Volume 224, number 2, 297–305 FEB 05307 November 1987

Extensive structural differences between genes for the α_1 and α_2 chains of type IV collagen despite conservation of coding sequences

Sirkka Liisa Hostikka and Karl Tryggvason

Biocenter and Department of Biochemistry, University of Oulu, 90570 Oulu, Finland

Received 30 September 1987

Analysis of the structure of the 3′-end of the human α_2(IV) gene demonstrated that the α_1(IV) and α_2(IV) genes have diverged extensively in spite of the apparent homology of the respective gene products. The NC-1 domain and the 3′-untranslated region are encoded by three exons in the α_2(IV) gene but five exons in the α_1(IV) gene. The two introns present in the NC-1 domain coding part of the α_2(IV) gene had the same location as two of the introns of the α_1(IV) gene. The junction exon in the α_2(IV) gene contains 53 bp coding for Gly-X-Y sequences whereas there are 71 bp in the α_1(IV) gene. Three other Gly-X-Y coding exons studied from the human α_2(IV) gene have sizes that differ from corresponding exons in the α_1(IV) gene and only one intron location matches here between the two genes. None of the exons studied has 54 bp or multiples thereof.

Collagen; Basement membrane; Evolution; (Human)

1. INTRODUCTION

The collagen molecules are composed of three α chains which associate to form a triple-helical conformation. The triple-helical region of the α chains is characterized by Gly-X-Y repeats where X and Y frequently represent proline or hydroxyproline [1].

To date, at least 12 distinct collagen types have been identified. Based on the characteristics of the proteins the different collagen types can be divided into three groups [1]: group 1 collagens contain large α chains (M_r > 95000) with a continuous helix; group 2 collagens have large (M_r > 95000) interrupted helices and group 3 collagens contain small α chains with discontinuous helices. Group 1 collagens consist of the fibrillar types I–III as well as the less characterized types V and XI. Group 2 contains types IV and VI–VIII which do not form fibrils and which are characterized by interruptions in the Gly-X-Y repeat sequences. Type IV collagen, the best characterized type of this group, is composed of two chains, α_1(IV) and α_2(IV) [2]. The carboxyl-terminal end of these chains is a globular structure termed NC-1, the rest being a discontinuous collagen sequence. The amino acid sequence of the NC-1 domains of both chains from man [3–5] and mouse [6–8] sources has been reported. About 70% of the amino acid sequence for the collagenous part of the human [3,9,10] and about 30% of the mouse [5,11,12] α_1(IV) chain have been reported and about 35% of the mouse α_2(IV) chain [13,14]. Sequence data have demonstrated that the NC-1 domains of the mammalian α_1(IV) and α_2(IV) chains share considerable interchain homology (63%) in the amino acid se-

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

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

quence [5]. A type IV collagen-like α chain has also been identified in *Drosophila* [15,16]. The sequence of the NC-1 domain of this chain is 59% identical with the α_1(IV) and α_2(IV) chains from both man and mouse [15,16]. In contrast, the overall homology between the amino acid sequence of the triple-helical domains of the two vertebrate type IV α chains is much lower or only about 22% when the invariant glycyl residues are excluded [5]. Similarly, the homology between the *Drosophila* α chain and the mammalian α_1(IV) and α_2(IV) chains is only 10 and 8.5%, respectively [15,16].

The division of collagen types into three groups appears to hold also remarkably well when considering the gene structure of respective collagen types. Thus, the exon structure of the genes for the α chains of the fibrillar collagens of types I–III and V is extremely homologous from avians to man [17–21]. Most of the Gly-X-Y coding exons have lengths of 54, 108 or 162 bp, the rest being 45 or 99 bp long. These observations have led to the hypothesis that the triple-helical domain of collagen may have evolved by tandem duplication of a 54 bp long ancestor element [17,19]. Studies on the genes for type IV collagen have shown the absence of 54 bp exons [22–26] which demonstrates that the genes for fibrillar and non-fibrillar collagens have diverged extensively if they derive from a common ancestor gene. The close relationship between the α_1(IV) and α_2(IV) genes is further signified by the finding that their gene loci are in the same region at the distal part of the long arm of chromosome 13 [27,28]. The genes for the other collagens of this group, types VI–VIII, have not yet been isolated. The gene for type IX collagen of group 3 has features in common with both the genes for group 1 collagens and the genes for type IV collagen. These features include 54 bp exons coding for Gly-X-Y repeats but also several other sizes and furthermore many glycine codons are split at the intron-exon junctions as found in the type IV collagen genes [29]. The gene for chick type X collagen, a short chain collagen found in hypertrophic cartilage, is very different in structure from other collagen genes because it does not contain intervening sequences [30].

