Biochem. J. (1978) **175**, 921–930
Printed in Great Britain

The Covalent Structure of Cartilage Collagen

AMINO ACID SEQUENCE OF RESIDUES 552–661 OF BOVINE α1(II) CHAINS

GILLIAN FRANCIS,* WILLIAM T. BUTLER
and JOHN EDWARD FINCH, JR.
Institute of Dental Research and the Department of Biochemistry, University of Alabama in Birmingham, University Station, Birmingham, AL 35294, U.S.A.

(*Received* 17 *April* 1978)

The covalent structure of the first 111 residues from the *N*-terminus of peptide α1(II)-CB10 from bovine nasal-cartilage collagen is presented. This region comprises residues 552–661 of the α1(II) chain. The sequence was determined by automated Edman degradation of peptide α1(II)-CB10 and of peptides produced by cleavage with trypsin and hydroxylamine. Comparison of this region of the α1(II) chain with the homologous segment of the α1(I) chain indicated a homology level of 85%, slightly higher than that of 81% reported for the *N*-terminal region of the α1(II) chain (Butler, Miller & Finch (1976) *Biochemistry* **15**, 3000–3006). The occurrence of two residues of glycosylated hydroxylysine was established at positions 564 and 603, the first present exclusively as galactosylhydroxylysine and the latter as a mixture of galactosylhydroxylysine and glucosylgalactosylhydroxylysine. Also, two residues at positions 648 and 657 were tentatively identified as glycosylated hydroxylysines. The amino acid sequences adjacent to the hydroxylysine residues so far identified in the α1(II) chain were compared with the homologous regions of the α1(I) and α2 chains, but no obvious prerequisite for hydroxylation could be seen. From comparison with the homologous sequence of the α1(I) chain, it appears that the α1(II)-chain sequence presented here contains three more amino acids than that reported for the α1(I) chain. This triplet would be interposed between residues 63 and 64 of the reported sequence of peptide α1(I)-CB7 from calf skin collagen. Data on the purification of the subpeptides and their amino acid compositions have been deposited as Supplementary Publication SUP 50087 (7 pages) at the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1978) **169**, 5.

Type-II collagen, the principal fibrillar component of cartilage, is composed of three α1(II) chains. The major component of each chain is involved in triple-helix formation and contains a glycine residue at every third position, whereas the *N*- and *C*-terminal regions do not display this type of sequence and are non-triple-helical (Miller, 1972; Butler *et al.*, 1977*a*).

A notable feature of type-II collagen is the high content of carbohydrate, which is linked to hydroxylysine residues either as Gal-Hyl or as Glc-Gal-Hyl and whose function remains unclear (Miller, 1971). Apart from this elevated level of glycosylation, the α1(II) chain displays a high degree of homology with both the α1(I) and α2 chains of type-I collagen and with the α1(III) chain of type-III collagen (Miller, 1976).

Studies are in progress to determine the entire covalent structure of the α1(II) chain of bovine nasal-cartilage collagen; recently Butler *et al.* (1977*b*) reported the amino acid sequence of residues 363–551 of this chain. [Numbering begins with the *N*-terminal triple-helical portion of the collagen α-chains (Hulmes *et al.*, 1973; Fietzek & Kühn, 1976).] The present paper is intended as a sequel to that work and investigates the first 111 residues of peptide α1(II)-CB10, a CNBr peptide that has a reported amino acid content of 316 residues (Miller & Lunde, 1973; Miller, 1972) and whose alignment in the chain has been established as being between peptides α1(II)-CB8 and α1(II)-CB9,7 (Miller *et al.*, 1973). These data extend the known sequence in the centre of the chain from residue 552 to residue 661.

Abbreviations used: Glc-Gal-Hyl, 2-*O*-α-D-glucopyranosyl-*O*-β-D-galactopyranosylhydroxylysine; Gal-Hyl, *O*-β-D-galactopyranosylhydroxylysine; Ptc, phenylthiocarbamoyl; Atz, anilothiazolinone.

* Present address: Department of Biochemistry, Imperial College of Science and Technology, London S.W.7, U.K.

Experimental

Materials

2-Aminonaphthalene-1,5-disulphonic acid was purchased from Aldrich Chemical Co. Inc., Milwaukee, WI, U.S.A., and 1-ethyl-3-(3-dimethylaminopropyl)carbodi-imide hydrochloride from Pierce Chemical Co., Rockford, IL, U.S.A. All other reagents used in the sequencer were obtained from Beckman Instruments Inc., Palo Alto, CA, U.S.A. Sephadex G-50s was purchased from Pharmacia, Piscataway, NJ, U.S.A., and Bio-Gel A-1.5m from Bio-Rad Laboratories, Richmond, CA, U.S.A. CM-cellulose (CM52) was obtained from Whatman Biochemicals Ltd., Maidstone, Kent, U.K. Trypsin (twice-crystallized) was purchased from Worthington Biochemical Corp., Freehold, NJ, U.S.A., and hydroxylamine hydrochloride from J. T. Baker Chemical Co., Phillipsburg, NJ, U.S.A.

Methods

Preparation of peptide α1 (II)-CB10. Bovine nasal cartilage previously extracted with 4.0M-guanidinium chloride to remove proteoglycans was digested with CNBr in the following manner. A portion (600mg) of material was suspended in 60ml of aqueous 70% formic acid and stirred at 24°C for 4h. Any insoluble material remaining after digestion was removed by centrifugation and the supernatant was desalted on columns of Sephadex G-25 (coarse grade) and freeze-dried. The resultant peptides were separated on CM-cellulose columns as described by Miller & Lunde (1973). Those fractions containing peptide α1(II)-CB10 and peptide α1(II)-CB9,7 were pooled, desalted and freeze-dried. Peptides α1(II)-CB10 and α1(II)-CB9,7 were separated from one another by gel filtration on a column (1.5cm × 85cm) of Bio-Gel A-1.5m equilibrated with 0.05M-Tris/HCl buffer, pH7.5, containing 2M-guanidinium chloride (Miller & Lunde, 1973).

