Soc. Exp. Biol. 24, 1040–1065.
Park, C. R., Morgan, H. E., Henderson, M. J., Regen, D. M., Cadenas, E., and Post, R. L. (1961), *Recent Prog. Horm. Res. 17,* 493–538.
Rodbell, M. (1964), *J. Biol. Chem. 239,* 375–380.
Rodbell, M. (1966), *J. Biol. Chem. 241,* 130–139.
Suzuki, F., Daikuhara, Y., Ono, M., and Takeda, Y. (1972), *Endocrinology 90,* 1220–1230.

Covalent Structure of Collagen: Amino Acid Sequence of Cyanogen Bromide Peptides from the Amino-Terminal Segment of Type III Collagen of Human Liver†

Jerome M. Seyer and Andrew H. Kang*

ABSTRACT: Human liver type III collagen was prepared by limited pepsin digestion, differential salt precipitation, and carboxymethylcellulose chromatography. Cyanogen bromide digestion of purified type III collagen chains yielded nine distinct peptides. Three peptides, α1(III)-CB3, α1(III)-CB7, and α1(III)-CB6, were isolated by carboxymethylcellulose chromatography and Sephadex G-50 SF gel filtration. Automated Edman degradation together with selective hydroxylamine cleavage and chymotrypsin and trypsin digestion enabled determination of their complete amino acid sequence. Compared with type I collagen, the data show tentative homology of α1(III)-CB3 with α1(I)-CB1, α1(I)-CB2, and α1(I)-CB4; α1(III)-CB7 with α1(I)-CB5; and α1(III)-CB6 with the amino-terminal portion of α1(I)-CB8. Close interspecies homology was found between the sequences presented here with 90 residues of α1(III)-CB3 and 26 of α1(III)-CB8 of calf aorta. The present study establishes the amino acid sequence of 229 residues near the amino terminus or nearly one-quarter of the type III collagen chains. The disaccharide, Glc-Gal, was convalently bound to hydroxylysine at a position corresponding to the same location in the α1(I) chain.

Collagen exists as a triple-stranded helix of three α chains, each containing over 1000 amino acid residues (Gallop et al., 1972; Traub and Piez, 1971). The most extensively studied so far is type I collagen consisting of two α1(I) chains and one α2 chain. A composite amino acid sequence of α1(I) constructed from CNBr peptides of chick, calf, and rat collagen has been tabulated (Hulmes et al., 1973; Gallop and Paz, 1975; Piez, 1977; Fietzek and Kuhn, 1976).

Recently, a genetically distinct collagen, type III, has been identified and found to exist with type I collagen in most soft connective tissues, such as human skin (Chung and Miller, 1974; Chung et al., 1974; Epstein, 1974), aorta (Fietzek and Rauterberg, 1975; Trelstad, 1974), lung (Hance et al., 1975; Seyer et al., 1976a), liver (Gay et al., 1975; Rojkind and Martinez-Palomo, 1976), and hypertrophic scar (Seyer et al., 1976b). Nine cyanogen bromide peptides from type III collagen of human skin and lung have been isolated and characterized (Chung et al., 1974; Seyer et al., 1976a). Partial amino acid sequences of three of these, α1(III)-CB4, α1(III)-CB5, and α1(III)-CB6, have been determined, which has enabled identification of their location within the α1(III) chains (Fietzek and Rauterberg, 1975). They were suggested to be homologous to α1(I)-CB3, α1(I)-CB7, and α1(I)-CB8, respectively.

This report represents a complete sequence analysis of three CNBr peptides, α1(III)-CB3, α1(III)-CB7, and α1(III)-CB6, which most probably comprise the amino-terminal one-quarter of human type III chains. The alignment of the three peptides within the collagen chain was tentatively suggested to be 3-7-6 by homology with the known α1(I) sequence. The α1(III)-CB7 was homologous to α1(I)-CB5 and contained a similar single Glc-Gal-Hyl residue at position 103 of the collagen polypeptide chain.

Materials and Methods

Preparation of Type III Collagen. Human cirrhotic liver was obtained from adult males after autopsy. Hepatic blood vessels were removed as much as possible and all subsequent operations were performed at 4 °C. The tissue was pulverized with a mechanical meat grinder, followed by brief homogenization in a Waring blender. The homogenate was subsequently extracted five times with 0.05 M Tris-HCl (pH 7.4), followed by five successive washes with distilled water in order to remove as much soluble noncollagenous material as possible.

Hepatic collagen was extracted by limited pepsin digestion (Chung and Miller, 1974) using 1 g of enzyme per 50 g wet weight at pH 2.8 (0.5 M acetic acid adjusted with formic acid). All operations were performed at 4 °C as previously described (Seyer et al., 1976a). Three successive pepsin digestions for 72 h enabled solubilization of nearly 80% of the total liver collagen. The collagens present in the three extracts were precipitated by dialysis against 0.01 M Na_2HPO_4 and harvested by centrifugation. The precipitate was redissolved in 0.5 M acetic acid and precipitated again by the addition of NaCl to a final concentration of 1 M. This precipitate was resolu-

† From the Connective Tissue Research Laboratory, Veterans Administration Hospital, and the Departments of Medicine and Biochemistry, University of Tennessee Center for the Health Sciences, Memphis, Tennessee 38104. *Received October 20, 1976.* This work was conducted under Veterans Administration Research Projects No. 4826-01, 4826-02, and 7820-01 and was supported in part by United States Public Health Service Grant AM-16506.

bilized in 0. M Tris[1]/1 M NaCl, pH 7.4, and type III collagen was selectively precipitated by a careful addition of NaCl to a final concentration of 1.7 M and collected by centrifugation. The remaining collagen in the supernate (predominately type I) was subsequently precipitated by increasing the NaCl concentration to 2.5 M. Type III collagen was dissolved in 0.1 M acetic acid, dialyzed exhaustively against the same solution, and lyophilized.