Here, we describe the exon-intron structure of the 3′-end of the human α_2(IV) gene and compare it with the human α_1(IV) gene and the type IV collagen gene from *Drosophila*.

2. MATERIALS AND METHODS

2.1. *Materials*

Restriction endonucleases, T_4 DNA ligase, DNA polymerase I (Klenow fragment), nick-translation kit and proteinase K were purchased from Boehringer Mannheim (FRG) and used according to the suppliers' recommendations. ^{32}P-labeled dCTP and ^{35}S-labeled dATP were provided by New England Nuclear (Boston, MA). The cloning vectors, M13 mp18, M13 mp19 and pUC18, were supplied by Boehringer Mannheim and nitrocellulose filters were obtained from Schleicher and Schuell. Site-specific oligonucleotide primers were synthesized with a DNA synthesizer (Applied Biosystems).

2.2. *DNA clones*

cDNA clones (HD-3, HD-4) used for the screening of the genomic library code for about 30% of the 3′-end of the mRNA and have been described previously [5]. The human genomic *Alu*I/*Hae*III library in λ Charon 4A was a kind gift from Dr Eric Fritsch. The genomic library was infected to *E. coli* LE392 and plated. Replicate filters were made and screened with the nick-translated probes according to standard procedures [31]. One positive clone was isolated and mapped by restriction endonuclease analysis and Southern blotting.

2.3. *DNA sequencing*

The genomic clone studied here was digested with suitable restriction endonucleases and the fragments were subcloned into the M13 vector [32] for sequencing. The M13 clones were sequenced using the Sanger dideoxynucleotide method [33] either with the 'universal primer' or site-specific oligonucleotide primers.

3. RESULTS

3.1. *Identification of α_2(IV) genome clones*

Our recently characterized human α_2(IV) cDNA clones, HD-3 and HD-4, that code for 2.2 kb of the nucleotide sequence of the 3′-end of the mRNA [5], were used to screen 300000 colonies of the human genomic λ phage library. One positive clone, termed G-2.1, containing 12200 bp of genomic DNA was purified and characterized by restriction endonuclease mapping and Southern

hybridization using subcloned fragments of the cDNA clones and genomic subclones as probes (fig.1). Several appropriate genomic fragments, isolated using *Xba*I which has suitable restriction sites in the cDNA clones and *Pst*I, *Bam*HI and *Eco*RI (see fig.1) were subcloned into M13 [32] for sequencing with the Sanger procedure [33]. The sizes of the exons and determination of their approximate location in the gene could be obtained directly by nucleotide sequencing, either using the universal primer or synthetic oligonucleotides that were designed based on the human sequence known from the cDNA. As can be seen in fig.1, the six most 3′ exons encoding about 25% of the mRNA are contained in about 10 kb of genomic DNA. This indicates that the gene is very large, as is the case for the human α_1(IV) gene [22,23].

3.2. *Exon structure*

The sequencing analysis established several distinct features of the α_2(IV) gene (fig.1). Exon 1 from the 3′-end is 1088 bp long and codes for the entire 3′-untranslated region (833 bp) and 255 bp coding for a part of the NC-1 domain. Exon 2 is 287 bp long and exon 3 is a fusion exon containing 139 bp coding for the NC-1 domain and 53 bp coding for the collagenous domain. Exons 4 (117 bp), 5 (147 bp) and 6 (99 bp) encode the carboxyl-terminal end of the collagenous domain.

