Volume 225, number 1,2, 188-194 FEB 05367 December 1987

Complete primary structure of the α_1-chain of human basement membrane (type IV) collagen

Raija Soininen*+, Tuula Haka-Risku*, Darwin J. Prockop+ and Karl Tryggvason*

*Biocenter and Department of Biochemistry, University of Oulu, SF-90570 Oulu, Finland and +Jefferson Institute of Molecular Medicine, Jefferson Medical College, Thomas Jefferson University, Philadelphia, PA 19107, USA

Received 16 October 1987

We have determined the primary structure of the α_1(IV)-chain of human type IV collagen by nucleotide sequencing of overlapping cDNA clones that were isolated from a human placental cDNA library. The present data provide the sequence of 295 amino acids not previously determined. Altogether, the α_1(IV)-chain contains 1642 amino acids and has a molecular mass of 157625 Da. There are 1413 residues in the collagenous domain and 229 amino acids in the carboxy-terminal globular domain. The human α_1(IV)-chain contains a total of 21 interruptions in the collagenous Gly-X-Y repeat sequence. These interruptions vary in length between two and eleven residues. The α_1(IV)-chain contains four cysteine residues in the triple-helical domain, four cysteines in the 15-residue long noncollagenous sequence at the amino-terminus and 12 cysteines in the carboxy-terminal NC-domain.

Amino acid sequence; Nucleotide sequence; Collagen; Primary structure; Basement membrane; (Human)

1. INTRODUCTION

Type IV collagen is the major structural component of basement membranes [1]. The molecule contains two distinct polypeptide chains, α_1(IV) and α_2(IV), with molecular masses of 185 and 170 kDa, respectively [2]. Three α-chains form a heterotrimer that apparently contains two α_1(IV)- and one α_2(IV)-chain [3] but a homotrimer of α_1(IV)-chains has also been described [4]. The molecule contains an about 350 nm long triple-helical collagenous domain with a large globule at the carboxy-terminal end [1]. In contradiction with the fibrillar collagens of types I, II, III and V [5], type IV collagen forms a network structure. Another significant difference from fibrillar collagens is that the triple helix of the type IV collagen molecule has imperfections in the collagenous Gly-X-Y- repeat sequence [1]. Interrupted triple helices are also found in collagen types VI, VII, VIII, IX and X that do not form fibrils [6–10].

A part of the amino acid sequence of the α_1(IV)-chain from several species has been determined by amino acid and nucleotide sequencing. The amino acid sequence of a 914-residue long pepsin fragment representing about two-thirds of the carboxy-terminal end of the collagenous domain [11] and the sequence of a 216-residue long fragment from the amino-terminus of the human α_1(IV)-chain has been determined by protein sequencing [12]. We and others [13,14] have determined the complete primary structure of the carboxy-terminal globular domain of the human α_1(IV)-chain by analysis of cDNA clones. For mouse, the structure of the globular domain and 511 residues from the carboxy-terminal end of the triple-helical region has been determined by a combination of protein and cDNA sequencing [15–17]. Recently, the

Correspondence address: K. Tryggvason, Biocenter and Department of Biochemistry, University of Oulu, SF-90570 Oulu, Finland

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under accession number Y00706

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

primary structure for the carboxy-terminal globular domain and a part of the collagenous region of the α_2(IV)-chain was deduced from cloned cDNAs for both human [18,19] and mouse [20–23] type IV collagen.

In the present study we describe the primary structure of the complete human α_1(IV)-chain based on amino acid sequences derived from cDNA clones.

2. MATERIALS AND METHODS

A human placental cDNA library in the λgt11 vector (Clontech, Palo Alto, CA) was screened with nick-translated probes according to standard techniques [24]. The initial probe was a 500 bp long *Pst*I-*Bam*HI fragment from the 5′-end of the cDNA clone HT-21 [13] that contains sequences for the carboxy-terminal end of human α_1(IV)-chain. Most of the 5′-fragments from isolated clones were used as probes in rescreening of the same library.

