Volume 227, number 1, 5–8 FEB 05472 January 1988

cDNA clones completing the nucleotide and derived amino acid sequence of the alpha 1 chain of basement membrane (type IV) collagen from mouse

Linda Wood, Nicole Theriault* and Gabriel Vogeli

Molecular Biology Research and **Biopolymer Chemistry Research, The Upjohn Company, Kalamazoo, MI 49001, USA*

Received 24 November 1987

Six cDNA clones add 3549 nucleotides to the DNA sequence of the alpha 1 chain of basement membrane (type IV) collagen. Thus the complete nucleotide and derived amino acid sequence of the alpha 1 type IV collagen with 5007 nucleotides coding for 1669 amino acids with a calculated M_r of 160 827 is known. The six cDNA clones cover the putative N-terminal signal peptide, the 7 S region and two thirds of the helical region extending into the previously published murine nucleotide sequence [(1986) Gene 43, 301]. The protein sequence for 289 amino acids of the helical region adjacent to the 7 S region has not previously been published for any species.

Basement membrane; Signal peptide; 7 S region, NC2 domain; Helical domain; Helical interruption

1. INTRODUCTION

Basement membranes are essential extracellular structures of animals forming histological and physiological barriers between many tissues [1]. The scaffold of the basement membrane is formed from type IV collagen molecules (review [2]) that cross-link head to head between the N-terminal (7 S) domains and tail to tail between the C-terminal (NC1) domains [3,4]. The aa sequence of the alpha 1 (IV) chain for the human 7 S region [5], but not for the N-terminus, and a partial aa sequence for the human [6] and the mouse [7] helical region has been published. cDNA and genomic clones for the alpha 1 (IV) chain have been characterized and reported ([8] and references therein, [9]). We report here DNA sequence data that complete the alpha 1 (IV) collagen chain.

Correspondence address: L. Wood, Molecular Biology Research, The Upjohn Company, Kalamazoo, MI 49001, USA

Abbreviations: aa, amino acid(s); cDNA, DNA complementary to mRNA; nt, nucleotide(s)

2. MATERIALS AND METHODS

The cDNA clones pCIV-1-C92, pCIV-1-C177, pCIV-1-C308, pCIV-1-C87 were isolated from the Engelbreth-Holm-Swan (EHS) cDNA library as in [4,8,10,11]. The cDNA clones pCIV-1-PE12 and pCIV-1-PE16 were isolated from a PYS cDNA library. Isolation of plasmid DNA, subcloning into Bluescribe vectors and sequencing with synthetic oligonucleotide primers (fig.1) was done as in [12,13].