Because of the close structural relationship between the two type IV collagen polypeptide chains it was of interest to compare the structure of the α_2(IV) gene studied here with the comparable region of the α_1(IV) gene we have previously described [22,23]. This comparison (fig.2) demonstrates that although highly related, the two genes have indeed developed differently. First of all, in the α_2(IV) gene the 3′-untranslated region and the NC-1 domain are coded for by only three exons of which exon 3 is a junction exon containing coding sequences also for a part of the collagenous domain. In contrast, the NC-1 domain of the α_1(IV) gene is encoded by five exons. This difference in structure between the genes is surprising because as much as 63% of the amino acids of the NC-1 domains of the two chains in both man [3,5] and mouse [6–8] are identical. The two genes could therefore be expected to have similar structure. Alignment of the exon sequences of the α_1(IV) and α_2(IV) genes showed, however, that after all there is a considerable intergene relationship (fig.2). Thus, the coding sequences at the boundaries of exons 1 and 2 in the α_2(IV) gene and exons 2 and 3 in the α_1(IV) gene as well as their

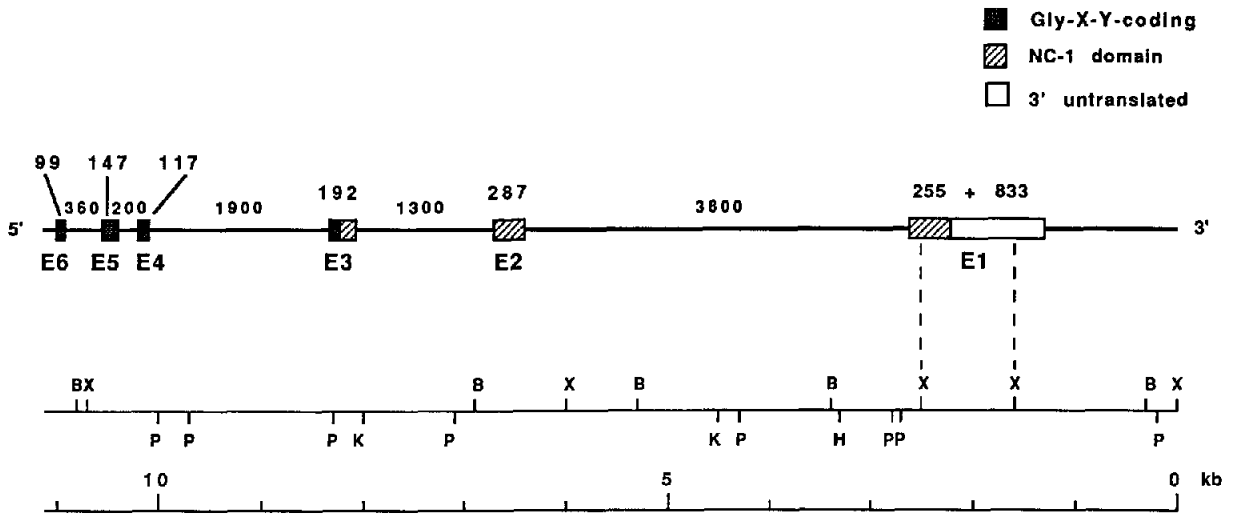

Fig.1. Restriction map, exon and intron sizes and locations in the α_2(IV) collagen coding G-2.1 genomic clone. Coding sequences are indicated by boxes. Sizes of exons (bp) are shown above the exons and the approximate sizes of introns are shown above the intron representing lines. Restriction endonuclease cleavage sites of *Bam*HI (B), *Hin*dIII (H), *Kpn*I (K), *Pst*I (P) and *Xba*I (X) in the G-2.1 clone are illustrated below the scheme of the genomic clone.