Preparation of CNBr Peptides. Purified type III collagen (1 g) was dissolved in 100 mL of 70% formic acid, and 2 g of CNBr (Pierce Chemical Co., Rockford, Ill.) was added to the N_2-flushed solution. Digestion was allowed to proceed for 4 h at 40 °C (Seyer et al., 1976a). Formic acid and CNBr were separated from the liberated peptides by gel filtration on Bio-Gel P-2 (100–200 mesh) (Bio-Rad Laboratories, Richmond, Calif.) with 0.02 M sodium citrate/0.02 M NaCl (pH 3.8) as the eluting solvent. The material in the excluded volume was warmed (45 °C) for 10 min to ensure denaturation, and subsequently applied directly to CM-cellulose columns (2.5 × 20 cm) equilibrated with the same buffer and maintained at 43 °C. Separation of the CNBr peptides was obtained with a linear gradient of NaCl from 0.02 to 0.16 M (total volume, 2000 mL). A flow rate of 250 mL/h was used and the effluent was continuously monitored at 230 nm. The fractions representing each peak were pooled, lyophilized, desalted on Bio-Gel P-2 in 0.1 M acetic acid and relyophilized.

The peptides α1(III)-CB3, α1(III)-CB6, and α1(III)-CB7[2] were further purified by Sephadex G-50 SF (Pharmacia) gel filtration. The samples (40 mg) were dissolved in 20 mL of 1 M $CaCl_2$/0.05 M Tris (pH 7.4), heated to 45 °C for 15 min to ensure denaturation prior to application on 4 × 110 cm columns equilibrated with 0.04 M sodium acetate (pH 4.8). The effluent was continuously monitored at 232 nm. A drop of tritiated water was used to mark the column volume (Piez, 1968).

Enzymatic Hydrolysis. Digestions with trypsin (Worthington TPCK-treated, three times crystallized) were performed in 0.2 M NH_4HCO_3 (pH 8.0). An enzyme/substrate ratio of 1:50 (wt/wt) was used. The enzyme was added as a 0.1% solution in the same buffer. Incubations were performed at 37 °C for 4 h and digestions were terminated by lyophilization after diluting the reaction mixture fivefold with cold distilled water. Chymotryptic digestions were performed in a similar manner, except that an enzyme (α-chymotrypsin, three times crystallized, Worthington) to protein ratio of 1:100 was used, and the incubation at 37 °C was confined to 2 h. Soybean trypsin inhibitor (Worthington) at a weight ratio to the substrate of 1:1000 was used to prevent any contaminating tryptic activity.

Maleylation of α1(III)-CB3 and Tryptic Digestion of the Product. The peptide α1(III)-CB3 (12 mg) was maleylated in 4 mL of 0.1 M sodium pyrophosphate buffer, pH 9.0, with a 20-fold molar excess relative to lysine of resublimed maleic anhydride (Butler et al., 1969). Maleic acid and salts were removed at the end of the reaction by chromatography on a 2.5 × 50 cm column of Sephadex G-25 equilibrated with 0.01 M NH_4OH, and the excluded protein was lyophilized. Tryptic digestion was performed as described above. Maleyl groups were subsequently removed from the lyophilized digestion products by heating in 4 mL of pyridine–acetic acid (1:10) at 60 °C for 6 h. The deblocked lysine-containing peptides were either sequenced directly or subjected to further trypsin digestion.

Hydroxylamine Cleavage of α1(III)-CB6. Cleavage with hydroxylamine (Eastman Organic Chemicals) of α1(III)-CB6 was performed at pH 9.0. The peptide (5 mg) was dissolved in 1 mL of H_2O and denatured for 15 min at 45 °C, and an equal volume of freshly prepared, cold 2 M NH_2OH in 1 M K_2CO_3 (pH 9.0) was added (Balian et al., 1971). The reaction was allowed to proceed for 90 min at 45 °C. For termination of the reaction, the solution was adjusted to pH 4.0 with HCl and applied directly to a Sephadex G-50 SF column (2.0 × 110 cm) equilibrated with 0.1 M acetic acid. No attempts were made to quantitate hydroxamate production.

Column Chromatography of Enzymatically Derived Peptides. Enzymatic digests were separated by one or more of the following separation procedures. Sephadex G-50 SF columns (2.0 × 110 cm) in 0.04 M sodium acetate (pH 4.8) were used for molecular-sieve chromatography. The effluent was monitored continuously at 230 nm and selected fractions were desalted on Bio-Gel P-2 columns (2.0 × 20 cm) in 0.1 N acetic acid. Phosphocellulose chromatography was performed on 1 × 6 cm columns of phosphocellulose (Whatman) equilibrated with 0.001 M sodium acetate (pH 3.8) at 43 °C. Samples of trypsin digests in 5 mL of the buffer were applied and elution was obtained using a linear gradient of NaCl from 0 to 0.1 M (flow rate 1.4 mL/min) over a total volume of 1000 mL. The column effluents were monitored at 230 nm and the peptide fractions were lyophilized and desalted as above on Bio-Gel P-2 using 0.1 M acetic acid as the eluent.

Automated peptide analyses of trypsin digests were performed on a 0.9 × 25 cm column of PA-35 resin (Beckman Instruments, Palo Alto, Calif.) at 60 °C, which enabled separation of the smaller tryptic peptides (Kang and Gross, 1970). An automatic analyzer (Technicon Instruments, Inc., Ardsley, N.Y.) equipped with a stream-split device allowed continuous monitoring of the ninhydrin reactivity of a portion of the effluent. Ninety percent of the column effluent was collected in a fraction collector. Separation was achieved utilizing a nine-chamber gradient starting with 0.02 M sodium citrate buffer, pH 3.1 (Technicon peptide methodology). Fractions were lyophilized and desalted utilizing 1 × 2 cm columns of Aminex 50-X8 (200–400 mesh) (Bio-Rad Laboratories, Richmond, Calif.) (Hirs, 1967).

