Amino Acid Sequence of the *N*-Terminal Non-collagenous Segment of Dermatosparactic Sheep Procollagen Type I

By HEILWIG ROHDE,* ELMAR WACHTER,† WILHELM J. RICHTER,‡ PETER BRUCKNER,*
ODDVAR HELLE§ and RUPERT TIMPL*

*Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, West Germany, †Institut für Physiologische Chemie, Physikalische Biochemie und Zellbiologie der Universität München, München, West Germany, ‡Ciba-Geigy, Zentrale Funktion Forschung, Basel, Switzerland, and §The Veterinarian College of Norway, Oslo, Norway

(Received 18 September 1978)

The non-collagenous *N*-terminal segment of type I procollagen from dermatosparactic sheep skin was isolated in the form of the peptide Col 1 from a collagenase digest of the protein. The peptide has a blocked *N*-terminus, which was identified as pyrrolid-2-one-5-carboxylic acid. Appropriate overlapping fragments were prepared from reduced and alkylated peptide Col 1 by cleavage with trypsin at lysine, arginine and *S*-aminoethyl-cysteine residues and by cleavage with staphylococcal proteinase at glutamate residues. Amino acid sequence analysis of these fragments by Edman degradation and mass spectrometry established the whole sequence of peptide Col 1 except for a peptide junction (7–8) and a single Asx residue (44), and demonstrated that peptide Col 1 consists of 98 amino acid residues. The *N*-terminal portion of peptide Col 1 (86 residues) shows an irregular distribution of glycine, whereas the *C*-terminal portion (12 residues) possesses the triplet structure Gly-X-Y and is apparently derived from the precursor-specific collagenous domain of procollagen. The central region of the peptide contains ten cysteine residues located between positions 18 and 73 and shows alternating polar and hydrophobic sequence elements. The regions adjacent to the cysteine-rich portion have a hydrophilic nature and are abundant in glutamic acid. The data are consistent with previous physicochemical and immunological evidence that distinct regions at the *N*- and *C*-termini of the non-collagenous domain possess a less rigid conformation than does the central portion of the molecule.

The various collagens in the body are synthesized in the form of precursors, the procollagens. These procollagens possess extra peptide segments connected to both the *N*- and *C*-termini of the constituent polypeptide chains (Fessler & Fessler, 1978; Prockop *et al.*, 1976). Truncated forms of procollagens with an apparently intact *N*-terminal segment but lacking the *C*-terminal segment accumulate in the skin of dermatosparactic animals (Lenears *et al.*, 1971; Becker *et al.*, 1976; Furthmayr *et al.*, 1972) and in foetal skin (Timpl *et al.*, 1975; Lenears & Lapiere, 1975). These procollagens have been convenient sources in previous studies to characterize chemical, physicochemical and immunochemical properties of the *N*-terminal segments (Becker *et al.*, 1976, 1977; Engel *et al.*, 1977; Nowack *et al.*, 1976; Rohde *et al.*, 1976; Rohde, 1977; Bruckner *et al.*, 1978; Hörlein *et al.*, 1978; Hörlein, 1978). The data showed that the precursor-specific regions at the *N*-terminus of type I and III procollagens consists of a rigid non-collagenous domain with a molecular weight of about 10000 and a collagenous triple-helical domain with a molecular weight of about 5000. Here we

report on the amino acid sequence of the non-collagenous domain from dermatosparactic procollagen type I. In the following paper (Rohde & Timpl, 1979) the various fragments described in the study were used to identify sequence-dependent antigenic determinants of the peptide.

Materials and Methods

Purification and chemical modification of procollagen peptide

The procollagen peptide Col 1, which is derived from the *N*-terminus of $\alpha 1(I)$ chain, was isolated from a collagenase digest of dermatosparactic sheep skin by chromatography on Sephadex G-50 and DEAE-cellulose (Becker *et al.*, 1976). Peptide samples were reduced with 0.02M-dithioerythritol in the presence of 8M-urea and subsequently alkylated with either an excess of sodium iodoacetate (Gollwitzer *et al.*, 1972) or with ethyleneimine (Gollwitzer *et al.*, 1974). Reagents were removed from modified peptides on columns of Bio-Gel P-2 equilibrated in 0.1M-formic acid.

Proteolytic degradation

Native or reduced and alkylated peptides were dissolved in 0.2M-ammonium bicarbonate and treated with 1-chloro-4-phenyl-3-L-toluene-*p*-sulphonamidobutan-2-one-treated trypsin (Worthington, Freehold, NJ, U.S.A.) for 16–24h at 37°C at an enzyme/substrate ratio of 1:100. In the case of aminoethylated peptide T3-4 the sample was dissolved in 0.2M-Tris/HCl (pH 8.0)/0.02M-CaCl₂ (10mg/ml) and after addition of 0.4mg of 1-chloro-4-phenyl-3-L-toluene-*p*-sulphonamidobutan-2-one-treated trypsin incubated for 8h at 50°C. Then the same amount of enzyme was added again and incubation continued for 16h. Digestions with staphylococcal proteinase V8 (Miles Laboratories, Slough, Berks., U.K.) were carried out in 0.05M-ammonium bicarbonate, pH 7.9, at an enzyme/substrate ratio of 1:100 for 16h at 37°C (Houmard & Drapeau, 1972). A similar procedure was used for treatment with Pronase B (Calbiochem), Digestion with thermolysin (Merck A.G., Darmstadt, Germany) has been described previously (Rauterberg *et al.*, 1972).

