Cross-link analysis of the C-telopeptide domain from type III collagen

W. HENKEL

Institut für Arterioskleroserecherche an der Universität Münster, Domagkstrasse 3, D-48149 Münster, Federal Republic of Germany

Several peptides were isolated from tryptic digests of insoluble calf aorta matrix by chromatography. Reductive pyridylethylation of a tryptic 15 kDa pool released fragments deriving from the C-terminus of type III collagen. A 50-residue peptide $T_C(III)$ was shown by sequence analysis to be the C-terminal peptide from the $\alpha 1(III)$ -chain, containing a helical and non-helical region of equal sizes. The peptide was further digested with collagenase to give $Col_C(III)$, comprising the complete C-terminal non-helical region of $\alpha 1(III)$ including a hydroxylysine in position 16^c. The peptide $T_C(III) \times T_N(III)$ was isolated, demonstrating covalent cross-linking between the C-terminal non-helical region of one type III molecule and the N-terminal helical cross-linking region of another. Its digestion with cy-

anogen bromide yielded the small fragments $\alpha 1(III)CB3B^*$ and $\alpha 1(III)CB3C$, confirming $T_N(III)$ as an N-terminal helical cross-link site. Sequence analysis of both $T_C(III) \times T_N(III)$ and its collagenase-derived cross-linked peptide $Col_C(III) \times T_N(III)$ established the 4D-staggered alignment of adjacent collagen III molecules. The cross-link structure of both peptides was mainly dihydroxylysinonorleucine with a small amount of hydroxylysinonorleucine, indicating that the lysine residues involved in formation of the cross-links are both hydroxylated. No pyridinoline or histidinohydroxylysinonorleucine cross-links were found within the non-reduced C-telopeptide region of type III collagen.

INTRODUCTION

The complete primary structure of human [2,3] and bovine [4–9] type III collagen is known on the basis of the sequence analysis of either cDNA species from type III procollagen or $\alpha 1(III)$ -derived peptides. The amino acid sequence of bovine $\alpha 1(III)$, forming the helical domain of the type III molecule, contains the N- and C-terminal helical cross-linking sites -Gly-Met-Hyl-Gly-His-Arg-Gly-Phe (residues 94–101) and Gly-Ile-Hyl-Gly-His-Arg-Gly-Phe- (residues 937–944), which are identical with the corresponding sites of the $\alpha 1(I)$ -chain (residues 85–92 and 925–932) [10]. These regions, located in $\alpha 1(III)CB7$ and $\alpha 1(III)CB9B$, have been shown to be involved in cross-linking [11]. The telopeptides of bovine type III collagen, deriving from the N- and C-termini, have also been sequenced [12–14] and shown to contain one Lys/Hyl residue. The involvement of the N-terminal non-helical region in cross-link formation was demonstrated by sequence analysis of the trifunctional cross-linked peptide $[Col_N(III)]_2 \times T_N(III)$ isolated from human leiomyoma [1,15].

So far the occurrence of lysine-derived cross-links from the C-terminal non-helical domain of the $\alpha 1(III)$ -chain has not been described. Here I present a cross-link analysis of this region for bovine aorta type III collagen.

METHODS

Materials

Bio-Gel P-100 (100–200 mesh), Bio-Gel P-2 (200–400 mesh), Bio-Gel A-1.5m (200–400 mesh) and AG 50W-X4 analytical grade cation-exchange resin (200–400 mesh) were from Bio-Rad (Richmond, CA, U.S.A.); Sephadex G-50(s) (200–400 mesh) was from Pharmacia (Uppsala, Sweden), and phosphocellulose P11

fibrous cation-exchanger was from Whatman (Maidstone, Kent, U.K.). Tosylphenylalanylchloromethane ('TPCK')-treated trypsin (EC 3.4.21.4) and 4-vinylpyridine were from Sigma (St. Louis, MO, U.S.A.). Bacterial collagenase (EC 3.4.24.3) was from Worthington (Freehold, NJ, U.S.A.) and sodium boro[³H]hydride from New England Nuclear Corp. (Boston, MA, U.S.A.). HPLC-grade reagents such as trifluoroacetic acid (TFA), heptafluorobutyric acid and isopropanol were from Fluka (Buchs, Switzerland), and acetonitrile and water were from Baker (Deventer, Holland). Reagents for amino acid analysis were from Pierce Co. (Rockford, IL, U.S.A.). All sequencer chemicals were from Applied Biosystems (Foster City, CA, U.S.A.). Calf aortas were delivered fresh from animals at slaughter.

Digestion of matrix and peptides

An insoluble matrix of calf aorta was prepared as outlined previously [1]. Approx. 77 g of matrix was obtained from 500 g of wet tissue. Digestion of the insoluble matrix with trypsin and peptides with bacterial collagenase was performed as described previously [1]. Cyanogen bromide cleavage of $T_C(III) \times T_N(III)$ followed the procedure of Bornstein et al. [16].

