Complete amino acid sequence of the *N*-terminal extension of calf skin type III procollagen

Andrea BRANDT,* Robert W. GLANVILLE, Dietrich HÖRLEIN,† Peter BRUCKNER,‡
Rupert TIMPL, Peter P. FIETZEK§ and Klaus KÜHN
Max-Planck-Institut für Biochemie, 8033 Martinsried bei München, Federal Republic of Germany

(Received 21 October 1983/Accepted 5 January 1984)

The *N*-terminal extension peptide of type III procollagen, isolated from foetal-calf skin, contains 130 amino acid residues. To determine its amino acid sequence, the peptide was reduced and carboxymethylated or aminoethylated and fragmented with trypsin, *Staphylococcus aureus* V8 proteinase and bacterial collagenase. Pyroglutamate aminopeptidase was used to deblock the *N*-terminal collagenase fragment to enable amino acid sequencing. The type III collagen extension peptide is homologous to that of the $\alpha 1$ chain of type I procollagen with respect to a three-domain structure. The *N*-terminal 79 amino acids, which contain ten of the 12 cysteine residues, form a compact globular domain. The next 39 amino acids are in a collagenase triplet sequence (Gly-Xaa-Yaa)_{*n*} with a high hydroxyproline content. Finally, another short non-collagenous domain of 12 amino acids ends at the cleavage site for procollagen aminopeptidase, which cleaves a proline–glutamine bond. In contrast with type I procollagen, the type III procollagen extension peptides contain interchain disulphide bridges located at the *C*-terminus of the triple-helical domain.

Intact procollagens, with both *N*- and *C*-terminal extension peptides, cannot be extracted directly from tissues. However, it is possible to extract collagen molecules with only *N*-terminal extensions intact (pN-collagens). Type III pN-collagen has been isolated from lathyrict rat skin (Byers *et al.*, 1974; Lanaers & Lapiere, 1975) and foetal-calf skin (Timpl *et al.*, 1975). During the biosynthesis of collagen the *N*- and *C*-terminal extension peptides are removed by specific peptidases (Tuderman *et al.*, 1978; Nusgens *et al.*, 1980). The freed *N*-terminal type III procollagen peptide is a trimer containing the extension peptides of the three α chains of the type III collagen triple helix, joined by a disulphide bridge (Nowack *et al.*, 1976). It contains a short collagen-like triple-

helical domain and a compact cysteine-rich *N*-terminal globular region, similar to those found in type I procollagen (Hörlein *et al.*, 1979). These structural features endow the extension peptide with unusual thermal stability and resistance to proteolytic attack (Bruckner *et al.*, 1978).

Different forms of the extension peptides can be immunologically detected in various normal and pathological body fluids (Rohde *et al.*, 1979a, 1983b). As the serum concentration of this peptide is elevated in some diseases, determination of its concentration in serum could have diagnostic applications.

The *N*-terminal propeptides of collagens have also been implicated in the control of collagen biosynthesis by a feedback-inhibition mechanism (Krieg *et al.*, 1978; Wiestner *et al.*, 1979). The extension peptide of type III pN-collagen inhibited the synthesis of type I and type III collagens in cultures of skin fibroblasts equally well, but did not affect the synthesis of type II collagen in cultures of chondrocytes. However, the synthesis of all three types of collagen was inhibited in cell-free translation systems, indicating that the collagen-type-specific inhibition is at least in part regulated by the cell membrane (Paglia *et al.*, 1979, 1981). These results suggested that the structures of the type I

* Present address: Institut für Pharmakologie und Toxikologie der Universität Marburg, Universität Lahnburg, 355 Marburg, Federal Republic of Germany.

† Present address: Biogen, 3 Route de Troinex, CH-1227 Genève, Switzerland.

‡ Present address: Laboratorium für Biochemie der Eidgenössischen Technischen Hochschule Zürich, Universitätsstrasse 16, CH-8006 Zürich, Switzerland.

§ Present address: Firma Karl Freudenberg, Zwischen den Dämmen, 6940 Weinheim, Federal Republic of Germany.

and type III extension peptides are similar but that that of type II is different.

The amino acid sequence of the *N*-terminal extension peptides of the $\alpha 1$ chain of type I pN-collagen isolated from dermatosparactic calf skin is known (Hörlein *et al.*, 1979). Those of dermatosparactic sheep skin (Rohde *et al.*, 1979b) and chick tendon (Pesciotta *et al.*, 1980) have been partially determined. The composition of the *N*-terminal propeptide of the $\alpha 2(I)$ chain from dermatosparactic sheep skin pN-collagen (Becker *et al.*, 1977) and the complete amino acid sequence of the *N*-terminal propeptide of the $\alpha 2(I)$ chain from chicken, deduced from DNA sequence data (Tate *et al.*, 1983), revealed as the most striking difference the absence of the *N*-terminal cysteine-rich globular domain from the pro $\alpha 2(I)$ chain. Preliminary structural data on the chick type II procollagen *N*-terminal extension indicate that it also differs considerably from that of type I procollagen (Curran & Prockop, 1982).

