Chemical Studies on the Cyanogen Bromide Peptides of Rat Skin Collagen. The Covalent Structure of α1-CB5, the Major Hexose-Containing Cyanogen Bromide Peptide of α1*

William T. Butler

ABSTRACT: After cleavage of α1 chains of rat skin collagen with cyanogen bromide, the majority of its hexose was found attached to a small peptide, α1-CB5, containing 37 amino acids.

The amino acid sequence of α1-CB5 was determined by proteolytic degradation into four small peptides followed by Edman degradation and further proteolysis of each of these four. The carbohydrate consisted of glucose and galactose in amounts approaching one residue each per peptide, attached as a disaccharide to the hydroxyl group of an NH_2-terminal hydroxylysine of α1-CB5. The amino acid sequence near this glycosylated hydroxylysine is rather unusual for collagen and suggests that α1-CB5 was derived from the same area of collagen as a previously studied glycopeptide. The structure of α1-CB5 illustrates current concepts of the amino acid sequence of collagen, being composed of triplets of amino acids of the form Gly-x-y and having alternating imino acid rich and poor sequences.

Collagen is a structural protein accounting for a large percentage of the total protein of the body and playing a variety of functional roles. The principal structure involved in the formation of the fibrillar network of collagen is a molecule of about 300,000 molecular weight composed of three helically intertwined polypeptide chains (α chains) of equal size. In order to develop an understanding of the basic biochemical phenomena relating to this very important material, such as the molecular interactions of collagen molecules in the formation of fibrils, our knowledge concerning the basic chemical topography of the molecule must be greatly enhanced.

The rather large size of the collagen molecule and the preponderance of its content of a few amino acids (glycine, proline, hydroxyproline, and alanine) have made it difficult to relate chemical studies performed on solutions of the macromolecule to exact positions or even relative locations in the molecule. A systematic approach to the study of the chemical nature of the collagen molecule has employed the following step-by-step procedures: (1) the use of lathyritic preparations of rat skin collagen to provide larger amounts of material with relatively few cross-links (Piez, 1967, 1968), (2) separation of heat-denatured collagen into α1 and α2 chains (Piez *et al.*, 1963; Piez, 1967), (3) cleavage of individual α chains with cyanogen bromide (CNBr) at the few methionyl residues followed by separation and characterization of the resultant peptides (Bornstein and Piez, 1965, 1966; Butler *et al.*, 1967; Miller *et al.*, 1969; Fietzek and Piez, 1969; Kang *et al.*, 1969; Lane and Miller, 1969), (4) determination of the linear array of the CNBr peptides from each α chain (Piez *et al.*, 1968, and in preparation), (5) determination of the chemical nature of individual peptides, *e.g.*, their amino acid sequences (Kang *et al.*, 1967; Bornstein, 1967).

Eight unique CNBr peptides resulting from cleavage of the α1 chain of rat skin collagen at seven methionyl residues have been purified and characterized (Butler *et al.*, 1967). Data were presented to show that these eight peptides represent the complete sequence of the α1 chain and that the two α1 chains of rat skin collagen are either identical or differ only slightly. The molecule may thus be represented by the formula $(\alpha 1)_2\alpha 2$. Some further chemical studies on individual CNBr peptides from rat skin collagen will be presented in this series of papers.

It was observed that only one of the CNBr peptides of the α1 chain contained significant amounts of hexose. This peptide, α1-CB5, composed of 37 amino acids and containing hexoses identified as glucose and galactose (Butler *et al.*, 1967), is the fourth CNBr peptide from the NH_2 terminus and represents residues 99–135 of the α1 chain of rat skin collagen (K. A. Piez *et al.*, 1969, in preparation). The carbohydrate of collagen is known to consist of mono- and disaccharides of galactose and glucosylgalactose bound glycosidically to hydroxyl groups of hydroxylysyl residues in the macromolecule (Butler and Cunningham, 1965, 1966; Cunningham *et al.*, 1967; Cunningham and Ford, 1968; Spiro, 1969). The present paper summarizes studies on the amino acid sequence of α1-CB5.

Materials and Methods

Preparation of α1 CNBr Peptides. The method of Bornstein and Piez (1966) was followed in preparing lathyritic rat skin collagen. The α chains from denatured collagen were isolated by chromatography on CM-cellulose columns (Piez *et al.*, 1963; Bornstein and Piez, 1966). After treatment of the α1 fraction with CNBr (Bornstein and Piez, 1966), the resultant peptides from α1 chains were separated by chromatography

* From the Institute of Dental Research and Department of Biochemistry, the University of Alabama in Birmingham, Birmingham, Alabama 35233. *Received July 22, 1969.* Part of this research was carried out while I was a staff fellow of the National Institute of Dental Research, Bethesda, Md. The remainder was sponsored by U. S. Public Health Service Grant DE-02670, from the National Institute of Dental Research.

