Proc. Natl. Acad. Sci. USA
Vol. 74, No. 1, pp. 84–86, January 1977
Biochemistry

Ordering of cyanogen bromide peptides of type III collagen based on their homology to type I collagen: Preservation of sites for crosslink formation during evolution

(amino-acid sequence/ α chain structure)

PETER P. FIETZEK*, HARTMUT ALLMANN*, JÜRGEN RAUTERBERG†, AND ELMAR WACHTER‡

* Max-Planck-Institut für Biochemie, Abt.f.Bindegewebsforschung, 8033 Martinsried b. München, Germany; † Institut für Arterioskleroseforschung, Westring 3, 4400 Münster, Germany; and ‡ Institut für Physiologische Chemie und Physikalische Biochemie der Universität, Goethestr. 33, 8000 München 2, Germany

Communicated by F. Lynen, October 12, 1976

ABSTRACT **The order of the cyanogen-bromide-derived peptides from α1(III) chains of pepsin-solubilized calf skin collagen was found to be 3A-3B-3C-7-6-1,8,2-4-5-9A-9B. The amino-acid sequences of the NH_2-terminal region of all peptides were determined by Edman's automated degradation procedure. The alignment of the peptides along the peptide chain was established by searching for the best homology between the partial sequences of the cyanogen bromide peptides from the α1(III) chain and the completely known sequence of the α1(I) chain. Characterization of three cyanogen-bromide-derived double peptides provided confirmation of the deduced order. A sequence Gly-Met-Hyl-Gly-His-Arg-Gly-Phe- was established near the NH_2-terminus and a sequence Gly-Ile-Hyl-Gly-His-Arg-Gly-Phe near the COOH-terminus of the α1(III) chain. Identical sequences have been found in the corresponding regions of the α1(I) chain. They include hydroxylysine, a site for intermolecular crosslink formation. Because these sequences are conserved during evolution of the collagen molecule, they are probably important for collagen structure and function.**

Determination of the primary structure of the chains of various types of collagen is a formidable undertaking since each chain contains more than 1000 amino acids (1). However, the complete sequence of amino acids in the α1 chain of type I collagen is known (2), and about two-thirds of the sequence of amino acids in the α2 chain is also established (3). In such studies the chains are digested with cyanogen bromide (CNBr), the order of the peptides along the chain is deduced, and the detailed sequence analysis of the various peptides is carried through to completion.

Work is now in progress to determine the sequence of α1(III) chains. Type III collagen was first detected in fetal skin (4), but since then it has been found in a variety of tissues from humans, calves, chickens, and rabbits (4–12). Since type III collagen is practically insoluble in neutral and acidic buffers, in all these investigations pepsin was used to solubilize this collagen. Chemical studies have established that it contains three identical α1(III) chains about the same size as α1(I) chains (5, 11, 13). Aggregates of aligned molecules (segment-long-spacing) have been produced and examined by electron microscopy (7, 14). These studies established that type I and type III collagen molecules are about the same length and have, except for three regions, a similar distribution of their charged polar amino acids. Recently, the peptides produced by CNBr cleavage of the α1(III) chains from humans (12, 15) and calves (16) have been isolated and their molecular weights and compositions established. Nine peptides were isolated from the human protein and ten from the calf. Peptide α1(III)-CB9B (CB, CNBr-derived peptide) was established as being the carboxyl-terminus of the molecule. Peptides 4, 5, and 6 were assigned to positions in the chain based on homology of their NH_2-terminal sequences with sequences occurring in the α1(I) chain (17).

Abbreviations: CNBr, cyanogen bromide; CB, cyanogen-bromide-derived peptide.

EXPERIMENTAL

Preparation of type III collagen from calf skin and the isolation of its CNBr peptides has been described in detail (16). Preparation of type III collagen, α1(III)-CB3, and α1(III)-CB8 from human skin followed the published procedure (15).

Sequence analysis was carried out automatically in a liquid-phase (Sequenator, model 890 from Beckman Instruments, Palo Alto, Calif.) and in a solid-phase sequenator (18, 19). The phenylthiohydantoin derivatives formed were identified by thin-layer and gas-liquid chromatography, and in some instances by mass spectrometry or after hydrolysis as amino acids on an amino-acid analyzer (20, 21).