In the present paper we report the complete amino acid sequence of the *N*-terminal extension peptide of type III pN-collagen isolated from foetal-calf skin and compare it with the corresponding sequences of bovine and chick type I pN-collagen *N*-terminal extension peptides.

Methods

Collagen and extension peptide

Foetal-calf skin type III pN-collagen was prepared as described previously (Timpl *et al.*, 1975). The free extension peptide was also isolated from extracts of foetal-calf skin as described previously (Glanville, 1982). This peptide is presumably present in the tissues or is liberated from type III pN-collagen by the specific type III N-proteinase during the extraction procedure.

Preparation of proteolytic fragments

A bacterial-collagenase-derived peptide Col 1-3 was prepared by treating type III pN-collagen with CLSPA collagenase (400 units/mg; Worthington Biochemical Corp.) at an enzyme/substrate ratio of 1:100 for 4 h at 37°C in 50 mM-Tris/HCl buffer, pH 7.5, containing 0.2 M-NaCl and 5 mM-CaCl₂. A second, shorter, collagenase-derived peptide Col 1 was prepared by cleaving reduced and aminoethylated type III pN-collagen or the extension peptide under the conditions described above. The digestion products were separated on a column (5 cm × 30 cm) of DEAE-cellulose (DE52; Whatman). The column was equilibrated at room temperature with 50 mM-Tris/HCl buffer, pH 8.6, containing 2 M-urea, and the peptides were eluted with a linear gradient of 0–0.45 M-NaCl in 4000 ml at a flow rate of 200 ml/h. The required

fraction was rechromatographed on a column (3.5 cm × 150 cm) of agarose A1.5 m (Bio-Rad Laboratories) equilibrated at room temperature with 50 mM-Tris/HCl buffer, pH 7.5, containing 1 M-CaCl₂ at a flow rate of 60 ml/h.

Staphylococcal proteinase (SP)-derived peptides were produced by degrading aminoethylated peptides Col 1-3 and Col 1. The enzyme (*S. aureus* V8 proteinase; 450 units/mg; Miles Laboratories) was added to the substrate dissolved in 0.2 M-NH₄HCO₃, pH 7.8, to give a final enzyme/substrate ratio of 1:40. After incubation at 37°C for 16 h, the peptide mixture was separated on a column (2 cm × 150 cm) of Sephadex G-50 (superfine grade) (Pharmacia) equilibrated with 0.2 M-NH₄HCO₃ at a flow rate of 20 ml/h. The smaller peptides were separated by using a combination of ion-exchange columns. Both phosphocellulose (P11; Whatman) in a column (1.3 cm × 12 cm) equilibrated at 40°C with 5 mM-sodium acetate buffer, pH 3.6, and DEAE-cellulose (DE52; Whatman) in a column (1.3 cm × 12 cm) equilibrated at 40°C with 50 mM-Tris/HCl buffer, pH 8.6, containing 2 M-urea were used. The peptides were eluted with a linear NaCl gradient, 0–0.1 M in 2000 ml for the phosphocellulose and 0–0.4 M in 800 ml for the DEAE-cellulose.

Tryptic peptides (T) were produced by cleaving the reduced and carboxymethylated extension peptide with trypsin (1-chloro-4-phenyl-3-tosylamido-butan-2-one-treated; 182 units/mg; Worthington Biochemical Corp.) in 0.2 M-NH₄HCO₃, pH 7.8, at 37°C for 4 h at an enzyme/substrate ratio of 1:40. The enzyme digest was fractionated on a Sephadex G-50 (superfine grade) column as described for the SP peptides above. The smallest-*M_r* fraction was rechromatographed on Bio-Gel P4 (200–400 mesh; Bio-Rad Laboratories). The column (2.5 cm × 110 cm) was equilibrated at room temperature with 0.1 M-acetic acid at a flow rate of 35 ml/h. The largest-*M_r* fraction was rechromatographed on DEAE-cellulose as described above.

In order to produce smaller tryptic peptides, aminoethylated peptide Col 1 was treated with trypsin in 0.2 M-Tris/HCl buffer, pH 8.0, containing 10 mM-CaCl₂ at 50°C for 6 h at an enzyme/substrate ratio of 1:40. Then the same weight of enzyme was added again, and the incubation was continued at 37°C for another 16 h. Initial fractionation of the products was carried out on Sephadex G-50 (superfine grade), and the fraction of interest was rechromatographed on phosphocellulose as described for the SP peptides.

