Biochem. J. (1984) **220**, 227–233
Printed in Great Britain

Sequence comparison of pepsin-resistant segments of basement-membrane collagen α1(IV) chains from bovine lens capsule and mouse tumour

Detlef SCHUPPAN,* Robert W. GLANVILLE,* Rupert TIMPL,* Saryu N. DIXIT† and Andrew H. KANG‡
**Max-Planck-Institut für Biochemie, D-8033 Martinsried, Federal Republic of Germany, †Lakeside Veterans Administration Medical Center and Department of Oral Biology, Northwestern University, Chicago, IL 60611, U.S.A., and ‡Department of Medicine and Department of Biochemistry, University of Tennessee Center for Health Sciences and Veterans Administration Medical Center, Memphis, TN 38104, U.S.A.*

(*Received 25 November 1983/Accepted 9 February 1984*)

The *C*-terminal peptic fragment P1 (about 518 amino acid residues) of bovine lens-capsule collagen α1(IV) chain was cleaved with CNBr and trypsin. The peptides were purified and characterized, allowing their ordering within the P1 fragment by comparison with a corresponding section of mouse collagen α1(IV) chain [Schuppan, Glanville & Timpl (1982) *Eur. J. Biochem.* **123**, 505–512]. About 67% of the sequence of bovine collagen fragment P1 was determined by Edman degradation. Comparison with the sequence of the corresponding mouse collagen fragment P1 showed 76% identity for positions Xaa and Yaa of the triplet structures Gly-Xaa-Yaa. Invariance was found for the positions of two non-triplet interruptions and of 3-hydroxyproline residues, pointing to the functional importance of these structures.

Collagen type IV is unique to basement membranes and consists of two distinct constituent chains, α1(IV) and α2(IV), with M_r values of about 180000. These chains are folded into two triple-helical segments of about 330 nm and 60 nm length and into two non-helical domains (Kühn *et al.*, 1981; Timpl *et al.*, 1981). The major triple helix contains several short interruptions of the triplet sequence Gly-Xaa-Yaa (Schuppan *et al.*, 1980; Glanville & Rauter, 1981), introducing several flexible regions into the triple-helical strand (Hofmann *et al.*, 1984). Some of these interruptions are readily cleaved by pepsin, giving rise to a variable number of chain fragments. Major fragments obtained from the α1(IV) chain of bovine lens-capsule collagen, for example, have M_r values of about 140000, 110000 and 50000 (Dixit, 1978; Dehm & Kefalides, 1978; Gay & Miller, 1979; Dixit & Kang, 1979). The smallest fragment corresponds in position and size to fragment P1 from the α1(IV) chain of a mouse tumour basement membrane collagen (Timpl *et al.*, 1979), for which about 70% of the sequence has been determined (Schuppan *et al.*, 1982). The CNBr-cleavage peptide pattern localized fragment P1 to the *C*-terminal portion of the major triple helix (Dixit & Kang, 1979; Schuppan *et al.*, 1980; Weber *et al.*, 1984).

The interstitial fibril-forming collagens types I, II and III have no interruptions in their triplet sequence, and show, in the Xaa and Yaa positions, a high interspecies identity of about 95% (Fietzek & Kühn, 1976; Bornstein & Traub, 1979). Similar comparative sequence data are not yet available for collagens type IV. In the present study we have determined more than half of the sequence of fragment P1 from bovine lens-capsule collagen type IV and have compared the data with the corresponding section of mouse collagen type IV (Schuppan *et al.*, 1982).