Chromatographic methods

Fragments prepared from peptide Col 1 were separated on columns of Bio-Gel P-10 (1.5cm × 108cm) or Bio-Gel P-4 (1.5cm × 120cm), both equilibrated in 0.2M-ammonium bicarbonate, pH 8.6. Salt was removed from the various peptide fractions by repeated freeze-drying. Chromatography on DEAE-cellulose was performed on columns (1–1.5cm × 10–15cm) equilibrated in 0.005M-Tris/HCl (pH 8.6)/8M-urea. Peptides were eluted from the columns by linear NaCl gradients and desalted on Bio-Gel P-2 (Becker *et al.*, 1976). Chromatography on Dowex 50 X2 (0.5cm × 10cm) followed the procedure of Press *et al.* (1966). Under these conditions peptides lacking basic groups pass through the column, whereas peptides bound to the column were eluted with aq. 1% NH₃.

Amino acid analysis

Peptides were hydrolysed with 6M-HCl under N₂ (20h, 110°C) and analysed on a Durrum D-500 amino acid analyser. Values for threonine and serine were corrected for loss during hydrolysis by factors 1.08 and 1.21 respectively (Rauterberg & Kühn, 1971). Tryptophan was determined by a spectrophotometric procedure (Edelhoch, 1967). Cysteine residues were quantified on the amino acid analyser as their respective *S*-carboxymethylated or *S*-aminoethylated derivatives.

Amino acid sequence analysis

Manual Edman degradation of peptides (100–300nmol) was carried out by a standard procedure

(Peterson *et al.*, 1972). Phenylthiohydantoin derivatives were identified by t.l.c. (Edman, 1970) or by high-pressure liquid chromatography (Zimmermann *et al.*, 1977). For solid-phase Edman degradation peptides (250–500nmol) were coupled to 3-aminopropyl-glass via acid amide bonds when these peptides lacked lysine and aminoethylcysteine residues. Peptides containing lysine or aminoethylcysteine residues located in the peptide chain or at the C-terminus were coupled to the *p*-phenylene diisothiocyanate derivative of 3-aminopropyl-glass via thiourea bonds. The amount of peptide covalently coupled to the glass was determined by amino acid analysis of a suitable sample after hydrolysis with 5.7M-HCl. Phenylthiohydantoin derivatives released by Edman degradation were identified by g.l.c. with a flame-photometric sulphur-sensitive detector and by quantitative multiple-ion mass spectrometry. The repetitive yield with the various peptides was in the range 94–96% (Wachter *et al.*, 1973; Machleidt & Wachter, 1977).

Treatment of peptides with carboxypeptidase A and/or B (Worthington), carboxypeptidase C (Roehm and Haas, Philadelphia, PA, U.S.A.) or carboxypeptidase Y (Boehringer, Tutzing, Germany) followed previously described procedures (Becker *et al.*, 1975; Fietzek & Kühn, 1975; Hayashi, 1976). Amino acids released by this treatment were quantified on the Durrum D-500 analyser. Hydrazinolysis of peptides was carried out for 16h at 80°C by the method of Braun & Schroeder (1967).

Mass spectrometry

Field-desorption and electron-impact mass spectra of the *N*-terminal peptides were obtained with a Varian MAT CH5 DF mass spectrometer equipped with a combined FD/FI/EI ion source. Tungsten-wire emitters (10μm) were activated at high temperatures in a commercially available Varian apparatus, and charged with solutions of samples in dimethyl sulphoxide by using Beckey's emitter-dipping technique. Trimethylsilylation for the electron-impact experiments was carried out by treating oligopeptide samples with *NO*-bis(trimethylsilyl)acetamide/pyridine (1:1, v/v) at 60°C for 2h.

Results

Tryptic fragments of reduced and S-carboxymethylated peptide Col 1

Peptide Col 1 with a size of about 100 amino acids contains three residues of arginine and four residues of lysine (Table 1). A tryptic digest of reduced peptide Col 1 could be resolved into several peaks on a Bio-Gel P-10 column (Fig. 1a). The first peak contained two peptides, T3-4 and T5, which could be

Table 1. Amino acid composition of peptide Col 1 and of its tryptic fragments (T) obtained from reduced and S-carboxymethylated peptide Col 1

Results are expressed as residues per peptide. A dash (—) denotes less than 0.2 residue. Numbers in parentheses refer to the amino acid numbers determined by sequence analysis.

	Col 1	T1	T2	T3-4	T5	T6	Sum of peptides T1-T6 (from sequence)
Cystine (half)	9.5 (10)	1.3 (1)	—	6.4 (6)	3.1 (3)	—	10
Aspartic acid	15.4 (15)	2.0 (2)	1.2 (1)	8.4 (9)	2.5 (2)	1.2 (1)	15
Threonine	6.9 (7)	1.3 (1)	—	—	4.9 (5)	1.0 (1)	7
Serine	1.7 (1)	—	—	—	1.2 (1)	—	1
Glutamic acid	20.1 (19)	9.9 (10)	—	2.6 (2)	7.2 (7)	—	19
Proline	10.2 (10)	2.2 (2)	—	2.8 (3)	3.6 (4)	1.0 (1)	10
Glycine	8.7 (9)	3.0 (3)	—	1.4 (1)	2.9 (3)	1.8 (2)	9
Alanine	1.4 (1)	—	—	1.1 (1)	—	—	1
Valine	9.8 (10)	2.2 (2)	—	4.8 (5)*	3.0 (3)	—	10
Isoleucine	2.7 (3)	0.9 (1)	—	1.8 (2)*	—	—	3
Leucine	3.4 (3)	0.9 (1)	—	1.8 (2)	—	—	3
Tyrosine	1.1 (1)	—	0.9 (1)	—	—	—	1
Histidine	1.1 (1)	—	1.0 (1)	—	—	—	1
Lysine	3.7 (4)	—	—	2.7 (3)	1.0 (1)	—	4
Arginine	2.7 (3)	1.0 (1)	1.0 (1)	—	—	1.0 (1)	3
Tryptophan	0.7 (1)	—	—	0.8 (1)	—	—	1
Total	100 (98)	24 (24)	4 (4)	35 (35)	30 (29)	6 (6)	98
Sequence position	1-98	1-24	25-28	29-63	64-92	93-98	
N-Terminal residue	<Glu	<Glu	Tyr	Asp	Val	Gly	

* Value after 72 h hydrolysis.