TABLE I: Amino Acid Compositions of α1-CB5 and Products Derived from its Proteolytic Digestion.[a]

Amino Acid	α1-CB5	C1	C2	T1	T2	T3
4-Hydroxyproline	2.9	2.9	–	2.8	–	–
Aspartic acid	3.1	3.0	–	1.0	1.0	0.9
Threonine	1.0	1.0	–	–	0.9	–
Serine	2.0	2.0	–	0.9	–	1.0
Glutamic acid	3.2	3.0	–	2.8	–	–
Proline	2.1	2.1	–	–	2.1	–
Glycine	12.7	10.3	2.0	5.0	3.4	2.1
Alanine	3.1	3.1	0.2	1.0	1.1	1.0
Leucine	1.0	1.0	–	–	–	1.1
Phenylalanine	0.9	–	0.7	–	–	–
Hydroxylysine	1.2	0.2	0.9	–	–	–
Lysine	1.8	1.6	–	–	1.0	0.9
Histidine	1.0	–	1.0	–	–	–
Arginine	1.0	–	1.1	–	–	–
Homoserine[b]	1.0	1.0	–	1.0	–	–
Total	37	31[c]	6	15	9	7

[a] Residues per peptide. A dash indicates the entire absence of an amino acid or that its content was less than 0.1 residue/peptide. The nomenclature is explained in the text. [b] Includes homoserine lactone. [c] In calculating the total, the value for hydroxylysine was added to that of lysine, since partial hydroxylation of these lysine sidechains has occurred (Butler, 1968).

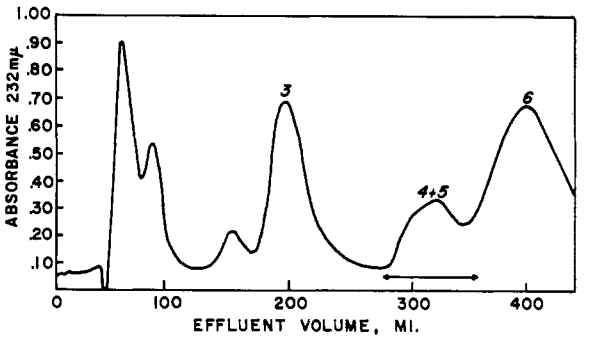

FIGURE 1: Chromatography of 300 mg of CNBr-treated lathyritic rat skin collagen on CM-cellulose at 40°. Elution was with a linear gradient formed from 760 ml each of 0.02 M sodium citrate–0.04 M NaCl (pH 3.6) (starting buffer), and 0.02 M sodium citrate–0.14 M NaCl (pH 3.6).

on CM-cellulose columns as described by Butler *et al.* (1967). It was found advantageous to limit the size of the samples to 80–100 mg and to elute peptides from the column with a linear gradient between 760 ml each of 0.02 M sodium citrate–0.04 M NaCl (pH 3.6) and 0.02 M sodium citrate–0.14 M NaCl (pH 3.6). An alternative method of isolating several of the CNBr peptides from α1 was cleavage of whole collagen with CNBr followed by CM-cellulose chromatography of this digest by the usual procedure (Figure 1). This method resulted in higher yields since chromatographic steps necessary for preparation of α chains were bypassed.

Preparation of α1-CB5. A small peptide peak appearing on the CM-cellulose chromatograms of CNBr-cleaved α1 chains (Figure 1) was found to contain two peptides (α1-CB4 and α1-CB5) of about the same size and having about the same net charge. These small peptides were separated from each other and from larger contaminating peptides by gel filtration on Sephadex G-50 (Butler *et al.*, 1967). Only partial separation of these two peptides was initially achieved; however, by combining proper fractions from several chromatographic runs and subjecting this sample to gel filtration on the same Sephadex G-50 column, a rather pure preparation of α1-CB5 (as evidenced by amino acid analysis, see Table I) was obtained. An alternative method used for purification of α1-CB5 was chromatography of the peak containing the two small peptides on phosphocellulose as described by Kang *et al.* (1969).

Enzymatic Hydrolysis of Peptides. Hydrolysis with chymotrypsin (twice crystallized; Worthington Biochemical Corp.) was effected by incubating 0.5–2.0 μmoles of peptide in 2 ml of 0.2 M NH_4HCO_3 (adjusted to pH 8.0 with 2 N HCl) with an enzyme–substrate molar ratio of 1:50. After 4 hr at 25° the reaction was stopped by the addition of 1 drop of glacial acetic acid.

Trypsin hydrolysis was performed in 2 ml of 0.1 M Tris buffer, containing 0.001 M $CaCl_2$ (pH 8.0). The peptide (0.5–2.0 μmoles) was incubated at 25° with 5% (w/w) of trypsin (Worthington, three-times crystallized) for 18 hr and the reaction was terminated by ion-exchange chromatography of the products (see below).

Digestion of peptides with collagenase (Worthington; highly purified) was accomplished by dissolving 0.3–1.0 μmole of peptide in 1 ml of 0.1 M Tris buffer containing 0.01 M $CaCl_2$ (pH 7.4), and adding collagenase at an enzyme–substrate molar ratio of 1:10. After 8 hr at 37° 1 drop of glacial acetic acid was added to the mixture to stop the reaction. When further digestion of a peptide was desired, an additional quantity (see above) of collagenase was added and digestion was continued for 18 hr.