Amino Acid Analysis. Samples were hydrolyzed in constant-boiling HCl at 108 °C for 24 h under an atmosphere of N_2. Analyses were performed on an automatic analyzer (Beckman Instruments, Palo Alto, Calif.) using a single-column method previously described (Kang, 1972). No correction factors were used for losses of the labile amino acids or for the incomplete release of valine. Analysis of the disaccharide-linked amino acid, Glc-Gal-Hyl, was carried out after 2 N NaOH hydrolysis of α1(III)-CB3 at 110 °C for 24 h (Askenasi and Kefalides, 1972). The hydrolysate (5 mg/0.5 mL) was diluted tenfold with H_2O, neutralized with HCl, and applied directly to the amino acid analyzer.

Edman Degradation. Automatic Edman degradations were performed with a Beckman Sequencer, Model 890C, according to the principles described by Edman and Begg (1967). Either the Slow Protein-Quadrol (072172C) or Slow Peptide-DMAA (071472) program of Beckman Instruments was employed. Small peptides were treated with 2-amino-1,5-naphthalenedisulfonic acid in the presence of *N*-ethyl-*N'*-(3-dimethylaminopropyl)carbodiimide to decrease extraction from the

[1] Abbreviations used are: CM-cellulose, carboxymethylcellulose; Pth, phenylthiohydantoin; Tris, 2-amino-2-hydroxymethyl-1,3-propanediol.

[2] Nomenclature of CNBr peptides used was according to Chung et al. (1974). Tryptic, chymotryptic, and hydroxylamine-derived peptides were numbered in the order of their elution from the specific columns.

TABLE I: Amino Acid Composition of α1(III)-CB3 and Tryptic Peptides of Type III Collagen from Human Cirrhotic Liver.[a]

Amino Acid	α1(III)-CB3	C1	C2	T1	T2	T3	T4	T(1–4)	T(2–3)	Total[b]
Hydroxyproline	18.2 (18)	6.7 (7)	10.6 (11)	10.9 (11)	1.7 (2)	2.1 (2)	3.0 (3)	13.6 (14)	3.8 (4)	18
Aspartic acid	1.1 (1)	0.1	1.2 (1)	—	0.1	0.1	1.3 (1)	1.0 (1)	0.2	1
Threonine	1.3 (1)	0.9 (1)	—	0.9 (1)	—	—	0.1	0.9 (1)		1
Serine	5.4 (5)	2.6 (3)	1.8 (2)	4.6 (5)	0.1	—	0.2	4.9 (5)		5
Glutamic acid	5.4 (5)	0.2	5.4 (5)	3.3 (3)	0.1	0.2	2.0 (2)	5.3 (5)	0.1	5
Proline	10.8 (11)	4.3 (4)	6.7 (7)	9.3 (9)	0.9 (1)	1.1 (1)	0.4	9.2 (9)	2.2 (2)	11
Glycine	30.1 (30)	10.4 (10)	19.9 (20)	19.4 (19)	4.0 (4)	3.1 (3)	4.1 (4)	23.1 (23)	6.9 (7)	30
Alanine	5.4 (5)	1.9 (2)	3.1 (3)	4.4 (4)	1.1 (1)	0.2	0.4	3.9 (4)	0.8 (1)	5
Valine	0.6 (1)	0.8 (1)		0.6 (1)	—	—	—	0.6 (1)		1
Isoleucine	2.7 (3)	—	2.8 (3)	1.2 (1)	1.0 (1)	0.8 (1)	0.2	0.9 (1)	2.1 (2)	3
Leucine	2.0 (2)	0.9 (1)	1.2 (1)	0.9 (1)	0.2	0.9 (1)	0.3	1.1 (1)	1.2 (1)	2
Tyrosine	1.6 (2)	1.6 (2)	—	1.5 (2)	—	—	—	1.5 (2)	—	2
Phenylalanine	1.0 (1)	—	1.1 (1)	0.1	1.1 (1)	0.1	—	—	1.2 (1)	1
Hydroxylysine[c]	0.3 (0.2)		0.3	0.1	—	0.2	0.1	0.1	0.2	0.4
Lysine	2.0 (1.8)	0.2	1.8 (2)	1.2 (1)	—	0.7 (1)	0.3	0.7 (1)	0.9 (1)	2.2
Histidine	0.9 (1)	0.6 (1)	—	0.7 (1)	—	—	—	0.8 (1)	—	1
Arginine	3.2 (3)	—	2.8 (3)	—	—	0.2	3.1 (3)	2.9 (3)	—	3
Homoserine[d]	1.2 (1)	—	1.0 (1)	—	1.1 (1)	—			1.0 (1)	1
Total	92	32	60	59	11	9	13	72	20	92

[a] Values expressed as residues per peptide. A dash indicates the level was less than 0.1 residue per peptide. Numbers in parentheses indicate assumed integral values. [b] Total of T1, T2, T3, and T4. [c] Partial lysine hydroxylation was noted. [d] Includes homoserine lactone.

TABLE II: Amino Acid Composition of α1(III)-CB7 and Tryptic Peptides of Type III Collagen from Human Cirrhotic Liver.[a]

Amino Acid	α1(III)-CB7	T-1	T-2
Hydroxyproline	3.3 (3)	1.1 (1)	1.8 (2)
Aspartic acid	4.1 (4)	0.2	1.7 (2)
Threonine	1.0 (1)	0.9 (1)	0.2
Serine	—	—	
Glutamic acid	4.4 (4)	1.2 (1)	2.1 (2)
Proline	1.4 (1)	—	0.7 (1)
Glycine	12.4 (12)	3.1 (3)	4.8 (5)
Alanine	2.4 (2)	1.0 (1)	1.0 (1)
Valine	—	0.2	
Isoleucine		—	
Leucine	2.1 (2)	0.8 (1)	1.0 (1)
Phenylalanine	0.8 (1)	—	
Hydroxylysine[b]	2.6 (3)	0.8 (1)	0.2
Lysine	0.4	0.1	
Histidine	0.9 (1)	—	
Arginine	2.1 (2)	—	
Homoserine[c]	1.2 (1)	—	0.9 (1)
Glc-Gal-Hly[d]	0.6 (1)	—	
Total	37	9	15

[a] Values expressed as residues per peptide. A dash indicates the level was less than 0.1 residue per peptide. Numbers in parentheses indicate assumed integral values. [b] Partial hydroxylation of lysine was noted. [c] Includes homoserine lactone. [d] From 2 N NaOH hydrolysis.

reaction cup (Foster et al., 1973). Slight modifications included use of smaller quantities of the reagents, thus allowing identification of free Pth-glutamic and -aspartic acids, and degradation to the penultimate COOH terminus (Dixit et al., 1975). The homoserine peptides were treated initially with 0.1% NH_4OH for 10 min to convert the homoserine lactone to free homoserine. Ammonia was removed by lyophilization. The Pth amino acids were identified by either gas chromatography before and after trimethylsilylation (Pisano and Bronzert, 1969) or by high-pressure liquid chromatography (Zimmerman et al., 1973). The COOH-terminal residues were not identified but inferred from the amino acid composition of the peptides, specificities of either trypsin or CNBr, or subsequent overlapping sequence analysis.

