assumed from positive Ehrlich reaction.

peptide from the N terminus are given in Figure 4; the C-terminal portion was assigned on the basis of fragments obtained from enzymic and dilute acid hydrolyses.

The electrophoretic behavior of C9 indicated the presence of two amide groups, and this was confirmed by the finding of 2.27 moles of amide nitrogen/mole of the peptide. The assignment of the first of these amides was based on indirect evidence stemming from observation of the yield of N-terminal residue at each of the nine steps of the Edman degradation of the peptide. As seen in Figure 4, the yield remained fairly constant from the first through the sixth step, but at the seventh step it fell steeply to about 5% of the original. It remained, however, virtually constant at this lower value in the subsequent steps. We consider this sharp fall in yield to be due to a conversion of most of the peptide remaining after the sixth step into a nondegradable form. This form could be a β-aspartyl peptide produced from the normal one through the cyclic intermediate α,β-aspartyl peptide under the influence of the trifluoroacetic acid. Such a conversion would be favored more by N-terminal asparagine than it would be by N-terminal aspartic acid. It might also be influenced by the size of the adjoining residue. Some evidence for these conversions will be presented in the Discussion. In any event the results are compatible with the sequence His-Leu-Ala-Ile-Asp-Asp-Asn-Ala-Asp- for the N-terminal segment of C9.

Dilute acid hydrolysis of C9 gave, as one of the products, a peptide which had the composition Asp, 0.9 (1); Leu, 2.0 (2); Gly, 1.0 (1); Glu, 1.0 (1). Its behavior on electrophoresis indicated the presence of one amide. This was confirmed by the isolation from a Pronase di-

gest of C9 of a dipeptide which had the composition Asp, 0.9 (1); Leu, 1.1 (1), but which was neutral at pH 6.4 and which therefore contained asparagine. Conjointly, an acidic peptide, of composition Gly, 0.8 (1); Glu, 1.1 (1); Leu, 1.1 (1), was isolated from both pepsin and subtilisin digests of C9. This peptide gave a yellow initial color with ninhydrin, indicating the presence of glycine at the N terminus, and released leucine (0.85 mole/mole of peptide in 6 hr at 38°) on incubation with carboxypeptidase A, thus confirming the sequence Gly-Glu-Leu. A peptide, of the composition Leu, 2.0 (2): Gly, 1.0 (1); Glu, 1.0 (1), was isolated from an *Aspergillus* peptidase digest of C9, indicating the sequence Leu-Gly-Glu-Leu. The above facts give for the C-terminal part of C9 the sequence Asn-Leu-Gly-Glu-Leu. The sequence of the entire peptide is therefore His-Leu-Ala-Ile-Asp-Asp-Asn-Ala-Asp-Asn-Leu-Gly-Glu-Leu.

Sequence of C11. The partial sequence derived from previous work was Gly-Lys-(His, Phe). The presence of phenylalanine at the carboxyl terminus, suggested by chymotryptic specificity, was confirmed by the observed release of 0.66 mole of phenylalanine/mole of peptide on incubation with carboxypeptidase A for 2 hr at 38°, with only traces of other residues. The sequence is therefore Gly-Lys-His-Phe.

Sequence of C12. The partial sequence derived from previous work was (Leu, Asp, Glu_2, Val)-Leu. The electrophoretic behavior of this peptide indicated the presence of two amide groups, and this was confirmed by the results of total digestion with leucine aminopeptidase, which released two residues of leucine, two of asparagine and glutamine combined, and one each of glutamic acid and valine. Four steps of Edman degradation gave, in succession and with good yield, leucine, aspartic acid (the amide group being lost by the alkaline regeneration), glutamic acid, and glutamic acid. From the leucine aminopeptidase results, one of the glutamic acid residues was known to be amidated. On short digestion with carboxypeptidase A, chiefly leucine was released, while on longer digestion, leucine, valine, and a trace of glutamic acid but not any glutamine, were released. Glutamine is known to be released by this enzyme, but not glutamic acid to any extent. If glutamine were next to the valine, one would expect some of it to be released, even with the presence of an adjacent refractory residue (glutamic acid). The sequence of C12 is therefore postulated to be Leu-Asn-Gln-Glu-Val-Leu.

Sequence of C20. The sequence of C20, (Asp, Asn)-Trp, was obtained from the study reported in the following paper (Klippenstein *et al.*, 1968). The assignment is Asp-Asn-Trp. Isolation of a minor peptide containing predominantly aspartic acid and tryptophan is consistent with this assignment, the zero electrophoretic mobility (Table II) indicating that the aspartic acid was amidated.

Discussion

The Usefulness of Gel Filtration in Peptide Separations. Gel filtration has not been generally used for separation of the total mixture of peptides from the enzymic digest of proteins, because of the obvious reason that such digests are too complex to be cleanly separated by this technique, and perhaps because of the success of ion exchangers in performing such separations. The results described in this paper show that the gel filtration technique can be used with advantage for an initial fractionation. Several of the major peptides were collected entirely in one fraction. Advantages of this procedure include, besides better recovery from the column and less damage to the peptides through milder conditions, smaller volume of eluate to deal with and elimination of the use of pyridine. Thus the effluent may be monitored directly by ultraviolet absorption, and the laborious ninhydrin assays can be circumvented.