Chromatographic separations

The 15 and 20 kDa fractions were fractionated by gel filtration on Bio-Gel P-100. Separation of the 20 kDa fraction from high-molecular-mass proteins was achieved by rechromatography on Bio-Gel Agarose A-1.5m. The pyridylethylated peptides and all collageneous digests were chromatographed on Sephadex G-50(s). The peptides $T_C(III)$ and $T_C(III) \times T_N(III)$ were finally purified by ion-exchange chromatography on phosphocellulose

Abbreviations used: DHLNL, dihydroxylysinonorleucine; HHL, histidinohydroxylysinonorleucine; HLNL, hydroxylysinonorleucine; TFA, trifluoroacetic acid. Nomenclature: $T_C(III) \times T_N(III)$ is a cross-linked peptide containing two polypeptide chains derived from $\alpha 1(III)$ of type III collagen. $T_C(III)$ is the trypsin-derived peptide containing the C-terminal non-helical region of type III collagen. $T_N(III)$ is the trypsin-derived peptide containing the helical N-terminal cross-linking region of type III collagen. When $T_C(III) \times T_N(III)$ is treated with collagenase it becomes $Col_C(III) \times T_N(III)$, as the $T_N(III)$ component remains unaltered; treated with cyanogen bromide it becomes $T_C(III) \times GlyHisArg$, as the $T_N(III)$ component contains two methionine residues. For the nomenclature of the cross-linked peptide $[Col_N(III)]_2 \times T_C(III)$ see [1].

P 11. The small peptides $\alpha 1(\text{III})\text{CB3B}^*$ and $\alpha 1(\text{III})\text{CB3C}$ were separated on Bio-Gel P-2. For conditions see the legends to the relevant figures. Reversed-phase HPLC of peptides was done with a Beckman HPLC system [17]. The peptides were applied to a high-pore reversed-phase HPLC column (Bio-Rad RP-318, 5 μm particle size, 300 Å pore size, 25 cm $\times$ 4.6 mm), equilibrated in solution A (0.1 % TFA). Elution was performed at room temperature with a linear gradient [0–60 % B over 60 min] of solution B [0.1 % TFA in 70 % (v/v) isopropanol] at a flow rate of 1 ml/min.

Reduction and S-Pyridylethylation of proteins

Peptides containing disulphide bonds were reduced with 2-mercaptoethanol and the resulting thiol groups were alkylated by treatment with 4-vinylpyridine [18]. The standard amino acid *S*- β -(4-pyridylethyl)-L-cysteine was prepared by the method of Cavins and Friedman [19].

Reduction with NaBH_4

Sodium boro[^3H]hydride (5 mCi; specific radioactivity 100 mCi/mM), corresponding to 1.9 mg of NaBH_4 , was dissolved in 1.9 ml of 0.05 M sodium phosphate buffer, pH 8.0. A 100 μl portion (0.1 mg of NaBH_4) of the solution was added to 1 mg of $\text{T}_c(\text{III}) \times \text{T}_n(\text{III})$. Reduction (1 h) was stopped by addition of one drop of 50 % acetic acid. The reduced peptide was desalted on AG 50W-X4 cation-exchange resin [17] and freeze-dried.

Amino acid analysis

Amino acid compositions of peptides were determined by automated amino acid analysis (Biotronik, model LC 6000 E) as outlined earlier [17]. Methionine was determined as homoserine. The dried hydrolysates of the peptides were treated with 0.1 M piperidine at 37 °C for 1 h, evaporated to dryness and immediately afterwards analysed with the standard programme.

Cysteine was determined as *S*- β -(4-pyridylethyl)-L-cysteine eluting behind arginine in the standard system. Identification of dihydroxylysinonorleucine (DHLNL) and hydroxylysinonorleucine (HLNL) was achieved by single-buffer elution with a 0.35 M sodium citrate buffer, pH 5.28, at 60 °C in combination with a split-stream device: one part of the effluent was submitted to ninhydrin detection, and the remainder was collected and monitored for ^3H in a scintillation counter (Packard) after mixing the single fractions (0.6 ml) with 9 ml of scintillation fluid (Instagel).

Amino acid sequencing

Amino acid sequences were determined with a Beckman liquid-phase spinning-cup sequencer 890 C fitted with P6 Auto-converters (Sequamat, Watertown, MA, U.S.A.). A protein Quadrol program (0.5 M Quadrol coupling buffer) was used (model 890 C program 122974) and the anilinothiazolinone amino acids were converted into phenylthiohydantoin using methanolic HCl. The phenylthiohydantoin analysis was performed with reversed-phase HPLC [20].

