DISCONTINUITIES IN THE TRIPLE HELICAL SEQUENCE GLY–X–Y OF BASEMENT MEMBRANE (TYPE IV) COLLAGEN

Detlef SCHUPPAN, Rupert TIMPL and Robert W. GLANVILLE
Max-Planck-Institut für Biochemie, D-8033 Martinsried, FRG

Received 12 May 1980

1. Introduction

Type IV collagens are major constituents of basement membranes and possess a fairly rigid triple helical conformation exceeding in length those found in interstitial collagens [1,2]. This triple helix makes interstitial collagens highly resistant to proteolytic cleavage except for the release of short non-helical segments from both ends of the α chains (M_r 95 000) [3]. In contrast, a variety of fragments ranging between M_r 27 000 and 140 000 could be produced by pepsin cleavage of type IV collagen from kidney glomeruli [4,5], placenta [6–9], lens capsule [10,11] and a tumor matrix [12]. This data suggested that the triplet sequence Gly–X–Y, which is considered to be essential for a stable triple helical conformation [13], is frequently interrupted.

This possibility was examined by sequence analysis of the pepsin fragment P1 (M_r 55 000) obtained from type IV collagen extracted from a mouse tumor [12]. P1 is a fragment of a longer chain (α (IV) or C chain) obtained from a variety of sources [5,6,9–11]. Previous studies have established that the first 8 residues at the N-terminus of peptide P1 do not resemble the Gly–X–Y type of sequence [12]. Here, results obtained for various fragments of P1 demonstrate that the triple helix of type IV collagen is interrupted by longer stretches of non-helical sequences as well as by deletions of single glycine residues. The data explain the sensitivity of type IV collagen to pepsin and other non-collagenolytic proteases which produce a number of well-defined, relatively stable intermediates, the size of which depends upon the conditions of the cleavage reaction [12,14].

2. Experimental

The pepsin fragment P1 of type IV collagen was obtained from a mouse tumor and purified as in [12]. This fragment was treated with mercaptoethanol [15] prior to cleavage with cyanogen bromide (CNBr) in 70% formic acid [16]. Further treatments of CNBr peptides included digestion with thermolysin [17] for 4 h at 37°C. Peptides were purified by chromatography on phosphocellulose [16], on Sephadex G-75s (column 2 × 140 cm) equilibrated in 0.03 M sodium acetate (pH 4.8) at 40°C, and on columns of Bio-Gel P-4 and P-2 equilibrated in 0.2 M ammonium bicarbonate (pH 8.5). For amino acid analysis on a Durrum D500 analyzer purified peptides were hydrolyzed with 6 M HCl (24 h, 110°C) under N_2.

Sequence analysis of peptides (50–200 nmol) was carried out by automated Edman degradation in a Beckman sequencer, model 890C [14,18]. Phenylthiohydantoin derivatives released during sequencing were identified by high-pressure liquid chromatography (HPLC) (Spectra Physics, model SP8000) using an isocratic elution system or by thin-layer chromatography (TLC).

3. Results

Pepsin fragment P1 (~540 amino acid residues, table 1) which comprises a major portion of the mouse type IV collagen α1 (IV) chain [12] was subjected to CNBr cleavage and the digest separated on Sephadex G-75 (fig.1). This allowed the partial separation and further purification of 9 individual peptides (CB1–CB9) varying in size from 3–140 amino acid residues [20]. The methionine content of P1 would indicate

Table 1
Amino acid composition of pepsin fragment P1 and its CNBr peptides CB5 and CB7 (residues/peptide)

	P1	CB5	CB7
3-Hyp	2.7	–	–
4-Hyp	68	8.3	1.8 (2)[b]
Asp	28	3.1	0.9 (1)
Thr	12	1.0	0.9 (1)
Ser	29	1.0	1.4 (2)
Glu	48	6.1	–
Pro	37	5.0	0.9 (1)
Gly	174	15.0	3.4 (3)
Ala	13	1.0	–
Val	19	1.4	0.7 (1)
Met	5.9	1.4[a]	0.3[a]
Ile	12	0.7	0.3
Leu	23	2.1	–
Tyr	1.0	–	–
Phe	14	1.1	0.5 (1)
His	4.6	–	0.6 (1)
Hyl	33	1.8	–
Lys	3.2	–	–
Arg	7.2	–	–
Total	535	48	12 (13)

