Chemical Studies on the Cyanogen Bromide Peptides of Rat Skin Collagen. Amino Acid Sequence of α1-CB3†

William T. Butler,* Sharon Ponds Underwood, and John Edward Finch, Jr.

ABSTRACT: The amino acid sequence of α1-CB3, containing 149 residues from the central portion of the α1(I) chain of rat skin collagen, is reported. This structure displays the repeating Gly-X-Y sequence seen throughout the helical portion of collagen α chains. The determination allows an interspecies comparison to that of α1-CB3 from calf collagen (Fietzek, P. P., Wendt, P., Kell, I., and Kühn, K., (1972), *FEBS* (*Fed. Eur. Biochem. Soc.*) *Lett. 26,* 74), the longest segment thus far available for such a comparison. Only five interspecies amino acid differences were noted. The two sequences were thus identical in 97% of the positions. On the other hand, a sequence identity level of only 80% has been noted for the first 75 residues of this sequence, when the α1(I) chain was compared to that from the α1(II) chain (Butler, W. T., Miller, E. J., Finch, J. E., Jr., and Inagami, T. (1974), *Biochem. Biophys. Res. Commun. 57,* 190).

During the past 6 years a number of studies from several laboratories have been concerned with the amino acid sequence of the α chains of collagen. Impetus to this problem came mainly from the isolation and characterization of CNBr[1] peptides which represent segments from known regions of these large polypeptide chains. Progress in this field, including that utilizing CNBr peptides, has recently been reviewed (Gallop *et al.*, 1972; Traub and Piez, 1971).

In previous publications we have reported the covalent structures of α1-CB4 and α1-CB5 from rat skin collagen (Butler, 1970; Butler and Ponds, 1971) and have described preliminary studies designed to delineate the structure of α1-CB3 from rat skin and dentin collagens (Butler, 1969; 1972). The latter peptide contains 149 amino acids and comprises residues 419–567 of the more than 1000 residues of the α1 chain. Recently the covalent structure of α1-CB3 from calf skin collagen was reported (Wendt *et al.*, 1972a; Fietzek *et al.*, 1972a). These studies consisted of isolation and characterization of the products of proteolysis of α1-CB3 with trypsin, chymotrypsin, and thermolysin, and of automated Edman degradation of α1-CB3, and some of the larger fragments derived therefrom.

We now report studies on the structure of 11 tryptic peptides derived from rat skin α1-CB3 (Butler, 1972) which, when compared to results obtained by the Munich group (Wendt *et al.*, 1972a; Fietzek *et al.*, 1972a), indicate the covalent structure of this peptide.

Materials and Methods

Tryptic Peptides of α1-CB3. The isolation of α1-CB3 and the subsequent digestion of this peptide with trypsin have been described (Butler, 1972). A modification of the method previously used for isolation of the tryptic peptides (Butler, 1972) has been employed in the latter stages of the present investigations. The modified fractionation scheme employed initial separation of peptide T6,[2] a large tryptic peptide (45 residues), from smaller fragments by molecular sieve chromatography on Sephadex G-50 (Figure 1). The mixture of smaller tryptic peptides, contained in the volume from 320 to 425 ml of this chromatogram, was further fractionated on a 0.9 × 150 cm column of Chromobeads, Type A resin (Technicon) (Figure 2). Peptides T5 and T11, which eluted together in the void volume of the latter column, were purified by chromatography on a CM-cellulose column (2.5 × 15 cm), which was eluted with 0.02 M sodium citrate buffer (pH 3.6) at 40°. Peptide T11 eluted in the void volume (40–50 ml) of this column, while peptide T5 was slightly retained and eluted at 120–140 ml. In order to separate peptides T9 and T10 (see Figure 2) the mixture from the Chromobeads A column was subjected to CM-cellulose chromatography utilizing a concave NaCl gradient in 0.02 M sodium acetate buffer (pH 4.8) (Figure 3).

Edman Degradation. This technique was performed essentially as described (Butler and Ponds, 1971). Extraction of reagents and of PTH-amino acids with butyl acetate, instead of ethyl acetate, following the coupling and cleavage steps of this procedure, reduced the amount of peptide extracted into the organic phase, and allowed more information to be obtained. In most cases a number of steps of Edman degradation could be performed on individual tryptic peptides, and unambiguous identification of the resultant PTH-amino acids could be obtained if 1.0–1.5 μmol of a peptide of high purity were used as starting material for the reaction. The PTH-amino acids were identified by thin-layer chromatography on flexible plates of silica gel (Eastman Organic Chemicals) that contained a fluorescent indicator. The chromatograms were developed in solvents D and E of Edman and Sjöquist (1956).

Enzymatic Cleavage of Tryptic Peptides. The procedures used to cleave peptides with collagenase (Worthington, highly purified) and with pepsin (Worthington, twice crystallized) were as described (Butler, 1970). Digestion of peptides with chymotrypsin (Worthington, twice crystallized) was accomplished by dissolving 0.5–1.0 μmol of peptide in 1 ml of 0.05 M Tris buffer (pH 7.4) and adding 1% (w/w) chymotrypsin. After incubation at 37° for 3 hr the reaction was stopped by

† From the Institute of Dental Research and Department of Biochemistry, The University of Alabama in Birmingham, Birmingham, Alabama 35294. *Received February 15, 1974.* This research was sponsored by U. S. Public Health Service Grant DE-02670 from the National Institute of Dental Research.

[1] Abbreviations used are: CNBr, cyanogen bromide; PTH, phenylthiohydantoin; PITC, phenyl isothiocyanate, Hse, homoserine.

[2] The tryptic peptides were initially named according to positions of elution from a Dowex 50 chromatogram (Butler, 1972). For clarity these names have been changed to correspond to the alignment of the tryptic peptides in α1-CB3 (Fietzek *et al.*, 1972a).

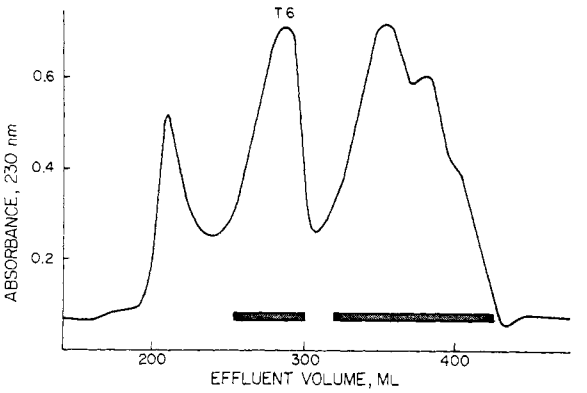

FIGURE 1: Gel filtration on Sephadex G-50 of a trypsin digest of approximately 50 mg of α1-CB3. The column (1.8 × 40 cm) was equilibrated and eluted with 0.2 M acetic acid. The two desired fractions (see text) were pooled as indicated by the bars.

