The Covalent Structure of Cartilage Collagen. Amino Acid Sequence of the NH_2-Terminal Helical Portion of the $\alpha 1$(II) Chain†

William T. Butler,* Edward J. Miller, and John Edward Finch, Jr.

ABSTRACT: The amino acid sequence of 162 residues from the NH_2-terminal region of bovine $\alpha 1$(II) is reported. Automated sequence analysis of chains from pepsin-treated type II collagen indicated the sequence and order of two CNBr peptides, $\alpha 1$(II)-CB2 and $\alpha 1$(II)-CB3, at the beginning of the repetitive triplet sequence of $\alpha 1$(II). The sequences of $\alpha 1$(II)-CB6, $\alpha 1$(II)-CB12, and 39 residues of $\alpha 1$(II)-CB11 were determined largely by automated Edman degradation. Comparative sequence data are reported which indicate that the level of homology between $\alpha 1$(I) and $\alpha 1$(II) chains in the NH_2-terminal region is about 80%. A similar level of homology was reported for the central portions of these chains (Butler, W. T., Miller, E. J., Finch, J. E., Jr., and Inagami, T. (1974), *Biochem. Biophys. Res. Commun. 57,* 190). The degree of *intraspecies* variability between chain types is thus greater than the *interspecies* variability for a single chain type. Within the sequence reported here, the $\alpha 1$(II) chain contains glucosylgalactosylhydroxylysine at three positions. The corresponding sequence of $\alpha 1$(I) contains only one glycosylated hydroxylysine with the other two positions occupied by lysyl residues.

The concept of genetic polymorphism for collagen was initially established in studies on chick sternal cartilage (Miller and Matukas, 1969; Trelstad et al., 1970; Miller, 1971, 1972). These studies showed that the majority of the collagen is comprised of three identical α chains, designated $\alpha 1$(II). This collagen with the chain composition, $[\alpha 1(II)]_3$, has now been shown to be prevalent in a number of cartilaginous tissues (for a review, see Miller, 1973) and is designated type II collagen. The more ubiquitous and previously well-characterized collagen with the chain composition, $[\alpha 1(I)]_2\alpha 2$, is now commonly referred to as type I collagen.

A third type of collagen was initially detected in studies on the CNBr[1] peptides released from infant dermis (Miller et al., 1971). The latter collagen (type III) is characterized by the chain composition $[\alpha 1(III)]_3$ (Chung and Miller, 1974; Chung et al., 1974; Epstein, 1974) and coexists with type I collagen in several tissues (Chung and Miller, 1974; Epstein, 1974; Trelstad, 1974; Butler et al., 1975). Evidence has also been reported that the collagen of basement membranes is comprised of yet additional types of chains, although their number and molecular organization have not yet been determined (Kefalides, 1971; Daniels and Chu, 1975).

The chains comprising types I, II, and III collagens have molecular weights of approximately 95 000 and exhibit amino acid compositions that are unlike other proteins, but similar to each other. Indeed, comparative sequence data on relatively short segments of $\alpha 1$(I) vs. $\alpha 2$ (Piez et al., 1972; Fietzek et al., 1972), $\alpha 1$(II) vs. $\alpha 1$(I) (Butler et al., 1974a), and $\alpha 1$(III) vs. $\alpha 1$(I) (Fietzek and Rauterberg, 1975) have established that these four polypeptides are homologous.

As part of a sustained effort to elucidate the complete covalent structure of the $\alpha 1$(II) chain, the present communication indicates the amino acid sequence of 162 residues, beginning with the repetitive triplet sequence at the NH_2-terminal portion of the chain from bovine nasal cartilage. The reported sequence is represented by the following CNBr peptides: $\alpha 1$(II)-CB2 (3 residues); $\alpha 1$(II)-CB3 (3 residues); $\alpha 1$(II)-CB6 (33 residues); $\alpha 1$(II)-CB12 (84 residues); and 39 residues at the NH_2 terminus of $\alpha 1$(II)-CB11 (Miller and Lunde, 1973). The linear order of all but two of the CNBr peptides ($\alpha 1$(II)-CB2 and $\alpha 1$(II)-CB3) along the $\alpha 1$(II) chain was established in previous studies (Miller et al., 1973). The locations of $\alpha 1$(II)-CB2 and $\alpha 1$(II)-CB3 have been determined in the present study following automated Edman degradation performed on $\alpha 1$(II) chains from which the nonhelical regions were removed by brief exposure of the native type II collagen to pepsin at low temperature.

Materials and Methods

Bovine Cartilage Collagen. Nasal septa were obtained from approximately 2-year-old cattle. The cartilaginous tissues were diced with a scalpel blade into small cubes (1 mm^3) and subsequently extracted with 15 volumes of 4.0 M guanidinium hydrochloride to remove proteoglycans and other noncollagenous components (Miller and Lunde, 1973). The tissues were then thoroughly rinsed with distilled water and lyophilized, prior to treatment with CNBr or pepsin.

Limited Pepsin Digestion of Cartilage Collagen. In order to obtain truncated $\alpha 1$(II) chains, bovine cartilage collagen was solubilized by a brief exposure to pepsin at 4 °C essentially as described (Miller, 1972). In the present studies the proteolysis with pepsin was further restricted by reducing the incubation time from 18 to 4 h. This alteration was introduced to obtain a greater degree of uniformity in the molecules solubilized by treatment with pepsin, and thereby obtain chains more suitable for automated sequence analysis.