Results

Isolation of α1(III)-CB3, α1(III)-CB6, and α1(III)-CB7. Relatively large quantities of type III collagen were obtained from human cirrhotic liver. The material in the 1.7 M NaCl precipitate was found to be greater than 95% type III in most cases by CM-cellulose and molecular-sieve chromatography and was therefore used directly for CNBr cleavage (Seyer et al., 1976a,b). The subsequent peptides were separated by CM-cellulose chromatography into nine distinct peptides indistinguishable from those previously reported[3] (Chung et al., 1974; Seyer et al., 1976a,b).

Three of the peptides α1(III)-CB3, α1(III)-CB6, and α1(III)-CB7 were further purified separately by Sephadex G-50 SF gel filtration. Rechromatography on a calibrated column of Sephadex G-50 SF indicated molecular weights of 9000, 4000, and 10 000 for α1(III)-CB3, α1(III)-CB7, and α1(III)-CB6, respectively. Their amino acid composition is shown in Tables I, II, and III. The peptide α1(III)-CB7 contained nearly 1 mol of Glc-Gal-Hyl per mol of peptide (Table II).

Amino Acid Sequence of α1(III)-CB3. A tryptic digest of α1(III)-CB3 was fractionated on Sephadex G-50 SF, which yielded two fractions of different apparent molecular weights.[3] The larger fraction (apparent molecular weight 5500) contained T1, which was shown to be homogeneous by subsequent phosphocellulose chromatography. The smaller fraction (apparent molecular weight 1100) was further separated into three peptides, T2, T3, and T4, by phosphocellulose chromatography.[3] The amino acid composition of these tryptic peptides is presented in Table I. The four peptides together account for the entire amino acid content of α1(III)-CB3. Since there are five basic residues in the parent peptide, one might have expected six tryptic peptides. However, two of the arginyl residues were followed by hydroxyprolyl residues (Figure 1), and

[3] Tables and figures containing these data are deposited as Supplementary Material.

TABLE III: Amino Acid Composition of α1(III)-CB6, Tryptic and Hydroxylamine Peptides of Type III Collagen from Human Cirrhotic Liver.[a]

	α1(III)-CB6	HA1	HA2	T1	T2	T3	T4	T5	T6	T7	T8	Total[b]
4-Hydroxyproline	13.8 (14)	9.4 (9)	4.8 (5)	4.6 (5)	2.1 (2)	4.0 (4)	2.2 (2)	0.1		0.8 (1)		14
Aspartic acid	5.1 (5)	4.1 (4)	1.1 (1)	1.0 (1)	1.0 (1)	1.0 (1)		0.2	2.2 (2)	0.1	0.2	5
Threonine	1.0 (1)	1.0 (1)	0.1	0.9 (1)	0.1							1
Serine	4.6 (5)	3.9 (4)	0.9 (1)	2.0 (2)	0.8 (2)	1.1 (1)						5
Glutamic acid	8.3 (8)	3.2 (3)	4.6 (5)	1.3 (1)	0.8 (2)	3.1 (3)		1.3 (1)	0.2	0.9 (1)	0.1	8
Proline	6.7 (7)	4.1 (4)	2.7 (3)	2.0 (2)	1.1 (1)	2.8 (3)				0.2	1.1 (1)	7
Glycine	34.4 (34)	20.6 (21)	13.2 (13)	8.1 (8)	6.3 (6)	10.2 (10)	4.1 (4)	1.2 (1)	1.9 (2)	2.1 (2)	1.2 (1)	34
Alanine	11.2 (11)	8.5 (8)	2.9 (3)	2.1 (2)	2.1 (2)	2.2 (2)	3.3 (3)	0.2	1.2 (1)	0.8 (1)		11
Valine	0.8 (1)	0.9 (1)	0.1		0.7 (1)							1
Isoleucine	1.0 (1)	0.2	0.7 (1)		0.1	0.8 (1)						1
Leucine	1.1 (1)	1.1 (1)	0.2				0.8 (1)					1
Phenylalanine	1.0 (1)	1.0 (1)	0.1	0.9 (1)	0.1							1
Hydroxylysine[c]	1.0 (1)	0.6 (1)	0.2	0.8 (1)	0.1	0.4						1
Lysine[c]	1.1 (1)	0.3	0.5 (1)	0.4	0.2	0.7 (1)						1
Histidine	0.8 (1)	0.3 (1)	0.6			0.9 (1)						1
Arginine	5.7 (6)	5.0 (5)	1.2 (1)		1.1 (1)		1.8 (2)		0.9 (1)	1.2 (1)	1.1 (1)	6
Homoserine[d]	1.1 (1)	0.1	0.9 (1)					0.8 (1)				1
Total	99	64	36	25	18	27	12	3	6	6	3	

[a] Values expressed as residues per peptide. A dash represents values less than 0.1 residue per peptide. Numbers in parentheses represent integral values. [b] Total of either HA1 and HA2 or T1–T8. [c] Partial lysine hydroxylation was noted. [d] Includes homoserine lactone.

were not cleaved by trypsin under the conditions of the experiment (Highberger et al., 1971).