Nucleotide sequencing was carried out using M13 cloning [25] and the dideoxynucleotide chain termination method [26] with the 'Sequenase' enzyme (United States Biologicals, Cleveland, OH). In most reactions dITP was used instead of dGTP which enabled reading through the compression areas common for collagen coding cDNAs. Primers used for the sequencing were either the 'universal primer' or specific 15-mer oligonucleotide primers synthesized in a DNA synthesizer (Applied Biosystems).

3. RESULTS AND DISCUSSION

3.1. *Isolation and characterization of cDNA clones*

Screening of the human placental cDNA library with the 5′-fragment from the α_1(IV)-coding cDNA clone HT-21 resulted in the isolation of several clones. Two clones, HP-2 and HP-3, containing inserts of 3.5 kb and 3.0 kb, respectively, were further characterized. The 5′-*Eco*RI fragment of HP-2 was then used as a probe for the isolation of clone HP-1, and to isolate the HP-11 cDNA clone. Partial restriction map of the clones is presented in fig.1. Three clones, HP-11, HP-2 and HP-3, cover the entire coding sequence for the human α_1(IV)-chain except for the signal peptide and the first 18 amino acids at the amino-terminus.

3.2. *Nucleotide sequence and primary structure*

*Eco*RI or *Bam*HI restriction fragments of cDNA clones were cloned into M13 vector and sequenced first with the universal primer and then with specific synthetic primers to obtain overlapping sequences. Coding sequences for regions of the α_1(IV)-chain not previously determined by amino acid sequencing were sequenced from both strands. The nucleotide sequence of the cDNA clones that reached 5′ from the previously reported HT-21 [13] and the deduced amino acid sequence is shown in fig.2. There are 5 differences between the sequence determined here and those determined previously by amino acid sequencing [11,12], assuming that amino acids marked with X are (hydroxy)lysines. These differences are (i) asparagine instead of aspartic acid at position 692, (ii) aspartic acid instead of tyrosine at position 810, (iii) lysine instead of proline at position 815, (iv) serine instead of hydroxylysine at position 887 and (v) lysine instead of proline at position 983.

The data provided the sequence of 295 amino acids not previously determined. The new sequence completed the structure of the whole chain. Accordingly, the results established that the α_1(IV)-chain contains 1642 amino acids and has a molecular mass of 157625 Da. The collagenous domain has 1413 residues and the carboxy-terminal globular domain has 229 amino acids [13]. In vitro translation of the mRNA specific to mouse type IV collagen cDNA resulted in the synthesis of a polypeptide with an apparent molecular mass of 165 kDa [27]. The apparent 185 kDa size [2] of the α_1(IV)-chain on SDS-PAGE can be attributed to the presence of carbohydrates [1].

The noncollagenous carboxy-terminal domain has usually been referred to as the NC-1 domain, because early results suggested that there would be a 10 kDa NC-2 domain within the triple helix [1]. However, the present data demonstrate that such a sizable NC-2 domain is not present in the α_1(IV)-chain and therefore, we refer to the carboxy-terminal globule as the NC (noncollagenous) domain.

The new sequence described here contains 5 imperfections in the Gly-X-Y repeat sequence and, therefore, the human α_1(IV)-chain contains a total