Pepsin digestion was in 1 ml of 0.01 M HCl. Digestion of 0.2–0.4 μmole of peptide was begun by addition of pepsin as a 0.3% solution in 0.02 M sodium citrate buffer (pH 3.6), at an enzyme–substrate molar ratio of 1:50. The reaction was stopped after incubation at 25° for 3–6 hr by freeze drying the products.

Ion-Exchange Chromatography. Products resulting from trypsin hydrolysis of a peptide were separated by ion-exchange chromatography on columns of Dowex 50-X4 (Bio-Rad Laboratories) at 40°. A 0.9 × 50 cm column was equilibrated with 0.2 M pyridine acetate, pH 3.1 (Schroeder, 1967). The sample, applied in 1–2 ml, was eluted at 30 ml/hr with a linear gradient formed from 300 ml each of 0.2 M pyridine acetate, pH 3.1 (starting buffer), and 2.0 M pyridine acetate, pH 5.0 (Schroeder, 1967). Effluent fractions were tested for ninhydrin-positive material by the method of Moore and Stein (1954).

Paper Chromatography. The products of enzymic hydrolysis of a peptide were isolated by preparative paper chromatography in butanol–acetic acid–water (12:3:5) using Whatman No. 1 paper. Guide strips were spotted on each side of the area containing the material to be eluted. Normally one of the strips was stained with a cadmium–ninhydrin stain (Blackburn, 1965) and the other with a collidine–ninhydrin

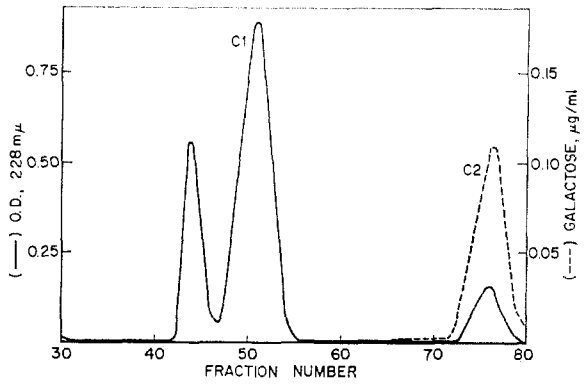

FIGURE 2: Gel filtration of the chymotryptic digest of α1-CB5 on a 1.8 × 140 cm column of Sephadex G-25. Elution was with 0.1 M acetic acid and 3.0-ml fractions were collected.

stain (Bennett, 1967). Appropriate areas with the desired peptides were cut from the paper and were eluted with water.

Paper Electrophoresis. Peptides were subjected to electrophoresis using a Gilson high-voltage electrophorator. Samples were spotted on sheets of Whatman No. 1 chromatography paper and run at pH 6.5 in pyridine acetate buffer (Butler and Cunningham, 1966) for 1 hr at 2000 V, and the dried sheets were then stained with collidine–ninhydrin (Bennett, 1967). The mobility of a peptide was compared with those of standards of arginine, glutamic acid, and asparagine to obtain information concerning its net charge at pH 6.5.

Amino Acid Analysis. Samples were hydrolyzed for 18 hr in constant-boiling HCl at 108° under nitrogen. Analyses were performed on a single-column automatic amino acid analyzer modified for high-speed analysis (Miller and Piez, 1966). No correction for losses of serine and threonine were made.

Hexose Analysis. Total hexose was measured by the method of Winzler (1955) and expressed as equivalents of galactose. The hexoses of α1-CB5 were identified by descending paper chromatography in butanol–pyridine–water (6:4:3). Before chromatography samples were hydrolyzed with 2 M HCl for 120 min at 108°. Hexoses were detected on paper chromatograms with aniline phthalate (Partridge, 1949).

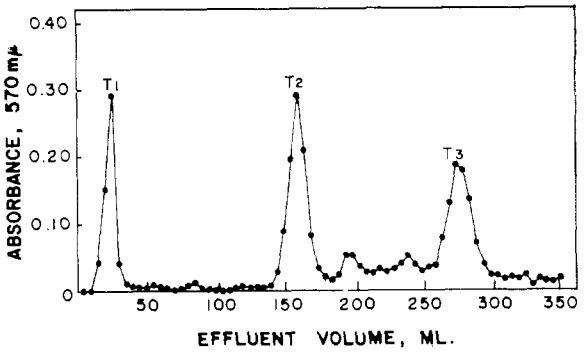

FIGURE 3: Ion-exchange chromatography of the tryptic digest of C1 (see Figure 2) on Dowex 50-X4 at 40°. Peptides were eluted from the column with a linear gradient formed from 300 ml each of 0.2 M pyridine acetate (pH 3.1) and 2.0 M pyridine acetate (pH 5.0).