Fig. 1. Purification of tryptic peptides prepared from reduced and S-carboxymethylated peptide Col 1 (a) Chromatography of the whole digest on a column of Bio-Gel P-10 (1.5 cm × 108 cm) equilibrated in 0.2 M-ammonium bicarbonate, pH 8.5. (b) Separation of fragments T5 and T3-4 on a DEAE-cellulose column (1 cm × 13 cm) equilibrated in 5 mM-Tris/HCl (pH 8.6)/8 M-urea and eluted with a linear gradient (200 ml/200 ml) from 0 to 0.3 M-NaCl. Numbers in parentheses in (a) refer to the relative amounts of peptide material (in percentages) recovered in the different pools of the column effluent.

separated on DEAE-cellulose (Fig. 1*b*). Peptide T3-4 emerged from DEAE-cellulose in two peaks, which, however, contained identical material as judged from amino acid composition and sequence analysis. A third peptide, T1, appeared in the shoulder of the first peak on Bio-Gel P-10 (Fig. 1*a*) and was found to be homogeneous after rechromatography on DEAE-cellulose. Two further peptides, T2 and T6, could be obtained in sufficient purity already after chromatography on Bio-Gel P-10.

The amino acid compositions of these five tryptic peptides when taken together (Table 1) account for the amino acid composition of peptide Col 1, including all the arginine and lysine residues. Each peptide could be obtained from Bio-Gel P-10 in approximately stoichiometric amounts (Fig. 1).

A peptide lacking arginine and lysine residues was not found in the digest. Reduced peptide Col 1 was therefore treated with carboxypeptidase C. A 24 h digest contained exclusively free arginine, which was released in about 30% yield. As shown below, this arginine originates from peptide T6, and a proline residue in the penultimate position of peptide T6 explains the low recovery.

Fragments produced by tryptic cleavage at S-aminoethylcysteine residues

Because of the difficulties met in sequence analysis (see below) and specific needs in the immunological

studies (Rohde & Timpl, 1979), further fragments were prepared from the tryptic peptides of peptide Col 1. An initial treatment of non-reduced peptide Col 1 with trypsin produced the same four cleavages as found for reduced peptide Col 1. Here, however, the three large tryptic peptides are held together by five disulphide bridges in a single core peptide that could be separated from the small peptides T2 and T6 on Bio-Gel P-10. Reduction and *S*-aminoethylation of the core peptide allowed the separation of the constituent fragments T1, T3-4 and T5 on DEAE-cellulose (Fig. 2*a*). Each of these peptides was then digested a second time with trypsin. This treatment produced a single cleavage in peptide T1 and in peptide T5. The fragments could be separated on Bio-Gel P-4 or P-10 columns. Amino acid and sequence analysis of these fragments (Table 2 and Fig. 5) clearly showed that the order of the fragments is T1a-T1b in peptide T1 and T5b-T5a in peptide T5 (Fig. 3).

Tryptic digestion of *S*-aminoethylated peptide T3-4 produced several fragments that could be separated by chromatography on Bio-Gel P-4 (Fig. 2*b*). One cleavage involved the lysine-aspartic acid bond joining the fragments T3 and T4. Apparently, cleavage occurred at a bond that was much more resistant to trypsin in the *S*-carboxymethylated variant of peptide T3-4. Three further cleavages at the *C*-terminus of *S*-aminoethylcysteine were observed in the T3 portion of the peptide, giving rise to

Fig. 2. Separation of tryptic peptides prepared from reduced and *S*-aminoethylated peptide Col 1
 (a) Chromatography on DEAE-cellulose of the reduced and alkylated chain constituents of the core peptide obtained from native peptide Col 1 after tryptic digestion. The chromatographic conditions were the same as in Fig. 1(b). (b) Separation of the fragments obtained by a second tryptic cleavage of peptide T3-4 on a Bio-Gel P-4 column (1.5 cm $\times$ 120 cm) equilibrated in 0.2 M ammonium bicarbonate, pH 8.6.

Fig. 3. Order of proteolytic fragments of peptide Col 1

The peptides were produced by cleaving S-carboxymethylated peptide Col 1 with trypsin (T) or with staphylococcal proteinase (SP). Further cleavages of the tryptic peptides with staphylococcal proteinase or of S-aminoethylated tryptic peptides with trypsin are shown in the third and fourth rows respectively. The last row shows selected fragments, produced by treatment with Pronase (Pr) or thermolysin (Tl). Numbers in parentheses refer to the numbers of amino acid residues in individual fragments.

the fragments T3a, T3b, T3c and T3d and in addition to an uncleaved fragment T3a-b (Table 2). The order of these fragments could be determined by sequence analysis as T3c-T3d-T3b-T3a-T4 (Figs. 3 and 5).