The alignment of the tryptic peptides were deduced from the following evidence. Only T2 contained a residue of homoserine, and therefore must be the COOH terminus. Automated Edman degradation of T1 yielded a sequence identical to that obtained for the intact α1(III)-CB3 (Figure 1). Therefore, T1 must be thg NH_2 terminus. The relative position of T2 and T3 was determined from isolation and characterization of tryptic peptides obtained from maleylated α1(III)-CB3. A tryptic digest of maleylated α1(III)-CB3 yielded two major peptides, one with an apparent molecular weight of 6500 and the other 2000, which could be separated by molecular-sieve chromatography on Sephadex G-50 SF (figure not shown). The larger peptide had an amino acid composition identical to the sum of that for T1 and T4 (Table I). The composition of the smaller peptide was consistent with it being the uncleaved (due to maleylation of lysine) T(2–3) (Table I). From these data, the alignment of the tryptic peptides was deduced to be T1-T4-T3-T2.

The complete amino acid sequence of α1(III)-CB3 was determined by automated Edman degradation of intact α1(III)-CB3, each of the tryptic peptides, and a chymotryptic peptide (see below) of α1(III)-CB3. The results are given in Figure 1. Automated sequenator degradation of T4, T3, and T2 established their amino acid sequence. The peptide T1, containing 59 residues, could be degraded through the first 44 residues (Figure 1). In order to determine the sequence of the COOH-terminal portion of T1, the following experiments were performed. Intact α1(III)-CB3 was digested with chymotrypsin and the products were separated on Sephadex G-50 SF.[3] Two peptides, C1 and C2, were obtained. Their amino acid composition is shown in Table I. Since C1 contains a residue of homoserine, the order of the chymotryptic peptides must be C2-C1. Edman degradation of C1 through 30 residues established the remaining sequence of T1.Each of the peptides was degraded at least twice. The yields of Pth amino acids at each step of degradation together with other details of the automated sequenator operation are documented in the Supplementary Material.[3]

Amino Acid Sequence of α1(III)-CB7. Automated amino acid sequence analysis was successful in determining 33 of the 37 residues of α1(III)-CB7[3] (Figure 2), greatly facilitating the analysis of its covalent structure. Trypsin digestion followed by phosphocellulose chromatography yielded only two clearly identifiable peptides, T1 and T2, which could be detected by absorbance at 230 nm.[3] Their amino acid composition is indicated in Table II. Their amino acid sequences were consistent with residues 14–22 and 23–37 (Figure 2). With the internal sequence determination of T2 and the above 33 residues of the intact peptide, the complete sequence of α1(III)-CB7 was established.

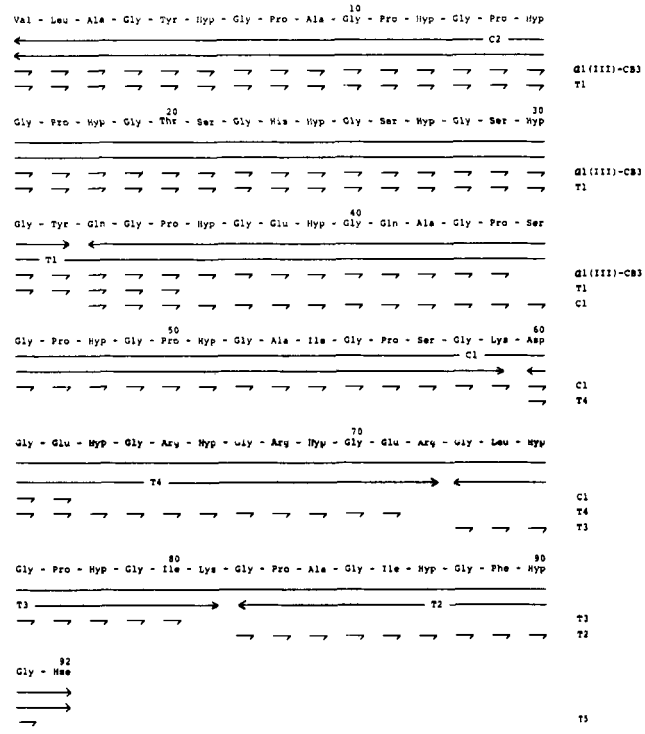

FIGURE 1: The complete amino acid sequence of α1(III)-CB3 of human liver type III collagen. The peptides isolated after trypsin and chymotrypsin digestion are indicated by long arrows (↔). Short horizontal arrows (→) indicate the extent of Edman degradation of each peptide.

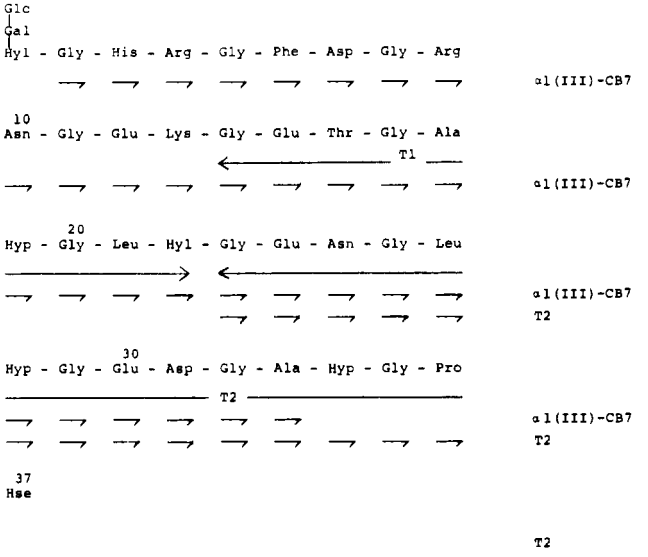

FIGURE 2: The complete amino acid sequence of α1(III)-CB7 of human liver type III collagen. The peptides isolated after trypsin cleavage are indicated by long arrows (↔). Short horizontal arrows (→) indicate the extent of Edman degradation of each peptide. Residue 1 was identified by subtractive Edman degradation. The disaccharide Glc-Gal is diagramatically shown attached to the δ hydroxyl group of hydroxylysine.

