AMINO ACID SEQUENCE OF THE N-TERMINAL NON-TRIPLE HELICAL CROSS LINK REGION OF TYPE III COLLAGEN

R. W. GLANVILLE and P. P. FIETZEK
Max-Planck-Institut für Biochemie, Abt. Bindegewebsforschung, D-8033 Martinsried, FRG

Received 27 September 1976

1. Introduction

The structural integrity of the collagen fibril and possibly the catabolism of the collagen molecule is determined to a large extent by the type and density of collagen intermolecular cross links. The major intermolecular cross links in type I collagen are those derived from lysine residues [1]. There are two structural requirements necessary for the formation of this type of cross link. First, the presence of a lysine or hydroxylysine residue in the N and/or C non-triple helical regions of collagen which have been enzymatically converted to an aldehyde form. Second, a lysine or hydroxylysine residue present in the triple helical region of the molecule in a position such that it aligns with the above described aldehyde when the molecules are in a quarter staggered arrangement as found in collagen fibrils [2].

The discovery of type III collagen presented, for the first time, a situation in which two types of collagen (types I and III), exhibiting different macromolecular organisations, are found in close association in the same tissue [3]. It has been established that, unlike type I, type III collagen contains cysteine residues which form interchain disulphide cross links [4]. Type III collagen in adult tissues is insoluble, but can be extracted from pepsin solubilised material. As the cysteine cross links are still present in pepsin solubilised type III collagen, this indicates that, like type I, the N- and C-terminal non-triple helical extensions of type III collagen also contain cross link sites that are removed by pepsin treatment.

It is known that the triple helical sequence of type I collagen has two regions, each of which contains a hydroxylysine residue, that are involved in the intermolecular cross link formation [2,5,6]. Recently, identical sequences in corresponding regions of the triple helix of type III collagen have been found [7] and therefore one requirement for the formation of lysine derived cross links in type III collagen is established. The second requirement, the presence of a lysine residue in the N and C terminal non-triple helical regions of type III collagen, has been indicated from amino acid analysis data [8].

In this publication the sequence of the N-terminal non-triple helical region of type III fetal calf skin collagen is described. The N-terminal cyanogen bromide peptide CB3A was isolated and cleaved with a protease from *Staphylococcus aureus*, chymotrypsin and trypsin. The peptides containing non-triple helical sequences were isolated using molecular sieve chromatography and sequenced. The N-terminal of CB3A was blocked, probably by a pyroglutamic acid residue. The non-triple helical sequence was found to be 14 residues long and has similar characteristics to that of type I collagen. It has a high content of hydrophobic amino acid residues, contains two tyrosine residues which are important components of type III collagen antigenic determinants [9] and has a lysine residue, a possible site for intermolecular cross link formation. Evidence indicating that the lysine residue is a site for intermolecular cross linking formation is discussed.

Address for correspondence: P. P. Fietzek, Max-Planck-Institut für Biochemie, Am Klopferspitz 2, D-8033 Martinsried, FRG

2. Methods

2.1. *Peptide CB3A*

Peptide CB3A was prepared from cyanogen bromide digests of type III collagen from fetal calf skin as recently described [8].

2.2. *Staphylococcal protease digestion of CB3A*

Staphylococcal protease digestion of CB3A was carried out in 0.2 M NH_4HCO_3–HCl buffer pH 7.8 [10]. The enzyme (protease from *Staphylococcus aureus*, strain V8, Miles Labs., England) was added to the peptide dissolved in buffer (3 mg/ml) to give an enzyme to substrate ratio of 1:30. Incubation at 37°C was continued for 16 h and then the mixture applied directly to a column of Sephadex G-50 superfine (2 × 150 cm) equilibrated at room temperature with 0.2 M NH_4HCO_3 solution at a flow rate of 18 ml/h.