RESULTS

Approx. 7 % of the dry weight of calf aorta is type III collagen [21,22]. Trypsin digestion of the denatured insoluble matrix, preserving telopeptides, released cross-linked peptides that were

Figure 1 Molecular sieve chromatography of collagen fragments

(a) Separation on Bio-Gel P-100 (100–200 mesh) of the tryptic digest of 4 g of insoluble matrix protein from calf aorta. The column (3.4 cm $\times$ 128 cm) was eluted with 0.05 M sodium acetate, pH 4.5, at 40 °C at a flow rate of 36 ml/h. (b) Chromatography on Sephadex G-50(s) (200–400 mesh) of the pyridylethylated 15 kDa fraction from Bio-Gel P-100. The 50 mg peptide mixture was applied to a column (2 cm $\times$ 150 cm), which was eluted with 0.05 M sodium acetate, pH 4.5, at 40 °C at a flow rate of 20 ml/h. (c) Isolation of the non-reduced 12 kDa fraction. Separation on Sephadex G-50(s) of fragments obtained by collagenase digestion of the 20 kDa fraction from (a). Conditions as for (b).

first fractionated on a Bio-Gel P-100 molecular sieve column (Figure 1a). The tryptic peptide $[\text{T}_c(\text{III})]_3$ was expected to contain approx. 150 amino acids and so the 15 kDa fraction from the P-100 column was reduced, treated with vinylpyridine and separated on Sephadex G-50(s) (Figure 1b). Pools 1–3 were purified by ion-exchange chromatography on phosphocellulose.

From its composition data, pool F3 was shown to contain peptide $\text{T}_c(\text{III})_8$, comprising a helical and an incomplete non-helical region of $\alpha 1(\text{III})$. The C-terminus of the peptide (Gly¹⁰⁰⁶–Ala¹⁰³⁷) coincides with the pepsin cleavage point Ala¹⁰³⁷ (Figure 2).

Pool F2 after chromatography on phosphocellulose proved to be heterogeneous. The different fractions represent the peptide $\text{T}_c(\text{III})_L$; sequence analysis yielded the 11-residue partial sequence Gly¹⁰⁰⁶–Ser–Glu–Gly–Ser–Hyp–Gly–His–Hyp–Gly–Glu¹⁰¹⁶ of $\alpha 1(\text{III})$ from calf aorta. $\text{T}_c(\text{III})_L$ (Gly¹⁰⁰⁶–Tyr¹⁰⁵²) contains the complete non-helical C-terminal region of $\alpha 1(\text{III})$ beyond the pepsin cleavage point of $\text{T}_c(\text{III})_8$ (Figure 2).

Pool F1 purified by ion-exchange chromatography was also resolved into several peaks, all containing the same cross-linked peptide $\text{T}_c(\text{III})_L \times \text{T}_n(\text{III})$. Three sequences were found: Gly¹⁰⁰⁶–Ser–Glu–Gly–Ser–Hyp–Gly–His–Hyp–Gly–Glu–Hyp–Gly–Pro¹⁰¹⁹, Gly⁸⁵–Pro–Ala–Gly–Met–Hyp–Gly–Phe–Hyp–Gly–Met⁹⁵ and Gly⁷⁶–Phe–Hyp–Gly–Pro–Hyp–Gly–Met⁸³. The first sequence corresponds to the 14 N-terminal residues of $\text{T}_c(\text{III})$ and the second to the 11 N-terminal residues of $\text{T}_n(\text{III})_8$ (Figure 2). The

Figure 2 Primary structure of $T_c(III) \times T_n(III)$

The fragments obtained by digestion with either collagenase or cyanogen bromide are represented. Col indicates the cleavage point of bacterial collagenase, T that of trypsin and CNBr that of cyanogen bromide. The non-helical and helical regions of the peptide are represented respectively by a straight line and sinusoidal curves. The component peptides $T_c(III)$ and $T_n(III)$, occurring as $T_c(III)_L$, $T_c(III)_S$, $T_n(III)_L$ and $T_n(III)_S$, are cross-linked by hydroxylysino-5-oxonorleucine. The half-arrows ($\rightarrow$) indicate the residues sequenced by automated Edman degradation. Partial sequences of $T_c(III)$ and $T_n(III)$, which were not sequenced, were taken from [9] and [4] respectively. Residue X within X-Gly-His-Arg stands for hydroxylysino-5-oxonorleucine.

formation of the 15-residue peptide (Gly⁸⁵-Arg⁹⁹) is correlated with the occurrence of Lys⁸⁴. The third sequence also is $T_n(III)$, starting differently from $T_n(III)_S$ at Gly⁷⁶. The 24-residue peptide (Gly⁷⁶-Arg⁹⁹) containing Hyl⁸⁴ is denoted by $T_n(III)_L$ (Figure 2). The heterogeneity of $T_c(III) \times T_n(III)$ is caused by both peptide chains being present in different forms.

The identity of $T_n(III)_S$ and the presence of a cross-link were confirmed by treating $T_c(III)_L \times T_n(III)_S$ with cyanogen bromide. After cleavage of $T_n(III)_S$, containing two methionine residues, the expected products $\alpha 1(III)CB3C$, $\alpha 1(III)CB3B^*$ and $T_c(III)_L \times XGlyHisArg$ were found (Figure 2). The five-residue peptide $\alpha 1(III)CB3B^*$, i.e. $\alpha 1(III)CB3B$ without Hyl/Lys, is formed by the occurrence of Lys⁸⁴.