The collagen solubilized by limited pepsin digestion was purified as described (Miller, 1972) and redissolved at a concentration of 10 mg/ml in 2.0 M guanidinium hydrochloride

† From the Institute of Dental Research and the Department of Biochemistry, the University of Alabama in Birmingham, Birmingham, Alabama 35294. *Received January 27, 1976.* 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; ANS, 2-amino-1,5-naphthalenedisulfonic acid; EDC, *N*-ethyl-*N'*-(dimethylaminopropyl)carbodiimide; Hse, homoserine; Hsl, hydroxylysine; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; CM, carboxymethyl.

(pH 7.5, 0.05 M Tris-HCl). The collagen was denatured by warming at 45 °C for 30 min, and the chains were recovered by chromatography on a standardized column of Bio-Gel A-5m (Chung et al., 1974).

Preparation of α1(II) CNBr Peptides. The initial separation of peptides on CM-cellulose after treatment of bovine nasal cartilage with CNBr has been described (Miller and Lunde, 1973). The protein in appropriate zones from several chromatograms was pooled and desalted. Peptide α1(II)-CB6 was subjected to gel filtration on 1.8 × 150 cm column of Sephadex G-50, equilibrated with 0.2 M acetic acid, to remove a small amount of α1(II)-CB4 which contaminated the preparation. Peptides α1(II)-CB11 and α1(II)-CB12, which coelute on CM-cellulose, were separated by gel chromatography on a 1.5 × 140 cm column of Bio-Gel A-1.5m by the procedure of Piez (1968). For further purification of α1(II)-CB11, the material from agarose was subjected to CM-cellulose chromatography in sodium acetate buffer (pH 4.8) as described (Butler et al., 1967).

Proteolytic Cleavage of CNBr Peptides. Peptides (1–3 μmol) were cleaved with trypsin by dissolution in 1 ml of 0.05 M Tris-HCl buffer (pH 7.4), 0.001 M $CaCl_2$, and incubating with 5% (w/w) of trypsin (Worthington Biochemical Corp., thrice crystallized) at 37 °C for 5 h. In the case of α1(II)-CB12 an additional 5% of tryspin was added and digestion continued for 18 h. Cleavage of α1(II)-CB6 with chymotrypsin was accomplished by incubating 6 mg of the peptide in 1 ml of the above buffer with 0.6 mg of chymotrypsin (Worthington, thrice crystallized). The method for cleavage with collagenase has been described (Butler, 1970).

Separation of Proteolytic Cleavage Products. Gel filtration experiments using Sephadex G-50 were performed as described (Butler et al., 1974b). Methods for separation of peptides on phosphocellulose were essentially as described (Miller, 1972); the columns were eluted with a linear gradient formed from 200 ml of each of 0.001 M sodium acetate buffer, pH 3.8 (starting buffer), and of the same buffer containing 0.15 M NaCl (limiting buffer). Small peptides resulting from trypsin or collagenase cleavage were separated on a 0.9 × 150 cm column of Chromobeads, type A resin (Technicon), utilizing pyridine acetate buffers (Butler et al., 1974b). The fractions (5 ml) were analyzed for ninhydrin-positive material with a Technicon auto-analyzer.

Edman Degradation. Automated Edman degradation was performed with a Beckman automatic sequencer (Model 890C) operated at 55 °C, and utilizing either the Slow Peptide-DMAA program (No. 071472) or the Slow Protein-Quadrol program (No. 042772) described in the Beckman Sequencer Manual. In one experiment the 0.1 M Quadrol procedure described by Brauer et al. (1975) was used. In general, repetitive yields of Pth-amino acids have been 95–96%. Pth-Amino acids were initially identified by gas–liquid chromatography (Pisano et al., 1972) with 4-ft U-shaped glass columns of 10% DC-560. Selected residues were also analyzed after trimethylsilylation. The identities of the Pth-amino acids were confirmed by thin-layer chromatography on flexible plate of silica gel (Inagami and Murakami, 1972). Pth-Arginine was identified by the phenanthrenequinone spot test, and Pth-histidine with Pauly's diazo reagent, as described in the Beckman Sequencer Manual.

In order to improve the retention of small peptides in the reaction cup of the sequencer, small peptides were modified by the attachment of ANS to the COOH-terminal amino acid, in the presence of the water soluble coupling agent EDC as described by Foster et al. (1973).

Subtractive Edman degradation was performed by the method of Balian et al. (1971).

Amino Acid Analysis. Analyses were performed either on a Beckman 120C amino acid analyzer modified for single-column analysis (Miller and Piez, 1966) or on a Beckman 119 automatic amino acid analyzer (Miller, 1972). Prior to analysis, samples were hydrolyzed at 108 °C for 24 h with constant-boiling 6 N HCl.