Hydroxylamine cleavage. Peptide α1(II)-CB10 was cleaved with hydroxylamine by using essentially the method of Balian *et al.* (1971). The procedure involved incubation of the peptide (2mg/ml) in a freshly prepared solution of 2M-hydroxylamine in 0.2M-K_2CO_3, pH9.0, at 45°C for 4h, followed by desalting and freeze-drying. Initial separation of peptides resulting from hydroxylamine cleavage of peptide α1(II)-CB10 was achieved on a column (1.4cm × 150cm) of Bio-Gel A-1.5m, equilibrated with 0.05M-Tris/HCl buffer, pH7.5, containing 2M-guanidinium chloride. The smallest hydroxylamine peptide, HA1, was subjected to gel chromatography on a column (1.8cm × 140cm) of Sephadex G-50s, equilibrated with 0.2M-acetic acid, for further purification.

Trypsin cleavage. Peptide α1(II)-CB10 (25mg) was dissolved in 25ml of 0.1M-Tris/HCl buffer, pH7.6, containing 1mM-$CaCl_2$. After addition of 6% (w/w) of trypsin, the mixture was stirred gently at 24°C for 24h and the digest freeze-dried. Tryptic digestion of peptide HA1, produced by hydroxylamine cleavage of peptide α1(II)-CB10, was performed in a similar manner. Initial separation of tryptic peptides of α1(II)-CB10 was achieved on a column (1.8cm × 140cm) of Sephadex G-50s equilibrated with 0.2M-acetic acid.

Ion-exchange chromatography. Phosphocellulose chromatography was used for separation of certain peptides produced by cleavage of peptide α1(II)-CB10 with trypsin or hydroxylamine. Resolution of peptides was achieved on a column (1.0cm × 10cm) of phosphocellulose by the method of Butler *et al.* (1977*c*), with a linear gradient from 0 to 0.2M-NaCl or from 0 to 0.3M-NaCl over a total volume of 1100ml.

Certain tryptic peptides were separated by CM-cellulose chromatography at 40°C in sodium acetate buffers, pH4.8, with a shallow concave gradient from 0 to 0.14M-NaCl as described by Butler *et al.* (1967). The gradient was formed by using a 1000ml beaker for that volume of starting buffer (0.02M-sodium acetate buffer, pH4.8). The limiting buffer, which flowed into and mixed with the starting buffer, consisted of 740ml of 0.02M-sodium acetate/0.14M-NaCl, pH4.8, in a 750ml Erlenmeyer flask.

Small tryptic peptides were separated by chromatography on a column (0.9cm × 150cm) of Chromobeads, type A resin (Technicon), and elution with a linear gradient formed from pyridine/acetate buffers as described by Butler *et al.* (1974). Portions of each fraction were analysed for ninhydrin-positive material with a Technicon autoanalyser.

Edman degradation. Sequential Edman degradation was performed with a Beckman automatic sequencer (model 890C) as described by Butler *et al.* (1977*b*).

The Ptc-amino acids were identified both by t.l.c. on silica-gel plates (Inagami & Murakami, 1972) and by g.l.c. with 1.22m (4ft) U-shaped glass columns packed with 10% DC-560. Certain residues were also subjected to trimethylsilylation before injection into the gas chromatograph. Ptc-arginine was identified by means of the phenanthrenequinone spot test. Identification of Ptc-isoleucine and Ptc-leucine was achieved by silylation before g.l.c. or by back-hydrolysis of the Ptc-amino acid in constant-boiling HCl at 150°C for 24h and subsequent analysis on the amino acid analyser.

Quantitative yields for some Ptc-amino acids were calculated by comparing the amount of peptide (determined by amino acid analysis) introduced into the sequencer cup with the peak heights on g.l.c. analysis. Repetitive yields were calculated by the method outlined in the Beckman Sequencer Manual. Ptc-amino acids of the same kind were used for these latter calculations because of the variability of yields in the conversion of Atz-amino acids into Ptc-amino acids.

To improve the retention of small peptides in the reaction cup of the sequencer, they were made to react with 2-aminonaphthalene-1,5-disulphonic acid in the presence of 1-ethyl-3-(3-dimethylaminopropyl)carbodi-imide hydrochloride. The method was that of Foster *et al.* (1973), as modified by Butler *et al.* (1977*b*).

Amino acid analysis. Samples were hydrolysed under N_2 at 108°C for 18–24h in constant-boiling HCl and were analysed on a Beckman 121 M automatic amino acid analyser by using a dual-column procedure (Butler *et al.*, 1977*c*).

Analysis of hydroxylysine glycosides. Peptides (50–100nmol) were hydrolysed in 2M-NaOH at 105°C for 24h in sealed alkali-resistant tubes. After adjustment of pH and volume (Butler *et al.* 1977*b*), each sample was analysed on a Beckman 119 amino acid analyser (Miller, 1972).

Results

N-Terminal sequence of peptide α1(*II*)-*CB*10

Intact peptide α1(II)-CB10 was subjected to automated Edman degradation by using the 0.1M-Quadrol procedure described by Brauer *et al.* (1975) and modified by Beckman (1976). Approx. 400nmol of peptide was used for the first attempt and the sequence of 20 of the first 21 amino acid residues was determined. The results are summarized in Table 1. The repetitive yield was calculated to be 97%, based on yields of Ptc-alanine at cycles 6 and 10. For the second run, 290nmol of material was used and a repetitive yield of 96% was obtained, again based on yields of Ptc-alanine at cycles 6 and 10. This time the sequence of 19 of the first 20 amino acid residues was obtained, and the residues at cycles 23, 24, 26 and 27 could also be identified as the Ptc derivatives of glycine, proline, glycine and alanine respectively. The residue at position 9 was identified as either leucine or isoleucine and was later confirmed as isoleucine by consideration of the amino acid content of peptide HA1-T2. In both cases a blank for which no Ptc-amino acid could be observed by either g.l.c. or t.l.c. was obtained at cycle 13. The results are shown diagrammatically in Fig. 1.