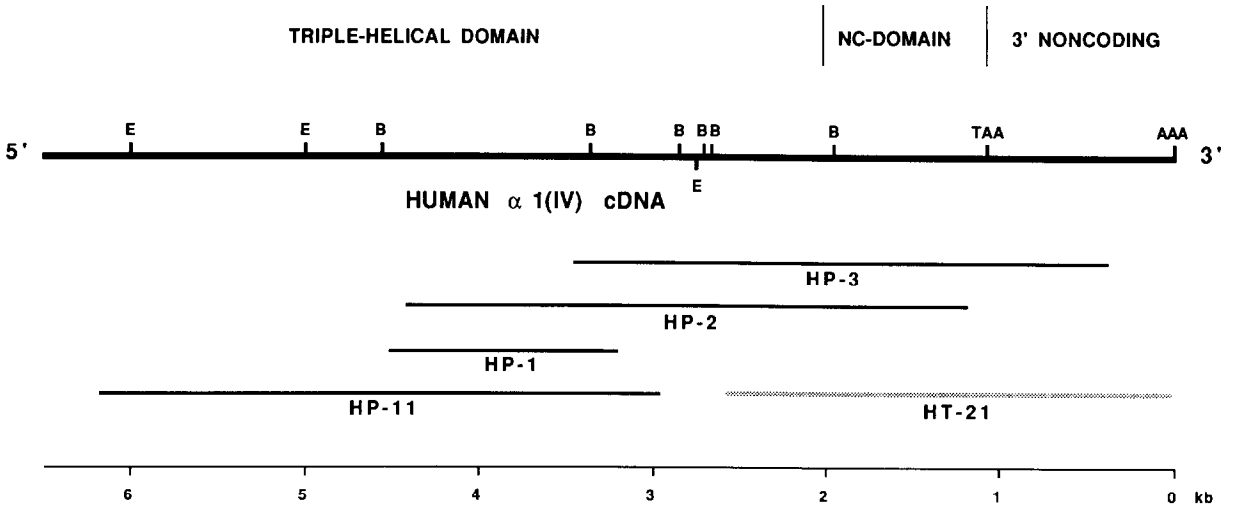


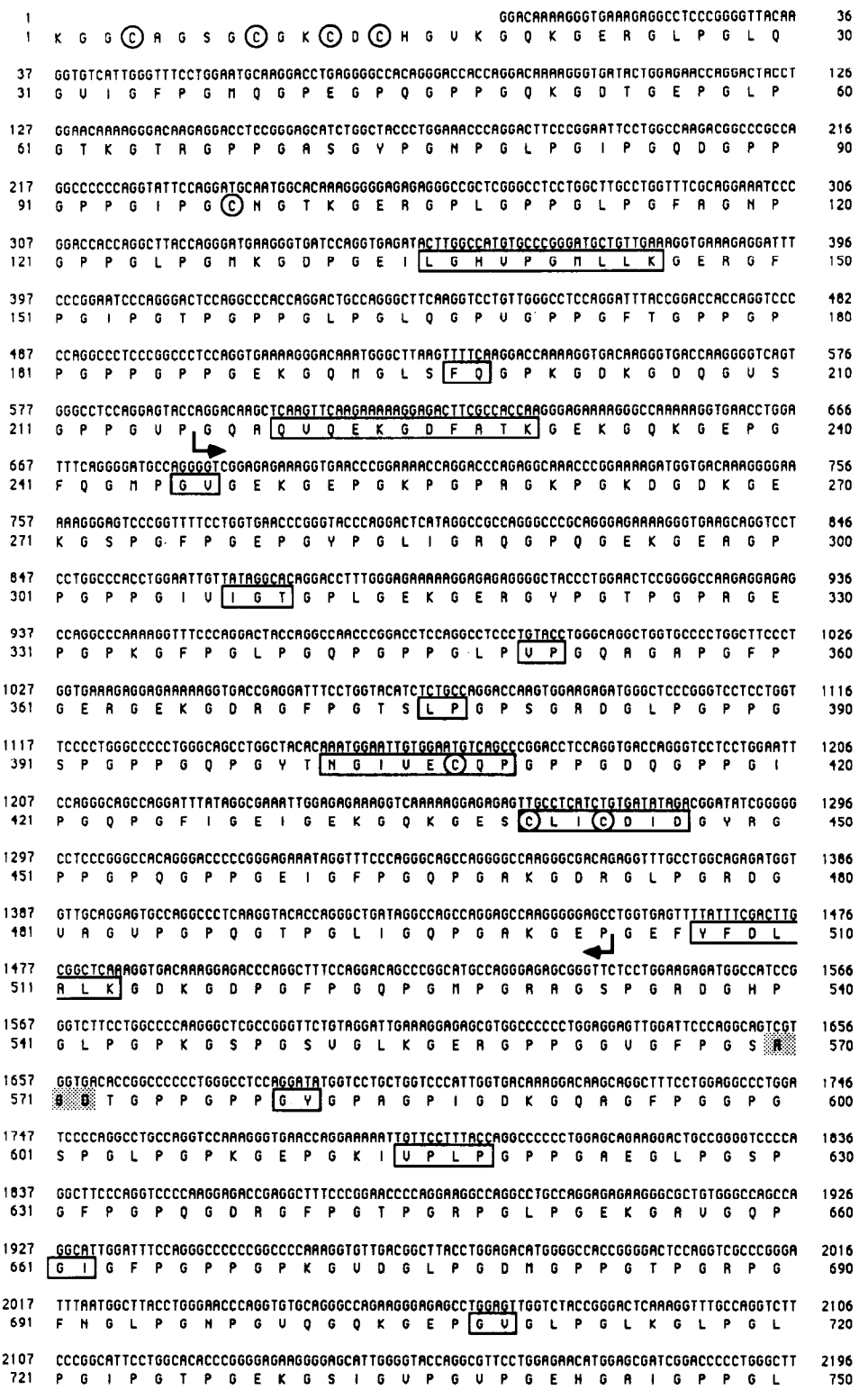
Fig.1. Overlapping cDNA clones coding for the α_1-chain of human type IV collagen. HT-21 has been described in [13]. Its 5'-fragment was used as a probe to isolate clones HP-2 and HP-3. *Eco*RI (E) and *Bam*HI (B) restriction sites are indicated. TAA is the site of translation termination and AAA indicates the poly(A) tail.