The bacterial-collagenase fragment Col 1 was deblocked by using pyroglutamate aminopeptidase (Boehringer, Mannheim, Germany). The peptide (200 nmol) was dissolved in 65 mM-sodium phosphate buffer, pH 7.4, and 1 mg of enzyme was

added. After 6 h at 37°C the digest was chromatographed on Sephadex G-75 (superfine grade) equilibrated with 0.2M-NH₄HCO₃.

Acid hydrolysis of the extension peptide

The extension peptide with intact disulphide bridges was hydrolysed in 0.25M-acetic acid in a sealed ampoule for 4 h at 110°C. The products were chromatographed on Sephadex G-50 (superfine grade) as described above.

Chemical modifications of peptides

Peptides were reduced with 20mM-dithioerythritol in the presence of 8M-urea and subsequently alkylated either with sodium iodoacetate (Gollwitzer *et al.*, 1972) or with ethyleneimine (Gollwitzer *et al.*, 1974). Reagents were removed by desalting on Bio-Gel P2 (100–200 mesh) equilibrated in 10mM-acetic acid.

Amino acid analysis

Peptides were hydrolysed with 6M-HCl under N₂ (20 h, 110°C) and analysed on a Durrum D500 amino acid analyser. Cysteine residues were determined as *S*-carboxymethylated or *S*-aminoethylated derivatives.

Amino acid sequence analysis

Sequence determinations were carried out in a liquid-phase sequencer (Beckman model 890B) with the Protein Quadrol Programme (no. 060275). All peptides were sequenced in the presence of Polybrene (Klapper *et al.*, 1978) and 0.1M-Quadrol. The amino acid phenylthiohydantoin derivatives formed were identified by using g.l.c. (Fietzek & Rexrodt, 1975) and high-performance t.l.c. (Glanville & Rauter, 1981) and more recently high-pressure liquid chromatography (Henkel & Glanville, 1982).

Results

The preparation of bacterial-collagenase-derived fragments is illustrated in Fig. 1. In this case the preparation of peptide Col 1-3 from type III pN-collagen is shown, but a similar chromatogram is obtained for the preparation of peptide Col 1 (Glanville, 1982). The amino acid compositions of peptide Col 1-3, peptide Col 1 and the extension peptide are shown in Table 1. The extension peptide, which is the *N*-terminal 130 amino acid residues of type III procollagen, is liberated by type III procollagen *N*-proteinase cleaving a proline-glutamine bond (positions 130–131 in Fig. 4). Peptide Col 1-3 contains the extension peptides plus 17 amino acid residues of type III collagen, as indicated in Fig. 4. Peptide Col 1 is 85 amino acid residues long and is located at the *N*-terminal end of

Fig. 1. Separation of peptide Col 1-3 from a bacterial-collagenase digest of foetal-calf skin type III pN-collagen. The digest (1 litre) was applied to the DEAE-cellulose column (5 cm × 30 cm), and the column was washed with equilibration buffer before application of the gradient at G (0–0.45 M-NaCl in 2 litres). For details see the text. The indicated pool from the DEAE-cellulose separation was further fractionated on agarose A1.5m (results not shown).

the extension peptides, as shown diagrammatically in Fig. 2 and in Fig. 4. All five intrachain disulphide bonds are located in the Col 1 region.

After peptide Col 1 had been deblocked with pyroglutamate aminopeptidase the peptide was sequenced and the *N*-terminal 21 residues were identified (Fig. 2).

Separation of the tryptic digest of carboxymethylated extension peptide is shown in Fig. 3(a). Peptides T1 and T3 were separated on Bio-Gel P4, and peptide T2 was purified on DEAE-cellulose (results not shown). Peptide T4 did not require further purification. The amino acid compositions of the tryptic peptides are given in Table 2. Peptides T2, T3 and T4 were sequenced, identifying 21, 11 and 26 positions respectively. Peptide T1 corresponded to the *N*-terminal 18 residues of peptide Col 1, but was blocked with an additional *N*-terminal pyroglutamic acid (pyrrolid-2-one-5-carboxylic acid) residue. The order of the peptides T1 and T2 was clear from the sequence of peptide Col 1 (Fig. 2).

Separation of a *Staphylococcus aureus*-proteinase digest of aminoethylated peptide Col 1-3 is shown in Fig. 3(b). This is a representative chromatogram, as the extent of cleavage at aspartic acid residues varied greatly with enzyme batch and peptide preparation. Cleavage at glutamic acid residues was more consistent. Unlabelled peaks are double peptides or fragments produced by additional cleavages at aspartic acid residues. Many of these peptides were also isolated and sequenced (results not shown). The peptides described in the present paper are the major products required to establish the complete sequence. Amino acid compositions of the major SP peptides are shown in Table 3. Peptides SP2, SP3, SP4, SP5 and SP7 were com-

Fig. 2. Schematic diagram showing the relative positions of the sequenced peptides

SP, *Staphylococcus aureus*-V8-proteinase-derived; T, tryptic; Col, bacterial-collagenase derived; TA, tryptic peptide produced by cleavage at aminoethylcysteine; HA, peptide released by mild acid hydrolysis. Type III CB3A indicates the beginning of the known sequence of type III collagen (Glanville & Fietzek, 1976). The scale at the top of the diagram indicates the residue number. The numbers below the peptides give the number of amino acids in the peptide. The thick bars represent sequenced regions.