Materials and methods

Purification of fragment P1 from bovine lens-capsule collagen IV and cleavage with CNBr and proteinases

Fragment P1 was isolated as a 50000-M_r component from a peptic digest of bovine lens-capsule collagen type IV (Dixit & Kang, 1979) and was further purified on CM-cellulose (Timpl *et al.*, 1979). It then appeared to be homogeneous by dodecyl sulphate/polyacrylamide-gel electrophoresis with a mobility identical with that of fragment P1 from a mouse tumour collagen type IV (Timpl *et al.*, 1979). Samples used in sequence analysis were further purified by high-pressure liquid chromatography (see below). Fragment P1 was reduced and cleaved with CNBr as previously described (Schuppan *et al.*, 1982). For cleavage with trypsin (Worthington Biochemical Corp.) (enzyme/substrate ratio 1:100,

w/w) peptides were dissolved in $0.2\,M$-NH_4HCO_3, pH 7.9, and incubated for 4h at 30°C. Digestion was stopped by addition of 7-amino-1-chloro-3-tosylamidoheptan-2-one ('TLCK'). Treatment with thermolysin (Merck) at an enzyme/substrate ratio of 1:100 (w/w) followed a previous protocol (Schuppan *et al.*, 1982).

Chromatographic methods

Digests of fragment P1 were initially separated on a column (2.2cm x 140cm) of Sephadex G-75 (superfine grade) equilibrated in $0.2\,M$-NH_4HCO_3, pH 7.9. Peptide material of separated peaks was then purified by chromatography on phosphocellulose (Schuppan *et al.*, 1982). CNBr-cleavage peptides CB3, CB4, CB5 and CB10b and tryptic peptides T4', T5 and T1aC were purified by using high-pressure liquid chromatography (model SP 8000; Spectra Physics) on a reversed phase Vydac TP RP-18 column. The column was equilibrated at 55°C with 0.1% trifluoracetic acid and eluted with a linear gradient of 0–40% (v/v) acetonitrile in the equilibration buffer.

Analytical methods and sequence analysis

Amino acid compositions were determined on a Durrum D-500 analyser from peptide samples hydrolysed in 6M-HCl containing 0.1% (v/v) 2-mercaptoethanol (24h, 110°C). M_r values were estimated by molecular-sieve chromatography (Sephadex G-75), with the use of CNBr-cleavage peptides of type I collagen for calibration, or were calculated from compositional data assuming a single residue of homoserine, lysine or arginine per peptide.

Edman degradation of peptides (20–150nmol) was performed in a Beckman sequencer (Glanville & Rauter, 1981). Residues released were identified by high-pressure liquid chromatography (Lottspeich, 1980).

Results

CNBr-cleavage and tryptic peptides of fragment P1

Fragment P1 was purified from a peptic digest of collagen type IV of bovine lens capsule (Dixit & Kang, 1979), and different samples of the fragment were cleaved with CNBr or trypsin. The size of the different peptides obtained and their order within fragment P1 is shown in Fig. 1. Since most of the peptides appeared related to peptides described for a mouse collagen type IV fragment P1 (Schuppan *et al.*, 1982), they were designated in accordance with the previous nomenclature.

CNBr cleavage produced nine peptides, which were initially separated on Sephadex G-75 (Fig. 2) and further purified by ion-exchange and reversed-phase chromatography. The peptides each showed a unique amino acid composition, which when taken together closely resembled the composition of fragment P1 (Table 1). Peptides CB2, CB3, CB4, CB5, CB7 and CB9 were identical within the limits of analytical error with similar peptides of mouse fragment P1 (Schuppan *et al.*, 1982). Peptides CB6-1, CB10a and CB10b were unique to the bovine fragment P1 owing to the substitution of a single methionine (position 368) and the addition of a new methionine residue (position 82) within the sequence. The bovine fragment P1 also lacked a tetrapeptide CB8, which is the *N*-terminal structure in the sequence of the mouse collagen type IV fragment P1.