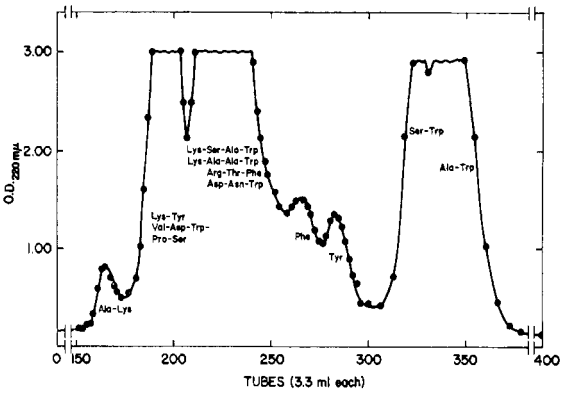

FIGURE 5: Gel filtration pattern of fraction 10 on a mixed-bed column of Sephadex G-15 and G10. The individual components in each peak were isolated and identified by preparative peptide mapping.

The separations on the G-50–G-25 column can be supplemented, for the smaller peptides, with a G-15–G-10 column. A good example is illustrated in Figure 5 which shows the elution pattern of fraction 10 on such a column. It is seen that this fraction is separated well into several components. The individual components in each peak were isolated by preparative peptide mapping for determination of their amino acid composition.

α–β Conversion of Aspartyl Peptides. We have noted such a conversion at two loci in the molecule, one based on direct observation, the other on inference. The direct evidence concerns the peptide C8. This peptide was isolated in high yield from both tryptic and chymotryptic digests. It was identified as a β-aspartyl peptide by its characteristic rust color with ninhydrin, its anodic migration at pH 1.9, and by its resistance to Edman degradation. The second residue in this peptide is glycine. The existence of a β-aspartyl bond in the intact protein is ruled out (apart from its being rather unlikely on the basis of present concepts of protein synthesis) by the results of amide nitrogen determination. This determination gave 10 amide groups/mole of subunit, of which only 9 are accounted for in all peptides other than A8. The β-aspartyl bond was therefore formed during or after the enzymic digestion. That this can occur under mild conditions is shown by the work of Haley *et al.* (1966), who found β-aspartylglycyl and β-aspartylalanyl peptides in enzymic digests of various proteins. With model peptides, conversion was shown