Table 1. Amino acid compositions of pN α 1(III) peptide Col 1-3, pN α 1(III) extension peptide, pN α 1(III) peptide Col 1, the tryptic peptide TA and the hydrolytic peptide AH

For details see the text. Numbers in parentheses are expected values calculated from the sequence. Abbreviation: N.D., not determined.

Amino acid	Peptide	Amino acid composition (mol of residue/mol of peptide)				
		Col 1-3	Extension peptide	Col 1	TA	AH
Hyp	...	9.9 (9)	9.1 (9)	—	—	—
Asp		16.4 (16)	15.0 (15)	12.6 (11)	5.8 (6)	3.2 (3)
Thr		3.2 (3)	3.3 (3)	2.7 (2)	—	2.5 (2)
Ser		8.4 (10)	7.3 (9)	4.9 (4)	—	—
Glu		16.6 (17)	13.7 (15)	12.3 (13)	3.2 (4)	—
Pro		15.8 (18)	16.1 (18)	12.4 (13)	2.8 (3)	1.8 (1)
Gly		25.0 (25)	20.1 (21)	8.5 (8)	1.5 (1)	4.3 (4)
Ala		6.7 (7)	4.2 (4)	4.5 (4)	—	—
Val		6.8 (7)	4.6 (5)	5.4 (5)	—	—
Cys		12.0 (12)	10.7 (12)	7.8 (10)	3.0 (3)	—
Ile		5.2 (7)	4.9 (6)	3.5 (4)	2.6 (3)	0.8 (1)
Leu		3.3 (3)	2.9 (3)	3.5 (3)	1.3 (1)	—
Tyr		3.9 (4)	1.0 (2)	1.0 (1)	—	—
Phe		0.7 (1)	1.2 (1)	0.8 (1)	1.5 (1)	—
His		0.9 (1)	1.1 (1)	0.9 (1)	—	—
Hyl		0.6 (3)	0.5 (2)	—	—	—
Lys		2.5 (3)	1.5 (2)	1.6 (2)	—	—
Arg		4.0 (3)	3.1 (3)	2.3 (2)	—	0.8 (1)
Trp		N.D. (1)	N.D. (1)	N.D. (1)	N.D.	N.D.
Total		141.9 (147)	121.3 (130)	84.7 (85)	21.7 (22)	12.4 (12)

Fig. 3. (a) Separation of the trypsin-derived peptides of the reduced and carboxylated extension peptide on Sephadex G-50 (superfine grade) and (b) separation of *Staphylococcus aureus*-V8-proteinase-derived peptides from reduced and aminoethylated peptide Col 1-3 on Sephadex G-50 (superfine grade)

For details see the text. *V*₀, Void volume. In (a) peptides T1 and T3 were separated on Bio-Gel P4, and peptide T2 was purified on DEAE-cellulose (results not shown). In (b) peptides SP6 and SP4 were re-chromatographed on the same column; peptides SP7, SP8 and SP3 and peptides SP2 and SP5 were separated on phosphocellulose (results not shown).

pletely sequenced. The first 28 positions of SP6 were identified (Fig. 4). The previously reported *N*-terminal end of type III collagen (Glanville & Fietzek, 1976) corresponded to the sequence of peptide SP8 plus the C-terminus of peptide SP7 (Fig. 2). Peptide SP1 was identified from its amino acid composition. From the sequences of peptides Col 1 and T2, peptides SP1, SP2, SP3 and SP4 could be ordered. Peptide T4 could then also be positioned, because its determined sequence overlapped that of peptide SP7 (Fig. 2). Peptide T3 also overlapped with peptide SP6 and so could be fixed in relation to peptide SP6, but the order of the two peptides SP5 and SP6 was not clear (Fig. 2).

Because of the high cysteine content of peptide Col 1, this peptide was aminoethylated and cleaved with trypsin under intensive conditions in order to fragment the peptide at arginine, lysine and aminoethylcysteine residues. After the preliminary fractionation of the digest on Sephadex G-50 (superfine grade), the required peptide TA was purified by using phosphocellulose (results not shown). Its amino acid composition is shown in Table 1. The amino acid sequence of this peptide fixed the order of peptides SP4, SP5 and SP6 (Fig. 2).