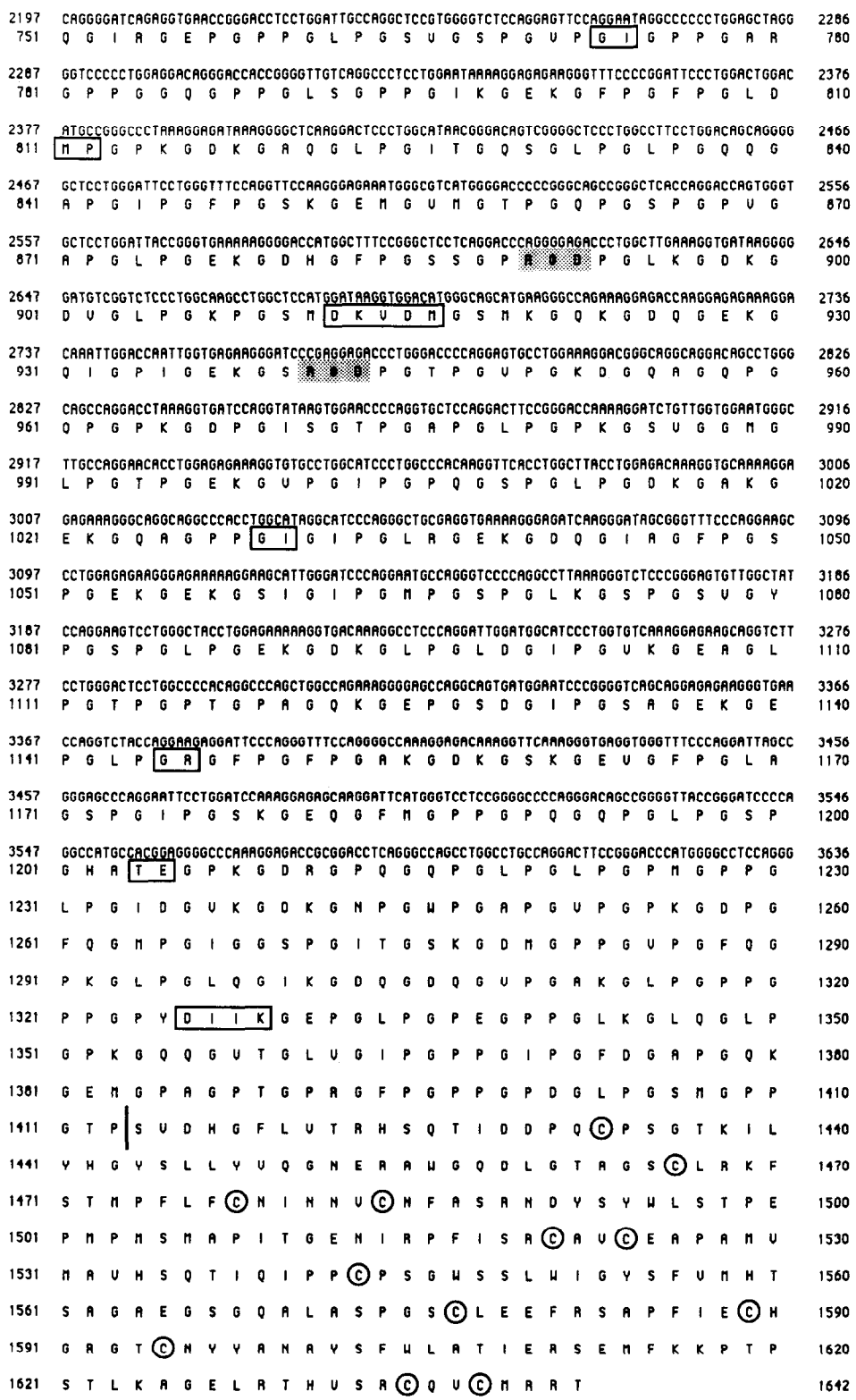
of 21 interruptions (fig.3). These interruptions vary in length between 2 and 11 residues (table 1). As can be seen in fig.3, the interruptions are scattered throughout the chain, although most of them are located towards the amino-terminus. In addition to the interrupted Gly-X-Y sequence, the α_1(IV)-chain has a 229-residue long noncollagenous sequence at the carboxy-terminus and a short 15-residue long noncollagenous sequence at the amino-terminus [12].