Limited tryptic digestion of the bovine fragment

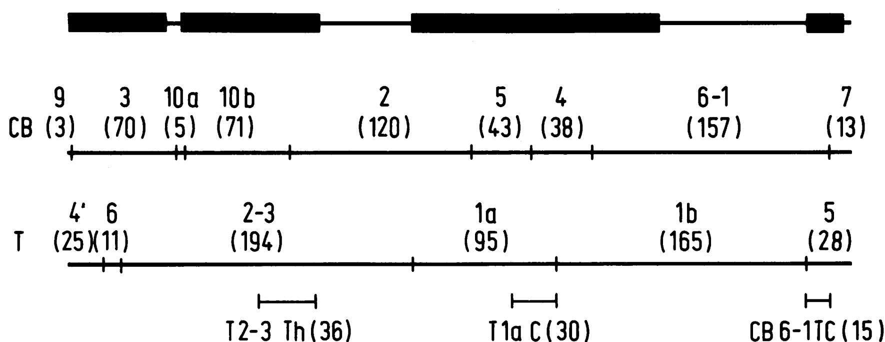

Fig. 1. *Size and order of CNBr-cleavage (CB) and tryptic (T) peptides of fragment P1 of bovine lens-capsule type IV collagen*
Numbers in parentheses refer to amino acid residues in individual peptides. The thick black bars indicate the segments for which the amino acid sequence was determined. Additional fragments used in sequencing were obtained by further digestion with trypsin or thermolysin (Th) and are shown underneath the tryptic peptides. Peptide CB10a was not aligned by sequencing.

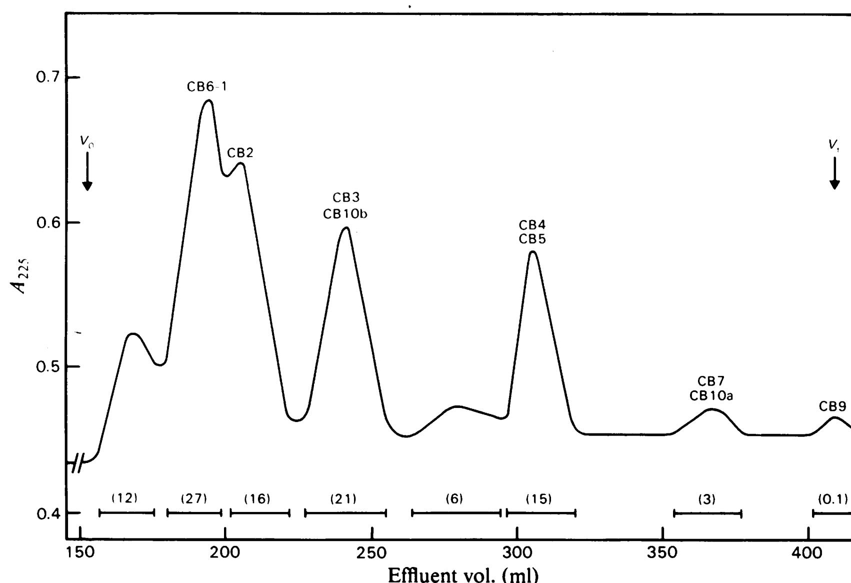

Fig. 2. *Chromatographic profile of CNBr-cleavage peptides of bovine fragment P1 on Sephadex G-75*
The column (2.2cm × 140cm) was equilibrated in 0.2M-NH_4HCO_3, pH 7.9. Horizontal bars indicate pools used for further purification of individual peptides. The relative amounts (in %) of peptide material in each pool are indicated in the parentheses. The void volume (V_0) and total volume (V_t) of the column are denoted by arrows.

Table 1. *Amino acid compositions of CNBr-cleavage peptides of fragment P1 of bovine lens-capsule collagen type IV*
Results are expressed as residues/peptide rounded off to the nearest integer, except for some hydroxylated residues. Numbers in parentheses refer to CNBr-cleavage peptides from a previous study (Dixit & Kang, 1979) that by composition resemble CNBr-cleavage peptides of fragment P1.