Furthermore, conversion of $T_c(III)$ and $T_c(III) \times T_n(III)$ to the corresponding collagenase-treated fragments $Col_c(III)$ and $Col_c(III) \times T_n(III)$ made possible the identification of the complete C-telopeptide from bovine $\alpha 1(III)$. The 26-residue sequence of $Col_c(III)$ consists of a non-helical domain apart from a triplet at the N-terminus (Figure 2). The 11 N-terminal residues agree with residues 1027–1037 of $\alpha 1(III)$ from calf skin [23]. The subsequent segment of 15 residues corresponds to residues 9^C–24^C of $\alpha 1(III)$ from aortic smooth muscle cells, as deduced by nucleotide sequence analysis of a bovine $\alpha 1(III)$ cDNA clone coding for the C-propeptide [14]. As shown by amino acid sequence analysis, Lys^{16C} is hydroxylated nearly completely. Sequence analysis of $Col_c(III) \times T_n(III)_S$ very clearly documented the identity of the cross-linked regions and gave direct evidence for the involvement of Hyl^{16C} in cross-linking (Figure 3): Gly¹⁰²⁷-Pro-Cys-Cys-Gly-Ala-Gly-Gly-Val-Ala-Ala-Ile[step 12]-Ala-Gly⁸⁵-Pro-Ala-Gly-Met-Hyp-Gly-Phe-Hyp-Gly-Met-Xaa⁹⁶-Gly-. No Lys/Hyl residue was present in position 96 (step 12), which is occupied by Lys/Hyl-derived cross-links.

Cross-link analysis of $Col_c(III) \times T_n(III)$ established the presence of dihydroxylysinonorleucine, which eluted as a ninhydrin-positive partly resolved double peak X, strongly labelled with tritium. Less than 10% hydroxylysinonorleucine was present (Figure 4).

The non-reduced cross-linking region of type III collagen was obtained by digestion of the 20 kDa fraction from Figure 1(a) with collagenase and separation of the fragments by gel filtration on Sephadex G-50(s) (Figure 1c). Sequence analysis of the 12 kDa pool, purified by reversed-phase HPLC, indicated cross-linking material of type III collagen: Gly¹⁰²⁴-Ala-Hyp-Gly¹⁰²⁷-Pro-Cys-Cys-Gly-Ala-, Gly¹⁰²⁷-Pro-Cys-Cys-Gly-Ala-Gly-Gly-Val- and Gly⁸⁸-Met-Hyp-Gly-Phe-Hyp-Gly-Met-. The first two sequences are variants of $Col_c(III)$ and the third sequence is a segment of $T_n(III)$. The main component should be $[Col_c(III) \times T_n(III)]_2 \times Col_c(III)$, on the basis of molecular mass (12 kDa) and composition data. Three cystine residues are consistent with the presence of three $Col_c(III)$ and two histidine or arginine residues with two $T_n(III)$. The collagenase-derived cross-linked peptide was obtained from a tryptic cross-linked peptide of equivalent structure. Reduction of the 20 kDa pool, yielding a molar ratio of 2 for the component peptides $T_c(III) \times T_n(III)$ and $T_c(III)$, is consistent with the occurrence of $[T_c(III) \times T_n(III)]_2 \times T_c(III)$ (Figure 6b).

DISCUSSION

The analysis of collagen III cross-linking structures, containing the C-terminus of $\alpha 1(III)$ as their basic component, is facilitated by the isolation of the C-telopeptide in its monomer non-cross-linked form. Its identification and quantitative determination contribute to the detection of cross-linking structures and can give an insight into their assembly. Reductive alkylation of the 15 kDa fraction from the trypsin-soluble aortic-collagen released peptide $T_c(III)_L$ and digestion with collagenase yielded $Col_c(III)$ consisting almost completely of a non-helical domain, i.e. the collagen III C-telopeptide.

The amino acid sequences of the C-telopeptides from collagen III and I, comprising approx. 25 residues, diverge considerably (Figure 5). A sequence divergence of approx. 77% was found for man, calf and chicken. The C-telopeptide divergence between different species is much more extensive within collagen III than

Figure 3 Amino acid sequence analysis of Col_c(III) × T_N(III)

The first 13 steps are shown, together with an analysis of a standard 1 nmol of phenylthiohydantoin, to illustrate the presence of a double sequence. The peaks in the standard and sample are identified by the single-letter codes for amino acids. P* denotes 4-hydroxyproline, which elutes as two isomers; NI denotes norleucine, which was used as an internal standard; PEC denotes pyridylethylcysteine.