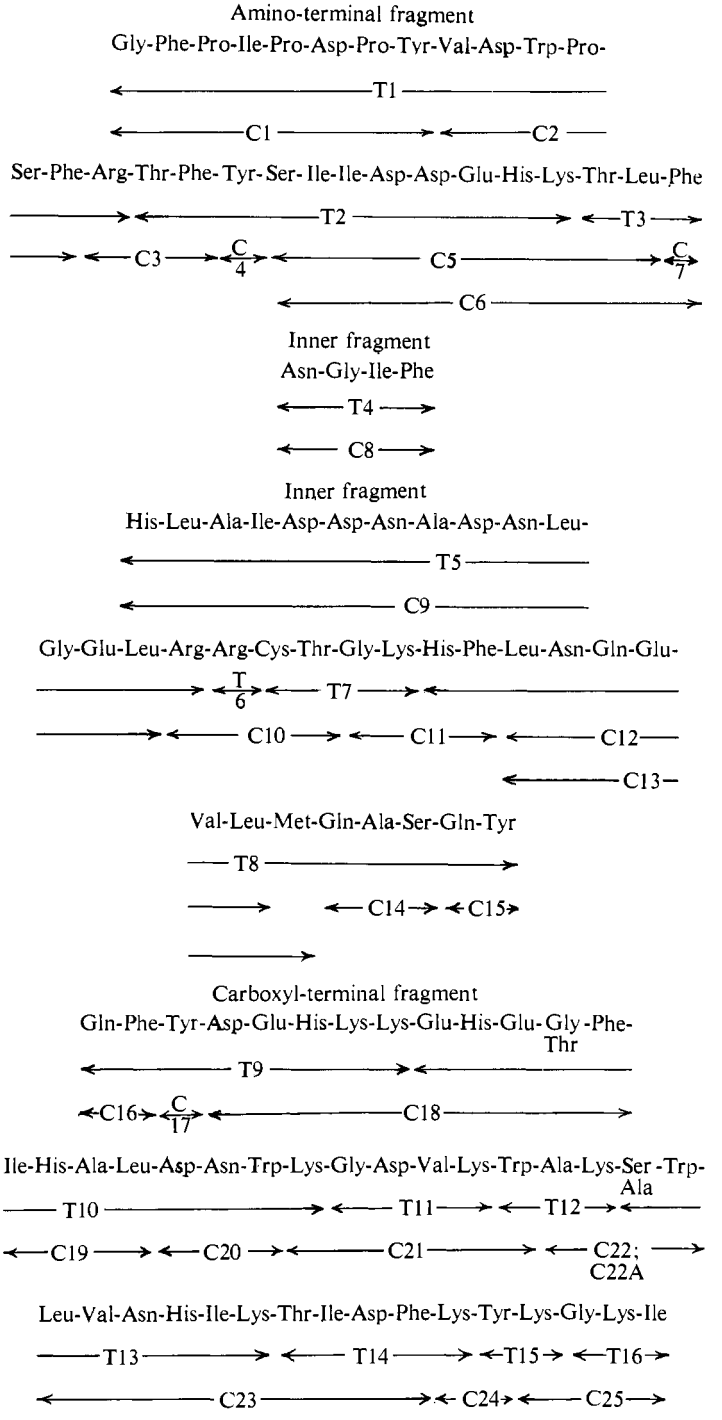

FIGURE 6: Placement of peptides within four fragments. Designation of tryptic peptides with respect to nomenclature used in previous work (Groskopf *et al.*, 1966b) is as follows: T1, TT-5ch4; T2, TT-9e1N; T3, T-27ch4; T4, T-21chl; T5, T-47elA1; T6, T-42chl; T7, TT-12elB1; T8, T-O/T-2ch2; T9, T-57elN; T10, TT-15elNch2; T11, T-40chl; T12, TT-13G3; T13, T-76chl; T14, T-45ch2elN; T15, T-59elB2; and T16, T-50ch2. Loci of amino acid variation in hemerythrin variants shown as Gly and Ser
Thr Ala
in carboxyl-terminal fragment.

to occur particularly readily in the case of asparaginyl-glycine. More recent work by Haley and Corcoran (1967) has shown conversion of an asparaginyl-glycine sequence into the β form in ribonuclease.

The presumed second case of conversion, which occurred on Edman degradation, involved an asparaginyl-alanine linkage. Konigsberg and Hill (1962) have shown that conversion of N-terminal aspartyl or asparaginyl residues into β-aspartyl linkage during the Edman degradation is minimized by the use of anhydrous trifluoroacetic acid and relatively low temperature for the cyclization rather than glacial acetic acid–HCl

and high temperature, and indeed, using a cyclization procedure which is similar to theirs we did not observe conversion in any other of our aspartyl or asparaginyl peptides. It seems therefore that the size of the adjoining residue influences the conversion and that it takes place appreciably only in case of residues with small side groups such as alanine and especially glycine.

Hemerythrin Subunit in Four Fragments. The results for the chymotryptic peptides can be combined with those on tryptic peptides to place all 113 residues of the hemerythrin subunit in sequence within four fragments, as shown in Figure 6. The order of the fragments and numbering of the tryptic and chymotryptic peptides stems from the work with chemically modified hemerythrin described in the following paper (Klippenstein *et al.*, 1968). The placement of the chymotryptic peptides within the fragments was guided to some extent by the work being done with the succinylated protein. Nonetheless, the results stand independently, and the work done in the two papers is mutually confirming.

The amino- and carboxyl-terminal fragments were identifiable from the work done previously with the tryptic peptides (Groskopf *et al.*, 1966b). The absence of data in the present paper required to place the 2 inner fragments in sequence was expected, since 3 of the 16 tryptic peptides had carboxyl termini formed from chymotrypsin-specific cleavages. Therefore no peptides that would contain these termini in the interior of a peptide could be expected to occur in a separate chymotryptic digest, and this was indeed the case.

The amino acid substitutions noted in the variant peptides C18 and C22 and C22A are discussed in the following article (Klippenstein *et al.*, 1968).

Acknowledgment

We are grateful to Dr. E. Margoliash for many helpful comments particularly during the early stages of this investigation.

References

Conway, E. J. (1962), Microdiffusion Analysis and Volumetric Error, 5th revised ed, London, Crosby, Lockwood and Son, p 98.

Edman, P., and Begg, G. (1967), *European J. Biochem. 1*, 80.

Fieser, L. F. (1957), Experiments in Organic Chemistry, 3rd ed, Boston, Mass., D. C. Heath, p 285.

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. (1966a), *Biochemistry 5*, 3779.

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

Haley, E. E., and Corcoran, B. J. (1967), *Biochemistry 6*, 2668.

Haley, E. E., Corcoran, B. J., Dorer, F. E., and Buchanan, D. L. (1966), *Biochemistry 5*, 3229.

John, W. D., and Young, G. T. (1954), *J. Chem. Soc.*, 24.

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

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

Klippenstein, G. L., Holleman, J. W., and Klotz, I. M. (1968), *Biochemistry 7*, 3868 (this issue; following paper).

Klotz, I. M., and Keresztes-Nagy, S. (1962), *Biol. Bull. 123*, 484.

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

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

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

Tsung, C. M., and Fraenkel-Conrat, H. (1965), *Biochemistry 4*, 793.

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