	Amino acid composition (residues/peptide)										
	CB6-1 (8)	CB2 (10)	CB3	CB4 (5)	CB5 (6)	CB7 (2)	CB9	CB10a (1)	CB10b (9)	Sum	Fragment P1
3-Hyp	1	–	–	–	0.5	–	–	–	–	1.5	1.5
4-Hyp	21	15	7	7	7	2	–	1	9	69	66
Asp	6	6	6	2	1	1	–	–	4	26	24
Thr	4	1	3	1	1	1	–	–	1	12	11
Ser	6	10	4	–	1	2	1	–	2	26	28
Glu	14	10	7	3	5	–	–	–	5	43	47
Pro	14	5	4	2	3.5	1	–	–	2	31.5	32.5
Gly	52	39	23	13	13	3	1	2	23	168	165
Ala	5	3	3	1	1	–	–	–	7	20	18
Val	5	3	2	–	1	1	–	–	1	13	14
Met*	1	1	1	1	1	–	1	1	1	8	5
Ile	3	3	1	–	–	–	–	–	5	12	14
Leu	8	6	1	3	2	–	–	1	3	24	25
Tyr	–	1	–	–	–	–	–	–	–	1	1
Phe	6	5	–	1	–	1	–	–	1	14	14
His	1	–	1	–	1	1	–	–	–	4	6
Hyl	9	9	5	2	1	–	–	–	6	32	32
Lys	–	1.5	1	–	–	–	–	–	–	3.5	4
Arg	1	1	1	1	2	–	–	–	1	7	8
Total	157	120	70	38†	41	13	3‡	5	71	518†	516

* As homoserine in the CB peptides.
† Including one tryptophan residue found by sequence analysis.
‡ Deduced from sequence.

P1 also produced a distinct peptide pattern (Fig. 3), which appeared complex owing to the partial cleavage at certain arginine and lysine residues. Most of the peptides were purified and showed a distinct amino acid composition (Table 2). Six peptides (T1a, T1b, T2-3, T4′, T5 and T6) were

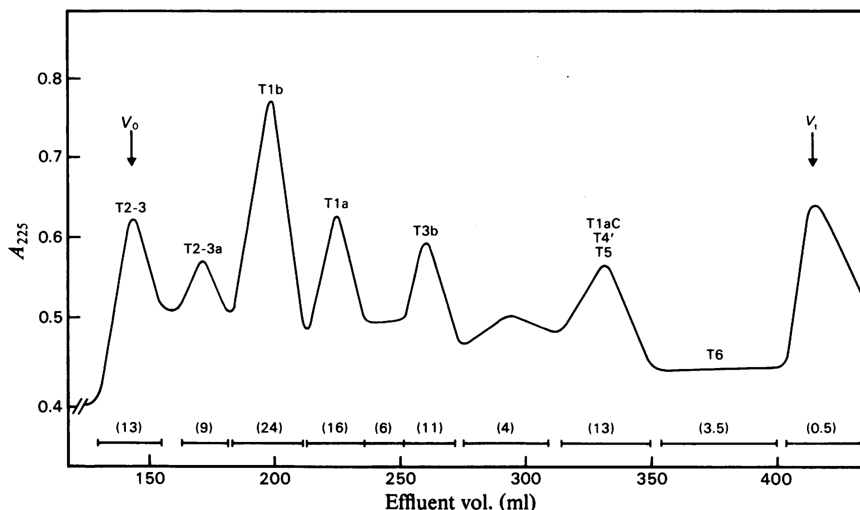

Fig. 3. *Chromatographic profile of tryptic peptides of bovine fragment P1 on Sephadex G-75*
The column (2.2cm x 140cm) was equilibrated in 0.2M-NH_4HCO_3, pH 7.9. Horizontal bars indicate pools used for further purification of individual peptides. The relative amounts (in %) of peptide material in each pool are indicated in the parentheses. The void volume (V_0) and total volume (V_t) of the column are denoted by arrows.

Table 2. *Amino acid compositions of tryptic peptides of fragment P1 of bovine lens-capsule collagen type IV*
The data include a minor cleavage product T1aC obtained from peptide T1a, a tryptic peptide CB6-1 TC from peptide CB6-1 and a thermolysin peptide T2-3Th prepared from peptide T2-3. Results are expressed as residues/peptide rounded off to the nearest integer, except for some hydroxylated residues.