within collagen I. A pairwise comparison between man/calf, man/chicken and calf/chicken yields values of 12%, 4% and 8% for collagen I, and 25%, 62.5% and 70.8% for collagen III. The different collagen III C-telopeptides are of amphipathic nature, i.e. they have a polar group (Glu^{15C}, Lys^{16C}) and two hydrophobic arms that can be thought to be stabilized by folding back on themselves. ¹H NMR studies of a synthetic type III C-telopeptide indicated a β -turn within the polar region [27], supporting the idea of a polar, more hydrophilic, region located on the molecule surface exposed to an aqueous environment and a non-polar region oriented to the interior of the fibril. The non-polar segment of the C-telopeptide (Gly^{2C}-Ala^{14C} and Ala^{17C}-Tyr^{24C}) has the potential to interact with the helix T_N(III)

hydrophobic region that is N-terminal to residue 96 (Gly⁸⁵-Met⁹⁵), indicating its role in fibril assembly. The polar region of the type III C-telopeptide, containing the conserved partial sequence Glu-Lys-Ala (Figure 5), is involved in cross-linking. The β -turn like fold of the C-telopeptide, putting Lys^{16C} into a conformation in which it could be a substrate for lysyl hydroxylase, should favour the synthesis of hydroxylysine, mainly found in Col_c(III) of the vessel wall. The cross-link forming hydroxylysine occupies the same position (16^C) within both Col_c(III) and Col_c(I), as shown for man, calf and chicken. Identical cross-linking domains within the non-helical as well as the helical regions of collagen III and I might allow formation of copolymers between both types of collagen. It is remarkable that the Col_c(III)

Figure 4 Cross-link analysis of $T_c(III) \times T_n(III)$

The acid hydrolysates of the reduced cross-linked peptide (**A, B**) and reduced bone tissue (**C**) for reference are compared by amino acid analysis. (**A**) Normal programme for amino acid analysis, applying stepwise elution with 0.07 M sodium citrate buffers. (**B, C**) Single-buffer elution at 60 °C with 0.35 M sodium citrate buffer, pH 5.28. Solid line, A_{570} ; broken line, 3H radioactivity.

regions, in contrast with the $Col_c(I)$ regions, do not contain any histidine, preventing $Col_c(III)$ from forming histidine-containing cross-links (see below).

Direct evidence for the involvement of Hyl/Lys from $Col_c(III)$ in covalent cross-linking was furnished by isolation of $T_c(III) \times T_n(III)$ and $Col_c(III) \times T_n(III)$. A sequence analysis of both peptides as well as isolation of $\alpha 1(III)CB3B^*$ and $\alpha 1(III)CB3C$, representing the $T_n(III)$ region, conclusively demonstrated that the C-terminal telopeptide of one $\alpha 1(III)$ -chain is connected with an N-terminal helical cross-link site of another. This is in agreement with the 4D-staggered arrangement of collagen III molecules within native fibres and is, after sequence analysis of $[T_n(III)]_2 \times T_c(III)$ [1], the second piece of direct chemical evidence (amino acid sequence) demonstrating this for type III collagen.

Cross-linking within $T_c(III) \times T_n(III)$ is dominated by dehydrodihydroxylysinonorleucine. The qualitative and quantitative pattern of the bivalent cross-links (90 % dehydro-DHLNL, 10 % dehydro-HLNL) agrees with that found in type III collagen of native bovine amnion tissue [28]. The dimer cross-

links are formed between the Hyl/Lys residues 16^C of $Col_c(III)$ and 96 of $T_n(III)$, taking into account the stereochemistry of the lysine residues, the chemical nature of the cross-links and the absence of a phenylthiohydantoin derivative of Hyl/Lys at the 16^C position. Another bivalent cross-link is assumed to be present in type III collagen, connecting two N-telopeptides of two different type III molecules and giving evidence for a registered alignment of type III molecules [1] (Figure 6c).

The trifunctional cross-link pyridinoline has not been found in the 12 kDa fraction assumed to contain $[Col_c(III) \times T_n(III)]_2 \times Col_c(III)$. This finding is unexpected in view of the dominant presence of DHLNL in $Col_c(III) \times T_n(III)$, which is thought to be the precursor of pyridinoline [29]. The deficiency of pyridinoline and simultaneously high levels of DHLNL seem to be typical of immature tissues such as a 3-month-old calf aorta and can be explained by the occurrence of DHLNL as an acid-labile Schiff base. It is suggested that during maturation of cross-links, the Schiff base is rearranged into the ketoamine hydroxylysino-5-oxonorleucine (Hyl 96-Hyl^{ald,16C}) reacting with a telopeptide hydroxyallysine of a third $\alpha 1(III)$ -chain (Hyl^{ald,16C}) or a vicinal ketoamine to form pyridinoline. Formation of pyridinoline is basically possible within collagen III, as shown for the collagen III N-telopeptide region by isolation of the trimeric cross-linked peptide $[T_n(III)]_2 \times T_c(III)$ of calf aorta and human leiomyoma (W. Henkel, unpublished work) (Figure 6c). The pathway of pyridinoline described is consistent with the packing mode of collagen III molecules (see below) and also the collagen I molecules of tendon [30]. Formation of pyridinoline should also proceed in collagen fibrils of skin and blood vessels tilted by about 18 ° with respect to the fibril axis, taking into account the potential of pyridinoline to cross-link either two or three molecules [17,31]. The occurrence of a pyridinoline linking three different molecules is inconsistent with the packing mode of skin collagen I molecules [32,33]. The small intermolecular displacement of neighbouring molecules within a layer precludes a vicinal arrangement of the two telopeptide hydroxyallysine residues located on different molecules. The packing mode is, however, consistent with a pyridinoline linking two molecules. Then the two telopeptide hydroxyallysine residues involved, both located on the same molecule, would be adjacent to one another and allow formation of pyridinoline. Indeed, the twisted collagen I fibrils of aortic tissue have been shown to contain pyridinoline [17]. There is no correlation between the occurrence of pyridinoline and D-periodicity within collagen fibrils. Pyridinoline has been found in