Edman Degradation. Subtractive Edman degradation essentially as described by Konigsberg (1967) was used to determine the amino acid sequence of a peptide. This method allowed reliable determination of sequences of three to six amino acids. An additional method employed toward the end of this study was identification of PTH[1] derivatives of amino acids. The derivatives were spotted on thin-layer sheets of silica gel containing a fluorescent indicator (Eastman Organic Chemicals) along with the appropriate standards. The chromatograms were developed initially in solvent E of Edman and Sjöquist (1956) and when necessary solvent D. The PTH-amino acids were then located by fluorescence quenching under an ultraviolet lamp.

Results

Sequential Degradation of α1-CB5. The glycopeptide α1-CB5 was first hydrolyzed with chymotrypsin and the products were separated by gel filtration on Sephadex G-25 (Figure 2). A relatively large peptide peak (C1) emerged just after the void volume and a second peak (C2) containing the hexose was eluted just before the salt. Amino acid analysis indicated that peptide C1 contained 31 of the original 37 amino acids (Table I). The remaining six amino acids, including the residues of hydroxylysine, histidine, and phenylalanine, were found in the hexose-containing peak C2 (Table I).

The second degradative step was trypsin hydrolysis of the two lysyl (or hydroxylysyl) bonds in C1. The resultant peptides were separated on a Dowex 50-X4 column as illustrated in Figure 3. Three peptide peaks (T1, T2, and T3) were found in the expected amounts. The two lysine side chains in peptide C1 have been enzymically hydroxylated to a minor extent; therefore, smaller amounts of peptides containing hydroxylysine instead of lysine have also been found. Further evidence for partial hydroxylation of the lysyl side chains is the observation that values for lysine were always slightly lower than integrals, and those of hydroxylysine slightly higher, on amino acid analysis of α1-CB5 and of C1 (Table I). These observations and their implications have been discussed elsewhere (Butler, 1968).

Alignment of Peptides. The four peptides T1, T2, T3 (Figure 3), and C2 (Figure 2) represent the complete sequence of α1-CB5, since a summation of their amino acid contents gives a composition the same as that of α1-CB5 (Table I). The problem of the alignment of the four peptides was solved by the following experiments and reasoning.

Cleavage of methionyl bonds with CNBr leaves homoserine (lactone) as the COOH-terminal amino acid (Gross, 1967). Regarding the products of chymotryptic cleavage of α1-CB5, the presence of homoserine in C1 established that it was derived from the COOH terminus, and C2 from the NH_2 terminus of α1-CB5.

The hexaglycopeptide C2 contained an arginyl bond which was susceptible to trypsin hydrolysis (see Table I). When α1-CB5 was hydrolyzed with trypsin without prior cleavage with chymotrypsin, a peptide was found that was eluted from the Dowex 50 column in the region expected for peptide T3 (see Figure 3). Its amino acid composition was: aspartic

[1] Abbreviation used is: PTH, phenylthiohydantoin.

TABLE II: Subtractive Edman Degradation of C2.[a]

	Step				
Amino Acid	0	1	2	3	4
Glycine	2.04	1.96	1.33	1.27	1.08
Alanine	0.18	–	–	–	–
Phenylalanine	0.72	0.77	0.76	0.76	0.80
Hydroxylysine	0.95	0.05	–	–	–
Lysine	0.14	–	–	–	–
Histidine	1.00	1.19	1.06	0.59	0.39
Arginine	1.08	0.84	0.94	0.80	0.42
Sequence		Hyl	Gly	His	Arg

[a] The amino acid composition after each step is expressed as residues of amino acid per peptide. The amino acid removed at each step is indicated below the appropriate column. The composition of the original peptide is in the first column (step 0).

TABLE III: Subtractive Edman Degradation of T3.[a]

	Step				
Amino Acid	0	1	2	3	4
Aspartic acid	0.94	0.95	1.00	0.94	0.51
Serine	1.00	0.15	0.10	–	–
Glycine	2.08	2.13	1.56	1.53	1.30
Alanine	1.00	1.06	1.11	1.06	1.00
Leucine	1.12	0.94	0.89	0.43	0.26
Lysine	0.89	0.68	0.65	0.72	0.60
Sequence		Ser	Gly	Leu	Asx

[a] Explanations are in Table II.

acid, 1.10; serine, 1.05; glycine, 2.99; alanine, 1.00; leucine, 0.98; phenylalanine, 0.90; lysine, 0.97. The additional residues of phenylalanine and glycine indicate that the peptide must represent an overlapping sequence involving peptides C2 and T3, since the only phenylalanine residue of α1-CB5 was found in C2 (Table I). These observations indicate that T3 is peptide 2 in the alignment. A second consideration leading to the same conclusion is drawn from the current concept that most of the primary structure of collagen is composed of triplets of amino acids of the form Gly-x-y (Hanig and Nordwig, 1967). The only one of the remaining three peptides that, when placed adjacent to C2, allows this triplet form to be assumed is T3 (see below). Since T1 contains the homoserine residue of peptide α1-CB5, it must be peptide 4 in the alignment, allowing the deduction that T2 is peptide 3. In summary, the alignment of the four peptides from α1-CB5, beginning at its NH_2 terminus, is C2-T3-T2-T1.