Only the absence of additional amino acids between peptides T3 and T4 remained to be confirmed. Inspection of the sequence in Fig. 4 showed that there are no suitable proteolytic cleavage points in this region. However, two aspartic acid-hydroxyproline bonds are present at positions 87-88 and 99-100. These should be labile to mild

Table 2. Amino acid composition of the tryptic peptides from the pNa1(III) extension peptide
For details see the text. Numbers in parentheses are expected values calculated from the sequence.

		Amino acid composition (mol of residue/mol of peptide)					
Amino acid	Peptide ...	T1	T2	T3	T4	Sum of T1-T4	Extension peptide
Hyp		—	—	3.0 (3)	6.4 (6)	9.4 (9)	9.1 (9)
Asp		1.8 (2)	8.7 (9)	1.0 (1)	3.1 (3)	14.6 (15)	15.0 (15)
Thr		—	1.8 (2)	—	1.0 (1)	2.8 (3)	3.3 (3)
Ser		1.6 (2)	1.7 (2)	—	3.6 (5)	6.9 (9)	7.3 (9)
Glu		3.7 (4)	7.1 (9)	—	2.1 (2)	12.9 (15)	13.7 (15)
Pro		—	11.9 (13)	1.1 (1)	3.3 (4)	16.3 (18)	16.1 (18)
Gly		3.9 (3)	7.2 (5)	3.9 (4)	8.9 (9)	23.9 (21)	20.1 (21)
Ala		2.2 (2)	2.1 (2)	—	—	4.3 (4)	4.2 (4)
Val		1.1 (1)	3.5 (4)	—	—	4.6 (5)	4.6 (5)
Cys		1.0 (1)	9.0 (9)	—	2.0 (2)	12.0 (12)	10.7 (12)
Ile		—	4.2 (4)	1.0 (1)	1.0 (1)	6.2 (6)	4.9 (6)
Leu		1.1 (1)	2.4 (2)	—	—	3.5 (3)	2.9 (3)
Tyr		0.7 (1)	—	—	— (1)	0.7 (2)	1.0 (2)
Phe		—	1.0 (1)	—	—	1.0 (1)	1.2 (1)
His		0.9 (1)	—	—	—	0.9 (1)	1.1 (1)
Hyl		—	0.6	—	—	0.6 (2)	0.5 (2)
Lys		—	1.0 (2)	—	—	1.0 (2)	1.5 (2)
Arg		1.0 (1)	1.4 (1)	1.1 (1)	—	3.5 (3)	3.1 (3)
Trp		—	— (1)	—	—	— (1)	— (1)
Total		19.0 (19)	63.6 (66)	11.1 (11)	31.4 (34)	126.1 (130)	121.3 (130)

Table 3. *Amino acid compositions of the Staphylococcus-proteinase peptides of pNaI(III) peptide Col 1-3*
 For details see the text. Numbers in parentheses are expected values calculated from the sequence. Abbreviation: N.D., not determined.

Amino acid	Peptide	...	SP1	SP2	SP3	SP4	SP5	SP6	SP7	SP8	Amino acid composition (mol of residue/mol of peptide)		Sum of SP1-SP8	Col 1-3
Hyp								8.2 (9)					8.2 (9)	9.9 (9)
Asp				3.2 (3)			1.9 (2)	4.2 (4)		1.1 (1)			17.4 (16)	16.4 (16)
Thr								2.0 (2)	1.0 (1)				3.0 (3)	3.2 (3)
Ser				1.7 (2)		1.8 (2)		2.8 (3)	1.9 (2)	1.3 (1)			9.5 (10)	8.4 (10)
Glu			(3)	1.2 (1)	2.0 (2)	2.3 (2)	2.0 (2)	4.3 (4)	2.9 (3)				17.7 (17)	16.6 (17)
Pro					1.3 (2)		2.5 (3)	12.1 (11)	1.1 (2)				17.0 (18)	15.8 (18)
Gly				3.0 (3)		1.2 (1)	1.9 (1)	12.4 (14)	2.2 (2)	4.3 (4)			25.0 (25)	25.0 (25)
Ala				2.1 (2)				2.5 (2)		2.8 (3)			7.4 (7)	6.7 (7)
Val				0.9 (1)	1.6 (2)	1.0 (1)		1.1 (1)		1.5 (2)			6.1 (7)	6.0 (7)
Cys				1.0 (1)	2.6 (3)	1.9 (2)	1.0 (1)	3.4 (4)	1.0 (1)				10.9 (12)	12.0 (12)
Ile					1.5 (1)	1.2 (2)	0.8 (1)	2.5 (2)		0.9 (1)			6.9 (7)	5.2 (7)
Leu				1.1 (1)		1.1 (1)	0.9 (1)						3.1 (3)	3.3 (3)
Tyr				0.8 (1)					1.7 (2)	1.0 (1)			3.5 (4)	3.9 (4)
Phe													0.8 (1)	0.7 (1)
His				0.8 (1)			0.8 (1)						0.8 (1)	0.9 (1)
Hyl													0.8 (1)	0.6
Lys					0.8 (1)			0.8 (1)					2.2 (3)	2.5 (3)
Arg				1.1 (1)				0.5		0.9 (1)			3.4 (3)	4.0 (3)
Trp					N.D. (1)			2.3 (2)					(1)	(1)
Total		(3)	16.9 (17)		12.3 (13)	14.8 (15)	11.8 (12)	59.1 (59)	13.0 (14)	13.8 (14)			141.7 (147)	141.9 (147)