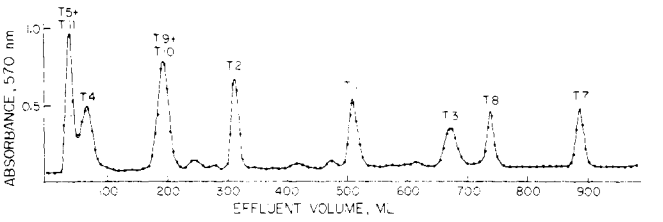

FIGURE 2: Ion-exchange chromatography of the small tryptic peptides of α1-CB3, found in the large included peak of the Sephadex G-50 chromatogram (see Figure 1). The 0.9 × 150 cm column of Chromobeads A was equilibrated with 0.2 M pyridine acetate (pH 3.1) at 24°. It was then eluted with 50 ml of the pH 3.1 buffer and finally with a linear gradient formed from 750 ml each of 0.2 M pyridine acetate (pH 3.1) and of 2.0 M pyridine acetate (pH 5.0). Fractions of 5 ml were collected and analyzed for ninhydrin-positive material with a Technicon Auto-Analyzer.

freezing the sample or by separating the products by gel chromatography.

Amino Acid Analysis. This technique was performed on samples that were hydrolyzed for 18 hr in constant boiling HCl at 108°. The hydrolysates were then analyzed with a Beckman 120C amino acid analyzer, modified for single column, high speed analysis (Miller and Piez, 1966). No corrections for losses, or incomplete release, of amino acids during acid hydrolysis were made.

Results

Character of the Tryptic Peptides of α1-CB3. The four chromatographic steps outlined in Materials and Methods allowed complete separation of 11 tryptic peptides from α1-CB3. Amino acid analysis of these peptides revealed that they were of high purity, since in all cases, calculation of the number of residues per peptide, from the observed yield of each amino acid (in μmoles), yielded values equal or close to integrals (Table I). A summation of the number of residues of each amino acid agreed with the composition reported for α1-CB3 (Butler *et al.*, 1967) in all cases, except for hydroxyproline, proline, and glycine (Table I). The computation utilizing the tryptic peptides was more accurate, as evidenced by amino acid sequence studies (see below). The size of the tryptic peptides ranged from a tripeptide (T7) to one containing forty-five residues (T6).

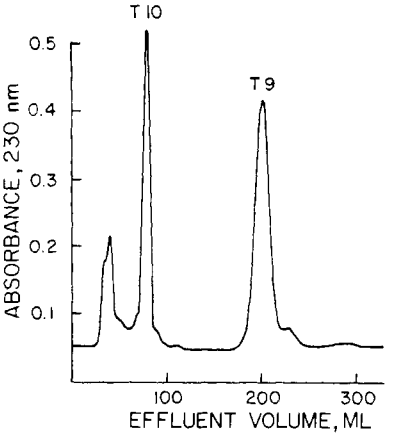

FIGURE 3: Separation of peptides T9 and T10 (see Figure 2) by CM-cellulose chromatography. The column (2.5 × 15 cm) was equilibrated with 0.02 M sodium acetate (pH 4.8) at 40°. Elution was with a concave gradient formed from 1000 ml of 0.02 M sodium acetate (pH 4.8) and 740 ml of 0.02 M sodium acetate–0.14 M NaCl (pH 4.8).

The Alignment of the Tryptic Peptides. Several experiments were performed to elucidate the alignment of the 11 peptides. First, uncleaved α1-CB3 was subjected to Edman degradation. Starting with 1.5 μmol of the peptide the following NH_2-termi-

TABLE I: Amino Acid Composition of the Tryptic Peptides of α1-CB3.[a]

Amino Acid	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	Total	α1-CB3[b]
4-Hydroxyproline	0.9	1.2	0	2.0	1.1	4.9	0	1.1	0.9	1.0	3.0	16	14
Aspartic acid	0	0	0	0	1.0	1.0	0	0	0	4.0	1.0	7	7
Threonine	0	0.9	0	0	0	0	0	0	0	0	1.0	2	2
Serine	0	0	0	0	0.1	1.8	0	0	0	0	1.0	3	3
Homoserine	0	0	0	0	0	0	0	0	0	0	1.0	1	1
Glutamic acid	0	1.0	1.1	0	3.0	6.1	1.0	1.0	0.9	0	2.0	16	16
Proline	1.1	0	0	2.0	2.2	4.8	0	0	3.1	0	0,2	13	14
Glycine	2.1	2.9	1.0	4.9	6.2	15.5	1.0	2.0	4.1	4.1	6.9	50	51
Alanine	0	1.1	0.9	2.1	5.0	5.2	0	0	1.1	2.1	2.9	20	20
Valine	0	0	0	1.9	0	0.9	0	0	0.9	0	0	4	4
Leucine	0	0	0	0	0	2.0	0	0	0	0	1.0	3	3
Phenylalanine	1.0	0	0	0	0	0.9	0	0.9	0	0	0	3	3
Lysine	1.0	1.0	0	1.0	0	1.0	0	0	0	0.9	0	5	5[c]
Arginine	0	0	1.0	0	1.0	1.0	1.0	1.0	1.0	0	0	6	6
Total	6	8	4	14	19	45	3	6	12	12	20	149	149

[a] Values are expressed as residues per peptide. A value of zero indicates that the level was less than 0.1 residue per peptide. [b] From Butler *et al.* (1967). [c] Total of lysine plus hydroxylysine.