Further fragments were prepared from *S*-aminoethylated peptides T3a-b or T3-4 by digestion with either Pronase or thermolysin respectively. Each proteinase produced a complex mixture of fragments that could be partially resolved by chromatography on Bio-Gel P-4 and DEAE-cellulose (results not shown). Two fragments were found useful for sequence analysis and characterized by amino acid analysis (Table 2). Pronase liberated a tripeptide T3Pr that included the single glycine residue of peptide T3-4. The 17-residue fragment T3T1 was obtained after thermolysin treatment and apparently comprised the *C*-terminal portion of peptide T3-4.

Order of the tryptic peptides

The order of the tryptic peptides in the Col 1 segment (Fig. 3) was determined by characterization of fragments produced from *S*-carboxymethylated peptide Col 1 by digestion with staphylococcal proteinase. The identification of fragments overlapping the tryptic peptides was facilitated by initially producing fragments from tryptic peptides T1, T3-4 and T5 with staphylococcal proteinase. These fragments were purified on Bio-Gel P-4 and on DEAE-cellulose (results not shown) and characterized by amino acid analysis (Table 3). Cleavage of peptide T1 yielded two peptides, of which the smaller one (T1SP2) with the composition Gly₁Glx₆ had a blocked *N*-terminus, since it did not bind to a Dowex 50 X2 column. A second treatment of peptide T1SP2 with staphylococcal proteinase at a higher enzyme/substrate ratio (1:20; 24 h at 37°C) produced in about 60% yield two fragments that could be separated on Bio-Gel P-4. One fragment (T1SP2a) had the composition Gly₁Glx₃. As judged by its behaviour on Dowex 50 X2, the other fragment, T1SP2b, was derived from the *N*-terminus of the molecule and consisted entirely of glutamic acid. Digestion of peptide T3-4 produced two fragments, whereas cleavage of peptide T5 produced five fragments. The sum of the amino acid compositions of the various staphylococcal-proteinase fragments resembled the composition of the original tryptic peptides (Table 3). The order of these fragments is shown in Fig. 3.

Overlapping peptides were obtained by cleaving non-reduced peptide Col 1 with staphylococcal proteinase. Chromatography of the digest on Bio-Gel P-10 (Fig. 4) separated a large disulphide-bonded peptide (SP1,2) from several smaller peptides. Amino acid analysis of the smaller peptides showed that only a single peptide, Col 1SP3, did not resemble a fragment characterized before in the digests of the

Table 2. Amino acid composition of tryptic fragments produced from reduced and *S*-aminoethylated peptide Col 1
Data on peptides T2 and T6 are not included, since they are identical with those given in Table 1. Results are expressed as residues per peptide. A dash (—) denotes less than 0.2 residue. Numbers in parentheses refer to the amino acid numbers determined by sequence analysis.

	T1a	T1b	T3a-b	T3a	T3b	T3c	T3d	T3T1	T3Pr	T4	T5a	T5b
Cystine (half)	1.1 (1)	—	2.3 (3)	1.0 (1)	1.8 (2)	0.9 (1)	1.0 (1)	2.7 (3)	—	1.0 (1)	1.7 (2)	0.7 (1)
Aspartic acid	1.0 (1)	1.0 (1)	6.2 (6)	3.1 (3)	3.0 (3)	1.3 (1)	—	5.0 (5)	1.1	2.0 (2)	1.4 (1)	0.9 (1)
Threonine	1.2 (1)	—	—	—	—	—	—	—	—	—	4.0 (4)	1.0 (1)
Serine	—	—	—	—	—	—	—	—	—	—	1.2 (1)	0.3
Glutamic acid	8.4 (9)	1.2 (1)	1.3 (1)	1.1 (1)	0.2	—	1.0 (1)	1.5 (1)	0.2	—	6.4 (6)	1.2 (1)
Proline	1.8 (2)	—	—	—	0.4	—	—	1.2 (1)	—	1.1 (1)	3.3 (3)	1.0 (1)
Glycine	2.1 (2)	1.0 (1)	1.2 (1)	0.2	0.8 (1)	—	—	0.6	0.7	—	3.1 (3)	—
Alanine	—	—	—	—	—	—	—	0.6 (1)	—	0.9 (1)	—	—
Valine	1.0 (1)	0.9 (1)	2.6 (3)	0.7 (1)	1.4 (2)	1.9 (2)	—	0.6 (1)	0.7	—	2.1 (2)	0.8 (1)
Isoleucine	1.0 (1)	—	1.0 (1)	0.6 (1)	0.4 (1)	—	—	0.4 (1)	—	—	—	—
Leucine	—	0.9 (1)	2.0 (2)	0.9 (1)	0.8 (1)	—	0.9 (1)	2.0 (2)	0.2	—	—	—
Lysine	—	—	1.0 (1)	0.6 (1)	0.2	0.7 (1)	—	1.7 (2)	—	0.9 (1)	—	—
Arginine	—	0.9 (1)	—	—	—	—	—	—	—	—	—	—
Tryptophan	—	—	—	—	—	—	—	—	—	—	—	—
Total	17 (18)	6 (6)	17 (18)	9 (9)	8 (9)	8 (8)	3 (3)	17 (17)	3	6 (6)	23 (23)	6 (6)
Sequence position	1-18	19-24	40-57	49-57	40-48	29-36	37-49	47-63	44-46	58-63	70-92	64-69
<i>N</i> -Terminal residue	<Glu	Val	Val	Asp	Va	Asp	Gln	Leu	Asx	Asp	Cys	Val

Table 3. Amino acid composition of fragments produced from Col 1 and from tryptic peptides T1, T3-4 and T5 by treatment with staphylococcal proteinase
Results are expressed as residues per peptide. A dash (—) denotes less than 0.2 residue. Numbers in parentheses refer to the amino acid numbers determined by sequence analysis.