Amino Acid Sequence of C2 (Residues 1–6). Four steps of subtractive Edman degradation (Table II) established the partial sequence Hyl-Gly-His-Arg-(Gly,Phe). Since the peptide was obtained after degradation of α1-CB5 with chymotrypsin, the COOH-terminal residue is most probably phenylalanine (Hill, 1965). To confirm the sequence at the COOH-terminal end, the phenylalanine-containing peptide which represents an overlap between C2 and T3 was subjected to subtractive Edman degradation. The sequential removal of glycine and phenylalanine was observed. The sequence of C2 is therefore: Hyl-Gly-His-Arg-Gly-Phe.

Amino Acid Sequence of T3 (Residues 7–13). The partial sequence Ser-Gly-Leu-Asx-(Gly,Ala)Lys was deduced from Edman degradation (Table III) and from knowledge of the specificity of the trypsin (Hill, 1965). The remainder of the sequence was proved by cleavage of the peptide at the leucyl bond with pepsin. One of the peptides isolated from the hydrolysate displayed R_F 0.12 in butanol–acetic acid–water and the following composition: aspartic acid, 0.96; glycine, 1.08; alanine, 1.02; lysine, 0.94. Two steps of Edman degradation of this tetrapeptide sequentially removed residues of aspartic acid and glycine. Since it migrated as a neutral compound on electrophoresis at pH 6.5 it must contain aspartic acid rather than asparagine. These observations established the following sequence for T3: Ser-Gly-Leu-Asp-Gly-Ala-Lys.

Amino Acid Sequence of T2 (Residues 14–22). Edman degradation (Table IV) established the partial sequence Gly-Asx-Thr-(Gly_2,Pro_2,Ala)-Lys for T2. Again it is assumed that trypsin hydrolysis leaves lysine as the COOH-terminal residue. To establish the remainder of the sequence, T2 was hydrolyzed with bacterial collagenase, an enzyme known to cleave certain peptide bonds in collagen, leaving NH_2-terminal glycyl residues (see Gallop and Seifter, 1966). Paper chromatography of the hydrolysate yielded four ninhydrin-staining components with R_F values and amino acid compositions given in Table V. The COOH-terminal sequence Gly-Pro-Lys was obtained by one-step Edman degradation of CL1. Edman degradation (two steps) of peptide CL4 resulted in the stepwise removal of glycine and proline, establishing the sequence Gly-Pro-Ala. Peptide CL2 moved as a neutral compound on electrophoresis at pH 6.5, indicating the presence of asparagine rather than aspartic acid in its sequence. These data indicate that the amino acid sequence of T2 is: Gly-Asn-Thr-Gly-Pro-Ala-Gly-Pro-Lys.

Amino Acid Sequence of T1 (Residues 23–37). In experiments employing thin-layer chromatography of PTH-amino acids after each step of Edman degradation, the following derivatives were identified: step 1, PTH-glycine; step 2,

TABLE IV: Subtractive Edman Degradation of T2.[a]

	Step			
Amino Acid	0	1	2	3
Aspartic acid	1.03	0.94	0.31	0.21
Threonine	0.89	0.91	0.83	0.37
Proline	2.15	2.17	2.13	1.91
Glycine	3.37	2.40	2.47	2.34
Alanine	1.11	1.08	1.11	1.04
Lysine	0.96	0.63	0.56	0.43
Sequence		Gly	Asx	Thr

[a] Explanations are in Table II.

TABLE V: Amino Acid Compositions and Sequences of Peptides[a] Isolated after Digestion of T2 with Collagenase.

Amino Acid	Peptide CL1 (0.14)	CL2 (0.19)	CL3 (0.26)	CL4 (0.39)
Aspartic acid	–	0.94	0.87	–
Threonine	–	0.90	0.67	
Proline	0.99	–	1.16	1.09
Glycine	1.13	1.15	2.23	1.04
Alanine	0.12	–	1.18	0.87
Lysine	0.88	–	–	–
Sequence	Gly-Pro-Lys	Gly-Asn-Thr	Gly-Asn-Thr-Gly-Pro-Ala	Gly-Pro-Ala

[a] Peptides were isolated by paper chromatography in butanol–acetic acid–water (12:3:5). Numbers in parentheses refer to R_F values. The sequences listed below each column were either directly or deductively determined (see the text). See Table I for details concerning the compositions.

TABLE VI: Subtractive Edman Degradation of T1-CL3. [a]

Amino Acid	Step 0	1	2	3
4-Hydroxyproline	0.91	0.83	0.82	0.85
Aspartic acid	1.07	1.02	1.10	0.74
Glutamic acid	1.05	1.01	0.48	0.45
Glycine	2.03	1.24	1.42	1.20
Alanine	0.96	0.91	1.08	1.08
Sequence		Gly	Glx	Asx

[a] Explanations are in Table II.