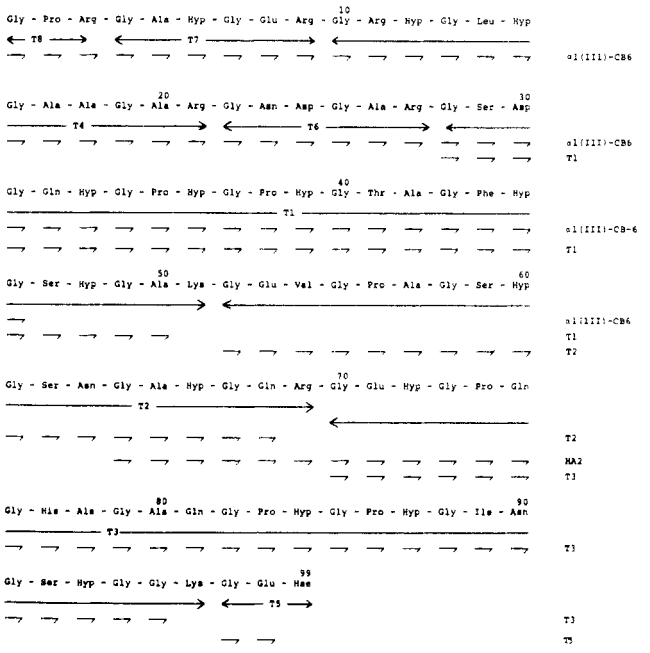

FIGURE 3: The complete amino acid sequence of α1(III)-CB6 of human liver type III collagen. The trypsin and hydroxylamine-derived peptides isolated are indicated by long arrows (↔). Short horizontal arrows (→) indicate the extent of Edman degradation of each peptide degraded.

A major difficulty was encountered in the identification of residue 1 from α1(III)-CB7. Trace amounts of Pth-Hyl were identified by high-pressure liquid chromatography. However, the yield at step 2 (Pth-Gly) was nearly 100-fold greater. Attempts to identify the NH_2-terminal Pth derivative by gas chromatography, thin-layer chromatography, or amino acid analysis after hydrolysis with HI were unsuccessful, yielding a complete blank for that position. The intact peptide was therefore degraded manually for one step of Edman degradation. Amino acid analysis of the remaining peptide residue yielded one less hydroxylysine. In the same manner, after hydrolysis with 2 N NaOH, the peptide was found to no longer contain the disaccharide Glc-Gal-Hyl previously found in intact α1(III)-CB7. By deductive reasoning, therefore, residue 1 was tentatively identified to be Glc-Gal-Hyl.

Amino Acid Sequence of α1(III)-CB6. Fractionation of tryptic digests of α1(III)-CB6 on Sephadex G-50 SF, phosphocellulose, and PA-35 (Beckman) enabled isolation of eight distinct tryptic peptides.[3] Although there is a total of eight basic residues in α1(III)-CB6, one of the arginyl residues was followed by a residue of hydroxyproline (residue 11, see Figure 3), and apparently was not cleaved by trypsin (Highberger et al., 1971). Their amino acid composition is presented in Table III.

Several experiments were performed to determine the alignment of the tryptic peptides. First, intact α1(III)-CB6 was subjected to automated Edman degradation, and the sequence of residues 1 through 61 was determined (Figure 3). These results and further sequence studies on individual tryptic peptides showed that the first five tryptic peptides in the alignment were T8-T7-T4-T6-T1. Since T5 contains the homoseryl residue, it must be located at the COOH terminus. The relative order of the two remaining peptides, T2 and T3, was determined from experiments with hydroxylamine cleavage.

Fractionation of the products of hydroxylamine reaction of α1(III)-CB6 on Sephadex G-50 SF yielded three peptide peaks.[3] The first peak eluting at the position of the starting material was apparently uncleaved peptide. Each of the remaining two peaks, HA1 and HA2, eluted as homogeneous material when subjected to phosphocellulose chromatography. Their amino acid composition is shown in Table III. The smaller fragment, HA2, contained a residue of homoserine, indicating it to be the COOH terminus. Sequential degradation of the HA2 through the first 12 residues, together with the sequence analyses of T2 and T3, indicated the order of these two peptides to be T2-T3. Thus, the alignment of the eight tryptic peptides of α1(III)-CB6 was deduced to be T8-T7-T4-T6-T1-T2-T3-T5.

Determination of amino acid sequences of the individual tryptic peptides in an automated sequenator then established the complete sequence of α1(III)-CB6. The results are summarized in Figure 3. The amounts of specific peptides used, the number of useful cycles degraded, and the observed yields of Pth amino acids at each step of degradation are documented in the Supplementary Material.[3]

Discussion

The covalent structure of three CNBr peptides of the type III collagen chain of human cirrhotic liver was determined. These results are presented in Figure 4 together with a comparison of the α1(I) collagen chain (residue 1 → 238). Although complete alignment of the CNBr peptides of α1(III) chain has not been established, these results allow tentative positioning at the NH_2 portion of the molecule. The amino acid sequence of α1(III)-CB3 suggests homology of this peptide with α1(I)-CB1, α1(I)-CB2, and α1(I)-CB4, by comparison of the results obtained in this study with the previously reported sequence for the α1(I) peptides (Bornstein, 1969; Butler, 1969; Butler and Ponds, 1971; Fietzek and Kuhn, 1975). Residue 1 of α1(III)-CB3 therefore corresponds to residue 11 of the α1(I) chain. It was assumed that a portion of the nonhelical NH_2-terminal segment of α1(III) (residues 1–10 of α1(III)) was removed by pepsin digestion used during the preparation (Bornstein et al., 1966). The peptide α1(III)-CB7 was located to follow α1(III)-CB3 by its sequence homology with α1(I)-CB5. This alignment would place the Glc-Gal-Hyl at identical