Table 1. *Summary of identification and recoveries of Ptc-amino acids in the various cycles of automated Edman degradation of peptide* α1(*II*)-*CB*10

The positive identification of a Ptc-amino acid is denoted by the name of the parent amino acid under the identification method. A dash (—) indicates that, although the method was utilized, identification of the Ptc-amino acid was not achieved. Recoveries are given as a percentage of the starting peptide material. The amounts of Ptc-amino acids were calculated from peak heights on g.l.c. compared with those of standard Ptc-amino acids. Solvent XM is xylene/methanol (8:1, v/v). Identities of the Ptc-amino acids were aided by spraying dried thin-layer chromatograms with ninhydrin (Inagami & Murakami, 1972).

		G.l.c.		T.l.c.	Recovery
Cycle	Conclusion	Non-silylated	Silylated	(solvent XM)	(%)
1	Hyp	Hyp		Hyp	
2	Gly	Gly		Gly	46
3	Glu		Glu	Glu	
4	Arg*	—	—	—	
5	Gly	Gly		Gly	30
6	Ala	Ala		Ala	32
7	Ala	Ala		Ala	32
8	Gly	Gly		Gly	47
9	Leu/Ile	Leu/Ile		Leu/Ile	28
10	Ala	Ala		Not run†	29
11	Gly	Gly		Gly	45
12	Pro	Pro		Pro	16
13	Gal-Hyl‡	—	—	—	
14	Gly	Gly		Gly	43
15	Asp		Asp	Asp	
16	Arg*	—	—	—	
17	Gly	Gly		Gly	43
18	Asp		Asp	Asp	
19	Val	Val		Val	20
20	Gly	Gly		Gly	44
21	Glu		Glu	Not run†	

* Identified by phenanthrenequinone spot test.
† Cycles 10 and 21 were not analysed by t.l.c.
‡ See the text for assignment of Gal-Hyl to this position.

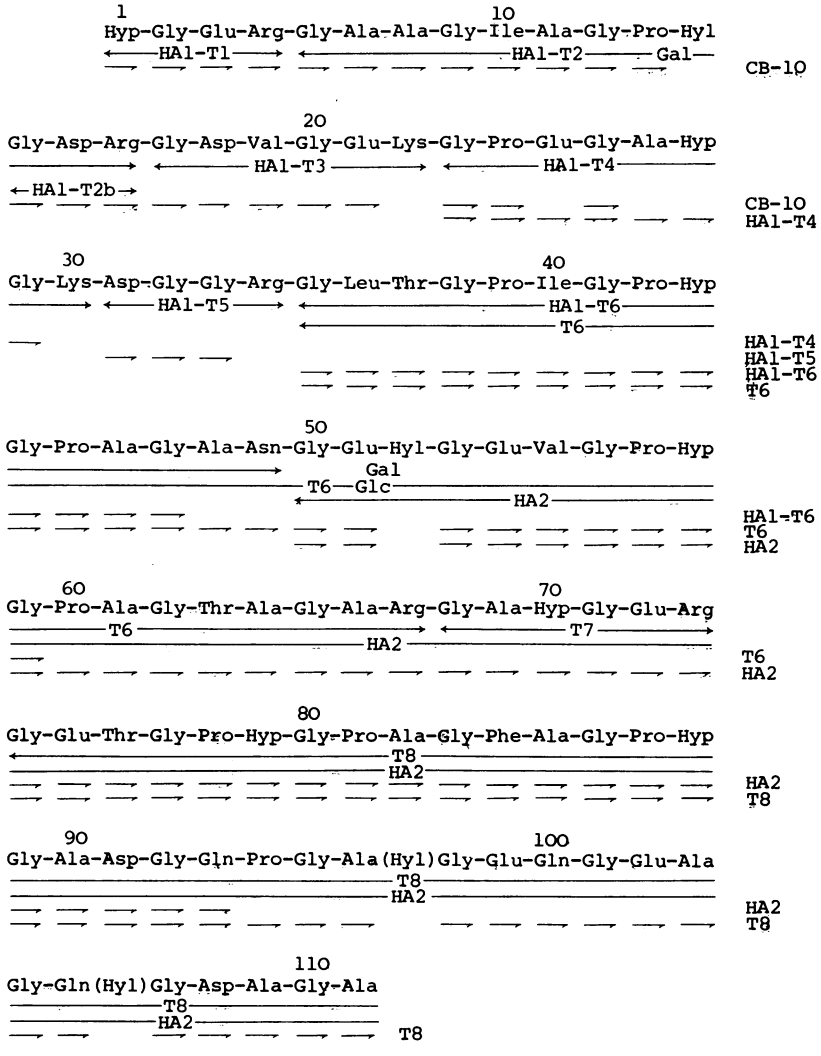

Fig. 1. *Amino acid sequence of the first* 111 *residues of peptide* α1(*II*)-*CB*10 *from bovine nasal-cartilage collagen*
The half-arrows (⟶) represent residues identified during Edman degradation. Residue 9 was initially identified as either leucine or isoleucine and was confirmed as isoleucine after consideration of the amino acid composition of peptide HA1-T2. Residue 13 was not identified by Edman degradation, but was shown to be hydroxylysine after acid hydrolysis of peptide HA1-T2. Alkaline hydrolysis of peptide HA1-T2 identified this hydroxylysine residue as being completely in the form of Gal-Hyl. No Ptc-amino acid could be identified at residue 52, but this was shown to be a hydroxylysine residue after hydrolysis of peptide T6. Analysis of an alkaline hydrolysate of peptide T6 showed that this hydroxylysine residue was present as both Glc-Gal-Hyl and Gal-Hyl. No Ptc-amino acid could be identified at residues 96 or 105, and glycosylated hydroxylysine residues have been tentatively assigned to these positions (see the text). The identities of leucine and isoleucine residues at positions 36 and 40 were confirmed by back-hydrolysis of their Ptc derivatives and subsequent identification on the amino acid analyser.

