Proposed alignment of helical interruptions in the two subunits of the basement membrane (type IV) collagen

Gabriel Vogeli, Elizabeth Horn+, Jane Carter* and Paul S. Kaytes

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

Received 1 August 1986

We have isolated a cDNA clone for part of the α_2 type IV collagen (pCIV-2-176). Deoxynucleotide sequence analysis shows that this clone codes for 439 amino acids from the helical domain adjacent to the C-terminal globular domain of the α_2 (IV) chain. By aligning the deduced amino acid sequence of the α_2 (IV) chain with the published sequence for the α_1 (IV) chain, we find that all interruptions in the α_1 (IV) chain coincide with an interruption in the α_2 (IV) chain. Additonal interruptions in the α_2 (IV) chain exist, however, three out of the four analysed only slightly disturb the collagen triple helix.

Mouse *Collagen* *DNA sequencing* *Helical interruptions* *Basement membrane* *Alignment*

1. INTRODUCTION

Basement membranes are essential histological and physiological barriers in animals and humans [1,2]. Type IV collagen that forms the scaffold of the basement membrane [3] has two α_1 (IV) and one α_2 (IV) subunit [4]. The central part of each subunit consists of a Gly-X-Y repeat, that unlike the interstitial collagens, is interrupted in the α_1 (IV) collagen chain [5]. To determine whether such helical interruptions also exist in the α_2 (IV) collagen chain, and to analyse their position relative to the α_1 (IV) chain [5–8], we characterized a partial cDNA clone for the α_2 (IV) chain.

2. MATERIALS AND METHODS

The cDNA clone pCIV-2-176 was isolated from a mouse Engelbreth-Holm-Swarm (EHS) tumor cDNA library (in the *Pst*I site of pBR322) [7] by hybridizing the cDNA library [9] at low stringency (e.g. 0.9 M NaCl, 30% formamide at 42°C for 16 h, washed in 15 mM NaCl, 1.5 mM Na citrate at 25°C) with pCIV-1-225, an α_1 (IV) cDNA clone [7]. The hybridization signal from pCIV-2-176 was removed at high stringency (washing at 51°C). Standard protocols [10] were used to prepare, subclone and sequence the DNA.

3. RESULTS AND DISCUSSION

pCIV-2-176 was isolated from a cDNA library [7] prepared from RNA isolated from the Engelbreth-Holm-Swarm (EHS) tumor, tissue that is rich in basement membrane components [11]. We hybridized the cDNA library under non-restrictive conditions with labelled DNA from pCIV-1-225, an α_1 (IV) cDNA clone [7]. We argued that under these conditions cDNA clones related to collagen will hybridize, and that some of the positive colonies will code for the α_2 (IV) collagen chain. The identity of pCIV-2-176 as an α_2 (IV) collagen clone was established by sequence comparison (Kuhn, K. and Timpl, R., personal communication). pCIV-2-176 also overlaps at its