PTH-glutamic acid; step 3, PTH-hydroxyproline; step 4, PTH-glycine; step 5, PTH-serine; step 6, PTH-hydroxyproline. The results from separate experiments employing subtractive Edman degradation were consistent with these observations. Hydrolysis of T1 with collagenase and separation of the products by paper chromatography yielded only two peptides in reasonable amounts. One of the peptides had the composition: glycine, 1.15; glutamic acid, 0.95; homoserine, 1.04; the other peptide contained the remainder of the amino acids of T1. Edman degradation of the former removed a residue of glycine. Following a second step of Edman degradation PTH-glutamine was isolated and identified. Electrophoresis indicated that this peptide was uncharged at pH 6.5. These experiments allow formulation of the partial amino acid sequence Gly-Glu-Hyp-Gly-Ser-Hyp-(Gly$_2$,Hyp,-Glx,Asx,Ala)Gly-Gln-Hse for T1.

Although only one bond of T1 was readily cleaved by collagenase, a longer period of incubation with an additional quantity of the enzyme (see Materials and Methods for details) resulted in the further degradation of T1. Following such a treatment the products of collagenase digestion of T1 were partially separated by ion-exchange chromatography on Dowex 50-X4. Besides the two peptides observed in previous experiments (see above), three additional peaks of ninhydrin-positive material were found. Amino acid analysis indicated that two of these peaks contained mixtures of small peptides; they were not further studied. Another peptide, T1–CL3, was isolated in 40% yield from a third peak. It migrated as a single, negatively charged compound on electrophoresis at pH 6.5 at a rate 0.43 relative to a glutamic acid standard. Amino acid analysis gave integral values for its composition (Table VI); it was therefore judged to be rather pure. Results from subtractive Edman degradation (Table VI) indicated the partial sequence Gly-Glx-Asx-(Gly,-Ala,Hyp) for T1–CL3. Assuming that glycine occupies every third position in collagen (see Hannig and Nordwig, 1967) and that hydroxyproline is found in position 3 of the collagen triplet (the repeating sequence Gly-x-y, with glycine designated in position 1) (see Bornstein, 1967), the probable sequence of T1–CL3 is Gly-Glx-Asx-Gly-Ala-Hyp. However, since the assumptions imposed here are of a tenuous nature, the amino acid sequence of residues 10–12 of T1 are uncertain at this time. The negative charge of T1–CL3 at pH 6.5 as indicated by high-voltage electrophoresis (see above) indicates that the carboxyl group of either the aspartic acid or glutamic acid side chain (or both) is free. The partial sequence of T1 is thus: Gly-Glu-Hyp-Gly-Ser-Hyp-Gly-Glx-Asx-(Gly,Ala,-Hyp)Gly-Gln-Hse.

Bornstein (1967) has shown that certain prolines in α1-CB2 (the second CNBr peptide from the NH_2 terminus of the α1 chain of rat skin collagen) are incompletely hydroxylated. Thus any given sequence containing hydroxyproline would have an analgous sequence with proline replacing hydroxyproline. Since peptide T1 was never found to contain proline (see Table I), it is concluded that the three hydroxyprolyl residues of α1-CB5 are completely hydroxylated.

The Hexose of α1-CB5. α1-CB5 contained about 0.7 mole each of glucose and galactose per mole of peptide (Butler *et al.*, 1967). The hexose-containing peptide C2 (see Figure 2) was found in 70–80% of the amounts of C1. It contained about 2 moles of hexose/mole of peptide, chromatographically identified as glucose and galactose. These results suggest that glycosylation of the hydroxylysine of this sequence is only partially complete but approaches stoichiometric amounts of each hexose. A corresponding peptide void of hexose has not been found. Hydrolysates of the other peptides of α1 were also examined qualitatively for hexoses. Though trace amounts were found with large peptides (α1-CB7 and α1-CB8, see Butler *et al.*, 1967), the quantities of hexoses present suggested amounts far below molar ratios of hexose to peptide. Thus α1-CB5 represents the only site in the α1 chain of soluble rat skin collagen that is glycosylated to any large extent.

Discussion

The partial amino acid sequence of α1-CB5 as indicated by the studies reported here is illustrated in Figure 4. Its amino acid sequence at residues 1–4 and its carbohydrate content indicate that α1-CB5 was derived from the same

portion of the collagen molecule as that of a glycopeptide from guinea pig skin collagen (Butler and Cunningham, 1966). It was therefore assumed that the mode of attachment and the structure of the disaccharide prosthetic group are the same as that of the guinea pig skin glycopeptide (Butler and Cunningham, 1966; Cunningham and Ford, 1968). Thus the structure depicted in Figure 1 shows a glucosyl-galactose moiety attached to the δ-hydroxyl group of the NH_2-terminal hydroxylysine.

The sequence around the site of attachment of the hexose is an unusual one for collagen, containing the rare amino acids methionine, hydroxylysine, and histidine. The biological hydroxylation of the lysine in this sequence to form the NH_2-terminal hydroxylysine is complete; yet other lysines which are candidates for this reaction are only hydroxylated to a very minor extent (Butler, 1968). The glycosylation of this hydroxylysine approaches completion though several other sites in the α1 chain apparently have smaller amounts of carbohydrate prosthetic groups (see The Hexose of α1-CB5 in Results). These considerations suggest that glycosylation of these other sites has occurred to a minor extent because of the scant hydroxylation of lysines in these areas. It therefore seems possible that one limiting factor in the addition of carbohydrate to collagen is merely the biological formation of hydroxylysine by enzymic hydroxylation of lysine side chains.