2.3. *Trypsin and chymotrypsin digestion of CB3A*

Trypsin and chymotrypsin digestion of CB3A and separation of the products was carried out as described above for the staphylococcal protease but using enzyme to substrate ratios of 1 : 50 and 1 : 200 and incubation times of 4 h and 1 h respectively (Trypsin: Worthington Biochem. Corp. TPCK-trypsin 212 U/mg. Chymotrypsin: Worthington Biochem. Corp., α-chymotrypsin 50 U/mg).

Small peptides from the Sephadex G-50 superfine separations were rechromatographed on a column (2.8 × 105 cm) of Bio-Gel P2 (Bio Rad 200–400 mesh) equilibrated at room temperature with 1% acetic acid at a flow rate of 20 ml/h.

2.4. *Sequence analysis*

Sequence analysis was carried out automatically in a liquid phase sequencer (model 890 from Beckmann Instruments, Palo Alto, Calif. USA). Experimental details were recently described elsewhere [11]. PTH derivatives were identified by thin layer chromatography and gas-liquid chromatography.

2.5. *Amino acid analysis*

Amino acid analysis of peptide hydrolysates (6 N HCl, 110 for 24 h under N_2) were carried out on a Durrum D 500 analyser (Palo Alto, Calif. USA).

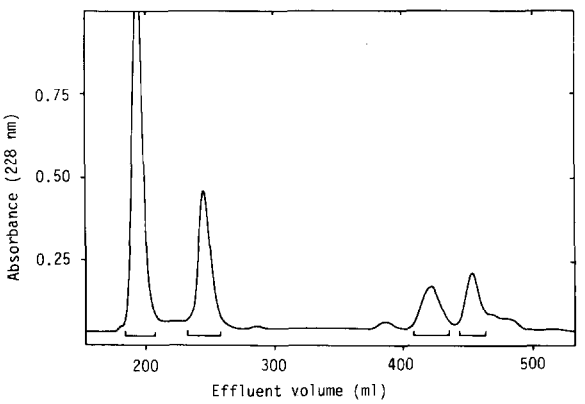

Fig.1. Molecular sieve chromatography of the staphylococcal protease digest of CB3A (9 mg) on a Sephadex G-50 superfine. The column (2 × 150 cm) was equilibrated, at room temperature, with a 0.2 M NH_4HCO_3 solution at a flow rate of 18 ml/h.

3. Results and discussion

Initial attempts to sequence intact CB3A failed because the N-terminal was blocked, probably by a pyroglutamic acid residue as is the case for type I collagen α-chains. This made it necessary to degrade the peptide in order to produce fragments that could be sequenced.

Staphylococcal protease, under the conditions used, cleaves specifically on the C-terminal side of glutamic acid residues. Digestion of CB3A with this enzyme and separation of the products, produced the peptide pattern shown in fig.1. Fraction a was shown, by amino acid composition and N-terminal sequence analysis, to contain the same material as fraction b*.

Fraction b was lyophylised, sequenced and was found to contain a peptide (fig.2, SP 2) with a N-terminal non-triple helical extension of 12 residues. From the 13th residue onwards it was identical to the known sequence of peptide CB3A, isolated from

* Peptides T1, C1 and intact CB3A isolated from salt soluble and pepsin solubilised type III collagen, all chromatographed as two peaks. However, when chromatography was carried out using sodium acetate buffer instead of ammonium bicarbonate these peptides chromatographed as single peaks. It appeared that the peptides aggregated in ammonium bicarbonate buffer and that the aggregation was not caused by the non-triple helical sequences.