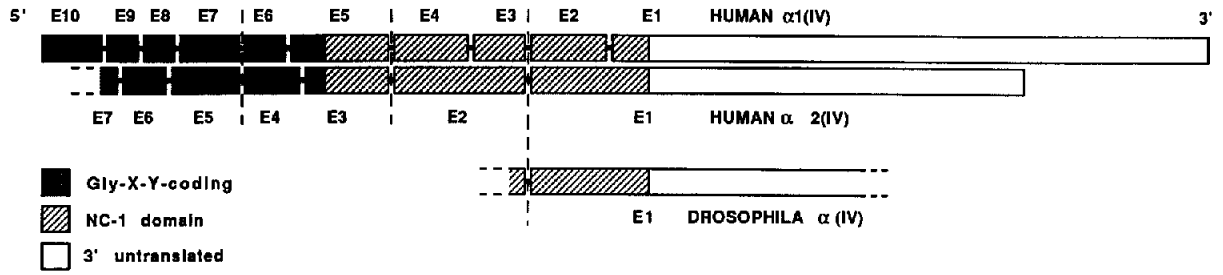

Fig.2. Comparison of the exon structure at the 3'-ends of the human genes for type IV collagen α_1 and α_2 chains. Exons are indicated by boxes and introns by the interconnecting lines. Exon (E) numbering from the 3'-end is shown. Interrupted vertical lines illustrate intron positions that are identical in both genes. Locations of the other intervening sequences do not match. An intron location in the type IV collagen gene of *Drosophila* [16] that is common with a mammalian location is also illustrated below the mammalian genes.

deduced amino acid sequence matched (fig.2). Similarly, the coding sequences at the boundaries of exons 2 and 3 of the α_2(IV) gene and exons 4 and 5 of the α_1(IV) gene matched (figs 2,3). The differences involve two extra introns and a longer 3'-untranslated region in the α_1(IV) gene as well as nucleotide substitutions and presumably deletions resulting in replacements and loss of amino acids so that the NC-1 domain of the α_1(IV) chain has two amino acids more than that of the α_2(IV) chain [5]. The total divergence between the NC-1 domains of the two human chains based on pairwise comparison is 35%. One intron position has been identified in the NC-1 domain coding region of the *Drosophila* type IV collagen gene [16] and interestingly, it matches perfectly with a location conserved between the human α_1(IV) and α_2(IV) genes (fig.2).

Another peculiar feature when comparing the α_1(IV) and α_2(IV) genes is the structure of the exons coding for the Gly-X-Y repeats, since the sizes of the exons compared here are all different (figs 1–3). The junction exon of the α_2(IV) gene has 53 nucleotides coding for such a Gly-X-Y repeat whereas the junction exon of the α_1(IV) gene has 71 bp. The next three 5'-exons of the α_2(IV) gene have sizes of 117, 147 and 99 bp whereas comparable exons in the α_1(IV) gene are 99, 129, 72 and 73 bp, respectively. Alignment of the Gly-X-Y repeat coding sequences of the genes established, however, structural similarities between the two genes. When calculated upstream from the beginning of the Gly-X-Y coding sequence the boundaries of exons 4 and 5 in the α_2(IV) gene and exons 6 and 7 from the α_1(IV) gene match (figs 2,3). However, as depicted in figs 2 and 3 the boundaries between exons 3 and 4 in the α_2(IV) gene and exons 5 and 6 in the α_1(IV) gene do not match nor do the boun-