TABLE II: Composition of Three Peptides Resulting from Hydrolysis of α1-CB3 with Pepsin.[a]

Amino Acid	Peptide P2	P3	P4
4-Hydroxyproline	3.1	1.8	1.8
Aspartic acid	1.1	—	—
Threonine	—	1.0	—
Serine	—	—	0.9
Glutamic acid	3.0	2.3	2.1
Proline	4.0	1.2	1.2
Glycine	10.8	6.6	4.3
Alanine	6.8	2.2	1.1
Valine	1.9	—	—
Leucine	0.1	—	—
Phenylalanine	—	0.9	1.0
Lysine	1.0	1.8	
Arginine	1.0	—	1.0
Total	33	17	12

[a] Compositions are given in residues per peptide. A dash indicates the complete absence of an amino acid or levels below 0.1 residue.

nal sequence was observed: Gly-Phe-Hyp-Gly-Pro-Lys-Gly-Thr-Ala-Gly-Glu-Hyp-Gly-Lys-Ala-Gly-Glu-X-Gly-Val-Hyp-Gly-Pro-Hyp-Gly. These results, along with amino acid compositions (Table I) and further sequence studies on the individual tryptic peptides (see below), show that the first four peptides in the alignment are T1-T2-T3-T4.

To obtain additional information concerning the alignment, fragments resulting from the hydrolysis of α1-CB3 with pepsin were isolated and characterized. Though these experiments were not completely successful, possibly due to differential reaction rates on a number of peptide bonds, three peptides were isolated in high yield by Dowex 50 chromatography (Butler, 1972). The compositions of these three peptides are given in Table II. The threonine and phenylalanine residues of peptide P3, along with two residues of lysine, strongly suggested that this peptide was an overlap between tryptic peptides T1, T2, and T3. Indeed the composition exactly matches that of the first 17 residues of α1-CB3, as determined by Edman degradation (see above).

Edman degradation of peptide P2 gave the partial sequence Arg-Gly-Val-Hyp-Gly-Pro-Hyp-Gly-Ala-Val showing that this portion of P2 is an overlap between the arginyl residue at the COOH-terminus of T3 and peptide T4. The remainder of the amino acids suggested strongly that peptide P2 also contained an overlap between tryptic peptides T4 and T5 (*e.g.*, the high level of alanine). Hydrolysis of peptide P2 with trypsin liberated a product which eluted in the void volume of the Dowex 50 chromatogram and which gave an analysis identical with that of peptide T5, except for the deletion of an arginine. These data indicate that peptide P2 is an overlap of peptides T3, T4, and T5, and that pepsin has cleaved the penultimate peptide bond of T5 (NH_2-terminal to its arginine residue).

Peptide P4 contained 11 amino acids known to arise from the NH_2-terminus of peptide T6 (see data on the cleavage of T6 with chymotrypsin, below), plus an arginine, suggesting that peptide P4 contained an overlap between T5 and T6. Edman degradation of P4 gave the sequence: Arg-Gly-Glu-Gln, confirming the suspicion that the NH_2-terminus of T6 was involved. Although the information does not lead to the definite conclusion that the arginine arises from the COOH-terminus of peptide T5 (several of the tryptic peptides not yet placed in the alignment contain COOH-terminal arginines), the conclusion is consistent with the alignment reported by Fietzek *et al.* (1972a).

Tryptic peptide T11 contained homoserine but no arginine or lysine and is thus the COOH-terminal peptide from α1-CB3. When α1-CB3 was treated with hydroxylamine, a bond in the sequence of peptide T10 was cleaved (Butler, 1969); one of the products of the reaction contained all of the amino acids of peptide T11 and nine residues of T10, but no amino acids from any of the other tryptic peptides. The data show that the alignment at the COOH-terminus of α1-CB3 is T10-T11.

The remainder of the alignment of tryptic peptides, indicated by the nomenclature, was derived from comparisons of the compositions and sequence data we have obtained for rat skin α1-CB3, with the data obtained for α1-CB3 from calf skin collagen (Wendt *et al.*, 1972a; Fietzek *et al.*, 1972a). Assurance that such a comparison results in the correct conclusion is gained by the close homology that exists between the sequence of the two peptides from rat and calf collagens.

Peptides T1, T2, and T3 (residues 1–18). Gly-Phe-Hyp-Gly-Pro-Lys-Gly-Thr-Ala-Gly-Glu-Hyp-Gly-Lys-Ala-Gly-Glu-Arg. These three peptides contained six, eight, and four residues, respectively (Table I). The sequence was determined by Edman degradation of uncleaved α1-CB3 (see above). Though residue 18 of α1-CB3 was not detected by Edman degradation, the composition of peptide T3 (Table I), and Edman degradation of peptide P2 (above), indicated it to be arginine. This sequence is identical with that from the calf with the exception that in rat collagen a threonine is found at position 8, while that in calf is alanine (Fietzek *et al.*, 1972a).

Peptide T4 (residues 19–32) contained 14 amino acids including a lysine and two valyl residues (Table I). Edman degradation indicated that the sequence of the first ten residues of peptide T4 was Gly-Val-Hyp-Gly-Pro-Hyp-Gly-Ala-Val-Gly. Cleavage of T4 with collagenase and separation of the products on Dowex 50 chromatography yielded four smaller peptides (Table III). Peptide T4-CL4 must have originated from the COOH-terminus of the tryptic peptide since it contained the lysine residue. Three steps of subtractive Edman degradation revealed that the sequence at the NH_2-terminus of T4-CL4 was Gly-Pro-Ala; since lysine is the COOH-terminal amino acid (known from the specificity of trypsin), the sequence of this pentapeptide must be Gly-Pro-Ala-Gly-Lys. The compositions of the other three peptides are consistent with the amino

TABLE III: Amino Acid Composition[a] of Peptides Resulting from Cleavage of T4 with Collagenase.

Amino Acid	Peptide T4-CL1	T4-CL2	T4-CL3	T4-CL4
4-Hydroxyproline	1.0	1.0		
Proline	0.9	0.1	0.8	1.0
Glycine	1.1	1.1	2.1	1.9
Alanine	0.1	0.1	2.3	1.1
Valine	—	0.9	0.9	—
Lysine	—	—	—	0.9
Position in peptide T4	4–6	1–3	7–12	10–14

[a] Residues per peptide. A dash indicates levels below 0.1 residue.