[a] As homoserine. A dash denotes <0.2 residues
[b] Values determined by sequence analysis

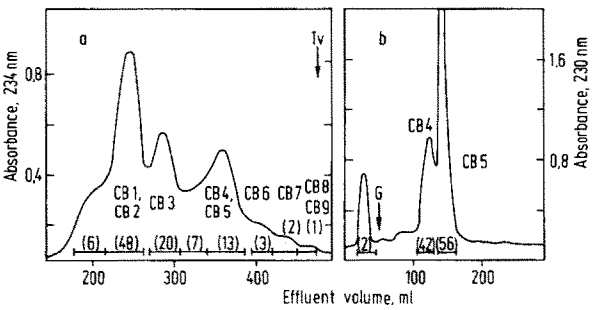

Fig.1. Fractionation of a CNBr digest of the pepsin fragment P1 on Sephadex G-75 (a) and separation of peptides CB4 and CB5 on phosphocellulose (b). The Sephadex G-75 column (2 × 140 cm) was equilibrated in 0.03 M sodium acetate (pH 4.8) at 40°C. The phosphocellulose column (1.5 × 10 cm) was equilibrated at 40°C in 0.001 M sodium acetate (pH 3.6) and eluted with a linear gradient (150/150 ml) from 0–0.4 M NaCl. Numbers in brackets refer to relative amounts of peptide material (in %) recovered in the particular peak. Peaks without designation contained incompletely cleaved CNBr peptides. Tv in (a) denotes the total column volume and G in (b) the start of the gradient.

the existence of 7 CNBr peptides but part of the methionine in basement membrane proteins is usually oxidized [21] and requires reduction [15] for efficient cleavage with CNBr [9,20]. Two of the CNBr peptides of P1 were subjected to more extensive sequence analysis. These included peptide CB5 which was separated from peptide CB4 on phosphocellulose (fig.1). Amino acid analysis (table 1) as well as molecular sieve chromatography showed that the peptide consists of ~48 amino acid residues including one homoserine. The second peptide CB7 contained 13 amino acid residues and was purified by chromatography on Bio-Gel P-4 and phosphocellulose. Since CB7 lacks homoserine it originates from the C-terminus of fragment P1.

Sequence analysis (fig.2) of P1 demonstrated as shown before [12], that the first 8 amino acid residues lack glycine in every third position but the continuing sequence was of the Gly–X–Y type. The positions of ~30 amino acid residues were determined by Edman degradation of peptide CB5 and of fragments obtained by thermolysine digestion which mainly cleaves at position 18–19 of CB5. This sequence began with six triplets Gly–X–Y, was interrupted by Val–Glu and then continued with the Gly–X–Y structure. A similar type of sequence has also been found for CNBr peptide CB4 (not shown).

The C-terminal peptide CB7 started, according to Edman degradation, with two Gly–X–Y segments followed by 5 residues of non-triplet nature. The entire sequence was completed by degradation of CB7 with carboxypeptidase Y (fig.3). The data showed

```
                          5                        10
P1   His-Val-Asp-Met-Gly-Ser-Met-Gly-Gly-Gln-
                                  15
     Gly-Gly-Asp-Gln-Gly-Glu-Gly-Gly-

                          5   +                    10
CB7  Gly-Pro-Hyp-Gly-Thr-Hyp-Ser-Val-Asp-His-
     Gly-Ser-Phe-OH

                                  15           +   10
CB5  Gly-Pro-Hyp-Gly-Pro-Gln-Gly-Gln-Hyp-Gly-
                                  15              20
     Leu-Hyp-Gly-Thr-Hyp-Gly-XXX-Pro-Val-Glu-
                                  25              30
     Gly-Pro-Pro-Gly-Asp-Gln-Gly-Pro-Gln-Gly-
```