Results

Sequence Analysis of Pepsin-Solubilized α1(II) Chains. The uncertainty regarding the positions of α1(II)-CB2 and α1(II)-CB3 in α1(II) led us to perform automated Edman degradation experiments on slightly shortened α1(II) chains (obtained after limited pepsin digestion of bovine cartilage collagen). Application of the Slow Protein-Quadrol program gave the following major sequence:

Gly-Val-Met-Gly-Pro-Met-Gly-Pro-Y[2]-Gly-Pro-Hyp

The results establish that α1(II)-CB2 (represented by residues 1 through 3, above) and α1(II)-CB3 (residues 4 through 6, above) occur at the beginning of the repetitive triplet structure of α1(II), immediately NH_2-terminal to α1(II)-CB6 (see below).

Smaller amounts of other Pth-amino acids were observed at almost every cycle during Edman degradation of the sample of truncated α1(II) chains. The yields of the Pth-amino acids obtained in lower amounts were less than 20% of that of the major ones. The minor levels probably reflect the presence of a portion of α1(II) represented by α1(II)-CB4 which, along with α1(II)-CB1, constitutes the nonhelical NH_2 terminus of the α1(II) chain (Miller et al., 1973) and is thus NH_2 terminal to α1(II)-CB2 and α1(II)-CB3.

The Sequence of α1(II)-CB6. The complete sequence of bovine α1(II)-CB6 was elucidated in the manner summarized diagrammatically in Figure 1.

Automated sequence analysis of intact α1(II)-CB6 was performed twice. Initially, 0.75 μmol of peptide was used with the Slow Peptide-DMAA program and, in a subsequent experiment, the 0.1 M Quadrol program was employed with 1.0 μmol of peptide previously reacted with ANS in the presence of EDC. The sequence determined for the first 24 residues was as follows:

Gly-Pro-Arg-Gly-Pro-Hyp-Gly-Pro-Ala-Gly-Ala-Hyp-
Gly-Pro-Gln-Gly-Phe-Gln-Gly-Asn-Pro-Gly-Glu-Hyp

Utilizing the quantities of Pth-proline derived from cycles 5, 8, and 21, the repetitive yields were calculated to be 96 to 97%.

When cleaved with trypsin, α1(II)-CB6 yielded two peptides in approximately equal amounts, T1 and T2, resolved by gel chromatography on Sephadex G-50. These tryptic peptides contained 3 and 30 amino acid residues, respectively (Table I). The tripeptide, containing arginine, confirmed the presence of arginine in the third position of α1(II)-CB6.

After incubation of α1(II)-CB6 with chymotrypsin, the products were separated by ion-exchange chromatography on phosphocellulose into two peaks. The sizes of these two peaks were essentially identical, indicating that the yields of the two peptides were approximately equimolar. These two chymotryptic peptides contained 17 (C1) and 16 (C2) amino acids,

[2] Although this residue was not unequivocally identified at this point in our studies, later analysis of α1(II)-CB6 (see below) showed that the amino acid at this site was arginine.

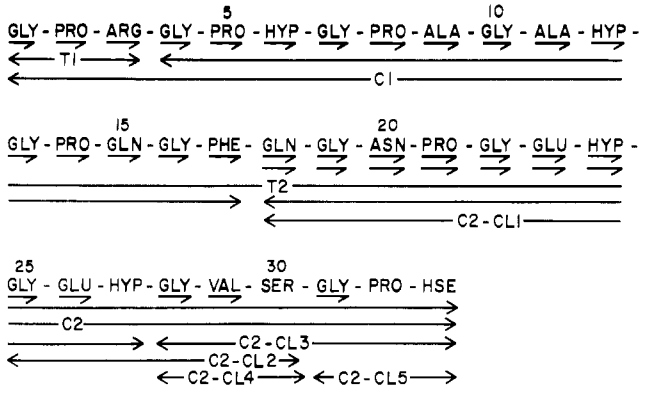

FIGURE 1: The amino acid sequence of α1(II)-CB6. The tryptic (T), chymotryptic (C), and collagenase (CL) peptides are indicated by long arrows (↔). The short half arrows (→) show the extent of Edman degradation of each peptide.

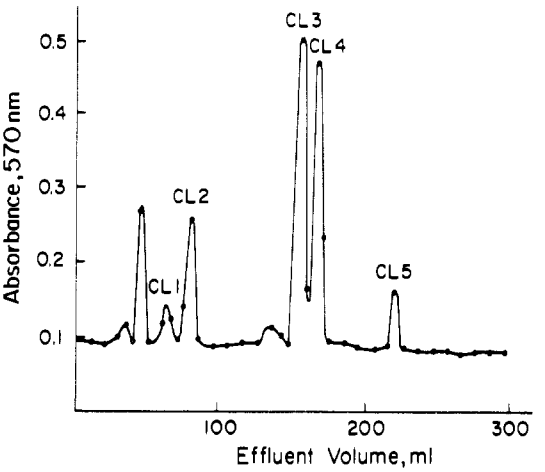

FIGURE 2: Ion-exchange chromatography (on Chromobeads A) of the small peptides resulting from treatment of α1(II)-CB6-C2 with collagenase by the method of Butler et al. (1974b).

TABLE I: Amino Acid Composition[a] of Products Obtained after Digestion of α1(II)-CB6 with Trypsin and Chymotrypsin.