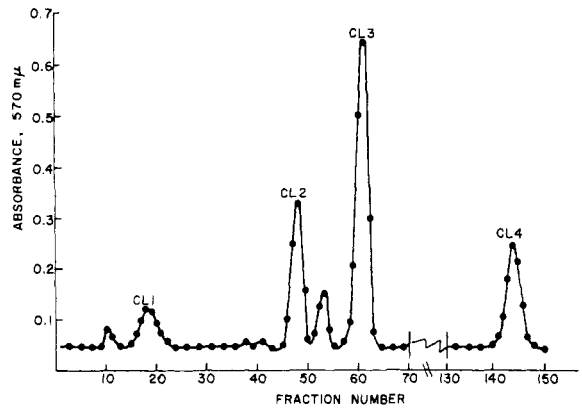

FIGURE 4: Ion-exchange chromatography of the collagenase digest of peptide T5 on a 0.9 × 150 cm column of Dowex 50-X4 at 24°. The elution conditions and assay were as described in Figure 2. Fractions of 5 ml were collected.

acid sequence of peptide T4 derived from Edman degradation. The sequence of this tetradecapeptide is thus: Gly-Val-Hyp-Gly-Pro-Hyp-Gly-Ala-Val-Gly-Pro-Ala-Gly-Lys.

Peptide T5 (residues 33–51) contained 19 amino acids including an arginine and an unusually large number (five) of alanyl residues (Table I). Edman degradation of peptide T5 indicated the partial sequence Asp-Gly-Glu-Ala-Gly-Ala-Gln-Gly-Ala-Hyp-Gly-Pro-Ala-Gly(Pro, Ala, Gly, Glx)Arg.[3] Dowex chromatography of a collagenase digest of peptide T5 (Figure 4) yielded three tripeptides and one heptapeptide with the compositions listed in Table IV. The material contained in fractions 51–54 did not originate from T5, since a relatively large quantity of serine was present in a digest of this peak. One of the tripeptides, T5-CL3, occurred in quantities about twice that of the other collagenase peptides, and thus occurs twice in the structure of peptide T5. Subtractive Edman degradation of peptide T5-CL3 sequentially removed residues of glycine and proline, showing that both sequences are Gly-Pro-Ala. One step of subtractive Edman degradation of peptide T5-CL4 removed a residue of glycine; thus the sequence at the COOH-terminus of peptide T5 must be Gly-Glx-Arg. Since the yield of ammonia from the hydrolysate of T5-CL4 was considerably lower than that of the amino acids, the second residue of this tripeptide must be glutamic acid, rather than glutamine. These data indicate that the amino acid sequence of peptide T5 is Asp-Gly-Glu-Ala-Gly-Ala-Gln-Gly-Ala-Hyp-Gly-Pro-Ala-Gly-Pro-Ala-Gly-Glu-Arg. This sequence is identical with that from calf skin except that position 9 (position 41 of α1-CB3) contains an alanyl residue in rat skin and a prolyl residue in calf skin collagen.

Peptide T6 (residues 52–56) contained 45 amino acids including a residue each of lysine, arginine, and phenylalanine (Table I). Edman degradation indicated the sequence Gly-Glu-Gln-Gly-Pro-Ala-Gly-X-Hyp-Gly-Phe-Gln-Gly-Leu-Hyp-Gly at the NH_2-terminus of peptide T6. To further characterize this large tryptic peptide, it was sequentially degraded with chymotrypsin and collagenase and the products were characterized by amino acid analysis and subtractive Edman degradation. Chymotrypsin cleaved only the Phe–Gln bond (positions 11 and 12 of T6) as indicated by the following data. After treatment of peptide T6 with chymotrypsin, two products were separated by gel filtration on Sephadex G-50. The larger of the two (C1) eluted in about the same position as peptide T6 (see Figure 1) and had the composition hydroxyproline 4.7; aspartic

[3] A single arginine or lysine in a tryptic peptide is placed at the COOH-terminus because of the specificity of trypsin.

TABLE IV: Amino Acid Composition[a] of Collagenase Peptides Derived from T5.

	Peptide			
Amino Acid	T5-CL1	T5-CL2	T5-CL3	T5-CL4
4-Hydroxyproline	—	0.8	—	—
Aspartic acid	1.0	—	—	—
Serine	—	—	—	—
Glutamic acid	1.9	—	—	0.9
Proline	—	—	1.0	—
Glycine	2.3	1.2	1.0	1.0
Alanine	2.1	1.0	0.9	—
Arginine	—	—	—	1.0
Total residues	7	3	3	3
Positions (residues) of T5	1–7	8–10	11–13 and 14–16	17–19

[a] Given in residues per peptide. A dash indicates a level below 0.1 residue.

acid, 1.1; serine, 0.9; glutamic acid, 3.9; proline, 3.6; glycine, 11.8; alanine, 3.8; valine, 0.9; leucine, 1.9; lysine, 0.8; and arginine, 0.9. The other chymotryptic peptide (C2) was well included in the gel and gave the following composition: hydroxyproline, 0.9; serine, 0.8; glutamic acid, 1.8; proline, 1.2; glycine, 4.0; alanine, 1.1; and phenylalanine, 0.8. From the sequence studies of T6, the overall composition of C2 and the occurrence of serine in peptide C2, it can be deduced that the latter peptide originated from the first 11 residues of T6, and that the unidentified amino acid at position 8 of T6 is serine.

Since the composition of peptide C1 was identical with that of the respective sequence from calf skin collagen (Fietzek *et al.*, 1972a), we strongly suspected that there were no differences in amino acid sequence between the rat and calf peptides at residues 63–96 of α1-CB3. To further test this hypothesis, peptide C1 was degraded with collagenase and the products were fractionated by ion-exchange chromatography on the Chromobeads A column (Figure 5). Peptides C1-CL1 and C1-CL2 were further purified by rechromatography on the same column, but elution was with 0.05 M pyridine acetate (pH 3.1). This procedure resulted in a stronger retention of C1-CL2 and complete separation of the two peptides. The other collagenase peptides of peptide C1 were of high purity, as indicated by integral values obtained on amino acid analysis and absence of significant levels of contaminating amino acids.

The conclusion concerning the complete sequence of peptide T6 is summarized in Figure 6. The alignment of the collagenase peptides was mainly by homology to the sequence of α1-CB3 of calf skin collagen (Fietzek *et al.*, 1972a). The position

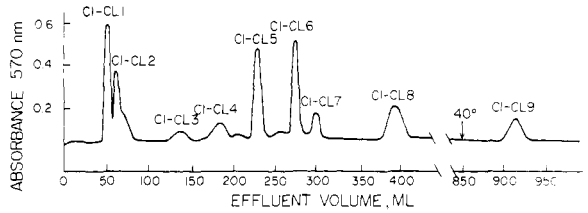

FIGURE 5: Separation of the small peptides resulting from collagenase digestion of peptide C1 on Chromobeads A. See Figure 2 for details. After collecting 850 ml of eluate, the temperature of the column was elevated to 40° in order to recover the very basic peptide C1-CL9. The graph is an artistic representation of ninhydrin-positive material found in the fractions.