Fig.2. Amino acid sequences of pepsin fragment P1 and its CNBr peptides CB5 and CB7. Stretches which lack the triplet structure Gly–X–Y are underlined. XXX denotes an unidentified residue; *partially hydroxylated proline

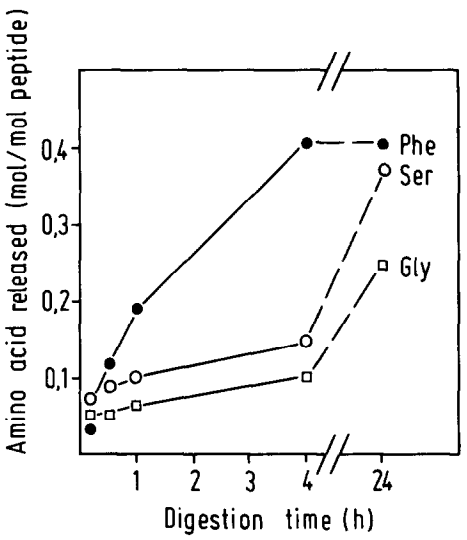

Fig.3. Kinetics of carboxypeptidase Y digestion of CNBr peptide CB7.

a rapid release of 0.4 residues of Phe in agreement with the low content of this amino acid in the peptide (table 1). This was followed by the liberation of serine and glycine indicating that CB7 has the C-terminal sequence Gly–Ser–Phe. Part of the peptide, however, may lack the ultimate amino acid residue indicating that pepsin has released fragment P1 from type IV collagen by either cleavage at the N- or C-terminal site of phenylalanine. Data similar to that for CB7 were obtained for P1 by kinetic analysis of carboxypeptidase Y digestion.

4. Discussion

The data demonstrated two types of discontinuities in the triplet sequence Gly–X–Y of the pepsin fragment P1 from the mouse α1 (IV) chain. These included two longer stretches of 7–8 amino acid residues located at both the N- and C-terminus of P1. A small interruption resembling a single deletion of glycine was found in the central portion of CNBr peptide CB5. Similar data was obtained for another pepsin fragment P2 ([12], unpublished) which presumably is a fragment of the mouse α2 (IV) chain.

The non-triplet sequences found at both ends of P1 resemble the type of structure found in the terminal, non-helical sequences of the α chains of interstitial collagens [3]. The sequence of the C-terminal CNBr peptide CB7 of P1 was completely elucidated. This peptide resembles in composition and size other C-terminal CNBr peptides for bovine lens capsule C′ and C fragments [10]. These chains are longer than P1 and may comprise in the case of C′ (M_r 140 000), the whole length of the triple helical domain of the α1 (IV) chain. Similarly, the triple helix of the mouse α1 (IV) chain [12] could be recovered as fragments P3A (M_r 72 000) and P1 (M_r 55 000). Taken together the data indicate that P1 is derived from the carboxyl end of α1 (IV) chain while C and C′ include the P1 segment but extend to the N-terminal region (fig.4). The non-helical sequence of peptide CB7 therefore resembles the C-terminal non-helical end of the triple helix in interstitial collagens and may perhaps continue into a procollagen-like structure [2,3]. The non-helical part at the N-terminus of P1, however, indicates a clear interruption of the triple helix in type IV collagen. Such an interruption explains the high sensitivity of this molecule to proteolytic attack resulting in a limited number of distinct fragments [4–11]. Similar sequence discontinuities have also been identified in human placenta type IV collagen [14].

The short discontinuity of the triple helical sequence as found for CNBr peptide CB5 resembles that observed in the collagenous segment of the complement component C1q [22]. Here, the A chain has a threonine inserted between two Gly–X–Y segments and the B chain has one glycine in the triple helical sequence substituted by alanine. These discontinuities are considered to be responsible for a bend in the triple helical strands of C1q as shown by electron microscopy [23,24]. It is still an open question as to