+ Present address: Oncor, Inc., Gaithersburg, MD 20879, USA

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

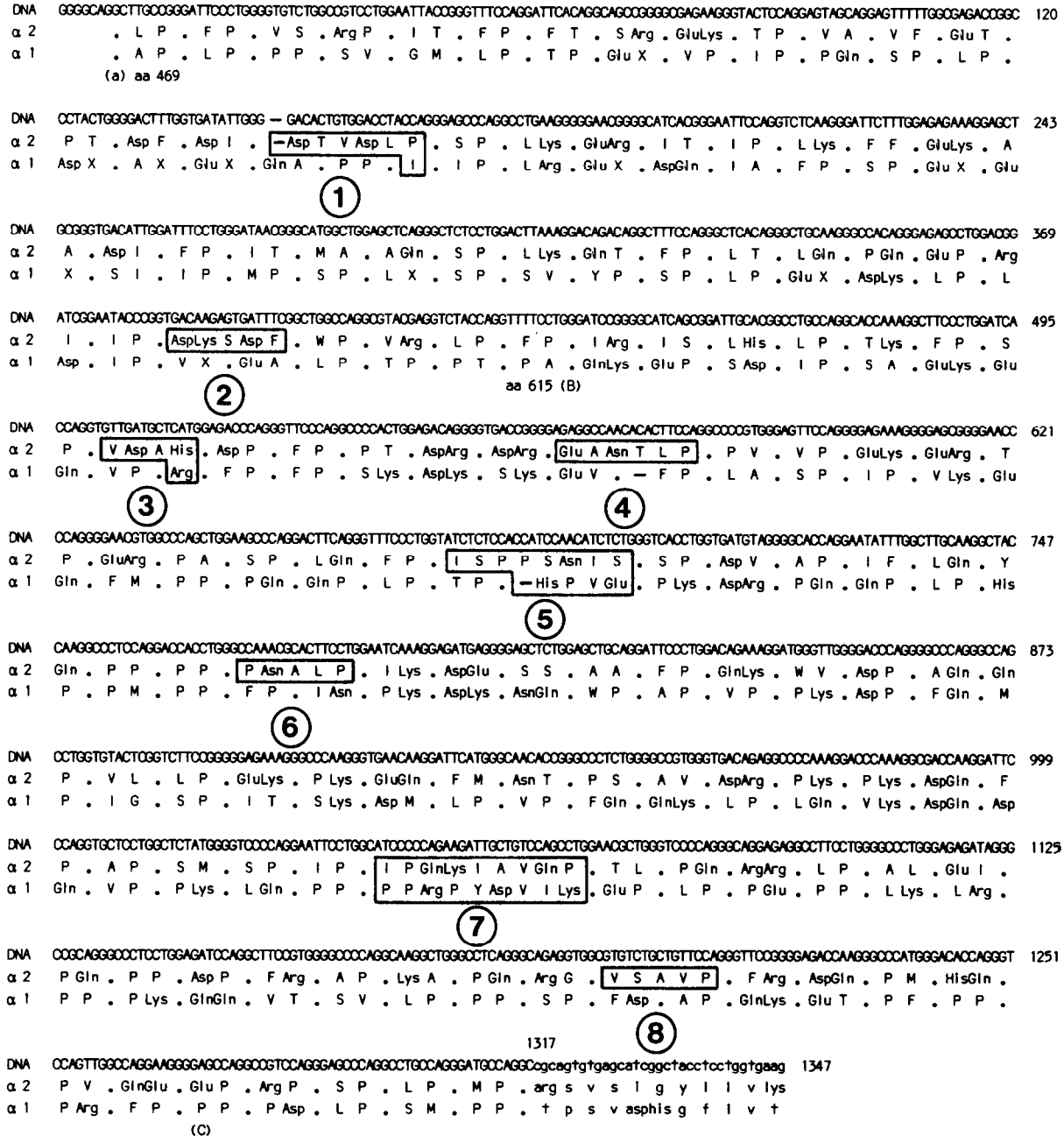

Fig.1. Deoxynucleotide sequence of pCIV-2-176, a cDNA clone for the α_2 (IV) collagen chain. The DNA sequence for the α_2 (IV) collagen chain is shown in the first lane (the numbering starts arbitrarily at the 5′-end of the clone), the derived amino acid sequence in the second lane. Line 3 shows the amino acid sequence for the α_1 (IV) collagen chain; from position A to B part of from the human α_1 (IV) sequence (amino acid no.469 to 615 [8]); followed by data from the mouse [7] to C and to the end [6]. The glycines essential for the typical collagen Gly-X-Y repeat are represented by a dot; the charged amino acids are written in the three-letter code; small letters indicate the C-terminal globular domain. The interruptions of the Gly-X-Y repeat in the α_2 (IV) chain are numbered. 'Missing' amino acids invoked to bring the Gly-X-Y repeat in phase after an interruption are marked by a dash.

N-terminal end with a published α_2 (IV) cDNA clone [12].

The DNA sequence from pCIV-2-176 for 1317 nucleotides and the deduced amino acid sequence for 439 amino acids of the helical Gly-X-Y repeat adjacent to the C-terminal globular domain are shown in fig.1. The border between the helical portion and the C-terminal domain is considered the nucleation point for triple helix formation in collagen [2]. Since pCIV-2-176 contains this area we tested the alignment of the amino acid sequence of the α_2 (IV) subunit with the published sequence of the α_1 (IV) subunit [6–8] starting with the last glycine in the last Gly-X-Y repeat of both subunits (see fig.1).

This portion of the α_2 (IV) collagen chain contains 8 interruptions, the α_1 (IV) chain only four. All interruptions seen in the α_1 (IV) chain (nos 1,3,5,7) that lack complete Gly-X-Y repeat units are also present at the same relative position in the α_2 (IV) chain. Three interruptions only found in the α_2 (IV) chain consist of a simple substitution for glycine in the Gly-X-Y repeat (nos 2,6,8), and one interruption (no.4) has an additional amino acid.