RESULTS AND DISCUSSION

To establish the order of the CNBr-derived peptides, we determined the sequence of the NH_2-terminal end of each peptide. The results are summarized in Table 1. The amino-acid sequence of α1(III)-CB3A was completed and will be reported elsewhere (22). The complete sequences of peptides α1(III)-CB3B and α1(III)-CB3C were elucidated. In addition, the amino-acid sequence of NH_2-terminal regions of peptides that contain one uncleaved methionine residue and represent double peptides [α1(III)-CB(1,8,2–4), α1(III)-CB(4–5), and α1(III)-CB(9A–9B)] were determined.

The order of the peptides was deduced by comparing the determined sequences of type III with those of type I and looking for homologous regions (Table 1), and from the molecular weights, amino-acid compositions, and NH_2-terminal sequences of three double peptides (16).

Order of peptides along calf α1(III) chain

On the basis of homology to α1(I) chains, CNBr peptides 7, 6, 4, 5, and 9B were positioned in the peptide chain (Table 1 and Fig. 1). Peptide 9B is the COOH-terminal peptide, since it was the only one without homoserine. Peptide 9A preceded α1(III)-CB9B, since a peptide was isolated that was identified as a double peptide α1(III)-CB(9A-9B) on the basis of amino-acid composition and NH_2-terminal amino-acid sequence. The positioning of peptide 5 following α1(III)-CB4 was confirmed by the isolation, characterization, and sequence analysis of a double peptide α1(III)-CB(4-5) and from renaturation studies (7) on α1(III)-CB5. Peptide 3A and α1(III) chains were found to have an identical sequence at the NH_2-terminus which established that α1(III)-CB3A arose from the NH_2-terminal end of α1(III) chain. When α1(III)-CB3A and α1(III)-CB7 are aligned to the α1(I) chain by homology (22), a gap remains

Table 1. Amino-acid sequence of the NH_2-terminal regions of various CNBr-derived peptides from calf and human type III skin collagen

Peptide*	Size†	Amino-acid sequence
α1(III)-CB3A	79	Ile -Ala -Gly -Tyr -Hyp-Gly -Pro -Ala -
α1(I) (1–3)		Gly -Ile -Ser -Val -Pro -Gly -Pro -Met
α1(III)-CB3‡	94	Val -Gly-Gly -Leu -Ala -Gly -Tyr -Hyp-
α1(III)-CB3B	6	Hyl -Gly-Pro -Ala -Gly -Met
α1(I) (75–80)		Arg -Gly-Leu -Hyp-Gly -Thr -
α1(III)-CB3C	6	Hyp-Gly-Phe -Hyp-Gly -Met
α1(I) (81–86)		Ala -Gly-Leu -Hyp-Gly -Met
α1(III)-CB7	37	Hyl -Gly-His -Arg -Gly -Phe -Asp -
α1(I) (87–93)		Hyl -Gly-His -Arg -Gly -Phe -Ser -
α1(III)-CB6	95	Gly -Pro -Arg -Gly -Ala -Hyp-Gly -Glu -Arg -Gly -Arg -Hyp-
α1(I) (124–135)		Gly -Pro -Arg -Gly -Leu-Hyp-Gly -Glu -Arg -Gly -Arg -Hyp-
α1(III)-CB1,8,2	184	Gly -Pro -Ala -Gly -Ile -Hyp-Gly -Ala -Hyp-Gly -Leu -Ile -Gly-Ala-Arg-Gly-Pro-Hyp-
α1(III)-CB8‡	126	Gly-Ala-Arg-Gly-Pro-Hyp-
α1(III)-CB4	149	Gly -Phe-Hyp-Gly -Pro -Lys -Gly -Asn -Asp -
α1(I) (403–411)		Gly -Phe-Hyp-Gly -Pro -Lys -Gly -Ala -Ala -
α1(III)-CB5	241	Hyp-Gly-Glu -Arg -Gly -Gly -Hyp-Gly -Gly -Hyp-
α1(I) (552–561)		Hyp-Gly-Glu -Arg -Gly -Ala -Ala -Gly -Leu -Hyp-
α1(III)-CB9A	133	Hyp-Gly-Ala -Arg -Gly -Ser -Hyp-Gly -Pro -Gln -
α1(III)-CB9B	95	Gly -Ile -Hyl -Gly -His -Arg -Gly -Phe -Hyp-Gly -Asn-Hyp-Gly-Ala-Hyp-
α1(I) (925–939)		Gly -Ile -Hyl -Gly -His -Arg -Gly -Phe -Ser -Gly -Leu-Gln -Gly-Pro-Hyp-

