Covalent Structure of Collagen: Amino Acid Sequence of Chick Skin Collagen α1(I)-CB6B†

S. N. Dixit, J. M. Seyer, A. H. Kang,* and J. Gross

ABSTRACT: The amino acid sequence of the peptide α1(I)-CB6B, the carboxyl-terminal peptide containing 91 residues of α1(I) chain of chick skin collagen, is described. The amino acid sequence was determined by automated Edman degradation of intact peptide plus its tryptic and thermolytic peptides. The comparison of this sequence with the corresponding segment of previously published α1(I)-CB6 of calf skin collagen showed a sequence identity of 90%.

Substantial progress has been made on the determination of the primary structure of different types of collagen chains in the last several years. Type I collagen, the most extensively studied collagen, consists of two α1(I) chains and one α2 chain. The complete amino acid sequence of α1(I) chain can be constructed from the composite amino acid sequence of CNBr peptides of chick, calf, and rat skin collagens (Hulmes et al., 1973; Gallop & Paz, 1975; Fietzek & Kuhn, 1976; Piez, 1976). However, a complete amino acid sequence of α1 chains from a single species has not been completed. The work on the determination of the amino acid sequence of α2, α1(II) (Dixit et al., 1977a,b; Butler et al., 1976, 1977), and α1(III) (Fietzek & Rauterberg, 1975; Seyer & Kang, 1977, 1978) chains is in an advanced stage.

Our laboratories have been committed to complete the primary structure of type I collagen. We have published previously the amino acid sequences of the CNBr peptides α1(I)-CBO, 1, 2, 3, 4, 5, 6A, and 7 (Kang & Gross, 1970; Dixit et al., 1975a,b; Kang et al., 1975; Highberger et al., 1975) of chick skin collagen. This report on the amino acid sequence of α1(I)-CB6B together with the amino acid sequence of α1(I)-CB8 (Highberger et al., manuscript under preparation) will establish the complete amino acid sequence of α1(I) chain of chick skin collagen.

Materials and Methods

Preparation of α1(I)-CB6B. Purified chick skin collagen and the α1(I) chains were prepared from the lathyritic chicks by procedures described elsewhere (Kang et al., 1969a). The peptide α1(I)-CB6B was purified by a combination of phosphocellulose and molecular sieve chromatography (Kang et al., 1969b) from the CNBr digest of α1(I) chain (Bornstein & Piez, 1966).

Enzymatic Hydrolysis. Digestion with trypsin (treated with L-1-tosylamido-2-phenylethyl chloromethyl ketone, Worthington) was carried out in 0.2 M Tris/1 mM $CaCl_2$, pH 7.6, for 4 h at 37 °C. A 1:50 molar ratio of enzyme/substrate was used. Thermolysin (3× crystallized and lyophilized, grade A, Calbiochem) digestion was done for 15 min in 0.2 M NH_4HCO_3, pH 8.0, containing 1 mM $CaCl_2$ at 37 °C. The enzyme/substrate molar ratio was 1:100. At the end of incubations, the samples were lyophilized.

Column Chromatography. The tryptic and the thermolytic peptides were initially fractionated on a column (2.0 × 120 cm) of Sephadex G-50S (Pharmacia) equilibrated with 0.04 M sodium acetate, pH 4.8, and eluted at a flow rate of 12 mL/h. Ion-exchange chromatography was performed on a 1 × 6 cm column of phosphocellulose (P-11, Whatman), equilibrated with 0.001 M acetate, pH 3.8, at 44 °C. After the application of the sample, the column was eluted with a linear gradient of NaCl from 0 to 0.3 M over a total volume of 500 mL. The effluents are continuously monitored at 230 nm in a Gilford spectrophotometer equipped with a flow cell. Appropriate fractions were pooled, lyophilized, and desalted on Bio-Gel P-2, 200–400 mesh (Bio-Rad Laboratories), using 0.1 M acetic acid as the eluent.

Small peptides were further fractionated on a column (0.9 × 25 cm) of PA-35 resin (Beckman) using a Technicon peptide analyzer and a nine-chamber varigrade gradient of sodium citrate (Technicon). By the use of a split-stream device, a portion of the effluent was continuously monitored for ninhydrin reactivity. The appropriate fractions were pooled, desalted on Dowex-50, and lyophilized.