Hydroxylamine peptides of α1(*II*)-*CB*10

Cleavage with hydroxylamine yielded a minimum of four products. Two of the hydroxylamine peptides, HA1 and HA2, were purified by gel chromatography on Bio-Gel A-1.5m, followed by either gel filtration on Sephadex G-50s (peptide HA1) or by ion-exchange chromatography on phosphocellulose (peptide HA1). Sequence analysis of peptide HA1 indicated that it was derived from the *N*-terminal segment of peptide

α1(II)-CB10. That peptide HA2 was derived from a portion of peptide α1(II)-CB10 adjacent to peptide HA1 was shown by sequence analysis of tryptic peptide T6 (discussed below).

To obtain further information about the sequence near the *N*-terminal region of peptide α1(II)-CB10, peptide HA1 was hydrolysed with trypsin and the resultant fragments were separated by ion-exchange chromatography on the column of Chromobeads A. In this way six tryptic peptides, representing residues 5–49, were obtained (Fig. 1). Peptide HA1-T1 could not be located, presumably because of the low colour yield with ninhydrin due to the *N*-terminal hydroxyproline. The composition of peptide HA1-T2 showed that the unidentified residue at position 13 (see Fig. 1) yielded hydroxylysine after acid hydrolysis; after alkaline hydrolysis and amino acid analysis, peptide HA1-T2 yielded Gal-Hyl but no Glc-Gal-Hyl or hydroxylysine. The composition of peptide HA1-T3 showed that position 22 of peptide α1(II)-CB10 was occupied by a lysine residue.

Peptide HA1-T4 was subjected to seven cycles of automated Edman degradation by using the Slow Peptide-DMAA program (Butler *et al.*, 1977*b*). Approx. 500nmol of peptide was made to react with 2-aminonaphthalene-1,5-disulphonic acid in the presence of 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide. A repetitive yield of 75% was obtained, based on the yields of Ptc-glycine at cycles 1 and 4 of the analysis. The sequence obtained was: Gly-Pro-Glu-Gly-Ala-Hyp-Gly. Both Ptc-proline and Ptc-hydroxyproline were identified in cycle 6 and thus the proline in this position is incompletely hydroxylated; it is depicted as 'Hyp' in the sequence (Fig. 1). The composition of peptide HA1-T4 and the specificity of trypsin indicated that the *C*-terminal amino acid is lysine.

By comparison of this sequence with the residues previously identified for positions 23, 24, 26 and 27 of peptide α1(II)-CB10, peptide HA1-T4 could be positioned as shown in Fig. 1. Since peptide HA1-T6 contained neither lysine nor arginine, it must have been derived from the *C*-terminus of peptide HA1, and, by deduction, peptide HA1-T5 was placed in positions 31–34.

Peptide HA1-T5 was subjected to one cycle of automated Edman degradation after reaction with 2-aminonaphthalene-1,5-disulphonic acid in the presence of 1-ethyl-3-(3-dimethylaminopropyl)carbodi-imide, with the Slow Peptide-DMAA program. The amount of peptide used was 420nmol and Ptc-aspartic acid was identified in cycle 1. From the amino acid composition and the known specificity of trypsin, the following sequence could be deduced: Asp-Gly-Gly-Arg. Since a Gly-Gly sequence is unusual for collagen, this was confirmed by isolation of this peptide from a tryptic digest of peptide α1(II)-CB10 by chromatography, first on Sephadex G-50s to remove large peptides, and then on the column of Chromobeads A. A peak was again obtained in the area of elution of peptide HA1-T5, and the material in this peak gave an amino acid composition identical with that of peptide HA1-T5. This peptide (630nmol) was subjected to three cycles of automated Edman degradation as above. Ptc-aspartic acid was identified in cycle 1 and Ptc-glycine at cycles 2 and 3. From the composition and consideration of trypsin specificity, the complete sequence Asp-Gly-Gly-Arg was obtained, confirming the above deduction.

Peptide HA1-T6 was subjected to 13 cycles of automated Edman degradation with the Slow Peptide-DMAA program. The peptide (780nmol) was coupled to 2-aminonaphthalene-1,5-disulphonic acid in the presence of 1-ethyl-3-(3-dimethylaminopropyl)carbodi-imide and a repetitive yield of 91% was obtained, based on yields of Ptc-glycine at cycles 1 and 10. A summary of the recoveries of Ptc-amino acids in the various cycles of Edman degradation is given in Table 2. The sequence of the first 13 residues of peptide HA1-T6 was thus established as: Gly-Leu-Thr-Pro-Ile-Gly-Pro-Hyp-Gly-Pro-Ala-Gly. Since the amino acid analysis of the peptide showed that it was slightly contaminated, the sequence of the *C*-terminal region of the peptide could not be deduced by consideration of its amino acid content alone.

Peptide HA1-T2b was obtained at a yield of approx. 20% those of peptides HA1-T3, HA1-T4, HA1-T5 and HA1-T6. Consideration of its amino acid composition along with this low yield led to the conclusion that peptide HA1-T2b arose from the partial tryptic cleavage of peptide HA1-T2, suggesting

Table 2. *Summary of identification and recoveries of Ptc-amino acids*
See Table 1 for details.

Cycle	Conclusion	G.l.c. (non-silylated)	T.l.c. (solvent XM)	Recovery (%)
1	Gly	Gly	Gly	27
2	Leu*	Leu/Ile	Leu/Ile	19
3	Thr	Thr/Pro	Thr	
4	Gly	Gly	Gly	29
5	Pro	Pro	Pro	14
6	Ile*	Leu/Ile	Leu/Ile	
7	Gly	Gly	Gly	15
8	Pro	Pro	Pro	8
9	Hyp	Hyp	Hyp	
10	Gly	Gly	Gly	11
11	Pro	Pro	Pro	
12	Ala	Ala	Ala	
13	Gly	Gly	Gly	

* Identification achieved by back-hydrolysis of Ptc derivative and subsequent analysis on the amino acid analyser.

that the hydroxylysine residue at position 12 of peptide α1(II)-CB10 may also occur to a minor extent in the non-glycosylated form or possibly as a non-hydroxylated lysine residue.