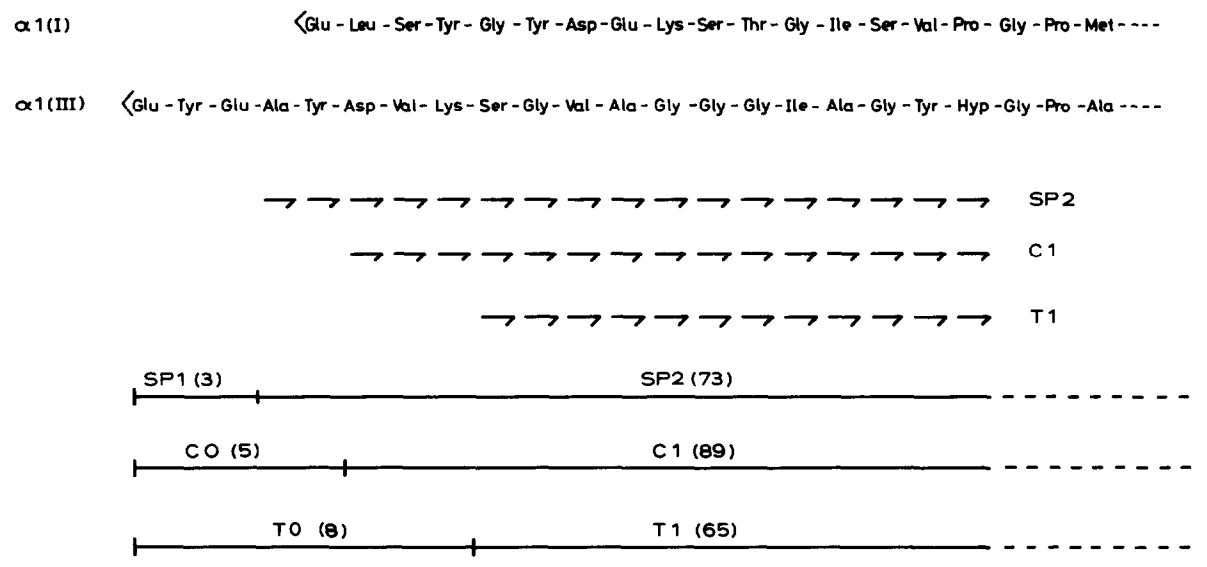

Fig.2. Amino acid sequence of the N terminal region of α1 (III) from fetal calf skin collagen and for comparison the N terminal region of α1 (I) from calf skin collagen [2]. The sequences are aligned so that there is maximum homology between the triple helical regions of the chains (not shown). The rows of arrows (——→) indicate the residues sequenced in the peptide indicated to their right. The lower schematic representation of the α1 (III) sequence shown, indicates the cleavage points of the enzymes used. the designations of the peptides so produced and in brackets the number of amino acid residues per peptide. The enzymes used were: staphylococcal protease (SP), trypsin (T), chymotrypsin (C).

pepsin solubilised type III collagen [12]. Fraction c (fig.1) contained a peptide from the C-terminal triple helical region of CB3A whose sequence is known [12]. Fraction d (fig.1) from its amino acid analysis and size, appeared to be a tripeptide (Tyr, GlX_2) (fig.2 SP 1). From the specificity of the staphylococcal protease used, Glu was expected to be at the C-terminal of the tripeptide and because CB3A was blocked, pyroglutamic acid at its N-terminal. The deduced sequence was therefore <Glu–Tyr–Glu. This is in agreement with other recently published data [14].

The complete sequence as shown in fig.2 was confirmed by sequencing the largest peptides produced by the chymotrypsin and trypsin digestion of CB3A (fig.2, peptides C1 and T1 respectively). The N-terminal peptides T O and C O were also isolated and characterised by amino acid analysis in order to confirm the composition and size of SP 1.

Elucidation of the non-triple helical sequence revealed that the first two residues of pepsin solubilised type III collagen, Ile–Ala, are in fact part of the first triplet of the triple helix. The N-terminal sequences of the α1 (I) and α1 (III) chains shown in fig.2 are aligned so that there is maximum homology between the N-terminal triple helical regions of the α chains. This alignment was established from electron microscope studies of the cross striation banding patterns of segment long spacing crystallites and sequence data of each chain [12]. Using this alignment it can be seen that the α1 (III) chain is two triplets longer at its N-terminal than the α1 (I) chain. There are 14 N-terminal non-triple helical residues as compared to 16 found in type I collagen from calf skin (fig.2). The cleavage site for pepsin is a Gly–Ile bond. The non-triple helical extension of type III collagen exhibits the same characteristics as that of type I collagen. Both contain two tyrosine residues which are responsible for the high antigenicity of these regions [9,13], and both have a high content of hydrophobic amino acid residues.