Amino Acid Analysis. Samples were sealed under an atmosphere of nitrogen in 6 N HCl and hydrolyzed for 24 h at 110 °C. The amino acid analyses were performed on an automatic analyzer (Beckman 121) using a single column method (Kang, 1972). No correction factors were used for the losses of 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(I)-CB6B at 110 °C for 24 h (Askensasi & Kefalides, 1972; Miller, 1972). The hydrolysate was diluted tenfold with distilled water, neutralized with HCl, and analyzed directly on the amino acid analyzer.

Edman Degradation. Automated Edman degradations (Edman & Begg, 1967) were performed on a Beckman Sequencer (890C) employing the slow peptide-DMAA (071472) program of Beckman Instruments.

All peptides were modified at the carboxyl terminal by treatment with 2-amino-1,5-naphthalenedisulfonic acid in the presence of *N*-ethyl-*N'*-(3-dimethylaminopropyl)carbodiimide (Foster et al., 1973) with a minor procedural modification (Dixit et al., 1975a,b). The Pth[1]-amino acids were identified by high pressure liquid chromatography (Zimmerman et al.,

† From the Connective Tissue Research Laboratory, Veterans Administration Hospital, and Departments of Biochemistry and Medicine, University of Tennessee Center for the Health Sciences, Memphis, Tennessee 38104, and the Developmental Biology Laboratory, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114. *Received July 20, 1978.* This study was supported in part by National Institutes of Health grants (AM-16506, AM-20385, and AM-3564). This is publication no. 757 of the W. Lovett Memorial Group for the Study of Diseases Causing Deformities.

[1] Pth stands for phenylthiohydantoin.

0006-2960/78/0417-5719$01.00/0

TABLE I: Amino Acid Composition of α1(I)-CB6B and Its Tryptic Peptides of Chick Skin Collagen.[a]

	T1	T2	T3	T4	T5	T6	total	α1(I)-CB6B
3-Hydroxyproline	—	—	—	0.7 (1)	—	—	1	0.6 (1)
4-Hydroxyproline	—	3.2 (3)	0.9 (1)	1.8 (2)	—	3.8 (4)	10	9.9 (10)
Aspartic acid	—	—	—	1.9 (2)	—	1.0 (1)	3	3.2 (3)
Threonine	—	—	—	—	—	0.9 (1)	1	1.2 (1)
Serine	—	2.8 (3)	0.8 (1)	—	—	0.9 (1)	5	4.1 (4)
Glutamic acid	—	3.0 (3)	—	—	—	1.0 (1)	4	4.1 (4)
Proline	—	5.2 (5)	0.8 (1)	1.9 (2)	—	7.2 (7)	15	15
Glycine	1.0 (1)	10	4.2 (4)	5.3 (5)	1.0 (1)	8.8 (9)	30	30
Alanine	—	3.2 (3)	3.1 (3)	—	—	—	6	6.2 (6)
Valine	—	—	—	—	—	1.8 (2)	2	1.8 (2)
Isoleucine	—	—	—	1.0 (1)	—	—	1	0.8 (1)
Leucine	—	1.0 (1)	—	2.0 (2)	—	1.0 (1)	4	3.7 (4)
Phenylalanine	—	0.9 (1)	—	—	—	0.9 (1)	2	1.7 (2)
Hydroxylysine	0.8 (1)	—	—	—	—	—	0.8	0.7
Lysine	—	—	1.0 (1)	—	—	—	1	1.1
Histidine	0.8 (1)	—	—	—	—	—	1	0.8 (1)
Arginine	1.0 (1)	1.0 (1)	—	1.0 (1)	1.0 (1)	—	4	4.1 (4)
Glc-Gal-Hyl								0.2
Total	4	30	11	16	2	28	91	90

[a] Values expressed as residues per peptide. Residues present over 10 rounded off to the nearest whole number. A dash indicates the level was less than 0.1 residue per peptide. Numbers in parentheses indicate assumed integral values.

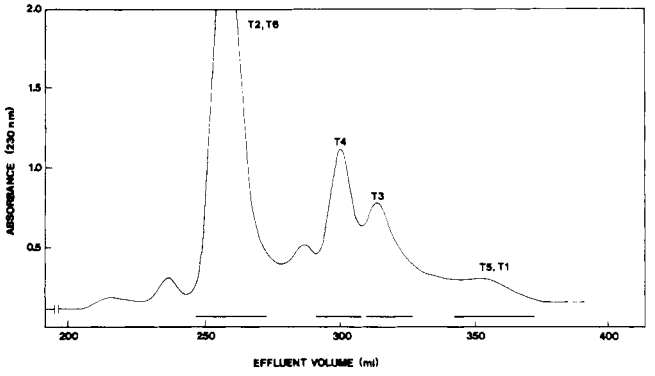

FIGURE 1: Sephadex G-50S chromatography of 30 mg of tryptic digest of α1(I)-CB6B. The eluent was 0.04 M sodium acetate, pH 4.8. Fractions indicated by the bars were collected and further separated by subsequent maneuvers.