	Peptide[b]			
Amino Acid	T1	T2	C1	C2
4Hyp	-	3.5	2.0	2.1
Asp	-	1.1	-	1.0
Ser	-	1.0	-	0.9
Hse	-	0.9	-	0.9
Glu	-	4.0	1.0	3.3
Pro	0.9	5.2	4.3	2.3
Gly	1.1	10.3	6.3	5.8
Ala	-	2.0	2.1	0.1
Val	-	1.0	-	1.1
Phe	-	0.9	0.7	-
Arg	1.0	-	1.0	-
Total:	3	30	17	16

[a] Values expressed as residues per peptide. Dashes indicate that values were less than 0.1 residue per peptide. [b] See text for nomenclature.

TABLE II: Compositions[a] and Subtractive Edman Degradation of the Collagenase Peptides of α1(II)-CB6-C2.[b]

	Hyp	Asp	Glu	Pro	Gly
		C2-CL1 (Residues 18-27 of α1(II)-CB6)			
	2.04	0.72	3.20	0.79	3.06
		(Not Sequenced)			
	Gly	Glu	Hyp	Val	Ser
		C2-CL2 (Residues 25-30)			
0	1.94	1.08	0.81	0.89	0.95
1	*1.15*	1.06	0.90	0.94	0.94
2	1.14	*0.30*	0.96	0.93	0.93
3	1.21	0.37	*0.00*	0.92	0.92
	Ser	Hse	Pro	Gly	Val
		C2-CL3 (Residues 28-33)			
	0.89	1.12	0.85	1.98	0.69
		(Not Sequenced)			
	Gly	Val	Ser		
		C2-CL4 (Residues 28-30			
0	1.03	0.90	0.97		
1	*0.24*	1.04	0.96		
2	0.30	*0.20*	1.00		
		Gly	Pro	Hse	
		C2-CL5 (Residues 30-33)			
	0	1.00	0.95	1.05	
	1	*0.30*	0.92	1.08	

[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. [b] See the text and Figures 1 and 2 for clarification of the nomenclature and source of the collagenase peptides.

respectively (Table I). The composition of C1 when compared with the data given above (e.g., two alanyl residues) showed that it was from the NH_2 terminus of α1(II)-CB6. This conclusion is also consistent with the specificity of chymotrypsin in cleaving peptide bonds involving the carboxyl group of phenylalanyl residues (Hill, 1965). Additionally, the homoserine in peptide C2 indicated its derivation from the COOH terminus of α1(II)-CB6. About 1.3 μmol of peptide C2 was coupled to ANS in the presence of EDC and subjected to automated Edman degradation, utilizing the Slow Peptide-DMAA program. The first six residues at the NH_2-terminal sequence of peptide C2 were established to be Gln-Gly-Asn-Pro-Gly-Glu, an observation consistent with the above conclusions. Only about 1% of the expected quantities of Pth-amino acids were recovered in this experiment, presumably because most of the peptide contained a blocked NH_2-terminal residue, resulting from cyclization of the glutaminyl residue.

In order to establish the structure of the final nine residues of α1(II)-CB6, 2 μmol of peptide C2 was digested with collagenase and the products were separated by ion-exchange chromatography (Figure 2). Five collagenase peptides were obtained with compositions indicating that they represented the complete sequence of C2. Peptides C2-CL2, C2-CL4, and C2-CL5 were subjected to subtractive Edman degradation. The results of the Edman degradations and the compositions of the five peptides are presented in Table II. The position of α1(II)-CB6 from which each collagenase peptide was derived, indicated by parentheses following each name, was deduced

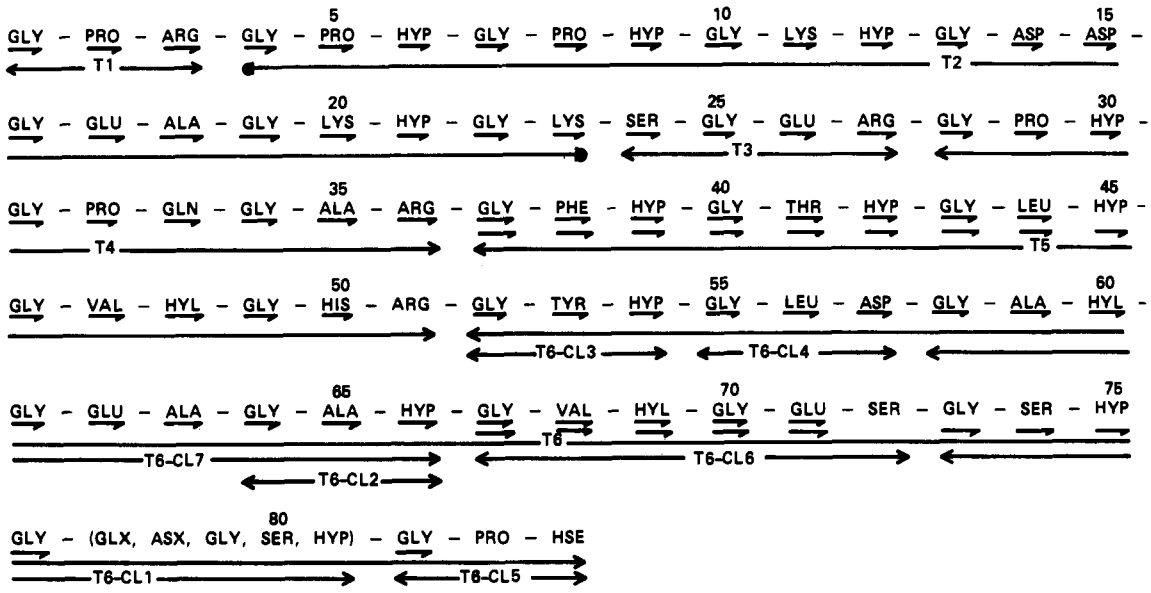

FIGURE 3: The amino acid sequence of α1(II)-CB12. Tryptic peptides (T) and collagenase peptides (CL) are designated by long arrows (↔) and Edman degradation by short half arrows (→). The hydroxylysyl residues at positions 48, 60, and 69 are glycosylated (see Results section).