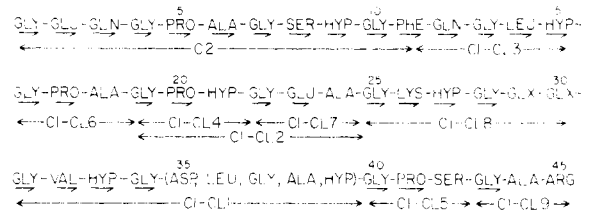

FIGURE 6: Partial amino acid sequence of peptide T6-C1. Edman degradations are marked by half-arrows; see text for details.

and sequence of C1-CL3 (residues 12–15 of C1) were known from the Edman degradation of peptide T6. Peptide C1-CL9 must be at the COOH-terminus of T6 (and C1) since the sequence studies reported below indicated that it was the only collagenase peptide with an arginine or a lysine which would be susceptible to trypsin cleavage (in α1-CB3). The other possibility, peptide C1-CL8, contained a Lys–Hyp (and Lys–Pro) bond which is resistant to trypsin hydrolysis (see Butler and Ponds, 1971).

A summary of the results of subtractive Edman degradation of the collagenase peptides of C1 can be found in Table V. As indicated in Table V, the ammonia liberated by hydrolysis of C1-CL2 and of C1-CL1 was well below stoichiometric levels, indicating that the residues at positions 23 and in the sequence from residues 35 to 39 were glutamic acid and aspartic acid, respectively (rather than glutamine and asparagine). The loss of lysine after the first cycle of Edman degradation of C1-CL8 is not unexpected, since the ε-amino group would be free to react with PITC. It is deductively reasoned that the lysine was removed after the second Edman cycle, since the levels of the other amino acids did not change. The occurrence of both hydroxyproline and proline in this sequence is due to incomplete hydroxylation of the proline, occurring during biosynthetic formation of hydroxyproline (Bornstein, 1967). Although the sequence of residues 35–39 of peptide T6 was not determined by our studies, the compositional data strongly suggest that this sequence is identical with that in the calf, namely Asp-Leu-Gly-Ala-Hyp.

Peptide T7 (residues 97–99). One step of subtractive Edman degradation on this tripeptide removed a residue of glycine. The amount of ammonia liberated by acid hydrolysis of peptide T7, as measured on the amino acid analyzer, was consistently much lower than integral values, indicating that glutamic acid rather than glutamine was present. The data therefore indicate that the sequence of T7 is Gly-Glu-Arg.

Peptide T8 (residues 100–105). Four steps of subtractive Edman degradation of this hexapeptide indicated the sequence to be Gly-Phe-Hyp-Gly-Glu-Arg. Glutamic, rather than glutamine, is specified because of the low yield of ammonia on amino acid analysis.

Peptide T9 (residues 106–117). Edman degradation and trypsin specificity indicated that the partial structure of peptide T9 was Gly-Val-Gln-Gly-Pro-Hyp-Gly-Pro-Ala-(Gly,Pro)Arg. In order to complete the determination of this sequence and to confirm the results from Edman degradation, peptide T9 was further degraded with collagenase. Four smaller peptides were isolated by Dowex chromatography in a manner similar to that used previously (see Figure 4). The compositions of these products (Table VI) were consistent with the results of Edman degradation cited above. The tripeptide T9-L64 was subjected to one step of subtractive Edman degradation; a residue of glycine was removed by this procedure indicating that the sequence at the COOH-terminus of T9 is Gly-Pro-Arg. The data demonstrate that the complete sequence of

TABLE V: Subtractive Edman Degradation of the Collagenase Peptides of C1.[a]

	Glu	Gly	Leu	Hyp			
C1-CL3 (residues 12–15)							
	0.99	1.04	0.99	0.98			
(Not subject to Edman degradation)							
		Gly	Pro	Ala			
C1-CL6 (residues 16–18)							
0		1.01	1.04	0.95			
1		*0.12*	0.95	1.04			
2		0.10	*0.13*	1.00			
		Gly	Pro	Hyp			
C1-CL4 (residues 19–21)							
0		1.05	0.93	1.01			
1		*0.23*	0.99	1.00			
2		0.20	*0.22*	1.00			
		Gly	Glu	Ala			
C1-CL7 (residues 22–24)							
0		0.98	1.01	1.01			
1		*0.19*	0.96	1.04			
2		0.21	*0.31*	1.00			
	Gly	Pro	Hyp	Glu	Ala	NH_3	
C1-CL2 (residues 19–24)							
	2.16	0.99	0.92	1.06	0.93	(0.48)	
(Not sequenced)							
		Gly	Lys	Hyp	Pro	Glu	
C1-CL8 (residues 25–30)							
0		2.02	0.98	0.69	0.38	2.02	
1		*1.01*		0.63	0.23	1.97	
2		1.02		0.68	0.21	2.00	
3		1.03		*0.26*	*0.05*	2.00	
4		*0.21*		0.10	0.00	2.00	
		Gly	Val	Hyp	Asp	Leu	Ala	NH_3
C1-CL1 (residues 31–39)								
0		3.05	0.94	1.84	1.01	1.02	1.02	(0.51)
1		*2.14*	0.92	2.20	1.01	0.95	1.13	
2		1.86	*0.20*	2.20	0.98	0.86	1.13	
3		2.10	0.12	*1.30*	0.97	0.97	1.03	
4		*1.35*		1.12	0.98	1.00	1.02	
		Gly	Pro	Ser				
C1-CL5 (residues 40–42)								
0		1.10	1.05	0.82				
1		*0.23*	1.04	0.96				
2		0.10	*0.42*	1.00				
		Gly	Ala	Arg				
C1-CL9 (residues 43–45)								
0		0.90	1.01	1.09				
1		*0.10*	0.95	1.05				

[a] Compositions for the various cycles of Edman degradation are given in residues per peptide. The number of the Edman cycle is indicated to the left of the computed values and the value for the residue removed at each cycle is italicized. For clarity the location of the collagenase peptide in peptide T6 (Figure 6) is given in parentheses.

TABLE VI: Amino Acid Composition[a] of the Collagenase Peptides of Peptide T9.