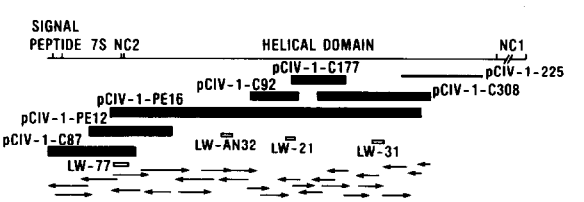
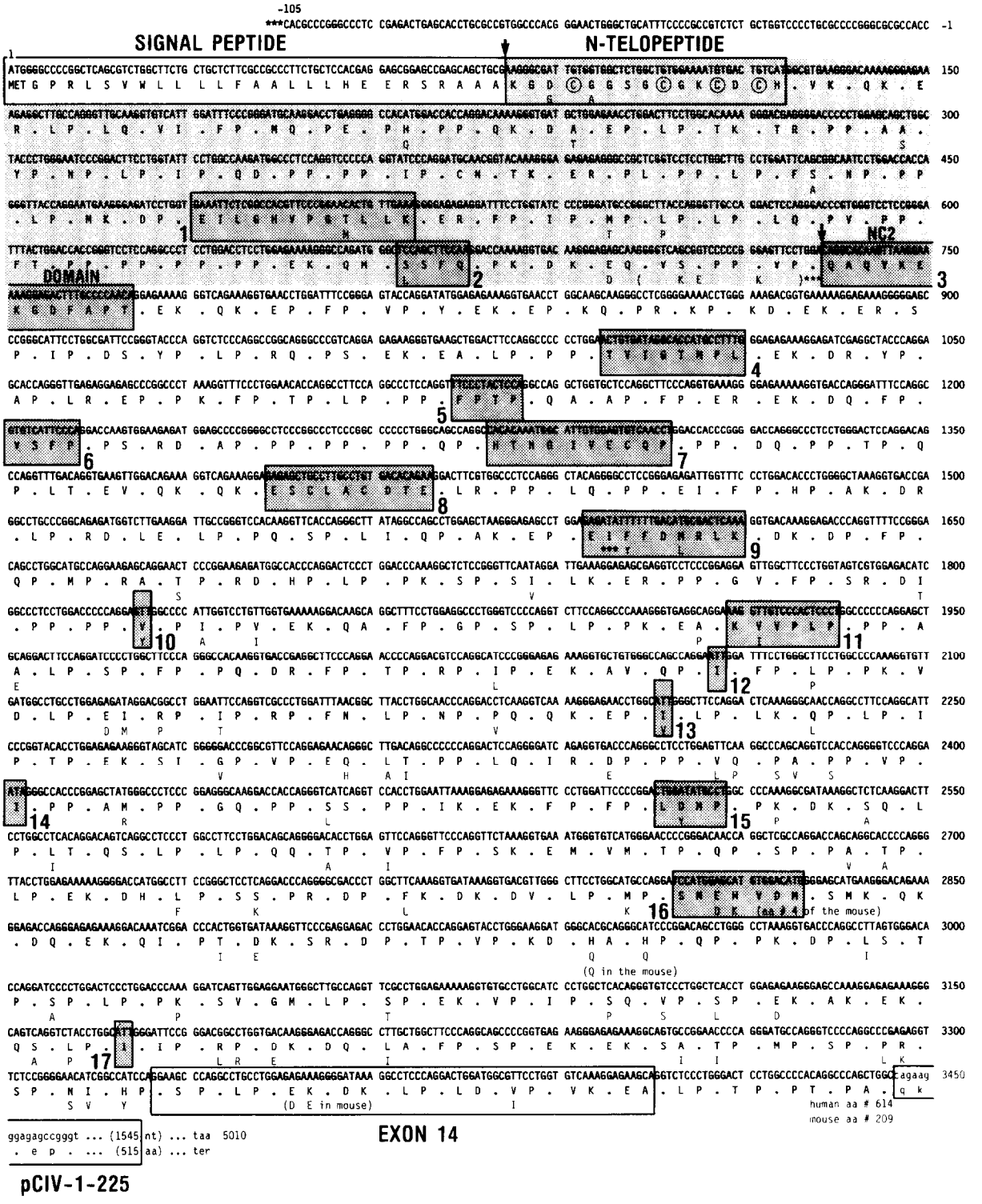


Fig.1. cDNA clones and sequencing strategy. Schematic representation of cDNA clones along the alpha 1 (IV) molecule with position of the synthetic oligonucleotides used to screen the libraries. LW-AN32 was derived from aa 13 to 22 of the published sequence [6], LW-21 and LW-77 were derived from our nt sequences and LW-31 was derived from the sequence of exon 14 [9]. The arrows indicate independent sequence determinations with synthetic 17-mer oligonucleotide primers.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50


SIGNAL PEPTIDE
N-TELOPEPTIDE
DOMAIN
NC2
EXON 14
pCIV-1-225
human aa # 614
mouse aa # 209
(Q in the mouse)
(D E in mouse)
(aa # 4 of the mouse)

3. RESULTS AND DISCUSSION

3.1. *Isolation of cDNA clones and sequence comparison*

The six cDNA clones (fig.1) cover a total of 3567 nt (see fig.2) including 105 nt of the 5′ untranslated region followed by 81 nt for the signal peptide, 651 nt for the 7 S region and 2730 nt for the helical region. The cDNA clones extend into the previously published sequences [8,14]. Thus the complete nt sequence of 5007 nt coding for the 1669 aa of the alpha 1 (IV) chain with a calculated M_r of 160827 is known. The three differences with the partial mouse aa sequence [7] are shown in fig.2; the homology between the aa sequence of the human and the mouse is 93% for the 7 S region [5] and 88% for the triple helix [6].

3.2. *Signal peptide and 7 S region*

The sequence of the signal peptide (fig.2) and the position of the possible signal peptidase cleavage site at nt 81 was determined by comparing our sequence with reported signal sequences [15]. The signal sequence corresponds to that found by P. Killen (personal communication). Thus Lys (nt 82) at the beginning of the 7 S region [5] is created by the signal peptidase and not by the proteolytic enzymes used to isolate the 7 S protein. The alpha 1 (IV) chain has no N-terminal globular domain, but starts with a telopeptide of 15 aa followed by the helical Gly-X-Y domain. The telopeptide contains 2 Lys and 4 Cys as reported [5] and is possibly the site of type IV collagen cross-linkage to form the basement membrane meshwork. The 7 S region has two interruptions in the Gly-X-Y repeat, ending at nt 732 with a large interruption of 13 aa, that sometimes is called NC2 domain [2]. This interruption probably provides the protease cleavage site for the isolation of 7 S protein from cross-linked type IV collagen.