The amino acid sequence represented by α1-CB5 is derived from a portion of the α1 chain of rat skin collagen beginning at residue 99 (K. A. Piez *et al.*, 1969, in preparation). In the formation of native fibrils, collagen molecules are oriented linearly in a specific staggered fashion. The staggering is non-integral (with regard to the length of the molecule) producing regions where molecules overlap (the so-called filled regions) and other regions void of material (the hole or open portions) (Hodge, 1967). A plot of the linear array and relative sizes of the CNBr peptides of α1 compared with one derived from electron microscopic observations indicates that α1-CB5 is in a hole region (K. A. Piez *et al.*, in preparation). Thus, the presence of the two bulky disaccharide side chains (assuming one for each α1 chain) would not be expected to impair fibril formation.

The collagen molecule is generally thought to be composed of alternating nonpolar regions, containing relatively large amounts of the imino acids proline and hydroxyproline, and of polar regions having relatively few imino acids (Gallop and Seifter, 1966). The amino acid sequence of peptide α1-CB5 (Figure 4) illustrates this alternating arrangement of imino acid rich and imino acid poor regions. Proline and hydroxyproline do not occur among the first seventeen residues of α1-CB5, nor in the last three; yet they occupy five of its seventeen remaining positions in the sequence. Complete exclusion of charged amino acids from the region containing proline and hydroxyproline is not observed, however, for residues of lysine (position 22) and glutamic acid (position 24) are found adjacent to proline and hydroxyproline, respectively.

Veis and Perry (1967) have isolated a component from periodate-treated dentin collagen that is apparently composed of a phosphoprotein attached to a degraded segment of the collagen backbone. In view of the fact that the component contained hydroxylysine that survived the periodate treatment in the same manner as the glycosylated hydroxylysine of the glycopeptide from guinea pig skin colagen (Butler and Cunningham, 1966), the authors proposed that the attachment of phosphoprotein to collagen may be *via* the carbohydrate. The composition reported for the phosphate-containing material suggests the possibility that it is derived from the same segment of the collagen molecule as α1-CB5 since it contains equimolar quantities of hydroxylysine, phenylalanine, and histidine. However, this tentative proposal has not at this time been tested by amino acid sequence studies on the phosphoprotein.

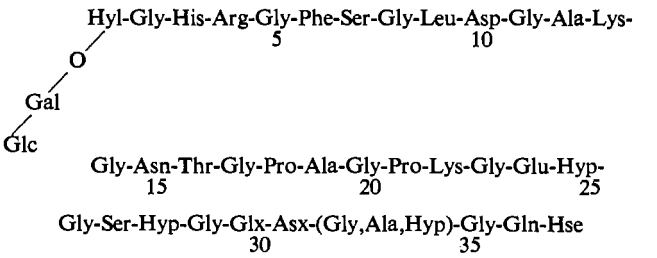

FIGURE 4: Proposed structure of α1-CB5. The disaccharide is diagrammatically shown attached to the δ-hydroxyl group of hydroxylysine at position 1 of the peptide.

Acknowledgments.

I wish to gratefully acknowledge the expert technical assistance of Miss Sharon L. Ponds and Mr. John Edward Finch, Jr., and the gracious contribution of artistic talent by Miss Jeanette Mills.

References

Bennett, J. C. (1967), *Methods Enzymol. 11*, 330.
Blackburn, S. (1965), *Methods Biochem. Anal. 13*, 1.
Bornstein, P. (1967), *Biochemistry 6*, 3082.
Bornstein, P., and Piez, K. A. (1965), *Science 148*, 1353.
Bornstein, P., and Piez, K. A. (1966), *Biochemistry 5*, 3460.
Butler, W. T. (1968), *Science 161*, 796.
Butler, W. T., and Cunningham, L. W. (1965), *J. Biol. Chem. 240*, 3449.
Butler, W. T., and Cunningham, L. W. (1966), *J. Biol. Chem. 241*, 3882.
Butler, W. T., Piez, K. A., and Bornstein, P. (1967), *Biochemistry 6*, 3771.
Cunningham, L. W., and Ford, J. D. (1968), *J. Biol. Chem. 243*, 2390.
Cunningham, L. W., Ford, J. D., and Segrest, J. P. (1967), *J. Biol. Chem. 242*, 2570.
Edman, P., and Sjöquist, J. (1956), *Acta Chem. Scand. 10*, 1507.
Fietzek, P. P., and Piez, K. A. (1969), *Biochemistry 8*, 2129.
Gallop, P. M., and Seifter, S. (1966), *Proteins 4*, 153.
Gross, E. (1967), *Methods Enzymol. 11*, 238.
Hannig, K., and Nordwig, A. (1967), *in* Treatise on Collagen, Vol. 1, Ramachandran, G. N., Ed., London, Academic, p 73.
Hill, R. L. (1965), *Advan. Protein Chem. 20*, 37.
Hodge, A. J. (1967), *in* Treatise on Collagen, Vol. 1, Ramachandran, G. N., Ed., London, Academic, p 185.
Kang, A. H., Bornstein, P., and Piez, K. A. (1967), *Biochemistry 6*, 788.