Fig. 4. Complete amino acid sequence of the N-terminal extension peptide of calf skin type III collagen
Col, bacterial-collagenase-derived; SP, *Staphylococcus aureus*-V8-proteinase-derived; T, tryptic; TA, tryptic peptide produced by cleavage at an aminoethylcysteine residue; AH, peptide produced by mild acid hydrolysis. →, Residues identified by automatic sequencing of the peptide shown to the left; —, known sequence of bovine type III collagen (Glanville & Fietzek, 1976).

acid hydrolysis (Inglis, 1983). After some preliminary experiments, dilute acetic acid was chosen, and the only peptide freed from the disulphide-bonded starting material was the required peptide (AH). The amino acid composition of this peptide is shown in Table 1. It was com-

pletely sequenced (Fig. 4) and contained the required overlap between peptides T3 and T4. The complete sequence of the 130-residue extension peptide is shown in Fig. 4, together with the known N-terminal non-helical sequence of type III collagen (Glanville & Fietzek, 1976).

Discussion

With the completion of the amino acid sequence of the *N*-terminal extension peptide reported here, 1176 consecutive amino acid residues of calf skin type III procollagen have now been sequenced. The complete amino acid sequence of the extension peptide is shown in Fig. 4.

The *N*-terminus of this peptide was blocked, and could be deblocked by using pyroglutamate aminopeptidase, indicating that the terminal amino acid is pyroglutamic acid. Partial removal of the second amino acid glutamine also occurred. Pyroglutamic acid residues were also found as the *N*-terminal amino acids in the corresponding peptides of calf type I (Hörlein *et al.*, 1979) and sheep type I (Rohde *et al.*, 1979b) procollagens.

The *N*-terminal 79 amino acids contain ten of the 12 cysteine residues. This non-collagenous domain contains only intrachain disulphide bridges and has a conformation and stability similar to that of the corresponding region from type I procollagen (Engel *et al.*, 1977; Bruckner *et al.*, 1978).

Positions 80–118 consist of 13 tripeptide units of the form (Gly-Xaa-Yaa)_{*n*}. In native type III procollagen and in the extension peptide and peptide Col 1-3, this region forms a triple-helical domain with a 'melting' temperature of 53°C (Bruckner *et al.*, 1978). This is significantly higher than that of the 1000-amino acid-residue-long major triple helix of type III collagen, which 'melts' at 39°C. The extreme thermal stability of the extension peptide helix cannot, however, be explained on the basis of its amino acid sequence. Secondary stabilizing effects, due to the cysteine-rich globular domain or to the interchain disulphide bridges at the C-terminal of the helix, may play a role. Unfortunately, it is not possible to isolate the native triple helix without the globular domain attached, in order to test this hypothesis. The cysteine residues at positions 118 and 121 form three interchain disulphide bridges between the three polypeptide chains. The cysteine residue at position 118 is probably incorporated into the triple-helical structure, as is the case at the C-terminal of type III collagen (Bruckner *et al.*, 1978).

Residues 127–129 represent a possible oligosaccharide-attachment site of the form Asn-Xaa-Ser, which is not present in the pN α 1(I) or pN α 2(I) chains. In dermatosparactic calf skins about one *N*-linked oligosaccharide unit per extension peptide was found, but in normal foetal-calf skin only one-tenth of this amount (Shinkai & Lapiere, 1983). The C-terminal 12 amino acids end with proline residue 130. The cleavage point for type III procollagen N-proteinase (Nusgens *et al.*, 1980) is therefore between proline and glutamine. Residues

131–147 are the *N*-terminal non-helical peptides of type III collagen (Glanville & Fietzek, 1976).