TABLE III: Amino Acid Composition of the Tryptic Peptides of Bovine Nasal Cartilage α1(II)-CB12.[a]

Amino Acid	T1	T2	T3	T4	T5	T6	T6a	Total[b]	α1(II)-CB12[c]
4-Hyp	-	2.7	-	1.0	2.3	3.7	1.4	10	9
Asp	-	2.1	-	-	-	2.1	1.1	4	4
Thr	-	-	-	-	1.0	-	-	1	1
Ser	-	-	0.8	-	-	2.8	2.6	4	4
Hse	-	-	-	-	-	0.8	1.0	1	1
Glu	-	1.4	1.0	1.0	-	3.2	2.4	6	7
Pro	1.0	3.1	-	1.8	0.6	1.5	1.6	8	8
Gly	1.0	6.8	1.2	3.1	5.4	11.1	5.8	28	28
Ala	-	1.0	-	1.2	-	3.4	-	5	5
Val	-	-	-	-	0.9	0.9	-	2	2
Leu	-	-	-	0.2	1.0	1.0	-	2	2
Tyr	-	-	-	-	-	0.8	-	1	1
Phe	-	-	-	-	1.0	-	-	1	1
Hyl	-	-	-	-	0.9	1.8	-	3	3
Lys	-	2.9	-	-	-	-	-	3	3
His	-	-	-	-	0.9	-	-	1	1
Arg	1.0	-	1.0	0.9	1.0	-	-	4	4
Total:	3	20	4	9	15	33	16	84	84

[a] Given as residues per peptide. A dash indicates that the level was less than 0.1 residue per peptide. [b] Summation of the nearest integrals of the residues of each peptide. The total for the individual amino acids is for the six tryptic peptides T1-T6 and does not include the composition of T6a. [c] From Miller and Lunde, 1973.

by the following considerations (diagrammatically shown in Figure 1). The composition of C2-CL1 (e.g., one aspartic acid), when compared with the sequence data above, showed it to be from the NH_2 terminus of C2. The fact that it contained the three glutamic acids of C2 indicated that C2-CL1 overlapped C2-CL2, which contained a single glutamic acid. The valyl and seryl residues of C2-CL2 showed that this hexapeptide overlapped C2-CL4. And finally peptide C2-CL3, known to be from the COOH terminus because of the homoserine residue, overlapped C2-CL4 and C2-CL5.

The residue at position 9 of peptide C2, which was removed at the second step of subtractive Edman degradation of peptide C2-CL2 (Table II) was deduced to be glutamic acid residue rather than glutamine since amino acid analysis of peptide C2 showed a content of only slightly over 2 mol of ammonia per mol of peptide. Since glutamine and asparagine residues were identified at positions 1 and 3, respectively, of peptide C2, there can be no other amide residue in this sequence.

Amino Acid Sequence of α1(II)-CB12. The sequence of 79 of the 84 residues of α1(II)-CB12 was established by Edman degradation of the peptide and by studies on the six tryptic peptides derived from it. The results are summarized in Figure 3.

Starting with 1.1 μmol of intact α1(II)-CB12, and utilizing the Slow Protein-Quadrol program of the automated sequencer, the sequence of 45 residues at the NH_2 terminus was established (Figure 3). The yield of Pth derivative at cycle 24 was quite low. Therefore, the unambiguous identification of Ser-24 in the sequence of α1(II)-CB12 was deduced from the amino acid composition of the tryptic peptide T3 (see below).

Proteolysis of α1(II)-CB12 with trypsin yielded principally six peptides with compositions listed in Table III. The large

TABLE IV: Composition[a] of Peptides Isolated after Cleavage of Peptide T6 with Collagenase.

Amino Acid	T6-CL1	T6-CL2	T6-CL3	T6-CL4	T6-CL5	T6-CL6	T6-CL7
4Hyp	1.7	1.0	1.0	–	–	–	1.1
Asp	1.0	–	–	1.0	–	–	–
Ser	1.8	0.2	–	0.2	–	0.9	–
Hse	–	–	–	–	0.8	–	–
Glu	1.1	–	–	0.3	–	1.1	1.1
Pro	0.4	–	–	–	1.0	–	–
Gly	3.0	1.1	1.0	1.3	1.0	2.0	2.9
Ala	–	1.0	–	–	–	–	2.8
Val	–	–	–	–	–	0.8	–
Leu	–	–	–	1.0	–	–	–
Tyr	–	–	0.8	–	–	–	–
Hyl	–	–	–	–	–	0.8	0.8
Total:	9	3	3	3	3	6	9
Recovery (%):	25	48	37	21	23	18	21
Position in α1(II)-CB12:[b]	73–81	64–66	52–54	55–57	82–84	67–72	58–66