	T1SP1	T1SP2	T3-4SP1	T3-4SP2	T5SP1	T5SP2	T5SP3	T5SP4	T5SP5	Col 1SP1	Col 1SP2	Col 1SP3
Cystine (half)	0.8 (1)	—	5.0 (5)	1.1 (1)	3.0 (3)	—	—	—	—	6.1 (6)	0.8 (1)	—
Aspartic acid	2.2 (2)	—	6.8 (7)	1.8 (2)	—	0.7 (1)	1.0 (1)	—	—	10.7 (10)	3.0 (3)	0.9 (1)
Threonine	1.0 (1)	—	—	—	0.2	1.0 (1)	1.9 (2)	1.9 (2)	—	0.9 (1)	1.1 (1)	2.6 (3)
Serine	—	—	—	—	—	0.2	1.0 (1)	—	—	—	—	—
Glutamic acid	4.3 (4)	6.0 (6)	2.4 (2)	—	1.4 (1)	1.1 (1)	1.9 (2)	1.2 (1)	2.1 (2)	6.3 (6)	1.0 (1)	1.3 (1)
Proline	1.8 (2)	—	2.2 (2)	0.9 (1)	1.6 (2)	0.4 (1)	0.3	1.1 (1)	—	4.0 (4)	2.1 (2)	1.8 (2)
Glycine	2.0 (2)	1.0 (1)	1.3 (1)	—	0.2	—	—	2.2 (2)	0.9 (1)	2.8 (3)	0.3	4.0 (4)
Alanine	—	—	—	1.1 (1)	—	—	—	—	—	—	1.0 (1)	—
Valine	1.9 (2)	—	4.8 (5)	—	0.9 (1)	0.6 (1)	—	0.9 (1)	—	6.1 (7)	0.9 (1)	1.0 (1)
Isoleucine	1.0 (1)	—	1.5 (2)	—	—	—	—	—	—	2.3 (3)	—	—
Leucine	1.0 (1)	—	1.0 (1)	0.9 (1)	—	—	—	—	—	1.8 (2)	0.9 (1)	—
Tyrosine	—	—	—	—	—	—	—	—	—	1.0 (1)	—	—
Histidine	—	—	—	—	—	—	—	—	—	1.0 (1)	—	—
Lysine	—	—	1.0 (1)	1.8 (2)	—	—	—	1.0 (1)	—	1.0 (1)	1.8 (2)	0.8 (1)
Arginine	1.0 (1)	—	—	—	—	—	—	—	—	1.8 (2)	—	1.0 (1)
Tryptophan	—	—	—	—	—	—	—	—	—	—	—	—
Total	17 (17)	7 (7)	27 (27)	8 (8)	7 (7)	5 (5)	6 (6)	8 (8)	3 (3)	47 (48)	13 (13)	14 (14)
Sequence position	8-24	1-7	29-55	56-63	69-75	64-68	79-84	85-92	76-78	8-55	56-68	85-98
N-Terminal residue	Gly	<Glu	Asp	Leu	Cys	Val	Ser	Thr	Gly	Gly	Leu	Thr

Fig. 4. Separation of fragments produced from native peptide Col 1 by digestion with staphylococcal proteinase on Bio-Gel P-10. The column was the same as in Fig. 1(a). Pool SP1,2 denotes a disulphide-linked core peptide in which two constituent chains (SP1 and SP2) represent fragments overlapping tryptic peptides.

tryptic peptides and demonstrated that it was composed of tryptic peptide T6 and the C-terminal portion of peptide T5, i.e. peptide T5SP4 (Table 3). Reduction of the disulphide-bonded peptide released three single-chain constituents that could be separated on Bio-Gel P-10. The smallest fragment resembled in composition fragment T5SP1. The other two fragments, Col 1SP1 and Col 1SP2, contained sufficient arginine and/or lysine to resemble overlapping peptides (Table 3). Amino acid analysis of fragment Col 1SP1 demonstrated that it was composed of the peptide T1SP1 from the C-terminus of peptide T1, of the tryptic peptide T2 containing the single histidine and tyrosine residues of peptide Col 1 and of the N-terminal fragment T3-4SP1. The other overlapping segment Col 1SP2 comprised the fragments T3-4SP2 and T5SP2. Thus the results show that the tryptic peptides of peptide Col 1 occur in the order T1-T2-T3-T4-T5-T6 (Fig. 3) and are in agreement with the sequence data.

Amino acid sequence analysis

Even though peptide Col 1 has a blocked N-terminus, a large portion of the sequence (positions 8-43 and 56-98) could be unequivocally determined by solid-phase Edman degradation of the overlapping fragments Col 1SP1, T3-4, Col 1SP2, T5, Col 1SP3 and T6 (Fig. 5). Each of the peptides except Col 1SP1

proved to be essentially homogeneous by sequence analysis. The major part of fragment Col 1SP1 (70%) showed, as determined by quantitative mass spectrometry of the phenylthiohydantoin derivatives, the sequence given in Fig. 5. It was apparently contaminated by a similar fragment that lacked the first four amino acid residues of peptide Col 1SP1, indicating a second cleavage at the Glu-Asp bond (11-12) by staphylococcal proteinase. The N-terminal sequence of peptide Col 1 (positions 4-7) could be in part elucidated by manual Edman degradation of peptide T1SP2a. The rest of the sequence (positions 1-3) was determined by mass spectrometry of peptide T1SP2b (see below).