	Amino acid composition (residues/peptide)									
	T1a	T1b	T2-3	T4′	T5	T6	Sum	T1aC	T2-3Th	CB6-1TC
3-Hyp	0.5	–	–	–	1	–	1.5	–	–	1
4-Hyp	13	24	27	–	5	3	72	4	4	3
Asp	4	7	8	2	2	1	24	1	1	1
Thr	1	4	3	2	1	1	12	–	1	–
Ser	5	6	12	3	2	–	28	–	4	1
Glu	9	15	15	4	–	–	43	4	3	–
Pro	4.5	13	10	1	3	–	31.5	3	2	1
Gly	32	54	62	8	8	4	168	10	11	5
Ala	1	6	9	–	–	–	16	–	1	–
Val	2	4	4	–	1	1	12	–	1	–
Met	2	1	3	1	1	–	8	1	1	1*
Ile	1	4	8	–	–	–	13	–	1	–
Leu	6	8	10	–	1	–	25	4	1	1
Tyr	–	–	1	–	–	–	1	–	–	–
Phe	4	6	2	–	2	–	14	–	1	1
His	1	1	1	–	1	–	4	–	–	–
Hyl	6	10	15	3	–	–	34	1	3	–
Lys	–	–	2	–	–	1	3	–	0.5	–
Arg	3	1	2	1	–	–	7	2	–	–
Total	95	165†	194	25‡	28	11	518†	30	36	15

* As homoserine.
† Including one tryptophan residue found by sequence analysis.
‡ Deduced in part from sequence.