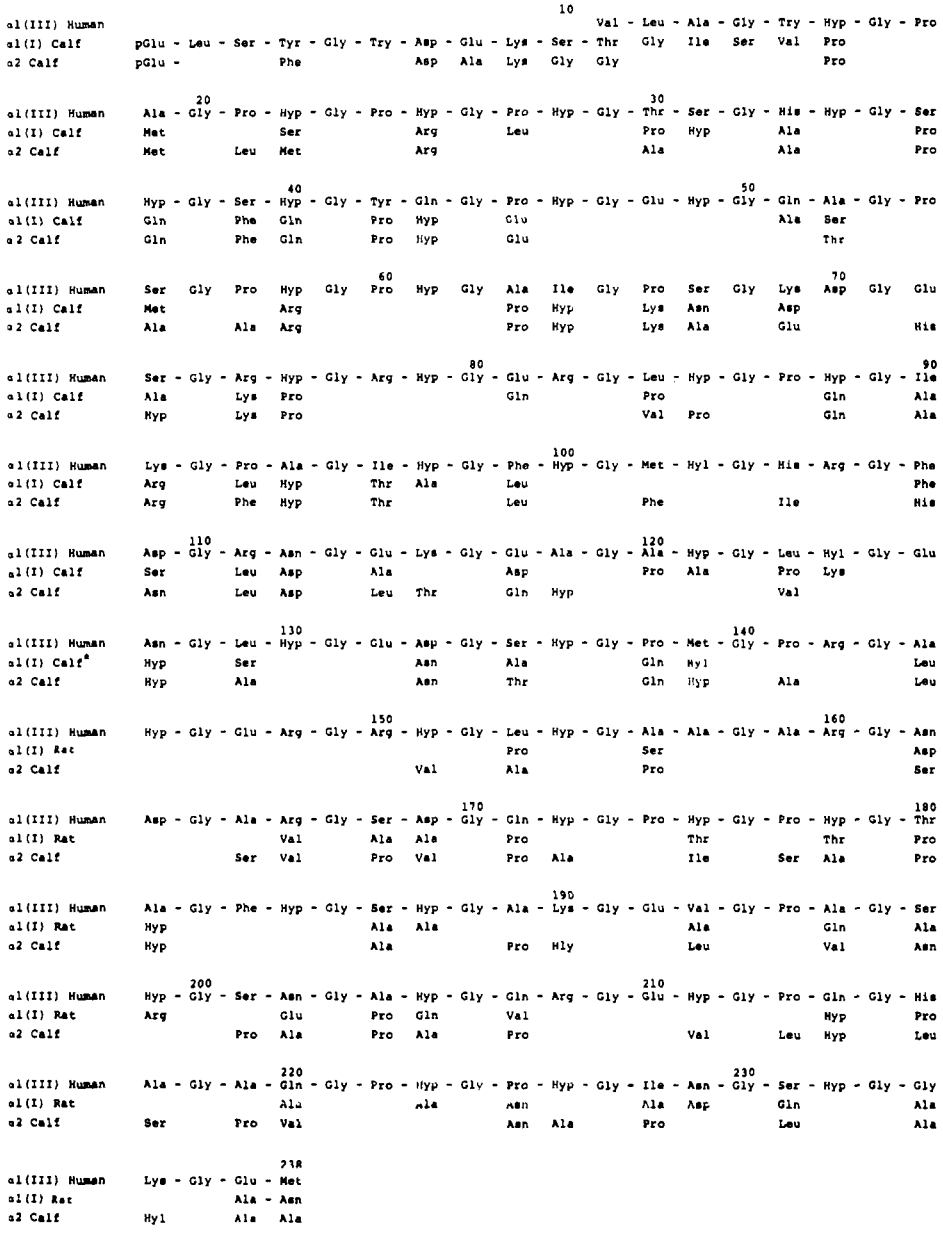

FIGURE 4: Comparison of the amino acid sequence of the NH_2-terminal portion of pepsin-solubilized human liver type III collagen with the identical regions from calf + rat $\alpha 1$(I) and calf $\alpha 2$ collagen chains. Pepsin cleavage removed residues 1 → 10 leaving $\alpha 1$(III)-CB3 as the NH_2-terminal CNBr peptide followed by $\alpha 1$(III)-CB7 and $\alpha 1$(III)-CB6. The type III collagen peptides presented correspond to $\alpha 1$(I)-CB1-2-4-5 and the NH_2-terminal portion of $\alpha 2$-CB4 of type I collagen chains. Identical residues in the same position of each chain are indicated by blank spaces, except of residues numbered 2,3,5,6,12,13, and 15 of the $\alpha 2$ chain, which are genetic deletions in the calf $\alpha 2$ collagen chain. The hydroxylysine at position 103 contains the disaccharide Glc-Gal in all three chains. The residues 1 → 169 are from calf skin (Fietzek and Kuhn, 1976) and 170 → 238 are from rat skin (Bornstein, 1969; Butler, 1969; Butler and Ponds, 1971), since the corresponding sequence of calf $\alpha 1$(I)-CB8 has not been established. The $\alpha 2$ sequence is from calf skin (Fietzek and Rexrodt, 1975).

positions in the two types of the collagen α chains. The peptide $\alpha 1$(III)-CB6 was located following $\alpha 1$(III)-CB7 by its homology with the NH_2-terminal portion of $\alpha 1$(I)-CB8. This location is consistent with that proposed earlier by Fietzek and Rauterberg (1975) on the basis of their partial sequence analysis of $\alpha 1$(III)-CB6 from bovine aorta.

The primary structure of $\alpha 1$(III), as obtained in this study, is consistent with several principles previously derived from studies of the type I collagen chains (Hulmes et al., 1973; Gallop and Paz, 1975; Piez, 1976). Glycine is present at every third residue in the Gly-X-Y triplet sequence, and hydroxyprolyl and hydroxylysyl residues are confined to the Y positions. However, there are several interesting features. There are two residues of tyrosine within the helical portion of the type III chains. A tyrosyl residue has been found only once in the helical region of collagen, i.e., type II collagen (Butler et al., 1976). Also noteworthy is the identity of the octapeptide sequence (Hyp-Gly-Met-Hyl-Gly-His-Arg-Gly) around the glycosylated hydroxylysyl residue of the type III chain with that found in other collagen chains (Hulmes et al., 1973; Gallop and Paz, 1975; Piez, 1976; Butler et al., 1976). This identical sequence is located in the same position in $\alpha 1$(I) and $\alpha 2$ and $\alpha 1$(II). The importance of this site as a possible helical-cross-link site (Gallop and Paz, 1975) and recognition site for galactosyl transferase has been discussed (Butler and Ponds, 1971).