A second problem in the sequence analysis arose by the inefficient yields in Edman degradation of peptides T3-4 or T3b beyond position 43 (numbering of peptide Col 1). Only traces of aspartic acid could be detected at position 44, which were presumably due to an overlap from positions 42 and 43. The amino acids in the following sequence (positions 45-57), starting with the single glycine residue of peptide T3-4, could be again identified, but the yield of phenylthiohydantoin derivatives was less than 10% of the yield found at position 43. The sequence until position 56, where fragment Col 1SP2 begins, could be in part confirmed by manual Edman degradation of fragment T3a (positions 49-57). Kinetic analysis of amino acids released from fragment T3a by

	$\langle$ Glu-Glu-Glu-Gly-Gln-Gln-Glu-Gly-Gln-Glu-Asp-Ile-Pro-Pro-Val-Thr-Cys-Val-Gln- 5 10 15 20									
T1SP2a:	→	→	→	→	→	→	→	→	→	→
SP1:	→									
	Asp-Gly-Leu-Arg-Tyr-His-Asp-Arg-Asp-Val-Trp-Lys-Pro-Val-Pro-Cys-Gln-Ile-Cys-Val- 25 30 35 40									
SP1:	→	→	→	→	→	→	→	→	→	→
T3-4:	→									
T3b:	→									
	Cys-Asp-Asp-Asx-Gly-Val-Leu-Cys-Asp-Asp-Val-Ile-Cys-Asp-Glu-Leu-Lys-Asp-Cys-Pro- 45 50 55 60									
T3-4:	→	→	→	→	→	→	→	→	→	→
T3b:	→	→	→	→	→	→	→	→	→	→
T3Pr:	→	→	→	→	→	→	→	→	→	→
T3T1:	→									
T3a:	→	→	→	→	→	→	→	→	→	→
SP2:	→									
	Asn-Ala-Lys-Val-Pro-Thr-Asp-Glu-Cys-Cys-Pro-Val-Cys-Pro-Glu-Gly-Glu-Glu-Ser-Thr- 65 70 75 80									
SP2:	→	→	→	→	→	→	→	→	→	→
T5:	→	→	→	→	→	→	→	→	→	→
	Thr-Asp-Gln-Glu-Thr-Thr-Gly-Val-Glu-Gly-Pro-Lys-Gly-Asp-Thr-Gly-Pro-Arg-OH 85 90 95 98									
T5:	→	→	→	→	→	→	→	→	→	→
SP3:	→	→	→	→	→	→	→	→	→	→
T6:	→	→	→	→	→	→	→	→	→	→

Fig. 5. Amino acid sequence of peptide Col 1
The various amino acids were identified after automated solid-phase Edman degradation (→), manual Edman degradation (→), carboxypeptidase treatment (→) or hydrazinolysis (↔). Mass-spectrometric analysis of peptides was used to determine positions 1-3 and to confirm positions 4-7.

various carboxypeptidases (Fig. 6) showed the sequence Asp-Glu-Leu-Lys and provided the required overlap to fragment Col 1SP2. The sequence of positions 46–48 was determined by digesting fragment T3b with carboxypeptidases A and B, which released after 60 min about 60% of the aminoethylcysteine and leucine and about 10% of the valine. Evidence that fragments T3b and T3a are adjacent to each other was provided by Edman degradation of the overlapping thermolysin peptide T3Tl (positions 47–63). In order to identify the amino acid residue in position 44 we treated the tripeptide T3Pr with hydrazine, which released 44% of the valine but only 8% of the glycine. Digestion with carboxypeptidases A and B liberated 36% of the valine and 6% of the glycine but no aspartic acid or asparagine. A subsequent hydrazinolysis of the digest increased the yield of free glycine to 33%, indicating that the peptide has the sequence Asx-Gly-Val (positions 44–46).

The sequence of the *N*-terminal amino acids (invariably of the Glx type) was established by mass-spectrometric techniques rather than by Edman

Fig. 6. Kinetics of the release of free amino acids from peptide T3a by the sequential action of carboxypeptidase B (CPB), carboxypeptidase A (CPA) and carboxypeptidase Y (CPY).

degradation. Field-desorption mass spectrometry, known to be highly valuable in the analysis of unaltered thermally labile oligopeptides (Beckey & Schulten, 1975; Winkler & Beckey, 1972; Asante-Poku *et al.*, 1975), demonstrated a molecular weight *M* of 387 for peptide T1SP2b (MH^+ and MNa^+ quasimolecular ions at m/z 388 and 410 respectively). This molecular weight corresponds to a tripeptide with the sequence <Glu-Glu-Glu (<Glu, pyrrolid-2-one-5-carboxylic acid) and clearly indicates the absence of any glutamine residues. The tripeptide was then subjected to trimethylsilylation to obtain direct sequence information by conventional electron-impact mass spectrometry. The main derivative formed was a tetratrimethylsilyl compound (molecular ion M^{++} at m/z 675). Pronounced *N*-terminal sequence ions were observed at m/z 156, 357 and 385, thus corroborating the sequence <Glu-Glu-Glu (Fig. 7).

Evidence that the tetrapeptide T1SP2a (Gly-Gln-Gln-Glu) follows the tripeptide T1SP2b in the sequence of peptide Col 1 was provided by field-desorption mass spectrometry of the heptapeptide T1SP2. The results indicated a molecular weight (nominal mass) *M* of 829 (MH^+ and MNa^+ ions at m/z 830 and 852 respectively), which agrees with a composition of <Glu₁Gly₁Gln₂Glu₃.