Sequence studies

Tryptic peptide T6. Cleavage of peptide α1(II)-CB10 with trypsin resulted in a relatively large number of peptides, which were initially separated by gel filtration on Sephadex G-50s. One of the larger tryptic peptides, T5, was further purified by phosphocellulose chromatography and shown by amino acid analysis to contain 33 amino acids. Peptide T6 was also shown to contain a mixture of Glc-Gal-Hyl and Gal-Hyl by alkaline hydrolysis and amino acid analysis; the amount of the disaccharide moiety was slightly greater than that of the monosaccharide; no unsubstituted hydroxylysine was detected. Peptide T6 was subjected to automated Edman degradation after reaction with 2-aminonaphthalene-1,5-disulphonic acid in the presence of 1-ethyl-3-(3-dimethylaminopropyl)carbodi-imide, and the sequence of 24 of the first 25 amino acid residues was obtained. The results of this analysis are summarized in Table 3 and diagramatically depicted in Fig. 1. At cycle 18, no Ptc-amino acid could be identified by either g.l.c. or t.l.c. The observation that peptides HA1-T6 and T6 have identical *N*-terminal sequences enabled the positioning of the latter peptide as shown in Fig. 1.

Peptide HA2. This peptide was subjected twice to automated Edman degradation by using the 0.1M-Quadrol procedure. In the first such analysis 280nmol of peptide was used and the sequence of 44 of the first 45 amino acid residues was determined. In the second analysis with 120nmol of peptide the sequence of 19 of the first 20 amino acid residues was obtained, with results identical with those of the first run. The sequence data for the first 45 residues of peptide HA2 are depicted in Fig. 1.

The results, together with previous information, established that peptide T6 overlapped the hydroxylamine peptides HA1 and HA2. Further, the amino acid, hydroxylysine glycoside and sequence analysis

Table 3. *Summary of identification and recoveries of Ptc-amino acids in the various cycles of automated Edman degradation of peptide T6*
See Table 1 for details.

Cycle	Conclusion	G.l.c. Non-silylated	G.l.c. Silylated	T.l.c. (solvent XM)	Recovery (%)
1	Gly	Gly		Gly	59
2	Leu*	Leu/Ile	Leu	Leu/Ile	
3	Thr	Thr/Pro	Thr	Thr	
4	Gly	Gly		Gly	56
5	Pro	Pro		Pro	21
6	Ile*	Leu/Ile	Ile	Leu/Ile	
7	Gly	Gly		Gly	51
8	Pro	Pro		Pro	22
9	Hyp	Hyp		Hyp	
10	Gly	Gly		Gly	44
11	Pro	Pro		Pro	25
12	Ala	Ala		Ala	
13	Gly	Gly		Gly	36
14	Ala	Ala		Ala	
15	Asn		Asn	Asn	
16	Gly	Gly		Gly	12
17	Glu		Glu	Glu	
18	Glc-Gal-Hyl†	—	—	—	
19	Gly	Gly		Gly	19
20	Glu		Glu	Glu	
21	Val	Val		Val	25
22	Gly	Gly		Gly	12
23	Pro	Pro		Pro	3
24	Hyp	Hyp		—	
25	Gly	Gly		Gly	

* Identification also achieved by back hydrolysis of Ptc derivative and subsequent analysis on the amino acid analyser.
† See the text for assignment of Glc-Gal-Hyl to this position.

obtained for peptide T6 and the sequence analysis for peptide HA2 showed that position 18 of peptide T6 was occupied by a glycosylated hydroxylysine. Thus, although position 52 of peptide α1(II)-CB10 (Fig. 1) is given as Glc-Gal-Hyl, not quite half is present as Gal-Hyl.

Tryptic peptide T8. Peptide T8 from the Sephadex G-50s chromatography of the tryptic hydrolysis of peptide α1(II)-CB10 was further purified by phosphocellulose chromatography. Its composition showed that peptide T8 contained 62 amino acids and was very pure. This peptide (130nmol) was subjected to automated Edman degradation by using the 0.1M-Quadrol procedure, and the sequence of 36 of the first 38 amino acid residues was established (Fig. 1). The results, when compared with the data for peptide HA2, allowed the placement of peptide T8 as shown in Fig. 1.

No Ptc-amino acid could be identified at cycle 24 or 33 of peptide T8; hydroxylysine residues have been tentatively assigned to these positions for the following reasons. First, previous experience has shown that, when blanks such as those obtained at cycles 23 and 33 are encountered, these residues are often elucidated to be glycosylated hydroxylysine residues (Butler *et al.*, 1976, 1977*b*). Secondly, consideration of the amino acid composition of peptide T8 revealed that it contained three residues of hydroxylysine, and thirdly, alkaline hydrolysis of peptide T8 and subsequent analysis on the amino acid analyser showed that it contained both Gal-Hyl and Glc-Gal-Hyl. The last piece of evidence in support of our assumption comes from comparison of the sequence of this peptide with that of the homologous region of peptide α1(I)-CB7 of calf-skin collagen, which possesses lysine residues at both these positions (Fietzek *et al.*, 1973). Previous work (Butler *et al.*, 1976, 1977*c*) has shown that positions occupied by lysine residues in the α1(I) chain are usually occupied by glycosylated hydroxylysine residues in the α1(II) chain.

Supplementary data

Detailed experimental evidence on the results of isolation procedures and compositions of the peptides utilized in the present study have been deposited as Supplementary Publication SUP 50087 (7 pages) at the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1978) **169**, 5.