As in all other N-terminal non-triple helical extensions that have been characterised, the type III extension contains a lysine residue. It is the 8th residue from the N-terminal and is not hydroxylated.

Although evidence for the presence of a lysine derived aldehyde (allysine) was not found in these studies, trypsin digests of insoluble calf skin have been found to contain a cross linked peptide from the N-terminal region of type III collagen, in which lysine is present presumably in the form of an aldol condensation product [14]. The requirements for lysine derived intermolecular cross link formation are therefore present in type III collagen. The N-terminal intermolecular cross link formed would be between allysine in the non-triple helical extention and hydroxylysine in the triple helical region giving rise to the acid labile dehydro-hydroxylysino–norleucine cross link. This assumes that the extracted type III collagen is representative of the whole type III collagen content of the tissue. However, as salt soluble type III collagen represents less than 5% of the total type III collagen present in fetal calf skin, it is possible that the lysine in the N-terminal non-triple helical region of insoluble type III collagen is hydroxylated, giving rise to the more stable hydroxylysino-5-ketonorleucine intermolecular cross link. This would explain the apparently contradictory data recently published [15,16] concerning the identity of the intermolecular cross links in fetal calf skin collagen when identified using sodium borohydride reduction methods.

Acknowledgements

The authors wish to thank Professor K. Kühn for his support and helpful discussions and Mr C. de Lorenzo and Miss G. Landrath for their excellent technical assistance. This work was supported by grants from the Deutscher Forschungsgemeinschaft (Sonderforschungsbereich 51, B/12) and the Arbeitsgemeinschaft Industrieller Forschungsvereinigungen e. V. (AIF). One of us, R. W. G. was financed by a Wellcome European Travelling Research Fellowship.

References

[1] Bailey, A. J., Robins, S. P. and Balian, G. (1974) Nature 251, 105–109.
[2] Fietzek, P. P. and Kühn, K. (1976) Int. Rev. Conn. Tiss. Res. 7, 1–59.
[3] Gay, S., Müller, P. K., Meigel, W. N. and Kühn, K. (1976) Der Hausarzt 27, 196–205.
[4] Chung, E. and Miller, E. J. (1974) Science 183, 1200–1201.
[5] Zimmermann, B. K., Timpl, R. and Kühn, K (1973) Eur. J. Biochem. 35, 216–221.
[6] Kang, A. H. (1972) Biochemistry 11, 1828–1835.
[7] Fietzek, P. P., Allmann, H., Rauterberg, J. and Wachter, E. (1976) Proc. Natl. Acad. Sci. USA (in press).
[8] Glanville, R. W. and Fietzek, P. P. (1976) Analytical Biochem. (submitted).
[9] Becker, U., Nowack, H., Gay, S. and Timpl. R. (1976) Immunology 31, 57–64.
[10] Houmard, J. and Drapeau, G. R. (1972) Proc. Natl. Acad. Sci. USA 69, 3506–3509.
[11] Fietzek, P. P. and Rexroth, F. W. (1975) Eur. J. Biochem. 59, 113–118.
[12] Fietzek, P. P., Allmann, H., Rauterberg, J. and Wachter, E. (1976) Eur. J. Biochem. (in press).
[13] Rauterberg, J., Timpl, R. and Furthmayr, H. (1972) Eur. J. Biochem. 27, 231–237.
[14] Becker, U., Fietzek, P. P., Nowack, H. and Timpl, R. (1976) Hoppe-Seyler's Z. Physiol. Chem. (in press).
[15] Bailey, A. J. and Sims, T. J. (1976) Biochem. J. 153, 211–215.
[16] Fujii, K., Tanzer, M. L., Nusgens, B. V. and Lapiere, C. M. (1976) Biochem. Biophys. Res. Commun. 69, 128–134.