C-telopeptides of $\alpha 1(I)$

1^C Ser Ala Gly Phe Asp Phe Ser Phe Leu Pro Gln Pro Pro Gln **Glu Lys Ala** His Asp Gly Gly Arg Tyr Tyr Arg Ala 26^C
 Gly Tyr Leu
 Gly
 man
 calf
 chicken

C-telopeptides of $\alpha 1(III)$

1^C Cys Gly Gly Val Gly Ala Ala Ala Ile Ala Gly Ile Gly Gly **Glu Lys Ala** Gly Gly Phe Ala Pro Tyr Tyr Gly 25^C
 Ala Gly Val Val Val Ala
 Gly Val □ □ □ □ Ser Leu Ala
 Gly Pro Val Tyr Gly Glu Arg
 man
 calf
 chicken

Figure 5 Amino acid sequences of C-telopeptides from $\alpha 1(I)$ and $\alpha 1(III)$

Sequences of calf and chicken are shown only when they differ from the human. Symbols: □, deletions in the chicken gene; ↓, amino acids absent from the human but present in the chicken chain. The C-telopeptide sequences of $\alpha 1(I)$ were taken from [24] and those from $\alpha 1(III)$ for man and chicken from [25] and [26]. The collagen III telopeptide sequence of calf was obtained by amino acid sequence analysis of $Col_c(III)$.

Figure 6 Proposed packing of type III molecules

(a) Representation of a collagen III molecule as a cylinder approx. 3000 Å long. The positions of the amino acid residues involved in intermolecular cross-linking are denoted. (b) Cross-linking structures within the dimer $[T_C(III) \times T_N(III)]_2 \times T_C(III)$. The structure of the peptide was confirmed by reduction, yielding a molar ratio of 2 for the component peptides $T_C(III) \times T_N(III)$ and $T_C(III)$. Hyl⁹⁶ and Hyl^{16C} of different $\alpha 1(III)$ -chains involved in cross-linking are indicated. The straight and jagged lines represent helical and non-helical regions of the $\alpha 1(III)$ -chain. The non-cross-linked $T_N(III)$ region is marked by a broken line. (c) Configuration of a tetramer as derived from cross-link analysis. The cross-links of the collagen III N-telopeptide are also considered. For details see the Discussion section. (d) Schematic illustration of a type III collagen filament. The collagen molecule is represented by an arrow, with the head being the C-terminus. The intermolecular cross-links denoted by ● correspond to those in (c).

tendon fibrils [34] with an axial *D* period of 67 nm and in collagen fibrils of aorta [17] or mineralized bone [35] with a decreased axial *D* period of 65 nm.

The absence of the trifunctional cross-link histidinohydroxylysinonorleucine (HHL) in the 12 kDa fraction might indicate that the C-telopeptide of collagen III cannot form any HHL, in contrast with the corresponding region of collagen I [32]. Formation of HHL in collagen III, equivalent to that in collagen I [36], would require an $\alpha 1(III)$ His¹⁰¹ residue and the bivalent cross-link dehydro-HLNL connecting $\alpha 1(III)$ Lys^{16C} and $\alpha 1(III)$ Hyl⁹⁶. Synthesis of HHL cannot be realized in collagen III because dehydro-HLNL occurs only in small amounts and $\alpha 1(III)$ His¹⁰¹ is not present. Formation of HHL, involving His^{18C}, as imaginable for type I collagen, must also be excluded, because

Col_C(III) does not contain any histidine (Figure 5). The failure to detect HHL in collagen III might be a direct consequence of the collagen III packing mode.