We find that our alignment of the two subunits positions the interruptions in the α_1 (IV) collagen chain opposite interruptions in the α_2 (IV) collagen chain. Our alignment is based upon (i) beginning the alignment at the nucleation point, and (ii) maximizing the length of the Gly-X-Y repeats between the interruptions. In addition, by using the human α_1 (IV) collagen sequence (fig.1 from the N-terminus to point B), we find that interruption no.1 in the α_2 (IV) of the mouse coincides with an interruption in α_1 (IV) from human. Our alignment of the two subunits is further supported by the fact that one interruption in the α_2 (IV) chain 73 amino acids further to the 5′-end, contained in a overlapping cDNA clone [12], is opposite an interruption in the α_1 (IV) chain [8] (not shown). For the above reasons, we assume that the register we have chosen for the presentation of our data (fig.1) is the most probable alignment.

The flexibility of the type IV collagen molecule seen in the electron microscope [13] can be explained by our finding that helical interruptions are opposite each other. Since the major interruptions (1,3,5,7) in the Gly-X-Y repeats of the two different chains of type IV collagen are in identical positions, we propose that the genes for the two collagen chains arose from the duplication of an ancestral gene which contained the common interruptions (1,3,5,7). We could then assume that the ancestral α_2 (IV) chain underwent further mutations in the course of evolution. We find that single base pair substitutions can account for the appearance of three (nos 2,4,8) of the 4 unique interruptions in the α_2 (IV) chain. In interruption no.2, Ser is substituted for Gly and in interruption nos 6 and 8, Ala is substituted for Gly. Either one of these substitutions will not affect the physical parameters of the collagen molecule substantially (Mao, B. and Vogeli, G., unpublished data from computer modelling). It is, however, also possible that the α_2 (IV) chain was the ancestral gene with many interruptions that were then repaired to form an α_1 (IV) subunit able to form a more stable triple helix. Further support for a common ancestral gene comes from the finding that some of the regions between the interruptions have a high homology to each other and that the C-terminal end of both chains has a high homology (Kaytes, P.S., unpublished).

During the submission of this manuscript a report of clones similar to ours appeared in the literature: Schwarz et al. (1986) Eur. J. Biochem. 157, 49–56.

ACKNOWLEDGEMENTS

We thank Dr R. Timpl and Dr K. Kuhn for amino acid sequencing data of the α_2 (IV) collagen; Boryeu Mao for advice on the 3 dimensional structure of collagen and Kathy Hiestand for help with the manuscript. The present work had been initiated by G.V. and E.H. at the Laboratory for Developmental and Molecular Biology, National Eye Institute, Bethesda, MD 20205.

REFERENCES

[1] Laurie, G.W. and Leblond, C.P. (1983) J. Histochem. Cytochem. 31, 159–163.

[2] Piez, K.A. and Reddi, A.H. (1984) Extracellular Matrix Biochemistry, Elsevier, Amsterdam, New York.

[3] Timpl, R., Wiedemann, H., Van Delden, V., Furthmayr, H. and Kuhn, K. (1981) Eur. J. Biochem. 120, 203–211.

[4] Tryggvason, K., Robey, P.G. and Martin, G. (1980) Biochemistry 19, 1284–1289.
[5] Schuppan, D., Timpl, R. and Glanville, R.W. (1980) FEBS Lett. 115, 297–299.
[6] Oberbaumer, I., Laurent, M., Schwarz, U., Sakurai, Y., Yamada, Y., Vogeli, G., Voss, X., Siebold, B., Glanville, R.W. and Kuhn, K. (1985) Eur. J. Biochem. 147, 217–224.
[7] Nath, P., Laurent, M., Horn, E., Sobel, M.E., Zon, G. and Vogeli, G. (1986) Gene, in press.
[8] Babel, W. and Glanville, R.W. (1984) Eur. J. Biochem. 143, 545–556.
[9] Vogeli, G., Horn, E., Laurent, M. and Nath, P. (1985) Anal. Biochem. 151, 442–444.
[10] Maniatis, T., Frish, E.F. and Sambrook, J. (1982) Cold Spring Harbor Laboratory, NY.
[11] Orkin, R.W., Gehron, P., McGoodwin, E.B., Martin, G.R., Valentine, T. and Swarm, R. (1977) J. Exp. Med. 145, 204–220.
[12] Kurkinen, M., Bernard, M.P., Barlow, D.P. and Chow, L.T. (1985) Nature 317, 177–179.
[13] Hofmann, H., Voss, T. and Kuhn, K. (1984) J. Mol. Biol. 172, 325–343.