Amino Acid	Peptide T9-CL1	T9-CL2	T9-CL3	T9-CL4
4-Hydroxyproline	1.0	—	—	—
Glutamic acid	—	1.0	—	—
Proline	1.0	—	0.9	1.1
Glycine	1.0	1.1	1.0	1.0
Alanine	0.2	0.3	1.1	0.2
Valine	—	0.9	—	
Arginine	—	—	—	0.9
Total residues	3	3	3	3
Positional residues of T9	4–6	1–3	7–9	10–12

[a] Values are expressed as residues per peptide. A dash indicates the level of amino acid was either nil or below 0.1 residue.

peptide 9 is Gly-Val-Gln-Gly-Pro-Hyp-Gly-Pro-Ala-Gly-Pro-Arg.

The amino acid sequence of this portion of α1-CB3 from rat skin is identical with that from calf skin collagen except we have noted that residue 3 of peptide T9 (residue 108 of α1-CB3) is glutamine, rather than glutamic acid, as reported by Fietzek *et al.* (1972a). In identifying the PTH-amino acid at this step of Edman degradation by thin-layer chromatography, we did note a small amount of PTH-glutamic acid, but the major spot was definitely PTH-glutamine.

Peptide T10 (residues 118–129) contained 12 amino acids, including four residues each of aspartic acid and glycine (Table I). Edman degradation indicated the partial amino acid sequence Gly-Asn-Asn-Gly-Ala-Hyp-Gly(Asx_2,Gly,Ala)Lys for T10. The loss of the first four residues was followed by subtractive Edman degradation as well as by identification of the PTH-amino acids. To complete the structure determination, T10 was cleaved with collagenase and the products were separated on the Chromobeads A column. Four well-separated peptide peaks were observed, each containing a highly pure collagenase peptide (Table VII). The compositions of these peptides were consistent with the above sequence data. Peptide T10-CL4 was subjected to subtractive Edman degradation as shown in Table VIII. The data indicate the complete structure of peptide T10 to be Gly-Asn-Asn-Gly-Ala-Hyp-Gly-Asx-Asx-Gly-Ala-Lys. Comparison of this structure to that of the appropriate region of calf collagen α1-CB3 (Fietzek *et al.*, 1972a) indicates that the two differ only at position 2 (the number is relative to peptide T10); instead of the asparagine in rat, the calf peptide contains alanine. The comparison also allows the speculation that the residues at positions 8 and 9 are asparagine and aspartic acid, respectively, and identical with those in calf collagen.

TABLE VII: Composition[a] of Peptides Isolated After Cleavage of Peptide T10 with Collagenase.

Amino Acid	Peptide T10-CL1	T10-CL2	T10-CL3	T10-CL4
4-Hydroxyproline	1.0	1.0	—	—
Aspartic acid	1.9	—	2.0	2.0
Glycine	2.0	1.0	1.0	2.0
Alanine	1.1	1.0	—	1.0
Lysine	—	—	—	1.0
Recovery, %	19	31	43	73
Positional residues of T10	1–6	4–6	1–3	7–12

[a] Given as residues per peptide. A dash indicates the absence of an amino acid or that its level was below 0.05 residue.

TABLE VIII: Subtractive Edman Degradation of a Collagenase Peptide of T10.[a]

	Gly	Asp	Ala	Lys
	T10-CL4 (residues 7–12 of T10)			
0	2.05	2.00	0.95	1.00
1	*1.26*	1.96	1.00	0.66
2	1.21	*0.92*	1.11	0.56
3	0.93	*0.00*	1.07	0.32
4	*0.36*	0.00	1.00	0.10

[a] For a clarification of the format, see Table V.

Peptide T11 (residues 130–149). Edman degradation yielded the following partial sequence for peptide T11: Gly-Asp-Y-Gly-Ala-Hyp-Gly-Ala-Hyp-Gly-X-Gln-Gly-Ala(Hyp,-Glx,Gly_2,Leu)Hse. To substantiate these results and provide additional data, peptide T11 (about 0.5 μmol) was hydrolyzed with collagenase and the products were separated by ion-exchange chromatography on Dowex 50 (for a similar chromatogram see Figure 4). The composition and yields of the four resultant peptides are given in Table IX.

The results were consistent with those from Edman degradation and clarified several other points about the structure of peptide T11. The composition of T11-CL1 indicated that the unidentified residue at step 3 of Edman degradation was threonine, while the composition of T11-CL3 suggested that the unknown residue at position 11 was serine. The high yield of pep-

TABLE IX: Amino Acid Composition[a] of Peptides Released After Collagenase Digestion of Peptide T11.

Amino Acid	Peptide T11-CL1	T11-CL2	T11-CL3	T11-CL4
4-Hydroxyproline	1.0	1.0	—	—
Aspartic acid	1.0	—	—	—
Threonine	0.9	—	—	—
Serine	—	—	1.0	—
Homoserine	—	—	—	—
Glutamic acid	—	0.1	1.0	1.0
Proline	—	—	—	—
Glycine	2.0	1.1	1.0	1.1
Alanine	1.0	0.9	—	—
Leucine	—	—	—	0.9
Recovery, μmol	0.28	0.75	0.32	0.20
Positional residues of T11	1–6	7–9 and 13–15	10–12	16–18

[a] Given as residues per peptide. A dash indicates the absence of an amino acid or that its level was below 0.1 residue.

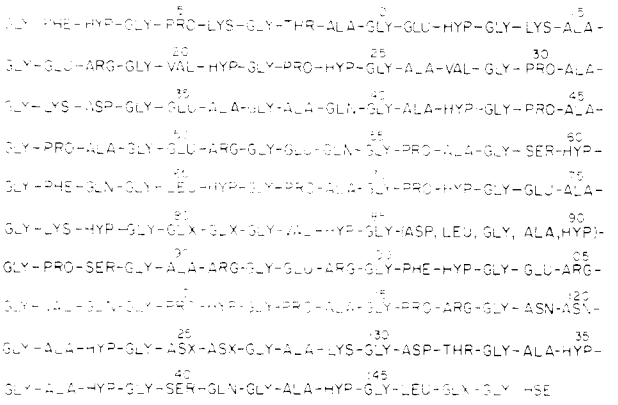

FIGURE 7: The amino acid sequence of α1-CB3 from rat skin collagen.

tide T11-CL2 was due to the occurrence of the sequence Gly-Ala-Hyp in the structure three times. The tripeptide originated mainly from residues 7–9 and 13–15, while the same sequence in residues 4–6 of T11 was obtained in good yield in the hexapeptide T11-CL1. Two steps of subtractive Edman degradation of T11-CL4 indicated the sequence Gly-Leu-Glx. For unknown reasons a peptide containing homoserine was not found in this collagenase digest. Based on the overall composition of peptide T11 (*i.e.*, 7 glycines and 1 homoserine, Table I) and on the fact that homoserine is the COOH-terminal amino acid, we conclude that the sequence Gly-Hse is at the COOH-terminus of T11. These data are consistent with the following sequence for peptide T11: Gly-Asp-Thr-Gly-Ala-Hyp-Gly-Ala-Hyp-Gly-Ser-Gln-Gly-Ala-Hyp-Gly-Leu-Glx-Gly-Hse.