There are several reports about type IV collagen promoting cell attachment (see [1]) and, therefore, one might expect the molecule to contain amino acid sequences recognized by membrane receptors for matrix components. An Arg-Gly-Asp sequence has been shown to play an important role in the adhesion of cells to components of the interstitial connective tissue such as fibronectin and type I collagen but also to vitronectin [28,29]. There are three such sequences in the human α_1(IV)-chain (fig.2) and one of them, Arg-Gly-Asp-Thr-Gly, is identical with a sequence present in the human α_2-chain of type I collagen that supports cell attachment [29]. It is possible that these tripeptide sequences within the α_1(IV)-chain are involved in mediating cell-matrix interactions.

The type IV collagen α-chains are linked to each other within the triple-helical domain by disulfide bonds whereas the cysteine residues in the NC domain only participate in intrachain bonds [30]. The results of this study demonstrate that, in addition to the 12 cysteine residues in the NC domain, there are four cysteine residues in the noncollagenous sequence at the amino-terminus and four within the triple-helical domain. One of the cysteine residues in the collagenous domain is in a Gly-X-Y sequence at location 98 (fig.2). There is one cysteine in interruption number 8 and two cysteines in in-

Fig.2. Complete amino acid sequence of the α_1(IV)-chain of human type IV collagen. First line: nucleotide sequence from the 5'-end of overlapping cDNA clones characterized in this study to the 5'-end of the previously reported sequence of the HT-21 cDNA [13] (fig.1). Second line: amino acid sequence deduced from the nucleotide sequence. The first 18 amino acids have been determined by amino acid sequencing [12], and amino acid residues number 1230–1642 are deduced from the nucleotide sequence of the cDNA clone HT-21 [13]. The numbering of nucleotide sequence starts at the first nucleotide in the cDNA clone HP-11 and the numbering of the amino acid sequence at the first amino acid of the published sequence [12]. The new 295-residue long sequence is located between the arrows and the junction of the collagenous and NC-domain is marked by a vertical bar. The cysteine residues are encircled and the imperfections in the Gly-X-Y repeat sequence of the collagenous domain are boxed. RGD sequences are indicated by shaded boxes.