Kang, A. H., Piez, K. A., and Gross, J. (1969), *Biochemistry 8*, 1506.
Konigsberg, W. (1967), *Methods Enzymol. 11*, 461.
Lane, J. M., and Miller, E. J. (1969), *Biochemistry 8*, 2134.
Light, A. L. (1967), *Methods Enzymol. 11*, 417.
Miller, E. J., Lane, J. M., and Piez, K. A. (1969), *Biochemistry 8*, 30.
Miller, E. J., and Piez, K. A. (1966), *Anal. Biochem. 16*, 320.
Moore, S., and Stein, W. H. (1954), *J. Biol. Chem. 176*, 367.
Partridge, S. M. (1949), *Nature 164*, 443.
Piez, K. A. (1967), *in* Treatise on Collagen, Vol. 1, Ramachandran, G. N., Ed., London, Academic, p 207.
Piez, K. A. (1968), *Ann. Rev. Biochem. 37*, 547.
Piez, K. A., Bladen, H. A., Lane, J. M., Miller, E. J., Bornstein, P., Butler, W. T., and Kang, A. H. (1968), *Brookhaven Symp. Biol. 21*, 345.
Piez, K. A., Eigner, E. A., and Lewis, M. S. (1963), *Biochemistry 2*, 58.
Schroeder, W. A. (1967), *Methods Enzymol. 11*, 351.
Spiro, R. G. (1969), *J. Biol. Chem. 244*, 602.
Veis, A., and Perry, A. (1967), *Biochemistry 6*, 2409.
Winzler, R. J. (1955), *Methods Biochem. Anal. 3*, 290.

Physical and Chemical Properties of a Protein Isolated from Red Cell Membranes*

S. L. Marchesi, E. Steers, V. T. Marchesi, and T. W. Tillack

ABSTRACT: A protein component of erythrocyte membranes has been extracted and purified from human red blood cell ghosts, and has been partially characterized. The protein was solubilized from the membranes in low ionic strength aqueous solutions containing ethylenediaminetetraacetate. Purification was achieved by gel filtration and was established by polyacrylamide gel electrophoresis. In neutral salt solutions, the purified protein polymerizes into two major species with $s_{20,w}$ values between 8 and 11 S. In 6 M guanidine the aggregates are dissociated into a single monomeric unit with an $s_{20,w}$ of 1.9 S. This subunit has a molecular weight of approximately 140,000 as measured by equilibrium ultracentrifugation and polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate. The number of peptides produced by cyanogen bromide cleavage is also consistent with monomeric units of 140,000–150,000. Amino acid analysis shows a predominance of glutamic acid (20 μM %) and the presence of cysteine (1 μM %).

We have described briefly in a recent publication (Marchesi and Steers, 1968) the extraction and purification of a protein from guinea pig red blood cell ghosts. The extraction procedure described consisted of dialysis of the ghosts against ATP[1] and β-mercaptoethanol at low ionic strength. The protein solubilized in this manner accounts for approximately 20% of the total membrane protein, has an ultraviolet absorption maximum at 282 mμ, and is free of lipid and carbohydrate. Other characteristics of this protein include formation of a single broad band on polyacrylamide gel electrophoresis and a single peak in free-boundary electrophoresis over a pH range of 6.5–9.8.

Immune serum from rabbits injected with the purified membrane protein forms a single precipitin band with the protein on Ouchterlony plates. No precipitin line forms when the immune serum is run against guinea pig plasma or red cell hemolysate, indicating that the protein is not derived from plasma or red cell contents, but is specific for the red cell membrane. The suggestion was made that this protein be named spectrin.

We have extended our study of spectrin to other species including horse, sheep, rabbit, and human. A comparison of the physical, chemical, and immunologic characteristics of the spectrin obtained from these species will be reported elsewhere (Tillack *et al.*, 1970). Spectrin obtained from human red cells was studied in more detail and the results are reported here.

Our purpose has been to define some of the properties of a purified membrane protein as a step toward understanding the structural role of proteins within membranes. Some important questions are: (1) the number and size of proteins in the membrane; (2) the characteristics of these proteins which make them especially suited to a structural role; and (3) the particular arrangement of these proteins which, together with associated lipid molecules, gives the membrane its characteristic stability, insolubility, and selective permeability.

Methods

Preparation of Erythrocyte Ghosts. Blood was collected in ACD (Fenwal; 75 ml of 0.8% citric acid, 2.2% sodium citrate, 2.45% dextrose for 500 ml of blood) from normal human

* From the Laboratories of Chemical Biology and Chemical Pathology, National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, Bethesda, Maryland 20014. *Received August 28, 1969.*

[1] The use of ATP in this sample is explained in the Discussion.