Discussion

The amino acid sequence of α1-CB3 from rat skin collagen, deduced from the studies described here, is depicted in Figure 7. The results extend the known sequence of the α1(I) chain of rat collagen to 567 residues from the NH_2-terminus (Kang *et al.*, 1967; Bornstein, 1967, 1969; Butler, 1970; Butler and Ponds, 1971; Balian *et al.*, 1971, 1972). The remaining sequence from this species which remains unreported is represented by two CNBr peptides, α1-CB7 (268 residues) and α1-CB6 (196 residues), and by a short, nonhelical segment at the extreme COOH-terminus (21 residues in calf skin collagen) (Rauterberg *et al.*, 1972). The latter sequences from calf skin collagen have been determined (Rexrodt *et al.*, 1973; Fietzek *et al.*, 1972b, 1973; Wendt *et al.*, 1972b; Rauterberg *et al.*, 1972); thus the complete sequence of the α1(I) chain from rat and/or calf collagen is known.

A comparison of the sequence of rat α1-CB3 with that of calf (Fietzek *et al.*, 1972) indicates five amino acid differences: that from rat contains threonine, alanine, glutamine, asparagine, and threonine at positions 8, 41, 108, 119, and 132, respectively, while in calf these positions are occupied by alanine, proline, glutamic acid, alanine, and alanine. All five interspecies differences apparently result from point mutations with high rates of acceptance (Dayhoff, 1972), and four of the five are consistent with the change of only one nucleotide in the triplet of DNA codons. The exception is the asparagine→alanine difference which is consistent with a two-base change.

The comparative data illustrate the principle (Bornstein, 1968) that the helical portion of the collagen molecule displays a high level of interspecies sequence identity. In the case of α1-CB3 from rat and bovine collagens, the sequences are identical in about 97% of the positions. Likewise a comparison of the sequences of α2-CB4 from rat and calf collagens revealed that they were identical in more than 90% of the positions (Fietzek *et al.*, 1972c).

In contrast to the high degree of interspecies homology, the relative number of interchain differences are slightly higher. Thus in a comparison of the segment from the α1(II) chain corresponding to the first 75 residues at the NH_2-terminus of α1-CB3, we have noted a level of identity of 80% (Butler *et al.*, 1974). The level of identity is much lower in certain restricted segments of amino acids (in some cases less than 50%), even when the invariant glycines are considered in the calculations. Implications of these findings are discussed elsewhere (Butler *et al.*, 1974). The higher level of interchain (compared to interspecies) variability was also evident from the studies of Fietzek *et al.* (1972c) and Piez *et al.* (1972) comparing the sequences of regions of the α1(I) and α2 chains.

The occurrence of subintegral levels of hydroxylysine in α1-CB3 from rat skin collagen (Butler *et al.*, 1967; Bornstein, 1969) indicates the partial hydroxylation of lysine in this sequence (Butler, 1968). The observations reported here do not allow the precise location of these hydroxylysines. However, studies on the tryptic peptides from α1-CB3 of rat dentin showed the lysines in T1 and T10 to be hydroxylated to high levels (Butler, 1972). We would therefore speculate that the lysines at positions 6 and 129 are partially hydroxylated in rat skin α1-CB3. Residue 129 was observed by Wendt *et al.* (1972a) to be hydroxylated to the level of 5%. It is also of interest that these positions in the α1(II) chain are occupied by glycosylated hydroxylysine (Butler *et al.*, 1974; W. T. Butler and E. J. Miller, unpublished observations).

References

Balian, G., Click, E. M., and Bornstein, P. (1971), *Biochemistry 10*, 4470.

Balian, G., Click, E. M., Hermondson, M. A., and Bornstein, P. (1972), *Biochemistry 11*, 3798.

Bornstein, P. (1967), *Biochemistry 6*, 3082.

Bornstein, P. (1968), *Science 161*, 592.

Bornstein, P. (1969), *Biochemistry 8*, 63.

Butler, W. T. (1968), *Science 161*, 796.

Butler, W. T. (1969), *J. Biol. Chem. 244*, 3415.

Butler, W. T. (1970), *Biochemistry 9*, 44.

Butler, W. T. (1972), *Biochem. Biophys. Res. Commun. 48*, 1540.

Butler, W. T., Miller, E. J., Finch, J. E., Jr., and Inagami, T. (1974), *Biochem. Biophys. Res. Commun. 57*, 190.

Butler, W. T., Piez, K. A., and Bornstein, P. (1967), *Biochemistry 6*, 3771.

Butler, W. T., and Ponds, S. L. (1971), *Biochemistry 10*, 2076.

Dayhoff, M. D., Eck, R. V., and Park, C. M. (1972), Atlas of Protein Sequence and Structure, Vol. 5, Silver Spring, Md., National Biomedical Research Foundation, p 89.

Edman, P., and Sjöquist, J. (1956), *Acta Chem. Scand 10*, 1507.

Fietzek, P. P., Kell, I., and Kühn, K. (1972c), *FEBS (Fed. Eur. Biochem. Soc.) Lett. 26*, 66.

Fietzek, P. P., Rexrodt, F. W., Hopper, K. E., and Kühn, K. (1973), *Eur. J. Biochem. 38*, 396.

Fietzek, P. P., Rexrodt, F. W. Wendt, P., Stark, M., and Kühn, K. (1972b), *Eur. J. Biochem. 30*, 163.

Fietzek, P. P., Wendt, P., Kell, I., and Kühn, K. (1972a), *FEBS (Fed. Eur. Biochem. Soc.) Lett. 26*, 74.

Gallop, P. M., Blumenfeld, O. O., and Seifter, S. (1972), *Annu. Rev. Biochem. 41*, 617.