* Numbers in parentheses refer to positions within the helical portion of the α1(I) chain (3).
† Size: residues per peptide (16).
‡ Peptide derived from human collagen.

between them that would accommodate 12 amino-acid residues, i.e., peptides 3B and 3C. Support for this assignment was obtained by sequence analysis. The triplet Gly-X-Y has been found throughout the helical portions of all the collagen α chains analyzed (3), and this unit is a prerequisite for forming the triplet helix (1). Peptide 3A ends in Gly-Met, an X position. Peptide 3B begins with Hyl-Gly, and α1(III)-CB3C with Hyp-Gly. Both therefore start with an amino acid from the Y position and end with an amino acid in the X position. Both peptides could connect to Hyl in the Y position, which is at the beginning of α1(III)-CB7. With the present data it cannot conclusively be decided whether the order is 3A-3B-3C-7 or 3A-3C-3B-7. However, the highest homology with the corresponding sequence in the α1(I) chain occurs with the order 3A-3B-3C-7 (Table 1).

Peptide 1,8,2, consisting of 184 amino-acid residues, has to be located between α1(III)-CB6 and α1(III)-CB4. This order was confirmed by the isolation of a double peptide α1(III)-CB(1,8,2-4). The peptide was identified as a double peptide by its amino-acid composition, molecular weight, and NH_2-terminal sequence. Peptide 1,8,2 and α1(III)-CB(1,8,2-4) have identical NH_2-terminal sequences.

It was concluded that the complete order of the CNBr peptides in the α1(III) chain from calf skin collagen is 3A-3B-3C-7-6-1,8,2-4-5-9A-9B.

Order of peptides of human α1(III) chains

The CNBr peptide patterns obtained from calf and human type III collagen show five major differences (12, 15, 16). Peptides 1 and 2, consisting of 12 and 40 amino-acid residues, respectively, are not present in calves, while the two hexapeptides, α1(III)-CB3B and α1(III)-CB3C, are not found in humans. In calves the region of peptide 9 contains one more methionine residue, giving rise to α1(III)-CB9A and α1(III)-CB9B. Peptides 3 and 8 are reported to consist of 109 and 126 residues, respectively, in humans, while α1(III)-CB3A and α1(III)-CB1,8,2 are reported to have 79 and 184 residues, respectively, in calves (15, 16, 22).

Since human α1(III)-CB3 is larger than the corresponding calf peptide, and peptides 3B and 3C are missing in humans, one might expect that methionine residues between 3A-3B and 3B-3C in calf collagen are substituted by another amino acid in human collagen. This would explain in part the difference in molecular weights. In order to find out if human α1(III)-CB3 is longer at its NH_2-terminal end than calf α1(III)-CB3A, we determined the sequence of the peptide and of the complete α1(III) chain from pepsin-solubilized human skin collagen. In contrast to the published data on human type III collagen (12, 15), the peptide was found to consist of 95 amino-acid residues, as determined by molecular sieve chromatography (J. Rauterberg, unpublished results). The sequence determined was identical in α1(III)-CB3 and α1(III) chains, and from position 4 onwards homologous to the sequence found in calf α1(III)-CB3A (see Table 1). The difference of some 58 amino-acid residues between peptide 8 from human and the corresponding peptide from calf collagen can be accounted for

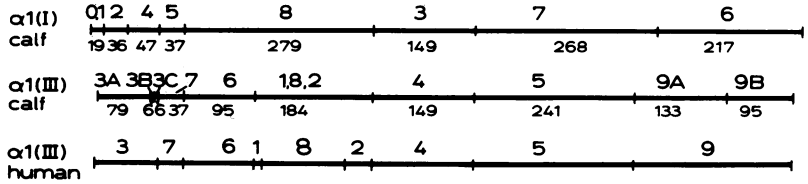

FIG. 1. The order of the CNBr-derived peptides of the α1(III) chain of calf and human collagens. For comparison, the order of the CNBr peptides of the α1(I) chain is also shown. Vertical lines indicate positions of methionine residues, numbers above the lines designate the peptides, and numbers under the lines correspond to the numbers of residues per peptide.