The present data also allow a comparison with the sequence of the homologous segment of the $\alpha 1$(I) and $\alpha 2$ chains from

other mammalian sources. Thus, α1(III)-CB3 contains 40 and 44 diffferences, respectively, when compared with the corresponding sequence of calf α1(I) and α2. There are 13 residue differences between α1(III)-CB7 and α1(I)-CB5 of calf, and 33 between α1(III)-CB6 and the corresponding segment of rat skin α1(I)-CB8. Although most substitutions are chemically conservative, some involve charged residues. However, the number of charged substitutions is relatively small; a total of 15 was noted in the portion of the α chain constituting approximately one-fourth of the length of the α chain.

The sequence of human α1(III)-CB6 obtained in the present investigation is identical with the 26 NH_2-terminal residues reported for calf aorta α1(III)-CB6 (Fietzek and Rauterberg, 1975). The interspecies sequence variation was somewhat greater with α1(III)-CB3 between human and calf (Fietzek and Kuhn, 1976). These included Ile → Leu substitution (position 12), Ala → Ser (positions 36, 39, and 55), Pro → Glu (position 81), Phe → Leu (position 84), and Met → Ile (position 90). The latter Met → Ile substitution accounts for the additional nine residues in human α1(III)-CB3 compared with the homologous peptide of calf aorta (Fietzek et al., 1976). This suggests that interspecies homology of α1(III) may be very close as is the case for α1(I) (Dixit et al., 1975). A relatively greater degree of interspecies sequence variations has been noted for α2 (Dixit et al., 1976).

The present data tentatively locate the first three peptides from the NH_2 terminus of the type III chain. It has been suggested that α1(III)-CB4 is homologous with α1(I)-CB3, α1(III)-CB5 is homologous with the NH_2-terminal portion of α1(I)-CB7, and α1(III)-CB9 is homologous with the COOH-terminal portion of α1(I)-CB7 and α1(I)-CB6 (Fietzek and Rauterberg, 1975). Preliminary sequence data obtained in our laboratory suggests that the remaining peptides, α1(III)-CB1, α1(III)-CB2, and α1(III)-CB8, are homologous to the COOH-terminal segment of α1(I)-CB8. Thus, tentatively, the alignment of the CNBr peptides of α1(III) can be written as α1(III)-CB3-7-6-1-8-2-4-5-9. However, additional studies will be needed to confirm the proposition.

Acknowledgments

The authors are grateful to Dr. J. M. Young of this Veterans Administration Hospital for pathologic examinations of tissue samples, and for valuable discussions. We also acknowledge the excellent technical assistance of T. Marshall, P. Hill, and E. Rowland.

Supplementary Material Available

Tables and figures containing additional data on peptide composition are given (10 pages). Ordering information is given on any current masthead page.

References

Askenasi, R. S., and Kefalides, N. A. (1972), *Anal. Biochem. 47,* 67.

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

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

Bornstein, P., Kang, A. H., and Piez, K. A. (1966), *Biochemistry 5,* 3803.

Butler, P. J. G., Harris, J. T., Hartley, B. S., and Leberman, R. (1969), *Biochem. J. 112,* 679.

Butler, W. R. (1969), *Biochemistry 8,* 44.

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

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

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

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

Dixit, S. N., Seyer, J. M., and Kang, A. H. (1976), *Eur. J. Biochem.* (in press).

Dixit, S. N., Seyer, J. M., Oronskey, A. O., Corbett, C., Kang, A. H., and Gross, J. (1975), *Biochemistry 14,* 1933.

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

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

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

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

Fietzek, P. P., and Rexrodt, F. W. (1975), *Eur. J. Biochem. 59,* 113.

Fietzek, P. P., and Kuhn, K. (1976), *Int. Rev. Conn. Tiss. Res. 7,* 1.

Fietzek, P. P., Kang, A. H., and Seyer, J. M. (1976), Handbook of Biochemistry and Molecular Biology, 3rd Ed., Vol. 3, Fasman, G. D., Ed., Cleveland, Ohio, CRC Press, p 474.

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

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

Gallop, P. M., and Paz, M. A. (1975), *Physiol. Rev. 55,* 418.

Gay, S., Fietzek, P. P., Remberger, K., Eder, M., and Kuhn, K. (1975), *Klin. Wochenschr. 53,* 205.

Hance, A. J., Bradley, K., and Crystal, R. C. (1975), *J. Clin. Invest. 57,* 102.

Highberger, J. H., Kang, A. H., and Gross, J. (1971), *Biochemistry 10,* 610.

Hirs, C. H. W. (1967), *Methods Enzymol. 11,* 386.

Hulmes, D. J. S., Miller, A., Parry, D. A., Piez, K. A., and Woodhead-Galloway, J. (1973), *J. Mol. Biol. 70,* 137.

Kang, A. H. (1972), *Biochemistry 11,* 1828.

Kang, A. H., and Gross, J. (1970), *Biochemistry 9,* 796.

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

Piez, K. A. (1976), in Biochemistry of Collagen, Ramachandran, G. N., and Redd, A. H., Ed., New York, N.Y., Plenum.

Pisano, J. J., and Bronzert, T. J. (1969), *J. Biol. Chem. 244,* 5597.

Rojkind, J., and Martinez-Palomo, A. (1976), *Proc. Natl. Acad. Sci. U.S.A. 73,* 539.

Seyer, J. M., Hutcheson, E. T., and Kang, A. H. (1976a), *J. Clin. Invest. 57,* 1498.

Seyer, J. M., Swann, D. A., Brickley, D. M., and Glincher, M. J. (1976b), *Clin. Res. 24,* A267.

Traub, W., and Piez, K. A. (1971), *Adv. Protein Chem. 25,* 243.

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

Zimmerman, C. L., Pisano, J. J., and Appella, E. (1973), *Biochem. Biophys. Res. Commun. 55,* 1220.