[a] Residues per peptide. A dash indicates levels below 0.1 residue. [b] See Figure 3.

tryptic peptide T6 was initially separated from smaller peptides by gel chromatography on Sephadex G-50 and finally purified by phosphocellulose chromatography. Peptides T2, T4, and T5, which emerged together on the Sephadex G-50 column, were purified by phosphocellulose chromatography, while peptides T1 and T3 were purified by ion-exchange chromatography on the Chromobeads column. The sequence analysis of the first 45 residues of α1(II)-CB12 (see above) established the alignment of the tryptic peptides T1, T2, T3, T4, and T5 (Figure 3) and the compositions of these peptides (Table III) supported this sequence analysis. A summation of the composition of the six tryptic peptides agreed very well with the composition of uncleaved α1(II)-CB12 (Miller and Lunde, 1973) so these fragments represent the complete sequence of the CNBr peptide. In addition to these peptides, a small amount of a seventh tryptic peptide T6a, eluting in the void volume of the Chromobeads column, was detected. The analysis of this peptide (Table III) indicated its origin from the COOH terminus of α1(II)-CB12 since it contained neither arginine nor lysine, but a residue of homoserine. Several possible reasons for the partial cleavage of the peptide bond joining residues 69 and 70 of α1(II)-CB12 can be offered. Position 69 may contain a small amount of lysine or of free (nonglycosylated) hydroxylysine, either of which could be susceptible to trypsin cleavage. A less likely explanation is that the glycosylated hydroxylysine in this position is slightly susceptible to cleavage by trypsin.

The sequence of peptide T5 was determined utilizing 1.0 μmol of peptide and the Slow-DMAA program of the automated sequencer (Figure 3). Arginine was known to be the COOH-terminal amino acid from trypsin specificity (Hill, 1965).

The hydroxylysine at position 48 of α1(II)-CB12 (Figure 3) was not determined by direct identification of the Pth derivative. The following data showed it to be glycosylated hydroxylysine. First, it has been our experience that, whenever such a residue is encountered in automated Edman degradation, a "blank" is observed at that cycle (Butler et al., 1974a). Cycle 12 of peptide T5 gave a "blank". Second, a residue of glucosylgalactosylhydroxylysine was identified after alkaline hydrolysis of T5 and chromatography of the hydrolysate on the amino acid analyzer (Spiro, 1969). And finally, the single glycosylated hydroxylysine of the α1(I) chain occurs in this homologous site (Butler, 1970).

The partial sequence of T6, the COOH-terminal tryptic peptide from α1(II)-CB12 was determined in the following manner. Edman degradation, utilizing 1.0 μmol of peptide and employing the Slow Peptide-DMAA program of the automatic sequencer indicated the sequence:

Gly-Tyr-Hyp-Gly-Leu-Asp-Gly-Ala-Y-Gly-Glu-Ala-
Gly-Ala-Hyp-Gly-Val-Y-Gly-Glu-Y-Gly

for the first 22 amino acids. Utilizing data for the alanines at positions 8 and 12, the repetitive yield was calculated to be 95%. After cleavage of peptide T6 with collagenase, seven peptides were isolated by ion-exchange chromatography on the Chromobeads A column. The compositions of these peptides are listed in Table IV and their derivation and sequences are shown in Figure 3. The compositions of peptides T6-CL3, T6-CL4, T6-CL7, T6-CL2, and T6-CL6 support the conclusion concerning the sequence analysis given above. The complete sequence of T6-CL6 (residues 67–72 of α1(II)-CB12) and the partial sequence of T6-CL1 (residues 73–81) were established by subtractive Edman degradation. The probable sequence of the latter portion of T6-CL1 (shown in parentheses) was deduced by comparison to the homologous α1(I) sequences of bovine and chick collagens (Fietzek and Kühn, 1975; Kang et al., 1975).

It should be noted that fractional levels of proline were observed in peptides T5, T6, and T6-CL1 (Tables III and IV). These nonintegral amounts of this amino acid undoubtedly reflect the incomplete hydroxylation of prolyl residues during the biosynthesis of hydroxyproline (Bornstein, 1967). After cycle 3 of subtractive Edman degradation of T6-CL1 (results given above), about 0.3 of a residue of proline and 0.7 residue of hydroxyproline were lost, indicating that residue 75 of α1(II)-CB12 is 70% hydroxylated. However, no other quantitative data for other partially hydroxylated prolines were obtained from the present studies.