Discussion

Sequence analysis and fragment patterns show that the *N*-terminal segment Col 1 of dermatosparactic sheep procollagen consists of 98 amino acid residues. This is in good agreement with a molecular weight of 10400 determined for peptide Col 1 by ultracentrifugation (Engel *et al.*, 1977). More than 90% of the sequence of peptide Col 1 could be established by using standard Edman degradation on appropriate overlapping peptide fragments. Two difficulties were encountered in sequence analysis, and one was due to a blocked *N*-terminus in peptide Col 1. The *N*-terminal sequence was elucidated by mass spectrometry as <Glu-Glu-Glu. Preliminary studies

Fig. 7. Formation of major fragments from the trimethylsilylated tripeptide T1SP2b by electron-impact mass spectrometry. Me_3Si denotes trimethylsilyl substituents.

on a similar fragment from dermatosparactic calf procollagen (C. Schwabe & D. Hörlein, personal communication) have also shown that the *N*-terminus of peptide Col 1 is occupied by pyrrolid-2-one-5-carboxylic acid and that in the following Glx-rich region glycine residues occur in positions 4 and 8. These data support our findings that fragments T1SP1, T1SP2a and T1SP2b comprise the whole *N*-terminal region of Col 1 (Figs. 3 and 5), and that no residues are missing between residues 7 and 8.

A further part in the sequence that could not be elucidated by Edman degradation is located in position 44. Fragment patterns and *C*-terminal analyses indicate that the amino acid in this position is either aspartic acid or asparagine. It is possible that carbohydrate chains linked to the Asx residue decreased the yield in further Edman degradation in spite of a low carbohydrate content reported previously for the Col 1 peptide (Furthmayr *et al.*, 1972). As an alternative explanation a cyclic peptide bond between asparagine and glycine, such as is known in other proteins (Bornstein, 1970; Bornstein & Balian, 1970), may block Edman degradation. Usually these cyclic peptide bonds can be cleaved with hydroxylamine. This treatment, however, consistently failed to produce fragments from both peptides T3-4 and T3a-b (H. Rohde & R. Timpl, unpublished work).

Previous studies have demonstrated that the Col 1 segment comprises only two-thirds of the whole precursor-specific region at the *N*-terminus of procollagen $\alpha 1(1)$ chain (Becker *et al.*, 1976; Rohde, 1977; Hörlein, 1978). The main portion of the remaining region apparently consists of a collagen-like sequence that is rich in glycine and hydroxyproline. This segment should have the triplet structure Gly-X-Y, since it was possible to renature the whole precursor-specific peptide to a triple-helical molecule (Engel *et al.*, 1977). Our present findings confirm previous reports (Becker *et al.*, 1976; Nowack *et al.*, 1976) that the non-collagenous portion of peptide Col 1 (positions 1–86) is located at the *N*-terminus of the precursor-specific region and that a part of the collagenous sequence (positions 87–98) is located at the *C*-terminus of peptide Col 1. Apparently bacterial collagenase is not able to release every triplet sequence Gly-X-Y in the procollagen peptides (Nowack *et al.*, 1976; Bruckner *et al.*, 1978; Hörlein *et al.*, 1978).

The ten cysteine residues, which all are involved in intrachain disulphide bridges (Furthmayr *et al.*, 1972; Becker *et al.*, 1976), are located in the central region (positions 18–73) of the non-collagenous domain of peptide Col 1. This region is particularly rich in aspartic acid and shows alternating polar and hydrophobic segments. The proline residues of peptide Col 1 are clustered around the first two and last four cysteine residues. Physicochemical studies (Engel *et al.*, 1977) have indicated that the central portion of peptide Col 1 has a rigid structure stabilized

by disulphide bonds and hydrophobic interactions and consists of β -structure and aperiodic structural elements. The terminal sequences of peptide Col 1 can be removed by proteinases without disturbing this rigid structure. The sequence data agree with these observations, showing that most of the hydrophobic amino acids are located in the disulphide-bonded region of peptide Col 1, whereas the terminal segments contain almost entirely hydrophilic amino acid residues. These data are also compatible with the assumption of a non-rigid conformation at both ends of the Col 1 peptide, and the change from the rigid to the more flexible portions of the molecule may be marked by proline residues. Immunochemical evidence on antigenic determinants that are shared by the native disulphide-bonded and the reduced unfolded forms of the peptide (Rohde *et al.*, 1976) supports this interpretation. As shown in the accompanying paper these antigenic determinants could be located to the terminal peptides T1SP2 and T6 respectively (Rohde & Timpl, 1979).

Large portions of the sequence of the homologous precursor-specific region in dermatosparactic calf procollagen have been elucidated (Hörlein, 1978; Hörlein *et al.*, 1978). As far as the established sequences of both Col 1 peptides allow a comparison (positions 8–43 and 56–98), a highly invariant structure is found. The only amino acid substitutions observed are located in positions 77 (Glu $\rightarrow$ Gln) and 80 (Thr $\rightarrow$ Pro), i.e. in the non-rigid portion of the molecule. The high degree of sequence homology is reflected by a complete immunological cross-reaction between the sheep and calf procollagen peptides (Rohde *et al.*, 1976).