Discussion

The amino acid sequence of the first 111 residues from the *N*-terminus of peptide α1(II)-CB10 is shown in Fig. 1. This region comprises residues 552-661 of the α1(II) chain. The Gly-X-Y sequence so far found to be common to the helical region of all collagen α-chains (Fietzek & Kühn, 1976; Miller, 1976) is again seen in this region of the α1(II) chain.

If the sequence presented here is aligned with the homologous regions of peptide α1(I)-CB7 from calf skin (Fietzek *et al.* 1973) and peptide α2-CB3 from chick skin (Dixit *et al.*, 1977*b*) so as to maximize the homology between the three chains, it is apparent that the α1(II) and α2 chains contain three more amino acid residues than does the reported sequence of the α1(I) chain. The relevant regions of the α1(I), α1(II) and α2 chains, numbered according to the reported sequence of peptide α1(I)-CB7, are shown in Fig. 2. There are two possibilities that could explain this discrepancy. Firstly, the published sequence of α1(I)-CB7 may be correct, in which case the α1(I) chain would actually be three residues shorter than the α1(II) and α2 chains. Alternatively, it is possible that previous reports (Fietzek *et al.*, 1973; Highberger *et al.*, 1975) have overlooked the presence of one Y-Gly-X triplet in the α1(I) chain. In our view the latter possibility seems more likely.

When the covalent structure reported in the present paper is compared with that of the homologous region

```
                                  50                  60
α1(I) Bovine     -Lys-Gly-Glu-Ala-Gly-Pro-Ser-Gly-Pro-Ala-
α1(II) Bovine    -Hyl-Gly-Glu-Val-Gly-Pro-Gyp-Gly-Pro-Ala-
α2 Chick         -Arg-Gly-Glu-Gly-Gly-Pro-Ala-Gly-Pro-Ala-

                    63                64                  70
α1(I) Bovine      Gly-Thr-            Ala-Gly-Ala-Hyp-Gly-Asp-Arg-
α1(II) Bovine     Gly-Thr-Ala-Gly-Ala-Arg-Gly-Ala-Hyp-Gly-Glu-Arg-
α2 Chick          Gly-Pro-Ala-Gly-Ala-Arg-Gly-Ile-Hyp-Gly-Glu-Arg-
```

Fig. 2. *Homologous sequences from collagen α1(I), α1(II) and α2 chains*
Numbering corresponds to that of peptide α1(I)-CB7 taken from Fietzek *et al.* (1973). The α1(II)-chain sequence represents residues 52–73 of peptide α1(II)-CB10 reported in the present paper (Fig. 1), and the α2 sequence corresponds to residues 114–135 of chymotryptic peptide C4 from peptide α2-CB3 (Dixit *et al.*, 1977*b*).

	87
α1(I) Bovine	Gly-Leu-Hyp-Gly-Met-Hyl-Gly-His-Arg-Gly-Phe-Ser
α1(II) Bovine	Gly-Leu-Hyp-Gly-Val-Hyl-Gly-His-Arg-Gly-Thr-Hyp
α2 Bovine	Gly-Leu-Hyp-Gly-Phe-Hyl-Gly-Ile-Arg-Gly-His-Asn
	99
α1(I) Bovine	Gly-Leu-Asp-Gly-Ala-Lys-Gly-Asp-Ala-Gly-Pro-Ala
α1(II) Bovine	Gly-Leu-Asp-Gly-Ala-Hyl-Gly-Glu-Ala-Gly-Ala-Hyp
α2 Bovine	Gly-Leu-Asp-Gly-Leu-Thr-Gly-Gln-Hyp-Gly-Ala-Hyp
	108
α1(I) Bovine	Gly-Pro-Ala-Gly-Pro-Lys-Gly-Glu-Hyp-Gly-Ser-Hyp
α1(II) Bovine	Gly-Ala-Hyp-Gly-Val-Hyl-Gly-Glu-Ser-Gly-Ser-Hyp
α2 Bovine	Gly-Ala-Hyp-Gly-Val-Hyl-Gly-Glu-Hyp-Gly-Ala-Hyp
	174
α1(I) Rat	Gly-Ala-Ala-Gly-Ala-Lys-Gly-Glu-Ala-Gly-Pro-Gln
α1(II) Bovine	Gly-Ala-Hyp-Gly-Ala-Hyl-Gly-Glu-Ala-Gly-Pro-
α2 Bovine	Gly-Ala-Hyp-Gly-Pro-Hyl-Gly-Glu-Leu-Gly-Pro-Val
	219
α1(I) Rat	Gly-Gln-Hyp-Gly-Ala-Lys-Gly-Ala-Asn-Gly-Ala-Hyp
α1(II) Bovine	Gly-Ile-Hyp-Gly-Ala-Hyl-Gly-Ser-Ala-Gly-Ala-Hyp
α2 Bovine	Gly-Leu-Hyp-Gly-Ala-Hyl-Gly-Ala-Ala-Gly-Leu-Hyp
	252
α1(I) Rat	Gly-Ala-Hyp-Gly-Pro-Lys-Gly-Asn-Ser-Gly-Glu-Hyp
α1(II) Bovine	Gly-Pro-Leu-Gly-Pro-Hyl-Gly-
α2 Bovine	Gly-Ala-Thr-Gly-Ala-Arg-Gly-Leu-Val-Gly-Glu-Hyp
	408
α1(I) Bovine	Gly-Phe-Hyp-Gly-Pro-Lys-Gly-Ala-Ala-Gly-Glu-Hyp
α1(II) Bovine	Gly-Phe-Hyp-Gly-Pro-Hyl-Gly-Ala-Asn-Gly-Glu-Hyp
α2 Chick	Gly-Phe-Hyp-Gly-Pro-Lys-Gly-Pro-Thr-Gly-Glu-Hyp
	420
α1(I) Bovine	Gly-Lys-Ala-Gly-Glu-Arg-Gly-Val-Hyp-Gly-Pro-Hyp
α1(II) Bovine	Gly-Lys-Ala-Gly-Glu-Hyl-Gly-Leu-Hyp-Gly-Ala-Hyp
α2 Chick	Gly-Lys-Hyp-Gly-Glu-Lys-Gly-Asn-Val-Gly-Leu-Ala
	531
α1(I) Bovine	Gly-Asn-Asp-Gly-Ala-Lys-Gly-Asp-Ala-Gly-Ala-Hyp
α1(II) Bovine	Gly-Thr-Asp-Gly-Pro-Hyl-Gly-Ala-Ala-Gly-Pro-Ala
α2 Chick	Gly-Pro-Asp-Gly-Asn-Lys-Gly-Glu-Hyp-Gly-Asn-Val
	564
α1(I) Bovine	Gly-Leu-Hyp-Gly-Pro-Lys-Gly-Asp-Arg-Gly-Asp-Ala
α1(II) Bovine	Gly-Ile-Ala-Gly-Pro-Hyl-Gly-Asp-Arg-Gly-Asp-Val
α2 Chick	Gly-Val-Hyp-Gly-Gly-Lys-Gly-Glu-Lys-Glu-Ala-Hyp
	573
α1(I) Bovine	Gly-Asp-Ala-Gly-Pro-Lys-Gly-Ala-Asp-Gly-Ala-Pro
α1(II) Bovine	Gly-Asp-Val-Gly-Glu-Lys-Gly-Pro-Glu-Gly-Ala-Pro
α2 Chick	Gly-Ala-Hyp-Gly-Leu-Arg-Gly-Asp-Thr-Gly-Ala-Thr

```
                                    603
α1(I) Bovine    Gly-Ala-Hyp-Gly-Asp-Lys-Gly-Glu-Ala-Gly-Pro-Ser
α1(II) Bovine   Gly-Asp-Val-Gly-Glu-Hyl-Gly-Glu-Val-Gly-Pro-Hyp
α2 Chick        Gly-Gly-Ala-Gly-Asp-Arg-Gly-Glu-Gly-Gly-Pro-Ala

                                    648
α1(I) Bovine    Gly-Gln-Hyp-Gly-Ala-Lys-Gly-Glu-Hyp-Gly-Asp-Ala
α1(II) Bovine   Gly-Gln-Pro-Gly-Ala-Hyl-Gly-Gly-Gln-Gly-Glu-Ala
α2 Chick        Gly-Glu-Hyp-Gly-Ala-Lys-Gly-Glu-Arg-Gly-Pro-Lys

                                    667
α1(I) Bovine    Gly-Asp-Ala-Gly-Ala-Lys-Gly-Asp-Ala-Gly-Pro-Hyp
α1(II) Bovine   Gly-Glu-Ala-Gly-Gln-Hyl-Gly-Asp-Ala-Gly-Ala
α2 Chick        Gly-Pro-Lys-Gly-Pro-Lys-Gly-Glu-Thr-Gly-Pro-Thr
```