The occurrence of hydroxylysine in the collagenase peptides T6-CL7 and T6-CL6 suggested that the "blanks" encountered at positions 9 and 18, respectively, in the sequence determination of peptide T6 (see above) were due to the presence of glycosylated hydroxylysines. Indeed, after alkaline hydrolysis

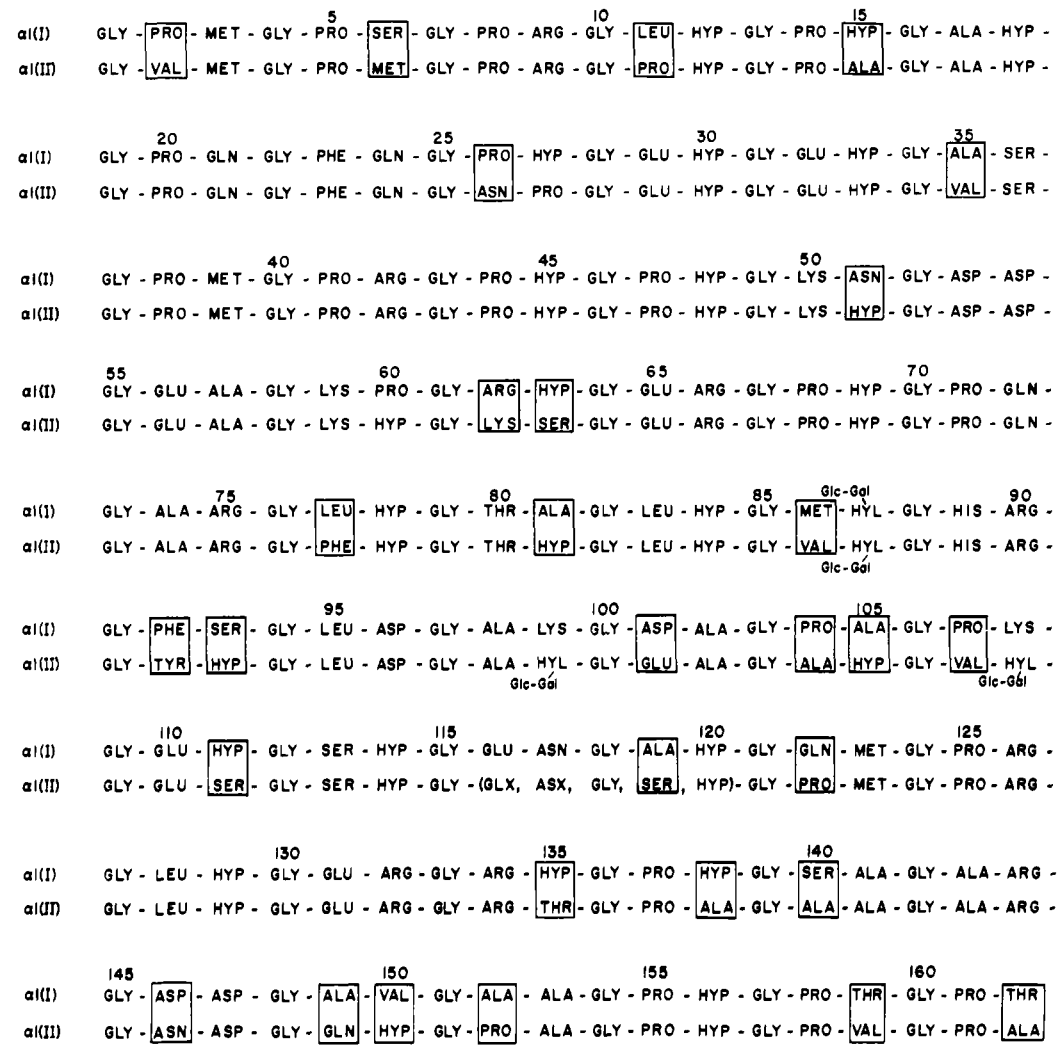

FIGURE 4: The comparative amino acid sequences of the NH_2-terminal helical portions of $\alpha 1$(I) and $\alpha 1$(II) chains. Positions where sequence differences result from a genetic variation are boxed. The numbering begins with the helical portion of the chains (the Gly-X-Y repeating sequence).

of the two collagenase peptides, glucosylgalactosylhydroxylysine and smaller amounts of galactosylhydroxylysine were detected on the amino acid analyzer. The ratios of the disaccharide to the monosaccharide component were approximately 4:1 for T6-CL7 and 5:1 for T6-CL6. Thus, we conclude that residues 60 and 69 of $\alpha 1$(II)-CB12 are principally glucosylgalactosylhydroxylysyls with lesser amounts of galactosylhydroxylysyl residues, occurring at these sites (see Figure 4).

The NH_2-Terminal Sequence of $\alpha 1$(II)-CB11. About 0.4 μmol of the peptide was subjected to automated sequence analysis utilizing the Slow Protein-Quadrol program. The analysis for 39 residues was run twice with essentially identical results. The following sequence was determined:

Gly-Pro-Arg-Gly-Leu-Hyp-Gly-Glu-Arg-Gly-
Arg-Thr-Gly-Pro-Ala-Gly-Ala-Ala-Gly-Ala-
Arg-Gly-Asn-Asp-Gly-Gln-Hyp-Gly-Pro-Ala-Gly-
Pro-Hyp-Gly-Pro-Val-Gly-Pro-Ala

Utilizing the levels of Pth-proline obtained at cycles 2 and 4, the repetitive yield was calculated to be 98%. When the levels of Pth-leucine and Pth-valine (cycles 5 and 36, respectively) were used, the repetitive yield was 95%.