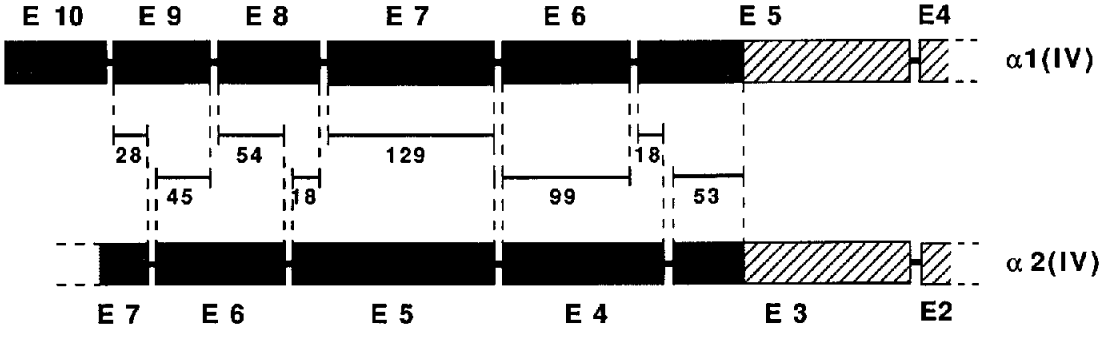

Fig.3. Illustration of common and mismatching exon-intron junctions in the 3'-end of the Gly-X-Y coding region of the human α_1(IV) and α_2(IV) genes. The reference point (0 bp) is the junction of the Gly-X-Y coding (dark box) and NC-1 domain (cross-hatched) coding sequences (see [5]). The positions of introns 4 and 6 in the α_1(IV) gene and introns 2 and 4 in the α_2(IV) gene, respectively, match perfectly. The lengths (bp) of mismatches between exons are shown.

daries of exons 5 and 6 from the α_2(IV) gene on the one hand and those of exons 7 and 8 from the α_1(IV) on the other. In both cases there is a mismatch of 18 bp (fig.3). Similarly, there is a 45 bp mismatch at the intron location between exons 6 and 7 in the α_2(IV) gene and exons 8 and 9 in the α_1(IV) gene (fig.3). On the other hand, the positions of introns between exons 6 and 7 in the α_2(IV) gene and exons 9 and 10 in the α_1(IV) gene have a 28 bp mismatch. Accordingly, although partially similar in structure, there has not been a strong evolutionary pressure to maintain the same sizes of Gly-X-Y coding exons in the two related type IV collagen genes. This is in clear contradistinction to the genes for the group I collagen genes, α_1(I), α_2(I), α_1(II) and α_1(III), that all have exactly the same exon size distribution of Gly-X-Y coding exons [17–20].

3.3. *Sequences of exon-intron junctions*

Analysis of the exon-intron junctions around exons 3–6 showed that all the Gly-X-Y exons studied here start with the second base of the codon for glycine (fig.4). Exon 2 that encodes sequences of the NC-1 domain also starts with the second base of a codon for glycine. In contrast, exon 1 starts with a complete codon CAC for histidine. The presence of split glycine codons appears to be typical for the type IV collagen genes, since present data [22,24,25] show that out of 13 exons reported to start with a codon for glycine, 10 start with the second base of the codon. Split codons for glycine are not found in the genes for fibrillar collagens (see [17–20]) which further demonstrates the evolutionary distance between the genes for type IV collagen and genes for fibrillar collagens.

4. DISCUSSION

The present results clearly established that the genes for the human α_1(IV) and α_2(IV) chains are

	5'										3'
EXON 6	tgttgcag	GG	CCC	AGG	GGT	...	...	GGA	TTC	CCT	G gtaagtga
99 bp		(Gly)	Pro	Arg	Gly			Gly	Phe	Pro	
EXON 5	ccttccag	GT	GCC	CCC	GGG	...	...	GGA	GAA	CCA	G gtagagtg
147 bp		(Gly)	Ala	Pro	Gly			Gly	Glu	Pro	
EXON 4	gcttgcag	GT	TTT	CGT	GGG	...	...	GGC	CAA	GAA	G gtgagtga
117 bp		(Gly)	Phe	Arg	Gly			Gly	Gln	Glu	
EXON 3	ctgcag	GT	GCA	CCA	GGC	...	...	CAG	GAC	CTG	G gtaggtac
192 bp		(Gly)	Ala	Pro	Gly			Gln	Asp	Leu	
EXON 2	ttccacag	GG	CTG	GCG	GGC	...	...	TTC	CTC	ATG	gtatgtgg
287 bp		(Gly)	Leu	Ala	Gly			Phe	Leu	Met	
EXON 1	ctgtacag		CAC	ACG	GCG	...	...	ATT	TTT	AAA	ccccgagt
1088 bp			His	Thr	Ala						
(255+833)											