	85	92
α1(I)	-Gly-Met-Hyl-Gly-His-Arg-Gly-Phe-	
α1(III)	-Gly-Met-Hyl-Gly-His-Arg-Gly-Phe-	
	925	932
α1(I)	-Gly- Ile -Hyl-Gly-His-Arg-Gly-Phe-	
α1(III)	-Gly- Ile -Hyl-Gly-His-Arg-Gly-Phe-	

FIG. 2. Amino-acid sequence around the intermolecular crosslink site (hydroxylysine) in the α1(I) and α1(III) chains. The numbers correspond to positions within the helical portion of the α1(I) chain (3).

by peptides 1 and 2. In order to clarify this question, we subjected α1(III)-CB8 from human skin collagen to automated sequence analysis. The results are included in Table 1. It was evident that this sequence was homologous to the sequence of the calf peptide, beginning at position 13. The sum of residues in positions 1–12 from the calf sequence is identical to the composition of human α1(III)-CB1 if methionine is substituted by isoleucine. This accounts in part for the difference in size. The remaining difference could be explained by assuming that in human collagen α1(III)-CB8 is followed by α1(III)-CB2, while in calf collagen this particular methionine residue is substituted. Fig. 1 shows a schematic representation of the order of the CNBr-derived peptides from calf and human α1(III) chains, which is in agreement with all data presently known.

The sequence data published earlier (17) on α1(III)-CB4, α1(III)-CB5, and α1(III)-CB6, the complete sequence of α1(III)-CB3A (22), and the data presented here account for about 23% of the sequence of the chain. In general, the same structural features are evident as described for type I collagen (3).

Analysis of the sequence of the α1(I) chain has shown that there is no evidence for long-range internal homology, except for one instance. The sequence of residues in positions 85–92 is nearly identical to that of positions 925–932 (Fig. 2). It is believed that the hydroxylysine residues in positions 87 and 927 are involved in formation of intermolecular crosslinks (23–25). Therefore, it is noteworthy that the amino-acid sequences of positions 85–92 and 925–932 are identical in both α1(I) and α1(III) chains (Fig. 2). The first 10 residues of peptide 9B are identical to positions 925–932 in the α1(I) chain, which therefore allowed the precise positioning of α1(III)-CB9B. It is highly unlikely that this similarity within the α chains and between α1(I) and α1(III) chains has arisen by chance. It is reasonable to conclude that these regions have a critical function in type I and III collagens. The same crosslink site in both types of collagen might even allow formation of crosslinks between the two different molecules.

We thank Prof. K. Kühn for his helpful discussions and Mrs. U. Habel and Miss G. Schneider for their excellent technical assistance. The project was supported by the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 51, Projekt B/12 and B/20) and the Arbeitsgemeinschaft Industrieller Forschungsvereinigungen (AIF).