Fig. 5 compares the extension peptides of calf type III procollagen, the α 1 chain of calf type I procollagen and the α 2 chain of chick type I procollagen, aligned to give maximum homology. The bond cleaved in both the pN α 1(III) and pN α 1(I) chains is between proline and glutamine, and in the pN α 2(I) between alanine and glutamine (position 144–145). The sequences to the left of the cleavage site in the pN α 1(I) and pN α 2(I) chains are identical for 19 out of 20 residues. The corresponding region of the pN α 1(III) chain contains 24 amino acid residues, including the two cysteine residues, which form interchain disulphide bridges. Only 12 of the 24 residues immediately to the left of the cleavage site in pN α 1(III) chain are identical with the corresponding region in pN α 1(I). Studies with synthetic peptides (Morikawa *et al.*, 1980) have indicated that the phenylalanine residue in the pN α 1(I) chain at position 142 is necessary for enzyme recognition. This residue is replaced by another bulky hydrophobic amino acid, namely tyrosine, in the pN α 1(III) chain. It has been shown that the different procollagen peptidases are specific either for type I or type III procollagen (Nusgens *et al.*, 1980). Comparing the linear sequences of the three chains, it is not obvious what determines this specificity. Alkylation of the cysteine residues in type I and type III procollagen under non-denaturing conditions does not affect the removal of the extension peptides by procollagen peptidase (Nusgens & Lapiere, 1979; Nusgens *et al.*, 1980), but unfolding the triple helix significantly decreases the rate of cleavage (Tuderman *et al.*, 1978). This indicates that the triple-helical region is involved in the recognition site, and, as the distance of the triple-helical domain from the cleaved peptide bond is different in type I and III procollagens, this may determine the enzymes' collagen-type-specificity. Sequences to the right of the cleavage point are not homologous. They vary in length and in distribution of charged and polar amino acids, and are probably not important for enzyme cleavage.

The highest homology (67%) is found in the compact region between the first and tenth cysteine residue (positions 18–74; Fig. 5), including invariance for all cysteine residues. The segments flanking the compact domain are more variable, and contain several deletions. The region from the *N*-terminus to the first cysteine residue in the pN α 1(III) chain is only eight amino acid residues long, but in the pN α 1(I) chain it is 17 amino acid residues long. These segments are very probably of a flexible nature, and positions 75–87 may provide a flexible link between the compact globular domain and the helix. Such a flexible link could be important for

Fig. 5. Comparison of the N-terminal extension peptides of the pNa1 chain of calf type III collagen [pNa1(III)] with those of the pNa1 chain of calf type I collagen [pNa1(I)] (taken from Hörlein *et al.*, 1979) and the pNa2 chain of chick type I collagen [pNa2(I)] (taken from Tate *et al.*, 1983)

The sequence of the pNa1(I) chain was used as reference sequence and the pNa1(III) and pNa2(I) chain sequences were aligned to it by using the ALIGN computer program of Orcutt *et al.* (1982). In order to optimize the alignment, the computer program introduced several gaps into each sequence. Thus the numbering in this Figure is not comparable with that of Fig. 4 or that used previously (Hörlein *et al.*, 1979). A mutation data matrix (Orcutt & Dayhoff, 1982) was used with a bias of 6 and a break penalty of 6 with 100 random runs. The ALIGN scores were as follows: pNa1(I) chain with pNa1(III) chain, 16.35; pNa1(I) chain with pNa2(I) chain, 13.75; pNa1(III) chain with pNa2(I) chain, 8.63. Identical residues are enclosed in boxes. The one-letter notation for amino acid residues (IUPAC-IUB Commission on Biochemical Nomenclature, 1969) is used; <Q denotes pyroglutamic acid.

allowing the positioning of the extension peptides on the surface of most, if not all, type III collagen fibrils in tissues (Fleischmajer *et al.*, 1981). It may also accommodate the physiological removal of the extension peptides, since a fragment similar to peptide Col 1 is the main form excreted into the urine (Rohde *et al.*, 1979a, 1983b).

The extremely high sequence homology between the first and the last cysteine residues suggests that the disulphide bond conformation and tertiary structure of the type I and type III procollagen globular domains are practically identical. This may explain why no differences in the biological activity of the N-terminal globular domains of type I and type III procollagens, with regard to their inhibitory effect on collagen synthesis, could be found (Wiestner *et al.*, 1979). However, distinct differences are found in their antigenic properties, which to a large extent depend on an intact conformation (Rohde *et al.*, 1976, 1983a).

We thank Miss Gudrun Landrath, Miss Sabina Rannenber and Mr. Wolfgang Straßhofer for their excellent technical assistance. This research was supported in part by grants from the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 51, B/22.