Kang, A. H., Bornstein, P., and Piez, K. A. (1967), *Biochemistry 6,* 788.
Miller, E. J., and Piez, K. A. (1966), *Anal. Biochem. 16,* 320.
Piez, K. A., Balian, G., Click, E. M., and Bornstein, P. (1972), *Biochem. Biophys. Res. Commun. 48,* 990.
Rauterberg, J., Fietzek, P., Rexrodt, F. W., Becker, V., Stark, M., and Kühn, K. (1972), *FEBS (Fed. Eur. Biochem. Soc.) Lett. 21,* 75.
Rexrodt, F. W., Hopper, K. E., Fietzek, P. P., and Kühn, K. (1973), *Eur. J. Biochem. 38,* 384.
Traub, W., and Piez, K. A. (1971), *Advan. Protein Chem. 25,* 243.
Wendt, P., Fietzek, P. P., and Kühn, K. (1972a) *FEBS (Fed. Eur. Biochem. Soc.) Lett. 26,* 69.
Wendt, P., von der Mark, K., Rexrodt, F., and Kühn, K. (1972b), *Eur. J. Biochem. 30,* 169.

Amino Acid Sequence of Ferredoxin from a Photosynthetic Green Bacterium, *Chlorobium limicola*†

Masaru Tanaka, Mitsuru Haniu, Kerry T. Yasunobu,* M. C. W. Evans, and Krishna K. Rao

ABSTRACT: The amino acid sequence of ferredoxin I from the photosynthetic green sulfur-reducing bacteria, *Chlorobium limicola,* was deduced to be: Ala-Leu-Tyr-Ile-Thr-Glu-Glu-Cys-Thr-Tyr-Cys-Gly-Ala-Cys-Glu-Pro-Glu-Cys-Pro-Val-Thr-Ala-Ile-Ser-Ala-Gly-Asp-Asp-Ile-Tyr-Val-Ile-Asp-Ala-Asn-Thr-Cys-Asn-Glu-Cys-Ala-Gly-Leu-Asp-Glu-Gln-Ala-Cys-Val-Ala-Val-Cys-Pro-Ala-Glu-Cys-Ile-Val-Gln-Gly. The protein consists of 60 amino acid residues and the molecular weight of the native ferredoxin was calculated to be 6923. A comparison of the sequences was made between *Chlorobium limicola* ferredoxin and the other ferredoxins which have already been sequenced. The photosynthetic bacterial ferredoxins appear to be intermediate in size when compared with the clostridial and plant ferredoxins and therefore the sequences of these ferredoxins are useful for extracting evolutionary data.

The amino acid sequences of ferredoxins from seven anaerobic fermentative bacteria (Tanaka *et al.,* 1966, 1971, 1973; Benson *et al.,* 1967; Tsunoda *et al.,* 1968; Rall *et al.,* 1969; Travis *et al.,* 1971) are known and these sequences are extremely homologous except for a species from *Desulfovibrio gigas* (Travis *et al.,* 1971). The amino acid sequence of ferredoxin from the purple sulfur photosynthetic bacterium *Chromatium,* although 26 amino acids longer than the clostridial ferredoxins, shows enough homology with the latter group to suggest a common ancestor for the two types of ferredoxins. The amino acid sequence of a ferredoxin from a green photosynthetic bacterium will be very useful in tracing the evolutionary history of anaerobic bacteria. We have now determined the sequence of the ferredoxin I, one of the two ferredoxins from *Chlorobium limicola* which was purified from the extracts of *Chloropseudomonas ethylicum. C. ethylicum* is now considered to be a mixed culture of *C. limicola* and a nonphotosynthetic bacterium (Gray *et al.,* 1972). We have therefore also prepared ferredoxin from a pure culture of *C. limicola* kindly supplied to us by Dr. J. Olson. The amino acid composition and the amino acid sequence of amino-terminal region and carboxyl-terminus of a ferredoxin from *C. limicola* and of a ferredoxin from *C. ethylicum* whose sequence we are reporting are the same.

Experimental Section

Materials. The bacteria was grown and the ferredoxin extracted as described by Rao *et al.* (1969). The ferredoxin was further purified by DEAE-cellulose column chromatography and gel filtration on Sephadex G-50. The purified protein had an A_{390}/A_{280} ratio of 0.77. Reagent grade chemicals were used and their sources have been described in previous publications (Tanaka *et al.,* 1971). Chymotrypsin was obtained from the Worthington Biochemical Corporation as three times crystallized preparation. Prior to the use, chymotrypsin was treated with L-1-tosylamido-2-lysylethyl chloromethyl ketone (Mares-Guia and Shaw, 1963). Thermolysin was purchased from Calbiochem.

Methods. Non-Heme Iron, Labile Sulfur, and Amino Acid Composition. Iron and inorganic sulfide content was determined by standard methods (Harvey *et al.,* 1955; Fogo and Popowsky, 1949; Lovenberg *et al.,* 1963) and was found to be 8 atoms each per molecule of ferredoxin assuming E_{390} = 30,000 mol^{-1} cm^{-1}. The amino acid composition of the protein and peptides was determined on acid hydrolysates in a Beckman-Spinco Model 120C automatic amino acid analyzer as described by Spackman *et al.* (1958). The instrument was equipped with high sensitivity cuvets and a 4–5 mV full scale range card.

NH_2- and COOH-terminal Residues and Sequence Determinations. The NH_2-terminal sequences of the Cm[1]-ferredoxin were determined by the Beckman-Spinco Model 890 protein/peptide sequencer utilizing the Protein Double Cleavage Pro-

† From the Department of Biochemistry-Biophysics, University of Hawaii, Honolulu, Hawaii 96822 (M. T., M. H., and K. T. Y.), and the Botany Department, University of London King's College, London, England (M. C. W. E. and K. K. R.). *Received January 31, 1974.* This project was supported by Grants GM 16784 and GM 16228 from the National Institutes of Health, the National Science Foundation (GB 18739 and GB 43448), and the Science Research Council of Great Britain.

[1] The abbreviations used are: Cm-, *S*-β-carboxymethylcysteinyl-; Cys(Cm), *S*-β-carboxymethylcysteine; PTH, phenylthiohydantoin; BPAW, 1-butanol-pyridine-acetic acid-water (60:40:12:48, v/v); BPW, 1-butanol-pyridine-water (50:50:50, v/v); and TLCK, L-1-tosyl- amido-2-lysylethyl chloromethyl ketone.