1. Traub, W. & Piez, K. A. (1971) "The chemistry and structure of collagen," *Adv. Protein Chem.* **25**, 243–352.
2. Fietzek, P. P. & Kühn, K. (1975) "Information contained in the amino acid sequence of the α1(I) chain of collagen and its consequences upon the formation of the triple helix of fibrils and crosslinks," *Mol. Cell. Biochem.* **8**, 141–157.
3. Fietzek, P. P. & Kühn, K. (1976) "The primary structure of collagen," *Int. Rev. Connect. Tissue Res.* **7**, 1–60.
4. Miller, E. J., Epstein, E. H., Jr. & Piez, K. A. (1971) "Identification of three genetically distinct collagens by cyanogen bromide cleavage of insoluble human skin and cartilage collagen," *Biochem. Biophys. Res. Commun.* **42**, 1024–1029.
5. Chung, E. & Miller, E. J. (1974) "Collagen polymorphism: Characterization of molecules with the chain composition $[\alpha 1(III)]_3$ in human tissues," *Science* **183**, 1200–1201.
6. Epstein, E. H., Jr. (1974) $[\alpha 1(III)]_3$ human skin collagen," *J. Biol. Chem.* **249**, 3225–3231.
7. Rauterberg, J. & Bassewitz, D. B. von (1975) "Electron microscopic investigations of type III collagen, segment-long-spacing crystallites of type III collagen from calf aorta and fetal calf skin," *Hoppe-Seyler's Z. Physiol. Chem.* **356**, 95–100.
8. Trelstad, R. L. (1974) "Human aorta collagens: Evidence for three distinct species," *Biochem. Biophys. Res. Commun.* **57**, 717–725.
9. Herrmann, H. & Mark, K. von der (1975) "Isolation and characterization of type III collagen from chick skin," *Hoppe-Seyler's Z. Physiol. Chem.* **365**, 1605–1612.
10. Gay, S., Fietzek, P. P., Remberger, K., Eder, M. & Kühn, K. (1975) "Liver cirrhosis: Immunofluorescence and biochemical studies demonstrate two types of collagen," *Klin. Wochenschr.* **53**, 205–208.
11. Fujii, T. & Kühn, K. (1975) "Isolation and characterization of pepsin-treated type III collagen from calf skin," *Hoppe-Seyler's Z. Physiol. Chem.* **356**, 1793–1801.
12. Epstein, E. H., Jr. & Munderloh, N. H. (1975) "Isolation and characterization of CNBr peptides of human $[\alpha 1(III)]_3$ collagen and tissue distribution of $[\alpha 1(I)]_2\alpha 2$ and $[\alpha 1(III)]_3$ collagens," *J. Biol. Chem.* **250**, 9304–9312.
13. Timpl, R., Glanville, R. W., Nowack, H., Wiedemann, H., Fietzek, P. P. & Kühn, K. (1975) "Isolation, chemical and electronmicroscopical characterization of neutral salt soluble type III collagen and procollagen from fetal bovine skin," *Hoppe-Seyler's Z. Physiol. Chem.* **356**, 1783–1792.
14. Wiedemann, H., Chung, E., Fujii, T., Miller, E. J. & Kühn, K. (1975) "Comparative electron-microscope studies on type III and type I collagens," *Eur. J. Biochem.* **51**, 363–368.
15. Chung, E., Keele, E. M. & Miller, E. J. (1974) "Isolation and characterization of the cyanogen bromide peptides from the α1(III) chain of human collagen," *Biochemistry* **13**, 3459–3464.
16. Rauterberg, J., Allmann, H., Henkel, W. & Fietzek, P. P. (1976) "Isolation and characterization of CNBr derived peptides of the α1(III) chain of pepsin solubilized calf skin collagen," *Hoppe-Seyler's Z. Physiol. Chem.*, in press.
17. Fietzek, P. P. & Rauterberg, J. (1975) "Cyanogen bromide peptides of type III collagen: First sequence analysis demonstrates homology with type I collagen," *FEBS Lett.* **49**, 365–368.
18. Fietzek, P. P. & Kühn, K. (1972) "Automation of the sequence analysis by Edman degradation of proteins and peptides," *Top. Curr. Chem.* **29**, 1–28.
19. Laursen, R. A. (1971) "Solid-phase Edman degradation, an automatic peptide sequencer," *Eur. J. Biochem.* **20**, 89–102.
20. Fietzek, P. P. & Rexrodt, F. W. (1975) "The covalent structure of collagen. The amino acid sequence of α2-CB4 from calf skin collagen," *Eur. J. Biochem.* **59**, 113–118.
21. Wachter, E., Machleidt, W., Hofner, H. & Otto, J. (1973) "Aminopropyl glass and its *p*-phenylene diisothiocyanate derivative, a new support in solid-phase Edman degradation of peptides and proteins," *FEBS Lett.* **35**, 97–102.
22. Fietzek, P. P., Allmann, H. & Rauterberg, J. (1976) "The amino acid sequence of the N-terminal region of pepsin solubilized type III collagen from calf skin," *Eur. J. Biochem.*, in press.
23. Kang, A. H. (1972) "Studies on the location of intermolecular cross-links in collagen. Isolation of a CNBr peptide containing δ-hydroxylysinonorleucine," *Biochemistry* **11**, 1828–1835.
24. Becker, U., Furthmayr, H. & Timpl, R. (1975) "Tryptic peptides from the cross-linking regions of insoluble calf skin collagen," *Hoppe-Seyler's Z. Physiol. Chem.* **356**, 21–32.
25. Dixit, S. N. & Bensusan, H. B. (1973) "The isolation of crosslinked peptides of collagen involving α1-CB6," *Biochem. Biophys. Res. Commun.* **52**, 1–8.