A first hint of the supramolecular assembly of type III molecules is the occurrence of $[T_C(III) \times T_N(III)]_2 \times T_C(III)$ (Figure 6b). The dimer, consisting of two 4*D*-staggered molecules, obviously contains at least two Hyl/Lys-derived cross-links arranged parallel to the disulphide bonds. Furthermore, several dimers can be thought of as being cross-linked by 0*D*-staggered disulphide bonds, considering their intermolecular function [37]. Finally, bearing in mind the results of the collagen III cross-link analysis, a tetrameric structure is proposed, as represented in Figure 6(c). Two sheets of in-register molecules, displaced against each other by 4*D*, are fixed in this arrangement by three strands of intermolecular bonds organized in a parallel manner. The disulphide bonds distinctive of type III collagen, present in duplicate at the C-terminal end of the triple helix, force the helical regions of collagen III molecules into a 'strict 0*D* stagger' arrangement. This alignment is consolidated by lysine-derived bivalent cross-links (Lys^{8N}-Lys^{8N}), connecting the N-telopeptides of different type III molecules. The 4*D*-staggered cross-links, deriving from either the N-telopeptide (Hyl^{8N}-Hyl^{8N}-Hyl⁹³⁹) or the C-telopeptide (Hyl^{16C}-Hyl⁹⁶) are believed to stabilize the potential for a strict 0*D*-staggered alignment of type III molecules within a layer. The organization of a collagen III filament, on the basis of the tetramer structure, is represented schematically in Figure 6(d). The proposed fibril structure is consistent with the configuration of pyridinoline, but incompatible with formation of HHL. The assembly of molecules has the implication that a His¹⁰¹ residue of $\alpha 1(III)$, essential for the formation of HHL, would have no function. This might be the reason why the histidine residue did not arise during evolution.

The packing mode of collagen III described in some respects is related to that of collagen I tendon. The in-register alignment of neighbouring molecules within layers [30] and the absence of HHL but the presence of pyridinoline could account for collagen III fibrils whose molecules are oriented along the fibril axis inducing a *D*-periodicity of 67 nm during the process of self-assembly. It would be of great interest to establish a model of electrostatic interactions between collagen III molecules, as has been done for collagen I molecules [33], and to ask whether there is a causal relation between the maximum stabilization of type III molecules and their orientation relative to the fibril axis. The possibility cannot be excluded that in a later stage of fibrillogenesis the type III molecules are reoriented by the presence of other matrix molecules such as type I collagen. So far there are no experimental results on the orientation and *D*-periodicity of type III fibrils in a tissue.

The considerations on the collagen III packing mode could shed light on formation of collagen fibrils in the vessel wall. Collagen fibrils of aorta with a tilting angle of about 17°, a shortened axial periodicity of 65 nm and small uniform diameters of about 50 nm [38] indicate the packing motif of collagen I skin [31]. The experimental results are explained by a packing mode of type I molecules which, different from that of collagen III, is distinguished by a 'near-register' that is a small intermolecular offset of neighbouring molecules within a layer and a tilting of the molecules against the fibril axis by about 16° [32]. This arrangement of molecules was computed to be the most stable [33] and is consistent with the configuration of HHL [32,33], which is suggested to occur in collagen I of the vessel wall. Furthermore the twisted fibrils of the aorta were shown to contain collagen III, as demonstrated by studies with monoclonal antibodies [39]. Finally there is experimental evidence that collagen III could regulate the diameter of collagen I fibrils [40].

Copolymerization of two similar but differently structured monomers might be the reason for the inhibitory influence of collagen III on the assembly of collagen I fibrils. The structural similarities of both monomers, as documented by, for example, identical helical and non-helical cross-link sites, allow the insertion of collagen III molecules into the growing type I fibril. However, during fibril growth two fibril systems are co-assembled the packing modes are believed to be different from each other. In addition there are differences in the monomer structure [2]. It is suggested that the divergent features of collagen III and collagen I fibrils, as present in the vessel wall, do bring about an incompatibility of both fibril systems, disturbing further fibril growth. There has also been discussion of whether there is a causal relationship between the presence of type III collagen and the twisted arrangement of collagen fibrils or a shorter *D* period, on the basis of data from the vessel wall and other connective tissues containing type I and type III collagen. However, there is experimental evidence indicating that collagen fibrils of aorta described are also formed in the absence of collagen III, as shown for a patient with type IV Ehlers–Danlos syndrome [41]. Studies on lamprey tissues seem to confirm that tilted fibrils and a reduced *D* period had already evolved before type III collagen was established during evolution [31].

I thank Dr. R.W. Glanville (Oregon Health Sciences University, Portland, OR, U.S.A.) for performing the amino acid sequence analyses.