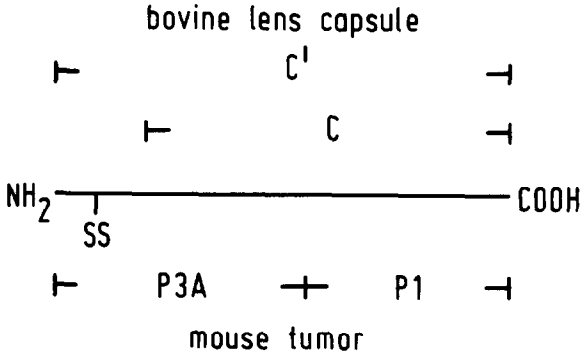

Fig.4. Scheme of the triple helical domain in α1 (IV) chain and tentative order of pepsin fragments obtained from bovine lens capsule [10] and mouse EHS sarcoma ([12], this report). SS denotes interchain disulfide bridges.

how an interrupted Gly–X–Y sequence affects the triple helical conformation and shape of type IV collagen. Available physical data [1,2] do not indicate that this protein is less rigid or stable than interstitial collagens which lack these sequence interruptions [3]. However, preliminary observations such as the biphasic nature of the melting profile of triple helical P1 [12] may reflect helix discontinuity. These kinds of sequences are also responsible for the continuous degradation of type IV collagen upon prolonged exposure to pepsin [12,14].

Acknowledgements

This study was supported by the excellent help of Mrs Liselotte Sieweck and Mrs Vera van Delden and by the financial aid of the Deutsche Forschungsgemeinschaft (projects Ti 95/4 and SFB 51 B36).

References

[1] Gelman, R. A., Blackwell, J., Kefalides, N. A. and Tomichek, E. (1976) Biochim. Biophys. Acta 427, 492–496.
[2] Timpl, R., Martin, G. R., Bruckner, P., Wick, G. and Wiedemann, H. (1978) Eur. J. Biochem. 84, 43–52.
[3] Fietzek, P. P. and Kühn, K. (1976) Int. Rev. Connec. Tiss. Res. 7, 1–60.
[4] Daniels, J. R. and Chu, G. H. (1975) J. Biol. Chem. 250, 3531–3537.
[5] Dixit, S. N. (1979) FEBS Lett. 106, 379–384.
[6] Glanville, R. W., Rauter, A. and Fietzek, P. P. (1979) Eur. J. Biochem. 95, 383–389.
[7] Bailey, A. J., Sims, T. J., Duance, V. C. and Light, N. D. (1979) FEBS Lett. 99, 361–366.
[8] Sage, H., Woodbury, R. G. and Bornstein, P. (1979) J. Biol. Chem. 254, 9893–9900.
[9] Kresina, T. F. and Miller, E. J. (1979) Biochemistry 18, 3089–3097.
[10] Dixit, S. N. and Kang, A. H. (1979) Biochemistry 18, 5686–5692.
[11] Gay, S. and Miller, E. J. (1979) Arch. Biochem. Biophys. 198, 370–378.
[12] Timpl, R., Bruckner, P. and Fietzek, P. (1979) Eur. J. Biochem. 95, 255–263.
[13] Traub, W. and Piez, K. A. (1971) Adv. Protein Chem. 25, 243–352.
[14] Rauter, A. and Glanville, R. W. (1980) submitted.
[15] Adelstein, R. S. and Kuehl, W. M. (1970) Biochemistry 9, 1355–1364.
[16] Bornstein, P. and Piez, K. A. (1966) Biochemistry 5, 3460–3473.
[17] Rauterberg, J., Timpl, R. and Furthmayr, H. (1972) Eur. J. Biochem. 27, 231–237.
[18] Fietzek, P. P. and Rexrodt, F. W. (1975) Eur. J. Biochem. 59, 113–118.
[19] Hayashi, R. (1976) Methods Enzymol. 45, 568–587.
[20] Schuppan, D. (1979) Diploma Thesis, University of Munich.
[21] Hudson, B. G. and Spiro, R. G. (1972) J. Biol. Chem. 247, 4229–4238.
[22] Reid, K. B. M. (1979) Biochem. J. 179, 367–371.
[23] Knobel, H. R., Villiger, W. and Isliker, H. (1975) Eur. J. Immunol. 5, 78–82.
[24] Reid, K. B. M. (1977) Biochem. J. 161, 247–251.