References

- Becker, U., Helle, O. & Timpl, R. (1977) *FEBS Lett.* **73**, 197-200
- Bruckner, P., Bächinger, H. P., Timpl, R. & Engel, J. (1978) *Eur. J. Biochem.* **90**, 595-603
- Byers, P. H., McKenney, K. H., Lichtenstein, J. R. & Martin, G. R. (1974) *Biochemistry* **13**, 5243-5248
- Curran, S. & Prockop, D. (1982) *Biochemistry* **21**, 1482-1487
- Engel, J., Bruckner, P., Becker, U., Timpl, R. & Rutschmann, B. (1977) *Biochemistry* **16**, 4026-4033
- Fietzek, P. P. & Rexrodt, F. W. (1975) *Eur. J. Biochem.* **59**, 113-118
- Fleischmajer, R., Timpl, R., Tuderman, L., Raisher, L., Wiestner, M., Perlsh, J. S. & Graves, P. N. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7360-7364
- Glanville, R. W. (1982) in *Immunochemistry of the Extracellular Matrix*, vol. 1, (Furthmayr, H., ed.), pp. 43-53, CRC Press, Boca Raton
- Glanville, R. W. & Fietzek, P. P. (1976) *FEBS Lett.* **71**, 99-102
- Glanville, R. W. & Rauter, A. (1981) *Hoppe-Seyler's Z. Physiol. Chem.* **362**, 943-951
- Gollwitzer, R., Timpl, R., Becker, U. & Furthmayr, H. (1972) *Eur. J. Biochem.* **28**, 497-506
- Gollwitzer, R., Becker, U. & Timpl, R. (1974) *FEBS Lett.* **47**, 177-180

- Henkel, W. & Glanville, R. W. (1982) *Eur. J. Biochem.* **122**, 205–213
- Hörlein, D., Fietzek, P. P., Wachter, E., Lapiere, C. M. & Kühn, K. (1979) *Eur. J. Biochem.* **99**, 31–38
- Inglis, A. S. (1983) *Methods Enzymol.* **91**, 324–332
- IUPAC-IUB Commission on Biochemical Nomenclature (1969) *Biochem. J.* **113**, 1–4
- Klapper, D. G., Wilder, C. E. & Capra, J. D. (1978) *Anal. Biochem.* **85**, 126–131
- Krieg, T., Hörlein, D., Wiestner, M. & Müller, P. K. (1978) *Arch. Dermatol. Res.* **263**, 171–180
- Lenaers, A. & Lapiere, C. M. (1975) *Biochim. Biophys. Acta* **400**, 121–131
- Morikawa, T., Tuderman, L. & Prockop, D. J. (1980) *Biochemistry* **19**, 2646–2650
- Nowack, H., Olsen, B. R. & Timpl, R. (1976) *Eur. J. Biochem.* **70**, 205–216
- Nusgens, B. V. & Lapiere, C. M. (1979) *Anal. Biochem.* **95**, 406–412
- Nusgens, B. V., Goebels, Y., Shinkai, H. & Lapiere, C. M. (1980) *Biochem. J.* **191**, 699–706
- Orcutt, B. C. & Dayhoff, M. D. (1982) *Scoring Matrices (N.B.R. Report 820541-08710)*, National Biomedical Research Foundation, Washington
- Orcutt, B. C., Dayhoff, M. D. & Barker, W. C. (1982) *ALIGN Score Program (N.B.R. Report 820501-08710)*, National Biomedical Research Foundation, Washington
- Paglia, L., Wilczek, J., Diaz de Leon, L., Martin, G. R., Hörlein, D. & Müller, P. (1979) *Biochemistry* **18**, 5030–5034
- Paglia, L. M., Wiestner, M., Duchene, M., Ouellette, L. A., Hörlein, D., Martin, G. R. & Müller, P. K. (1981) *Biochemistry* **20**, 3523–3527
- Pesciotta, D., Silkowitz, M., Fietzek, P., Graves, P., Berg, R. & Olsen, B. (1980) *Biochemistry* **19**, 2447–2454
- Rohde, H., Becker, U., Nowack, H. & Timpl, R. (1976) *Immunochemistry* **13**, 967–974
- Rohde, H., Vargas, L., Hahn, E., Kalbfleisch, H., Bruguera, M. & Timpl, R. (1979a) *Eur. J. Clin. Invest.* **9**, 451–459
- Rohde, H., Wachter, E., Richter, W. J., Bruckner, P., Helle, O. & Timpl, R. (1979b) *Biochem. J.* **179**, 631–642
- Rohde, H., Bruckner, P. & Timpl, R. (1983a) *Eur. J. Biochem.* **135**, 197–202
- Rohde, H., Langer, I., Krieg, T. & Timpl, R. (1983b) *Collagen Relat. Res. Clin. Exp.* **3**, 377–379
- Shinkai, H. & Lapiere, C. M. (1983) *Biochim. Biophys. Acta* **758**, 30–36
- Tate, V. E., Finer, M. H., Boedtker, H. & Doty, P. (1983) *Nucleic Acids Res.* **11**, 91–104
- Timpl, R., Glanville, R. W., Nowack, H., Wiedemann, H., Fietzek, P. P. & Kühn, K. (1975) *Hoppe-Seyler's Z. Physiol. Chem.* **356**, 1783–1792
- Tuderman, L., Kivirikko, K. I. & Prockop, D. J. (1978) *Biochemistry* **17**, 2948–2954
- Wiestner, M., Krieg, T., Hörlein, D., Glanville, R. W., Fietzek, P. P. & Müller, P. K. (1979) *J. Biol. Chem.* **254**, 7016–7023