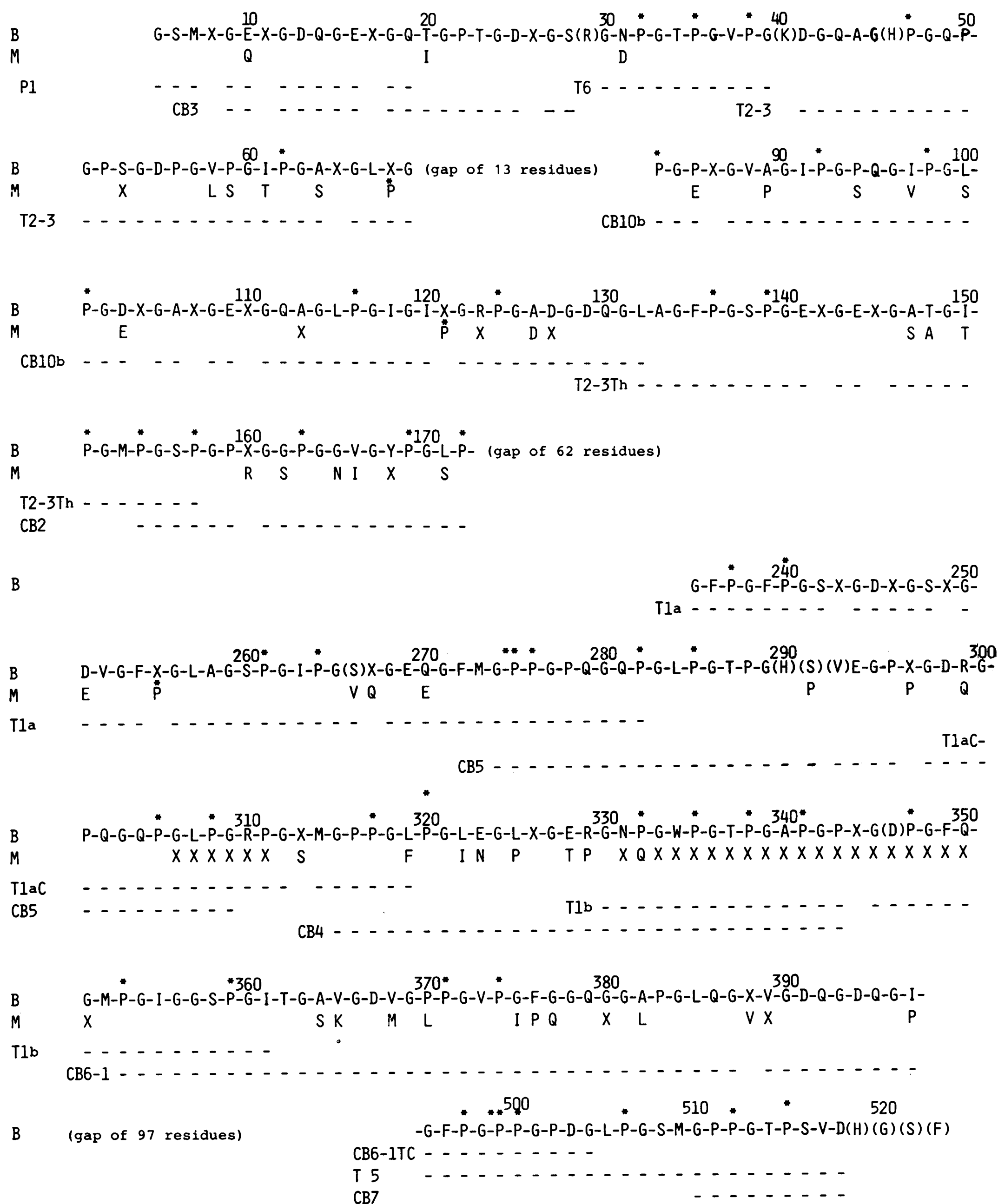

Fig. 4. *Partial amino acid sequence of fragment P1 of bovine type IV collagen α1(IV) chains (B)*
Sequences are expressed in the one-letter code (IUPAC–IUB Commission of Biochemical Nomenclature, 1969). One or two asterisks over P denote 4-hydroxyproline or 3-hydroxyproline respectively. The numbers above the sequence indicate the residue number in accordance with the tentative numbering for the mouse fragment P1 (Schuppan *et al.*, 1982). This numbering was corrected by subtracting five residues in two regions (positions 306–311 and 333–351) owing to the fact that two previous gaps in the sequence of mouse fragment P1 were now closed by overlapping sequences of bovine fragment P1. Amino acid substitutions found in the sequence of mouse fragment P1 (M) are listed in the second row. The sequence was determined by automated Edman degradation (–) with fragment P1 or fragments produced by cleavage with CNBr (CB), trypsin (T) or thermolysin (Th). Residues in parentheses could not be identified with certainty or are deduced from compositional data or a comparison with the mouse sequence. X denotes residues not identified, but are in most cases glycosylated hydroxylysine as deduced from compositional data.

Table 3. *List of peptides sequenced, their size, location in the P1 fragment and a summary of sequence data*
All peptides were sequenced by using a 0.1M-Quadrol protein program (Beckman number 122974) in the presence of 1 mg of Polybrene.

Peptide	Position (size)	Amount sequenced (nmol)	Positions identified
P1	5–522 (518)	10	15
CB3	8–77 (70)	20	18
T6	30–40 (11)	70	10
T2-3	41–234 (194)	100	29
CB10b	83–153 (71)	20	50
T2-3Th	132–167 (36)	15	27
CB2	154–273 (120)	100	19
Ta1	235–329 (95)	30	48
CB5	274–314 (41)	170	36
T1aC	300–329 (30)	30	20
CB4	315–352 (38)	60	29
T1b	330–494 (165)	45	32
CB6-1	353–509 (157)	150	45
CB6-1TC	495–509 (15)	70	10
T5	495–522 (28)	60	20
CB7	510–522 (13)	170	9

obtained in highest yields and accounted for the total mass of fragment P1. A comparison by composition and sequence analysis (see below) showed that only peptides T5 and T6 resembled identical peptides from mouse fragment P1 (Schuppan *et al.*, 1982). Bovine peptide T4′ was similar to the *N*-terminal mouse peptide T4 except for a *N*-terminal tetrapeptide sequence formerly identified as peptide CB8. Peptides T1a and T1b apparently resemble together mouse peptide T1 and arose by cleavage at an additional arginine residue (position 329, Fig. 4). A second additional arginine residue, at position 299, is responsible for small amounts of a further peptide T1aC originating from the *C*-terminus of peptide T1a. Lack of arginine in position 160 of the bovine sequence (Fig. 4) results in the double peptide T2-3. Shorter variants (T2-3a, lacking about 52 residues from the *C*-terminus) as well as larger variants (T6-2-3, T6-2-3a) were also found.