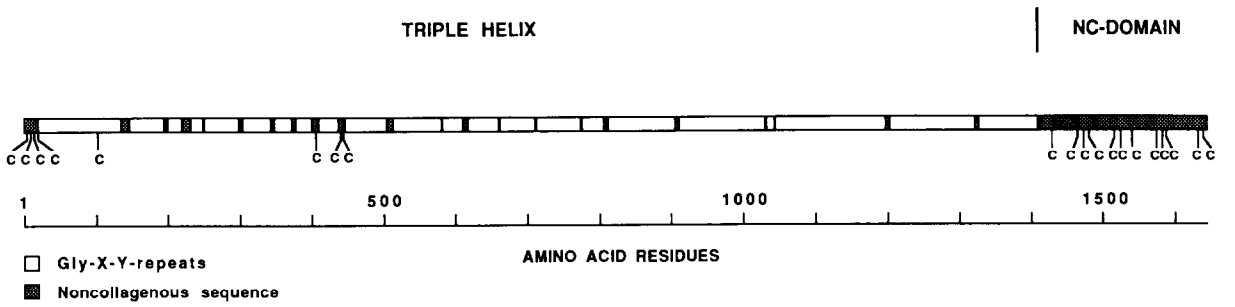


Fig.3. Primary structure of the α_1-chain of human type IV collagen. Interruptions of the collagenous (Gly-X-Y) sequence are marked by shaded boxes. Cysteine residues are indicated by C.

terruption number 9 (table 1). One or more of the last three cysteine residues must participate in interchain disulfide bonds because the carboxy-terminal fragment generated by type IV collagenase, containing about 75% of the length of the chain, is disulfide bonded [30].

The present data provide the sequence of the region of the α_1(IV)-chain that is cleaved by type IV collagenase [31]. The sequence of the cleavage site has not been determined but it has been measured from degradation fragments visualized by rotary shadowing to be 89 ± 10 nm from the amino-terminal end of the polypeptide chain [30]. This knowledge makes it possible to prepare synthetic peptides that might serve as enzyme inhibitors or substrates for the enzyme.

Table 1

Interruptions in the triple-helical $(Gly\text{-}X\text{-}Y)_n$ sequence of the human α_1(IV)-chain

Helix interruption	Amino acid residue number	Amino acid sequence
1	136– 145	LGHVPGMLLK
2	197– 198	FQ
3	220– 230	QVQEKGDFATK
4	246– 247	GV
5	308– 310	IGT
6	350– 351	VP
7	376– 377	LP
8	402– 409	NGIVECQP
9	440– 446	CLICDID
10	507– 513	YFDLRLK
11	580– 581	GY
12	615– 618	VPLP
13	661– 662	GI
14	708– 709	GV
15	773– 774	GI
16	811– 812	MP
17	912– 916	DKVDM
18	1029–1030	GI
19	1145–1146	GR
20	1204–1205	HATE
21	1326–1329	DIIK

ACKNOWLEDGEMENTS

We thank Raymond Boot-Handford for valuable comments and Pekka Kallunki for preparing the synthetic oligonucleotides used for the DNA sequencing. This work was supported in part by grants from the National Research Council for Natural Sciences, Academy of Finland and from NIH, AM-38818, AM-22051.

REFERENCES

[1] Timpl, R. and Dziadek, M. (1986) Int. Rev. Exp. Pathol. 29, 1–112.
[2] Tryggvason, K., Gehron Robey, P. and Martin, G.R. (1980) Biochemistry 19, 1284–1289.
[3] Trüeb, B., Gröbli, B., Spiess, M., Odermatt, B.F. and Winterhalter, K.H. (1982) J. Biol. Chem. 257, 5239–5345.
[4] Haralson, M.A., Federspiel, S.J., Martinez-Hernandez, A., Rhodes, R.K. and Miller, E.J. (1985) Biochemistry 24, 5792–5797.
[5] Miller, E.J. (1985) Ann. NY Acad. Sci. USA 460, 1–13.
[6] Jander, R., Rauterberg, J., Voss, T. and Von Bassewitz, D.B. (1981) Eur. J. Biochem. 114, 17–25.