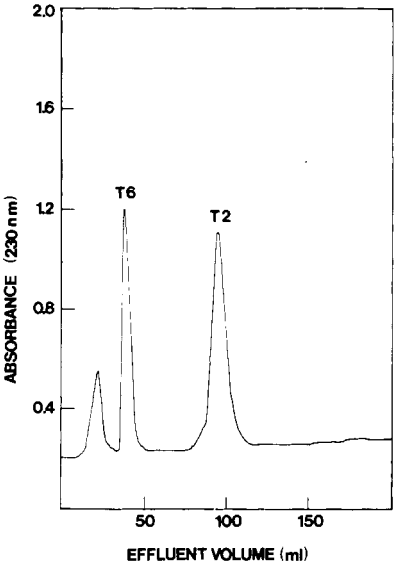

FIGURE 2: Phosphocellulose chromatography of the peptides in peak 1 (Figure 1). The sample was applied to a 1 × 6 cm column of phosphocellulose equilibrated with 0.001 M sodium acetate, pH 3.8, at 44 °C, and the peptides were eluted with a linear gradient of NaCl from 0 to 0.3 M over a total volume of 500 mL.

1973). The COOH-terminal residues were inferred from the amino acid composition of peptides, the known specificities of trypsin or by subsequent overlapping amino acid sequence analysis. The Pth-3-hydroxyproline was identified by comparing the elution position with the Pth derivative of authentic 3-hydroxyproline.[2]

Results

Isolation of Tryptic Peptides of α1(I)-CB6B. The tryptic digest of α1(I)-CB6B was chromatographed on Sephadex G-50S. The elution profile is presented in Figure 1. The first large peak contained tryptic peptides T2 and T6. The peptides T2 and T6 were separated by phosphocellulose chromatography as shown in Figure 2. The peaks labeled T4 and T3 (Figure 1) represented individual peptides and were further purified by phosphocellulose chromatography (figures not shown). The last peak from G-50S (Figure 1) was further fractionated on PA-35 (Beckman) on a Technicon analyzer. Tryptic peptide T5 elutes in the basic region. Peptide T1 was eluted with 0.5 N NaOH at the end of the gradient (figure not shown). The single hydroxylysine residue (position 1) of α1(I)-CB6B which is present in T1 is partially (20%) glycosylated. The amino acid composition of six tryptic peptides is presented in Table I. The sum of the residues of the peptides agrees within experimental error with the observed composition of α1(I)-CB6B.

Isolation of Thermolytic Peptides. The mixture of thermolysine-derived peptides was fractionated on Sephadex G-50S. The chromatographic pattern is shown in Figure 3. The peaks labeled Th2, Th4, Th3, and Th1 were pooled, desalted, and further purified by phosphocellulose chromatography. The amino acid composition of the isolated thermolytic peptides is presented in Table II. Sequence determination of Th3 was sufficient to deduce the alignment of tryptic peptides T4, T5 and T6.

Alignment of the Tryptic Peptides. The alignment of tryptic peptides T1, T2, and T3 was deduced by the amino acid se-

[2] Standard 3-hydroxyproline used in this study was a generous gift of Dr. R. L. Trelstad, Harvard Medical School.

TABLE II: Amino Acid Composition of Thermolytic Peptides Isolated from α1(I)-CB6B of Chick Skin Collagen.[a]

	Th1	Th2	Th3	Th4
3-Hydroxyproline	—	—	0.7 (1)	—
4-Hydroxyproline	—	3.7 (4)	1.8 (2)	3.8 (4)
Aspartic acid	—	1.1 (1)	1.0 (1)	1.0 (1)
Threonine	—	—	0.8 (1)	—
Serine	—	2.7 (3)	—	0.8 (1)
Glutamic acid	—	3.1 (3)	1.1 (1)	7.3 (7)
Proline	—	6.2 (6)	2.0 (2)	7.3 (7)
Glycine	2.2 (2)	13.3 (13)	6.1 (6)	8.2 (8)
Alanine	—	6.2 (6)	—	—
Valine	—	—	—	1.7 (2)
Isoleucine	—	—	1.0 (1)	0.8 (1)
Leucine	—	1.0 (1)	2.1 (2)	0.8 (1)
Phenylalanine	—	—	—	—
Hydroxylysine	0.8 (1)	—	—	—
Lysine	—	1.0 (1)	—	—
Histidine	0.8 (1)	—	—	—
Arginine	1.0 (1)	1.0 (1)	2.0 (2)	—
Total	5	39	19	25

[a] Values expressed as residues per peptide. A dash indicated the level was less than 0.1 residue per peptide. Numbers in parentheses indicated assumed integral values.