Two more peptides were prepared because of special requirements in the sequence analysis. Tryptic cleavage of peptide CB6-1 produced two peptides, which could be separated on Sephadex G-75. The composition of the smaller peptide CB6-1TC (Table 2) demonstrates its origin from the *C*-terminal end of peptide CB6-1. A thermolysin digest was prepared from peptide T2-3 (containing some variants) and purified on Bio-Gel P-10 and phosphocellulose. This allowed the isolation of a fragment T2-3Th (Table 2), which overlaps the junction between peptides CB10b and CB2 (Fig. 1).

Sequence analysis of fragment P1

About 67% of the amino acid sequence of bovine fragment P1 was determined by automated Edman degradation of fragment P1 and its fragments produced with CNBr, trypsin and thermolysin (Table 3). These sequences are compared in Fig. 4 with those determined previously for the mouse fragment P1 (Schuppan *et al.*, 1982). Bovine fragment P1 showed a distinct heterogeneity in its *N*-terminal sequence, indicating several cleavage sites during the release of the fragment P1 from the α1(IV) chain by pepsin. A major sequence could, however, be identified, which started with position 5 of the mouse sequence. This is compatible with the lack of peptide CB8 (sequence His-Val-Asp-Met in mouse fragment P1) in bovine fragment P1. Sequence determinations of the various peptides allowed ordering of most of the CNBr-cleavage peptides along the fragment P1, leaving only three gaps open (Figs. 1 and 4). The positions of some peptides (CB10b, T6, T2-3) for which no overlaps were obtained could be tentatively assigned by comparison with the sequence of mouse fragment P1. The position of peptide CB10a (Fig. 1) is still tentative (see the Discussion section).

Discussion

CNBr cleavage of the 50000-M_r fragment of bovine lens-capsule collagen α1(IV) chain produces nine peptides, which were homologous to CNBr-cleavage peptides of the α1(IV)-chain fragment P1 from a mouse tumour basement membrane (Schuppan *et al.*, 1982). Most of the CNBr-cleavage peptides from bovine fragment P1, including the *C*-terminal peptide CB7, have been previously identified in larger fragments of the α1(IV) chain (Dixit & Kang, 1979; see also Table 1), supporting a previous interpretation that fragment P1 comprises the *C*-terminal portion of the major triple-helical segment in the α1(IV) chain. All except one of the CNBr-cleavage peptides could be precisely localized within fragment P1 by sequence analysis of overlapping tryptic peptides or by comparison with the sequence of mouse fragment P1. The pentapeptide CB10a was tentatively placed between peptides CB3 and CB10b (sequence positions 78-82). This position appears very likely, since peptide CB10b is shorter by five amino acid residues when compared with mouse peptide CB10, and the composition of peptide CB10a (Gly, Leu, Hyp, Gly, Hse) closely matches the starting sequence Gly-Leu-Hyp-Gly-Ser- of mouse peptide CB10 (Schuppan *et al.*, 1982).

Analysis of the bovine fragment P1 allowed the precise positioning of peptides CB2, CB5, CB4 and CB6-1 by overlapping sequences, thus supporting the order of the same peptides in the mouse frag-

ment P1, which has been based on more indirect evidence (Schuppan *et al.*, 1982). On the basis of the sequence data for mouse and bovine fragments P1 only two gaps of about 19 residues (positions 216–234) and 79 residues (positions 416–494) still remain open in fragment P1. The data also show a high apparent degree of conservation of the position of non-triplet sequence elements, which exist at both ends of the P1 fragment and internally at positions 117–118 and 292–293. Both internal interruptions and the terminal non-triplet sequences are very similar in the sequences of bovine and mouse fragments P1. The *N*-terminal sequence of bovine fragment P1 lacks four residues, indicating a different sequence to the sequence His-Val-Asp-Met of mouse fragment P1, since it can be cleaved by pepsin. The invariance in position and sequence of the triplet interruptions indicate a certain functional importance. This could be related to the increased flexibility of the collagen type IV triple helix (Hofmann *et al.*, 1984), which may allow the adaptation of basement-membrane structures to special physiological requirements (Timpl *et al.*, 1981).