Fig. 3. *Comparison of the amino acid sequences adjacent to hydroxylated lysine residues in the α1(I), α1(II) and α2 chains of collagen*

Numbering of amino acid residues begins with the first glycine residue of the repeating Gly-X-Y triplet, and takes into account the extra triplet found in the α1(II) and α2 chains, so that the hydroxylysine residue at position 648 is that previously cited at position 645 in the α1(I) chain:

Chain	Residues	Reference
Bovine α1(I)	4–123	Fietzek & Kühn (1975)
	403–551	Fietzek *et al.* (1972)
	552–819	Fietzek *et al.* (1973)
Rat α1(I)	124–222	Balian *et al.* (1971)
	223–402	Balian *et al.* (1972)
Bovine α1(II)	1–162	Butler *et al.* (1976)
	163–362	W. T. Butler & J. E. Finch, Jr. (unpublished work)
	363–551	Butler *et al.* (1977*b*)
Bovine α2	7–327	Fietzek & Rexrodt (1975)
Chick α2	358–489	Dixit *et al.* (1977*a*)
Chick α2	490–695	Dixit *et al.* (1977*b*)

of the α1(I) chain, and allowance is made for placement of the extra triplet, 17 of the 111 amino acid residues are found to be different, resulting in an 85% homology. If the invariant glycine residues in each repeating triplet are disregarded, the degree of homology is decreased to 77%. These values are slightly higher than the corresponding ones of 81 and 73% calculated for the *N*-terminal regions of these chains (Butler *et al.*, 1976). Comparison of residues 552–661 of the α1(II) chain with the homologous region of the α2 chain shows an overall sequence homology of 66%, which is decreased to 49% when the invariant glycine residues are removed from the calculation.

When comparing the portion of the α1(II) chain described here with the homologous region of the α1(I) chain, it can be seen that in some segments the homology is strictly conserved, whereas in others of the same span there may be several differences. For example, between residues 34 and 48 every amino acid is invariant, but between residues 19 and 33 substitutions occur at five positions. An examination of the comparative sequences of residues 361–660 from several different α-chains has been made by Butler *et al.* (1977*b*), and the region considered in the present paper is included in that report.

The sequence presented here contains one leucine residue and one phenylalanine residue, and both these amino acids occupy the X-position of the Gly-X-Y repeating sequence, an observation also noted for the α1(I) and α2 chains (Fietzek & Kühn, 1976; Traub & Fietzek, 1976; Dixit *et al.*, 1977*a*).

Our studies establish two residues of glycosylated hydroxylysine at positions 564 and 603 of the α1(II) chain, and suggest two others at positions 648 and 657. The hydroxylysine at position 564 is present exclusively as Gal-Hyl, whereas that in position 603 is found as both Gal-Hyl and Glc-Gal-Hyl. We are unable to state at present the form of glycosylation at positions 648 and 657.

An interesting feature of the results reported here is the fact that a non-hydroxylated lysine residue in the Y-position of the Gly-X-Y repeating sequence was noted at position 573. Although this is not an unusual feature of the α1(I) and α1(III) chains, it is the first time it has been found in the α1(II) chain, and the reason for this non-hydroxylation is unclear at the present time.