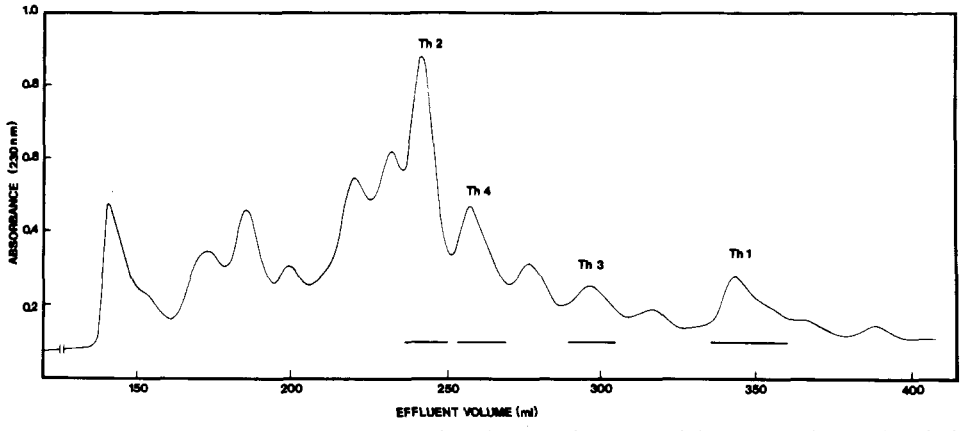

FIGURE 3: Sephadex G-50S chromatography of a thermolytic digest of α1(I)-CB6B. Conditions of chromatography were identical with those described in Figure 1. The fractions were pooled as indicated by the bars.

quence determination of intact α1(I)-CB6B through 40 residues (Figure 4). Peptide T6 must be the carboxyl-terminal peptide as it lacks a residue of lysine or arginine (Table I). That the tryptic peptide T5 precedes T6 and follows T4 was determined by the amino acid sequence of thermolytic peptide Th3 (Leu-Asn-Gly-Leu-Hyp-Gly-Pro-Ile-Gly-^{3}Hyp-Hyp-Gly-Pro-Arg-Gly-Arg-Thr-Gly-Glu), a 19-residue peptide composed of 14 residues from the COOH-terminal part of T4, the dipeptide T5 and the first three residues, Thr-Gly-Glu of T6. By exclusion, then, T4 must follow T3.

In summary, the alignment of tryptic peptides in α1(I)-CB6B is T1-T2-T3-T4-T5-T6. This is consistent with the tryptic peptide alignment in α1(I)-CB6-C2 of calf skin collagen reported earlier (Fietzek et al., 1972).

Internal Sequences of Tryptic Peptides. All peptides including α1(I)-CB6B were modified at carboxyl groups by treatment with 2-amino-1,5-naphthalenedisulfonic acid in the presence of *N*-ethyl-*N'*-(3-dimethylaminopropyl)carbodiimide. Intact α1(I)-CB6B was subjected to automated degradation through 40 residues. Tryptic peptides T2, T3, T4, and T6 and thermolytic peptide Th3 were degraded through the penultimate amino acid residue. The peptide Th4 was sequenced through 12 residues. Each peptide was analyzed at least twice. Figure 4 shows each peptide with the number of residues degraded. The repetitive yields of Pth derivatives were determined by calculating the amount of Pth-glycine. In each peptide every second glycine was used and the repetitive yields were in the range of 90–95%.

Discussion

This paper describes the amino acid sequence of α1(I)-CB6B, the carboxyl-terminal cyanogen bromide peptide of α1(I) chain of chick skin collagen containing 91 residues. The sequence presented shows the characteristics of collagen peptides derived from the helical portion of the molecule except at the short sequence of nonhelical region at the carboxyl end. Glycine is present at every third residue in the Gly-X-Y triplet. Phenylalanine at position 6 and leucines at positions 9, 48, 51 (Figure 4) present in the helical portion of the peptide are in position X of the triplet (Balian et al., 1971; Piez, 1976) as discovered earlier, although we found one phenylalanine in α2-CB3 (Dixit et al., 1977a) in Y position of the triplet. The X position in the triplet Gly-X-Y at residue 57 is occupied by 3-hydroxyproline with no proline present. This single residue of 3-hydroxyproline in α1(I) chain is unique and its role is not known at present. A single residue of 3-hydroxyproline at the identical site has been reported in calf α1(I) chain (Wendt et al., 1972). The octapeptide sequence Gly-Met-Hyl-Gly-His-