A comparison of the sequences of the bovine and mouse fragments P1 shows, as expected, invariance of glycine, but in 170 corresponding Xaa and Yaa positions analysed (amino acids in parenthesis and comparisons with X, Fig. 4, were not included) only 76% were identical. About half of the substitutions are chemically conservative, but the others include a change from a polar to a non-polar residue or vice versa. This is a surprisingly low interspecies homology (88%) when compared with about 100% homology found among interstitial collagens (Bornstein & Traub, 1979). For interstitial collagens it is thought that a distinct arrangement of polar and hydrophobic amino acid residues is required to allow their staggered lateral association (Hulmes *et al.*, 1973; Hofmann *et al.*, 1978).

Our data suggest lack of such a requirement for collagen type IV, which would be compatible with a network-like association of type IV collagen molecules with only short terminal segments being used for interactions (Timpl *et al.*, 1981).

Both sequences also contain two partially (50–90%) hydroxylated 3-hydroxyproline residues in identical positions (residue numbers 275 and 499). The surrounding sequences are rather similar, indicating specific structural requirements for the hydroxylation reaction (Schuppan *et al.*, 1982). The invariance in position could also suggest a distinct function for 3-hydroxyproline, which however, still remains to be elucidated.

The study was supported by the Deutsche Forschungsgemeinschaft, and by Grants AM 32749 and AM 16506 from the National Institutes of Health and by the Veterans Administration.

References

Bornstein, P. & Traub, W. (1979) in *The Proteins*, 3rd edn. (Neurath, H. & Hill, R. L., eds.), vol. 4, pp. 412–632, Academic Press, New York

Dehm, P. & Kefalides, N. A. (1978) *J. Biol. Chem.* **253**, 6680–6686

Dixit, S. N. (1978) *FEBS Lett.* **85**, 153–157

Dixit, S. N. & Kang, A. H. (1979) *Biochemistry* **18**, 5686–5692

Fietzek, P. P. & Kühn, K. (1976) *Int. Rev. Connect. Tissue Res.* **7**, 1–60

Gay, S. & Miller, E. J. (1979) *Arch. Biochem. Biophys.* **198**, 370–378

Glanville, R. W. & Rauter, A. (1981) *Hoppe-Seyler's Z. Physiol. Chem.* **362**, 943–951

Hofmann, H., Fietzek, P. P. & Kühn, K. (1978) *J. Mol. Biol.* **125**, 137–165

Hofmann, H., Voss, T., Kühn, K. & Engel, J. (1984) *J. Mol. Biol.* **172**, 325–343

Hulmes, D. J., Miller, A., Parry, D. A. D., Piez, K. A. & Woodhead-Galloway, J. (1973) *J. Mol. Biol.* **79**, 137–148

IUPAC–IUB Commission on Biochemical Nomenclature (1969) *Biochem. J.* **113**, 1–4

Kühn, K., Wiedemann, H., Timpl, R., Risteli, J., Dieringer, H., Voss, T. & Glanville, R. W. (1981) *FEBS Lett.* **125**, 123–128

Lottspeich, F. (1980) *Hoppe-Seyler's Z. Physiol. Chem.* **361**, 1829–1834

Schuppan, D., Timpl, R. & Glanville, R. W. (1980) *FEBS Lett.* **115**, 297–300

Schuppan, D., Glanville, R. W. & Timpl. R. (1982) *Eur. J. Biochem.* **123**, 505–512

Timpl, R., Bruckner, P. & Fietzek, P. (1979) *Eur. J. Biochem.* **95**, 255–263

Timpl, R., Wiedemann, H., van Delden, V., Furthmayr, H. & Kühn, K. (1981) *Eur. J. Biochem.* **120**, 203–211

Weber, S., Engel, J., Wiedemann, H., Glanville, R. W. & Timpl, R. (1984) *Eur. J. Biochem.* in the press