More sequence studies on *N*-terminal procollagen peptides from less closely related animal species such as the chick will be required to provide information on the possible biological importance of certain regions of the molecule on the basis of an apparent invariance in amino acid sequence. So far, previous suggestions on the role of *N*-terminal procollagen peptides as providing a triple-helical nucleus or registration site during biosynthesis of collagen (Prockop *et al.*, 1976; Fessler & Fessler, 1978; Becker *et al.*, 1976) could not be supported by experimental evidence (Engel *et al.*, 1977; Bruckner *et al.*, 1978). Other suggestions with regard to a feedback inhibition of collagen synthesis (Lichtenstein *et al.*, 1973) have been made (Krieg *et al.*, 1978) and they may provide a better understanding of the functions of the precursor-specific segments in procollagen.

We are grateful for the skilled technical assistance of Mrs. Vera van Delden and Mr. Paul Birrer. Support for these studies was obtained from the Deutsche Forschungsgemeinschaft (project Ti 95/3) and Sonderforschungsbereich 51, 3–15.

References

- Asante-Poku, S., Wood, G. W. & Schmidt, D. E. (1975) *Biomed. Mass Spectrom.* **2**, 121–125
- Becker, U., Furthmayr, H. & Timpl, R. (1975) *Hoppe-Seyler's Z. Physiol. Chem.* **356**, 21–32
- Becker, U., Timpl, R., Helle, O. & Prockop, D. J. (1976) *Biochemistry* **15**, 2853–2862
- Becker, U., Helle, O. & Timpl, R. (1977) *FEBS Lett.* **73**, 197–200
- Beckey, H. D. & Schulten, H. R. (1975) *Angew. Chem.* **87**, 425–438
- Bornstein, P. (1970) *Biochemistry* **9**, 2408–2421
- Bornstein, P. & Balian, B. (1970) *J. Biol. Chem.* **245**, 4854–4856
- Braun, V. & Schroeder, W. A. (1967) *Arch. Biochem. Biophys.* **118**, 241–252
- Bruckner, P., Bächinger, H. P., Timpl, R. & Engel, J. (1978) *Eur. J. Biochem.* **90**, 595–603
- Edelhoch, H. (1967) *Biochemistry* **6**, 1948–1954
- Edman, P. (1970) in *Protein Sequence Determination* (Needleman, S. B., ed.), pp. 211–255, Springer-Verlag, New York
- Engel, J., Bruckner, P., Becker, U., Timpl, R. & Rutschmann, B. (1977) *Biochemistry* **16**, 4026–4033
- Fessler, J., & Fessler, L. I. (1978) *Annu. Rev. Biochem.* **47**, 129–162
- Fietzek, P. P. & Kühn, K. (1975) *Eur. J. Biochem.* **52**, 77–82
- Furthmayr, H., Timpl, R., Stark, M., Lapiere, C. M. & Kühn, K. (1972) *FEBS Lett.* **28**, 247–250
- 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
- Hayashi, R. (1976) *Methods Enzymol.* **45**, 568–587
- Hörlein, D. (1978) Doctoral Thesis, University of Munich
- Hörlein, D., Fietzek, P. P. & Kühn, K. (1978) *FEBS Lett.* **89**, 279–282
- Houmard, J. & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 3506–3509
- Krieg, T., Hörlein, D., Wiestner, M. & Müller, P. K. (1978) *Arch. Dermatol. Forsch.* **263**, 171–180
- Lenears, A. & Lapiere, C. M. (1975) *Biochim. Biophys. Acta* **400**, 121–131
- Lenears, A., Ansay, M., Nusgens, B. V. & Lapiere, C. M. (1971) *Eur. J. Biochem.* **23**, 533–543
- Lichtenstein, J. R., Martin, G. R., Kohn, L. D., Byers, P. H. & McKusick, V. A. (1973) *Science* **182**, 298–300
- Machleidt, W. & Wachter, E. (1977) *Methods Enzymol.* **47**, 263–277
- Nowack, H., Olsen, B. R. & Timpl, R. (1976) *Eur. J. Biochem.* **70**, 205–216
- Peterson, J. D., Nehrllich, S., Oyer, P. E. & Steiner, D. F. (1972) *J. Biol. Chem.* **247**, 4866–4871
- Press, E. M., Piggot, P. J. & Porter, R. R. (1966) *Biochem. J.* **99**, 356–366
- Prockop, D. J., Berg, R. A., Kivirikko, K. I. & Uitto, J. (1976) in *Biochemistry of Collagen* (Ramachandran, G. N. & Reddi, A. H., eds.), pp. 163–273, Plenum Press, New York
- Rauterberg, J. & Kühn, K. (1971) *Eur. J. Biochem.* **19**, 398–407
- Rauterberg, J., Timpl, R. & Furthmayr, H. (1972) *Eur. J. Biochem.* **27**, 231–237
- Rohde, H. (1977) Doctoral Thesis, University of Munich
- Rohde, H. & Timpl, R. (1979) *Biochem. J.* **179**, 643–647
- Rohde, H., Becker, U., Nowack, H. & Timpl, R. (1976) *Immunochemistry* **13**, 967–974
- Timpl, R., Glanville, R. W., Nowack, H., Wiedemann, H., Fietzek, P. P. & Kühn, K. (1975) *Hoppe-Seyler's Z. Physiol. Chem.* **356**, 1783–1792
- Wachter, E., Machleidt, W., Hofner, H. & Otto, J. (1973) *FEBS Lett.* **35**, 97–102
- Winkler, H. U. & Beckey, H. D. (1972) *Biochem. Biophys. Res. Commun.* **46**, 391–398
- Zimmermann, C. L., Appella, E. & Pisano, J. J. (1977) *Anal. Biochem.* **77**, 569–573