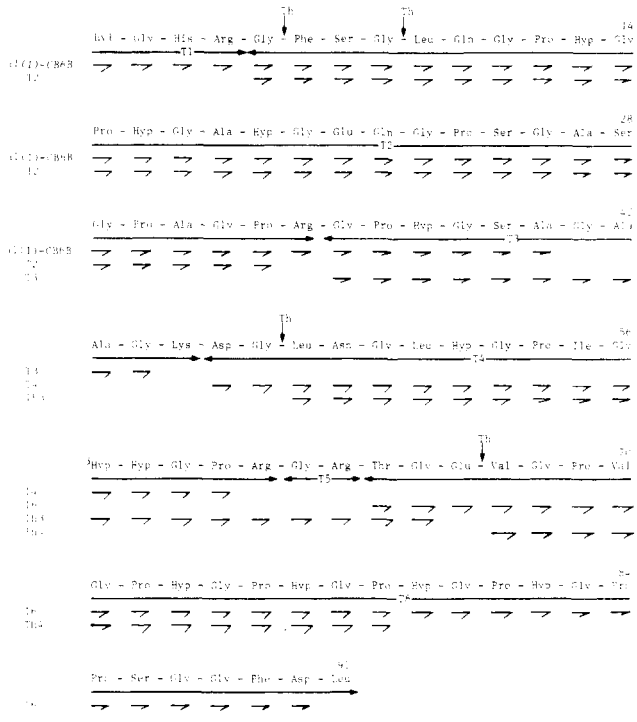

FIGURE 4: The amino sequence of α1(I)-CB6B of chick skin collagen. The inverted arrow (↓) indicates the sites of thermolytic cleavage. The tryptic peptides are indicated by long horizontal arrows (↔). The residues degraded by automated sequenator for each peptide are shown by short horizontal arrows (—).

Arg-Gly-Phe observed previously (residues 85-92) in α1(I) chain of chick (Kang et al., 1975), rat (Butler & Ponds, 1971), and in type III human collagen (Seyer & Kang, 1977) is repeated (residues 1-6, Figure 4, and the COOH-terminal residues Gly-Met of α1(I)-CB6A, Dixit et al., 1975b). This octapeptide in calf contains Ile in the place of Met. The biological significance of this unit is not clearly understood. The hydroxylysine residue (position 1, Figure 4) is 20% glycosylated (Glc-Gal).

The sequence presented here also provides an opportunity to compare the sequence with the homologous portion of α1(I)-CB6 of calf skin (Fietzek et al., 1972; Wendt et al., 1972) collagen. A total of seven substitutions were found. These are Ser → Ala (position 18, 42), Hyp → Ala (position 43), Asp → Glu (position 66), Ala → Val (positions 67 and 70), and Tyr → Phe (position 89). Thus, interspecies identity of over 92% is observed in this part of the α1(I) chain, which is in good agreement with the previously published data (Dixit et al., 1975a,b; Kang et al., 1975) on interspecies variation in α1(I) chain.

Acknowledgments

The authors' thanks are due to Ms. Gail Kirkpatrick, Mr. Tim Todd, and Ms. E. Rowland for excellent technical assistance.

References

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

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

Bornstein, P., & Piez, K. A. (1966) *Biochemistry 5,* 3460.

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

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

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

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

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

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

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

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

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

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

Fietzek, P. P., Rexrodt, F. W., Wendt, P., Stark, M., & Kuhn, K. (1972) *Eur. J. Biochem. 38,* 163.

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

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

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

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

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

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

Kang, A. H., Piez, K. A., & Gross, J. (1969a) *Biochemistry 8,* 1506.

Kang, A. H., Piez, K. A., & Gross, J. (1969b) *Biochemistry 8,* 3648.

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

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

Piez, K. A. (1976) *Biochemistry of Collagen* (Ramachandran, G. N., & Reddi, A. H., Eds.), Plenum Publishing Co., New York, N.Y.

Seyer, J. M., & Kang, A. H. (1977) *Biochemistry 16,* 1158.

Seyer, J. M., & Kang, A. H. (1978) *Biochemistry 17,* 3404.

Wendt, P., Mark, K. V. D., Rexrodt, F., & Kuhn, K. (1972) *Eur. J. Biochem. 30,* 169.