REFERENCES

- Henkel, W. and Glanville, R. W. (1982) *Eur. J. Biochem.* **122**, 205–213
- Ala-Kokko, L., Kontusaari, S., Baldwin, C. T., Kuivaniemi, H. and Prockop, D. J. (1989) *Biochem. J.* **260**, 509–516
- Janeczko, R. A. and Ramirez, F. (1989) *Nucleic Acids Res.* **17**, 6742
- Fietzek, P. P., Allmann, H., Rauterberg, J., Henkel, W., Wachter, E. and Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 809–820
- Dewes, H., Fietzek, P. P. and Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 821–832
- Bentz, H., Fietzek, P. P. and Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 833–840
- Lang, H., Glanville, R. W., Fietzek, P. P. and Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 841–850
- Dewes, H., Fietzek, P. P. and Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 851–860
- Allmann, H., Fietzek, P. P., Glanville, R. W. and Kühn, K. (1979) *Hoppe-Seyler's Z. Physiol. Chem.* **360**, 861–868
- Fietzek, P. P., Allmann, H., Rauterberg, J. and Wachter, E. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 84–86
- Nicholls, A. C. and Bailey, A. J. (1980) *Biochem. J.* **185**, 195–201
- Glanville, R. W. and Fietzek, P. P. (1977) *Anal. Biochem.* **80**, 282–288
- Glanville, R. W. and Fietzek, P. P. (1976) *FEBS Lett.* **71**, 99–102
- Stepp, M. A., Kindy, M., Franzblau, C. and Sonenshein, G. (1985) *Ann. N. Y. Acad. Sci.* **460**, 510–513
- Henkel, W., Rauterberg, J. and Glanville, R. W. (1979) *Eur. J. Biochem.* **96**, 249–256
- Bornstein, P. and Pietz, K. A. (1966) *Biochemistry* **5**, 3460–3473
- Henkel, W., Glanville, R. W. and Greifendorf, D. (1987) *Eur. J. Biochem.* **165**, 427–436
- Friedman, M., Krull, L. H. and Cavins, J. F. (1970) *J. Biol. Chem.* **245**, 3868–3871
- Cavins, J. F. and Friedman, M. (1970) *Anal. Biochem.* **35**, 489–493
- Lottspeich, F. (1980) *Hoppe-Seyler's Z. Physiol. Chem.* **361**, 1829–1834
- Barnes, M. J. (1985) *Collagen Relat. Res.* **5**, 65–97
- Ooshima, A. (1981) *Science* **213**, 666–668
- Glanville, R. W., Allmann, H. and Fietzek, P. P. (1976) *Hoppe-Seyler's Z. Physiol. Chem.* **357**, 1663–1665
- Bernard, M. P., Chu, M. L., Myers, J. C., Ramirez, F., Eickenberry, E. F. and Prockop, D. J. (1983) *Biochemistry* **22**, 5213–5223
- Chu, M. L., Weil, D., de Wet, W., Bernard, M., Sippola, M. and Ramirez, F. (1985) *J. Biol. Chem.* **260**, 4357–4363
- Yamada, Y., Kühn, K. and de Crombrughe, B. (1983) *Nucleic Acids Res.* **11**, 2733–2744
- Liu, X., Otter, A., Scott, P. G., Cann, J. R. and Kotovych, G. (1993) *J. Biomol. Struct. Dyn.* **11**, 541–555
- Cannon, D. J. and Davison, P. F. (1978) *Biochem. Biophys. Res. Commun.* **85**, 1373–1378
- Eyre, D. R. and Oguchi, H. (1980) *Biochem. Biophys. Res. Commun.* **92**, 403–410
- Hulmes, D. J. S. (1992) *Essays Biochem.* **27**, 49–67
- Brodsky, B., Bel Bruno, K. C., Hardt, T. A. and Eickenberry, E. F. (1994) *J. Mol. Biol.* **243**, 38–47
- Mechanic, G. L., Katz, R. P., Henmi, M., Noyes, C. and Yamauchi, M. (1987) *Biochemistry* **26**, 3500–3509
- Katz, E. P. and David, C. W. (1992) *J. Mol. Biol.* **228**, 963–969
- Fujimoto, D. (1980) *Biochem. Biophys. Res. Commun.* **93**, 948–953
- Otsubo, K., Katz, E. P., Mechanic, G. L. and Yamauchi, M. (1992) *Biochemistry* **31**, 396–402
- Yamauchi, M., London, R. E., Guenat, C., Hashimoto, F. and Mechanic, G. L. (1987) *J. Biol. Chem.* **262**, 11428–11434
- Cheung, D. T., Di Cesare, P., Benya, P. D., Libaw, E. and Nimni, M. E. (1983) *J. Biol. Chem.* **258**, 7774–7778
- Raspanti, M., Ottani, V. and Ruggeri, A. (1989) *Int. J. Biol. Macromol.* **11**, 367–371
- Keene, D. R., Sakai, L. Y., Bächinger, H. P. and Burgeson, R. E. (1987) *J. Cell Biol.* **105**, 2393–2402
- Romanic, A. M., Adachi, E., Kadler, K. E., Hojima, Y. and Prockop, D. J. (1991) *J. Biol. Chem.* **266**, 12703–12709
- Folkhard, W., Christmann, D., Geercken, W., Knörzer, E., Koch, M. H. J., Mosler, E., Nemetschek-Gansler, H. and Nemetschek, T. (1987) *Z. Naturforsch.* **42c**, 1303–1306