3.3. *Helical region and interruptions*

Fig.2 shows a helical region of 2730 nt; the protein sequence for 289 aa of the helical region adjacent to the 7 S region has not previously been published for any species. There are 17 interruptions of the Gly-X-Y repeat, including the 3 interruptions of the 7 S region. Five interruptions contain only one aa, four Ile, one Tyr. The other interruptions vary in size. Together with the 3 interruptions previously published [8], the complete alpha 1 type IV collagen chain contains 20 interruptions. These interruptions of the collagenous Gly-X-Y repeat are typical for type IV collagen [7,8], which are found in corresponding chains in equivalent positions and are implicated in the formation of the basement membrane meshwork [11].

4. CONCLUSIONS

Our data add 3549 nt to the published DNA sequence [8,14] for a total of 5007 nt for the complete coding region of the alpha 1 (IV) chain. The calculated M_r of the 1669 aa coding for the alpha 1 (IV) chain is 160827, in agreement with [2]. It is not known if the large interruption of 13 aa between the 7 S region and the helical domain, that sometimes is referred to as NC2 domain, is the site of protease attack during basement membrane remodeling. Such remodeling is implicated in the migration of macrophages and metastatic cancer cells during exudation [16].

Acknowledgements: We would like to thank Markku Kurkinen for the PYS cDNA library and Lee Giordano for the preparation of the bacterial supplies.

REFERENCES

[1] Laurie, G.W. and Leblond, C.P. (1983) J. Histochem. Cytochem. 13, 159–163.
[2] Piez, K.A. and Reddi, A.H. (1984) Extracellular Matrix Biochemistry, Elsevier, Amsterdam, New York.
[3] Yurchenco, P.D. and Furthmayr, H. (1984) Biochemistry 23, 1839–1850.
[4] Kaytes, P.S., Theriault, N.Y. and Vogeli, G. (1987) Gene 54, 139–144.
[5] Glanville, R.W., Qian, R., Siebold, B., Risteli, J. and

Fig.2. DNA and derived aa sequence of alpha 1 (IV) cDNA clones and comparison with the published aa sequence [5–7], only the aa that differ are shown below the aa of the murine sequence. The published aa of the 7 S region starts with Lys in position nt 82 and ends before the Gly in position nt 732. The published human aa of the helical region starts with Tyr in position nt 1600. The published murine aa sequence [7] starts at nt 2818, the three aa at nt 2944, 3346 and 3349 that differ are marked. The signal peptide is boxed. The arrows indicate the beginning and the end of the 7 S region which is shaded. The interruptions in the Gly-X-Y repeat are shaded and numbered. The published DNA sequence of exon 14 [9] and the beginning of the cDNA clone pCIV-1-225 [8] are boxed.

Kuhn, K. (1985) Eur. J. Biochem. 152, 213–219.
[6] Babel, W. and Glanville, R.W. (1984) Eur. J. Biochem. 143, 545–556.
[7] Schuppan, D., Glanville, R.W. and Timpl, R. (1982) Eur. J. Biochem. 123, 505–512.
[8] Nath, P., Laurent, M., Horn, E., Sobel, M.E., Zon, G. and Vogeli, G. (1986) Gene 43, 301–304.
[9] Sakurai, Y., Sullivan, M. and Yamada, Y. (1986) J. Biol. Chem. 261, 6654–6657.
[10] Vogeli, G., Horn, E., Laurent, M. and Nath, P. (1985) Anal. Biochem. 151, 442–444.
[11] Vogeli, G., Horn, E., Carter, J. and Kaytes, P.S. (1986) FEBS Lett. 206, 29–32.
[12] Maniatis, T., Frish, D.F. and Sambrook, J. (1982) Cold Spring Harbor Laboratory, NY.
[13] Zagursky, R.J., Baumeister, K., Lomax, N. and Berman, M.L. (1985) Gene Anal. Techn. 2, 89–94.
[14] Oberbaumer, I., Laurent, M., Schwarz, U., Sakurai, Y., Yamada, Y., Vogeli, G., Voss, T., Siebold, B., Glanville, R.G. and Kuhn, K. (1985) Eur. J. Biochem. 147, 217–224.
[15] Watson, M.E.E. (1984) Nucleic Acids Res. 13, 5145–5164.
[16] Liotta, L.A., Abe, S., Robey, P.G. and Martin, G.R. (1979) Proc. Natl. Acad. Sci. USA 76, 2268–2272.