[7] Benz, H., Morris, N.P., Murray, L.W., Sakai, L.Y., Hollister, D.W. and Burgeson, R.E. (1983) Proc. Natl. Acad. Sci. USA 80, 3168–3172.
[8] Sage, H., Trüeb, B. and Bornstein, P. (1983) J. Biol. Chem. 258, 13391–13401.
[9] Van der Rest, M., Mayne, R., Ninomiya, Y., Seidah, N.G., Greidien, M. and Olsen, B.R. (1985) J. Biol. Chem. 260, 220–225.
[10] Schmid, T.M. and Linsenmayer, T.F. (1985) J. Cell. Biol. 100, 598–605.
[11] Babel, W. and Glanville, R.W. (1984) Eur. J. Biochem. 143, 545–556.
[12] Glanville, R.W., Qian, R., Siebold, B., Risteli, J. and Kühn, K. (1985) Eur. J. Biochem. 152, 213–219.
[13] Pihlajaniemi, T., Tryggvason, K., Myers, J., Kurkinen, M., Lebo, R., Cheung, M.-C., Prockop, D.J. and Boyd, C.D. (1985) J. Biol. Chem. 260, 7681–7687.
[14] Brinker, J.M., Gudas, L.J., Loidl, H.R., Wang, S.-Y., Rosebloom, J., Kefalides, N.A. and Myers, J.C. (1985) Proc. Natl. Acad. Sci. USA 82, 3649–3653.
[15] Oberbäumer, I., Laurent, M., Schwarz, U., Sakurai, Y., Yamada, Y., Vogeli, G., Voss, T., Siebold, B., Glanville, R. and Kühn, K. (1985) Eur. J. Biochem. 147, 217–224.
[16] Nath, P., Laurent, M., Horn, E., Sobel, M.E., Zon, G. and Vogeli, G. (1986) Gene 43, 301–304.
[17] Schuppan, D., Glanville, R.W. and Timpl, R. (1982) Eur. J. Biochem. 123, 505–512.
[18] Hostikka, S.L., Kurkinen, M. and Tryggvason, K. (1987) FEBS Lett. 216, 281–286.
[19] Myers, J.C., Howard, P.S., Jelen, A.M., Dion, A.S. and Macarac, E.J. (1987) J. Biol. Chem. 262, 9231–9238.
[20] Schwarz, U., Schuppan, D., Oberbäumer, I., Glanville, R.W., Deutzmann, R., Timpl, R. and Kühn, K. (1986) Eur. J. Biochem. 157, 49–56.
[21] Vogeli, G., Horn, E., Carter, J. and Kaytes, P.S. (1986) FEBS Lett. 206, 29–32.
[22] Schwarz-Magdolen, U., Oberbäumer, I. and Kühn, K. (1986) FEBS Lett. 208, 203–207.
[23] Kurkinen, M., Condon, M.R., Blumberg, B., Barlow, D., Quinones, S., Saus, J. and Pihlajaniemi, T. (1987) J. Biol. Chem. 262, 8496–8499.
[24] Maniatis, T., Fritsch, E. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[25] Messing, J. (1983) Methods Enzymol. 101, 20–78.
[26] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[27] Kurkinen, M., Barlow, D.P., Helfman, D.M., Williams, J.G. and Hogan, B.L.M. (1983) Nucleic Acids Res. 11, 6199–6209.
[28] Ruoslahti, E. and Pierschbacher, M.D. (1986) Cell 44, 517–518.
[29] Pierschbacher, M.D. and Ruoslahti, E. (1984) Proc. Natl. Acad. Sci. USA 81, 5985–5988.
[30] Fessler, L.I., Duncan, K.G., Fessler, J.H., Salo, T. and Tryggvason, K. (1984) J. Biol. Chem. 259, 9783–9789.
[31] Salo, T., Liotta, L.A. and Tryggvason, K. (1983) J. Biol. Chem. 258, 3058–3063.