Discussion

The covalent structure of the NH_2-terminal region of the $\alpha 1$(II) chain, as determined by the studies presented here, is depicted in Figure 4 along with a comparison with that of the $\alpha 1$(I) chain. It should be noted that the sequence data for the first 123 residues of $\alpha 1$(I) presented here are from calf collagen (Fietzek and Kühn, 1975). The remaining 39 residues are for rat collagen (Balian et al., 1971) since this latter sequence ($\alpha 1$(I)-CB8) from bovine collagen has not yet been reported.

The comparative data of Figure 4 indicate that the $\alpha 1$(I) and $\alpha 1$(II) chains are identical in about 80% of the positions. A similar level of sequence identity was observed when the central portions of bovine $\alpha 1$(I) and $\alpha 1$(II) chains were compared (Butler et al., 1974a). The level of variability is significantly higher than when *interspecies* comparisons of $\alpha 1$(I) chains are made (Dixit et al., 1975). For example, when amino acid sequences of the central portion of $\alpha 1$(I) chains from chick skin collagen are compared with homologous regions from bovine and rat collagens, the levels of identity are 91 and 89%, respectively. Twenty-two of the 30 genetic differences shown in Figure 4 (denoted by boxes) involve changes of only one nucleotide in the triplet of DNA codons.

The levels of hydroxylysine and of hydroxylysine-linked carbohydrate in the $\alpha 1$(II) chain are relatively high, compared with those of $\alpha 1$(I) (Miller, 1971). The results reported here and elsewhere (Butler et al., 1974a) corroborate the earlier compositional data and indicate that every lysyl residue in the $\alpha 1$(II) chain which is a potential substrate for lysyl hydroxylase

is fully hydroxylated. In contrast, residues 99 and 108 in the α1(I) chain are not hydroxylated. Furthermore, each of the hydroxylysyl residues of the α1(II) chain is glycosylated, indicating that the extent of glycosylation in the latter chain is dependent largely on the availability of hydroxylysyl residues.

References

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

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

Brauer, A. W., Margolies, M. N., and Haber, E. (1975), *Biochemistry 14,* 3029.

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

Butler, W. T., Birkedal-Hansen, H., Beegle, W. F., Taylor, R. E., and Chung, E. (1975), *J. Biol. Chem. 250,* 8907.

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

Butler, W. T., Piez, K. A., and Bornstein, P. (1967), *Biochemistry 6,* 3371.

Butler, W. T., Underwood, S. P., and Finch, J. E., Jr. (1974b), *Biochemistry 13,* 2946.

Chung, E., and Miller, E. J. (1974), *Science 183,* 1200.

Chung, E., Keele, E. M., and Miller, E. J. (1974), *Biochemistry 13,* 3459.

Daniels, J. R., and Chu, G. H. (1975), *J. Biol. Chem. 250,* 3531.

Dixit, S. N., Kang, A. H., and Gross, J. (1975), *Biochemistry 14,* 1929.

Epstein, E. H., Jr. (1974), *J. Biol. Chem. 249,* 3225.

Fietzek, P. P., Kell, I., and Kühn, K. (1972), *FEBS Lett. 26,* 66.

Fietzek, P. P., and Kühn, K. (1975), *Eur. J. Biochem. 52,* 77.

Fietzek, P. P., and Rauterberg, J. (1975), *FEBS Lett. 49,* 365.

Foster, J. A., Bruenger, C. L., Hu, C. L., Albertson, K., and Franzblau, C. (1973), *Biochem. Biophys. Res. Commun. 53,* 70.

Hill, R. L. (1965), *Adv. Protein Chem. 20,* 37.

Inagami, T., and Murakami, K. (1972), *Anal. Biochem. 47,* 501.

Kang, A. H., Dixit, S. N., Corbett, C., and Gross, J. (1975), *J. Biol. Chem. 250,* 7428.

Kefalides, N. A. (1971), *Biochem. Biophys. Res. Commun. 45,* 226.

Miller, E. J. (1971), *Biochemistry 10,* 1652.

Miller, E. J. (1972), *Biochemistry 11,* 4903.

Miller, E. J. (1973), *Clin. Orthop. 92,* 260.

Miller, E. J., Epstein, E. H., Jr., and Piez, K. A. (1971), *Biochem. Biophys. Res. Commun. 42,* 1024.

Miller, E. J., and Lunde, L. G. (1973), *Biochemistry 12,* 3153.

Miller, E. J., and Matukas, V. J. (1969), *Proc. Natl. Acad. Sci. U.S.A. 64,* 1264.

Miller, E. J., and Piez, K. A. (1966), *Anal. Biochem. 16,* 320.

Miller, E. J., Woodall, D. L., and Vail, M. S. (1973), *J. Biol. Chem. 248,* 1666.

Piez, K. A. (1968), *Anal. Biochem. 26,* 305.

Piez, K. A., Balian, G., Click, E. M., and Bornstein, P. (1972), *Biochem. Biophys. Res. Commun. 48,* 990.

Pisano, J. J., Bronzert, T. J., and Brewer, H. B., Jr. (1972), *Anal. Biochem. 45,* 43.

Spiro, R. G. (1969), *J. Biol. Chem. 244,* 602.

Trelstad, R. L. (1974), *Biochem. Biophys. Res. Commun. 57,* 717.

Trelstad, R. L., Kang, A. H., Igarashi, S., and Gross, J. (1970), *Biochemistry 9,* 4993.