Thirteen hydroxylysine residues have so far been identified in the α1(II) chain. A comparison of the amino acid sequences around these residues and around the non-hydroxylated lysine residue at position 573 with the homologous regions of the α1(I) and α2 chains is shown in Fig. 3. The degree of lysine hydroxylation in the α1(II) chain is considerably higher than that in the α1(I) and α2 chains. It is possible that this difference is a result of differences in the amino acid sequence in those regions adjacent to the lysine residues. However, examination of Fig. 3 would seem to refute this, since lysine residues with almost identical surrounding sequences are hydroxylated in the α1(II) chain but not in the α1(I) or α2 chains.

The most distinctive feature of the sequences shown in Fig. 3 is the frequent occurrence of a glutamic acid or an aspartic acid residue in the X-position of the Gly-X-Y triplet immediately after a lysine or hydroxylysine residue. However, this acidic amino acid residue would not appear to be a significant factor in hydroxylation, since it can be found after both lysine and hydroxylysine residues and, conversely, in some cases is not present after either of these.

Another possible explanation for the differing hydroxylation could be due to the fact that the α1(II) chain is synthesized by a cell type different from that producing the α1(I) and α2 chains. However, since the α1(I) and α2 chains are produced by the same cell and in some instances show variation in their hydroxylation at homologous sites, this would appear doubtful.

Alternatively, since hydroxylation of lysine residues has been shown to be prevented by triple-helix formation (Kivirikko *et al.*, 1973; Ryhänen & Kivirikko, 1974), it is possible that hydroxylation could be controlled by the length of time for which the nascent collagen polypeptide chains remain attached to the ribosomes in the non-helical form.

Another possibility is that some system of subcellular compartmentalization of the hydroxylase enzyme system could be, in effect, controlling the degree of hydroxylation of lysine residues and accounting for the various extents of hydroxylation in the different α-chains.

We are grateful to Mr. Glen Bridges and Miss Virginia Wright for their technical assistance, and to Miss Janet Whitlow for preparation of the sequence diagrams. We thank Dr. Sayru Dixit, the University of Tennessee Medical School, for the copy of a manuscript of the sequence of peptide α2-CB3 before its publication. This research was sponsored by U.S. Public Health Service Grant no. DE-02670 from the National Institute of Dental Research.

References

Balian, G., Click, E. M. & Bornstein, P. (1971) *Biochemistry* **10**, 4470–4478

Balian, G., Click, E. M., Hermodson, M. A. & Bornstein, P. (1972) *Biochemistry* **11**, 3798–3806

Beckman (1976) Program no. 030176, *In Sequence*, May 1976, Beckman Instruments, Palo Alto

Brauer, A. W., Margolies, M. N. & Haber, E. (1975) *Biochemistry* **14**, 3029–3035

Butler, W. T., Piez, K. A. & Bornstein, P. (1967) *Biochemistry* **6**, 3771–3780

Butler, W. T., Underwood, S. P. & Finch, J. E., Jr. (1974) *Biochemistry* **13**, 2946–2953

Butler, W. T., Miller, E. J. & Finch, J. E., Jr. (1976) *Biochemistry* **15**, 3000–3006

Butler, W. T., Finch, J. E., Jr. & Miller, E. J. (1977*a*) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **36**, 680

Butler, W. T., Finch, J. E., Jr. & Miller, E. J. (1977*b*) *Biochemistry* **16**, 4981–4990

Butler, W. T., Finch, J. E., Jr. & Miller, E. J. (1977*c*) *J. Biol. Chem.* **252**, 639–643

Dixit, S. N., Seyer, J. M. & Kang, A. H. (1977*a*) *Eur. J. Biochem.* **73**, 213–643

Dixit, S. N., Seyer, J. M. & Kang, A. H. (1977*b*) *Eur. J. Biochem.* **81**, 599–607

Fietzek, P. P. & Kühn, K. (1975) *Eur. J. Biochem.* **52**, 77–82

Fietzek, P. P. & Kühn, K. (1976) *Int. Rev. Connect. Tissue Res.* **7**, 1–60

Fietzek, P. P. & Rexrodt, R. (1975) *Eur. J. Biochem.* **59**, 113–118

Fietzek, P. P., Wendt, P., Kell, I. & Kühn, K. (1972) *FEBS Lett.* **26**, 74–76

Fietzek, P. P., Rexrodt, F. W., Hopper, K. & Kühn, K. (1973) *Eur. J. Biochem.* **38**, 396–400

Foster, J. A., Bruenger, D. L., Hu, C. L., Albertson, K. & Franzblau, C. (1973) *Biochem. Biophys. Res. Commun.* **53**, 70–74

Highberger, J. H., Corbett, C., Kang, A. H. & Gross, J. (1975) *Biochemistry* **14**, 2872–2881

Hulmes, D. J. S., Miller, A., Parry, D. A. D., Piez, K. A. & Woodhead-Galloway, J. (1973) *J. Mol. Biol.* **79**, 137–148

Inagami, T. & Murakami, K. (1972) *Anal. Biochem.* **47**, 501–504

Kivirikko, K. I., Ryhänen, L., Antinnen, H., Bornstein, P. & Prockop, D. J. (1973) *Biochemistry* **12**, 4966–4971

Miller, E. J. (1971) *Biochemistry* **10**, 1652–1659

Miller, E. J. (1972) *Biochemistry* **11**, 4903–4909

Miller, E. J. (1976) *Mol. Cell. Biochem.* **13**, 165–192

Miller, E. J. & Lunde, L. G. (1973) *Biochemistry* **12**, 3153–3159

Miller, E. J., Woodall, D. L. & Vail, M. S. (1973) *J. Biol. Chem.* **248**, 1666–1671

Ryhänen, L. & Kivirikko, K. I. (1974) *Biochim. Biophys. Acta* **343**, 129–137

Traub, W. & Fietzek, P. P. (1976) *FEBS Lett.* **68**, 245–249