Fig.4. Nucleotide sequence at the intron-exon boundaries of exons 1–6 of the human α_2(IV) gene (see figs 1–3). Lowercase letters are intron nucleotide sequences. The derived amino acid sequence is shown below exon coding sequences. The broken line flanking the 5'-end of exon 3 illustrates incompleted sequence.

closely related. This is in agreement with the extensive homology of primary structure reported for the NC-1 domain of the respective polypeptide chains [3–8]. It is also apparent that the difference in gene structure between the α_1(IV) gene on the one hand and genes for group 1 collagens on the other [22–24] also holds for the α_2(IV) gene. However, the results of this study also demonstrate that the two type IV collagen genes differ quite substantially. In fact, the differences in the structure of the two related type IV collagen genes are more prominent than those between the genes for the different types of group 1 collagens (types I–III, V). These observations raise questions about the evolution of the type IV collagen genes, of collagen genes in general but in particular why so much more divergence has been tolerated in the development of type IV collagen genes that code for homologous polypeptides present in the same molecule as opposed to the genes for group 1 collagens whose gene structure is stringently conserved.

As described above, the region of the type IV collagen α chain genes studied here contained nine exons coding for α_1(IV) chain sequences but only six exons coding for the corresponding region of the α_2(IV) chain. Another striking finding was that three intron locations are precisely conserved between the two mammalian genes, three intron locations in the collagen coding region mismatch and two introns are completely lacking from the NC-1 domain coding region of the α_2(IV) gene (figs 2,3). Interestingly, one of the two common intron locations in the NC-1 domain coding region of the mammalian genes is exactly conserved in a type IV collagen gene of *Drosophila* ([16] fig.2). As yet it is not known whether the other common intron locations of the α_1(IV) and α_2(IV) genes are conserved in *Drosophila*. As far as age is concerned, it is apparent that the type IV collagen gene must be more than 700 million years old, since it was present prior to the branchpoint of insects and vertebrates [34]. The present findings also indicate that the mammalian α_1(IV) and α_2(IV) genes have developed from the ancestor gene through a gene duplication followed by an evolutionary process. Whether the gene duplication occurred before or after the radiation of insects and vertebrates is not completely sure. The finding of only one type IV collagen chain in *Drosophila* [15] indicates, however, that the duplication took place after the branching of vertebrates.

The presence of common and uncommon intron locations in the mammalian and *Drosophila* genes shows that the evolutionary process of individual genes has involved either both selective and random loss of introns or alternatively both deletions and insertions of intron sequences. According to one hypothesis [35,36], selective loss of introns can explain cases where intron positions are not conserved between members of several multigene families [37–41]. If this hypothesis were solely to explain the evolutionary process of type IV collagen genes, the ancestor gene would have had at least five exons coding for the NC-1 domain and 3′-untranslated region still present in the α_1(IV) gene but also two 18 bp, one 28 bp, one 45 bp and one 54 bp exon coding for collagenous domains that have later been differentially fused as illustrated in fig.5. The presence of such exons in the ancestor gene would be necessary to explain the generation of the randomly formed mismatching introns because their position differences are too large to be accounted for by 'intron sliding' [42]. According to this hypothesis introns 6, 8 and 10 in the ancestor gene (fig.5) must have been deleted from the α_2(IV) gene and introns 5, 9 and 11 from the α_1(IV) gene after the duplication of the ancestor gene since their loss was 'random'. The random loss of these introns also suggests that they are not functionally important. In contrast, intron 2 in the ancestor gene could possibly contain some important regulatory elements because there may have been a selective pressure to maintain this intervening sequence. For example, a functional enhancer sequence has been found in the first 5′ intron of the human α_2(I) gene (De Wet and De Crombrugghe, personal communication).

Another hypothesis that could explain the divergent evolution of the type IV collagen genes states that common features of related genes represent the structure of an ancestor gene when it was first duplicated and that introns could since then have been inserted randomly into the gene. Such a pattern has recently been proposed for the fibrinogen gene [43]. This hypothesis cannot be rejected in the case of the mismatched introns in the α_1(IV) and α_2(IV) genes. Actually, it has been shown that at least in the β-globin gene exons smaller than 24 or 14 bp are spliced inefficiently,

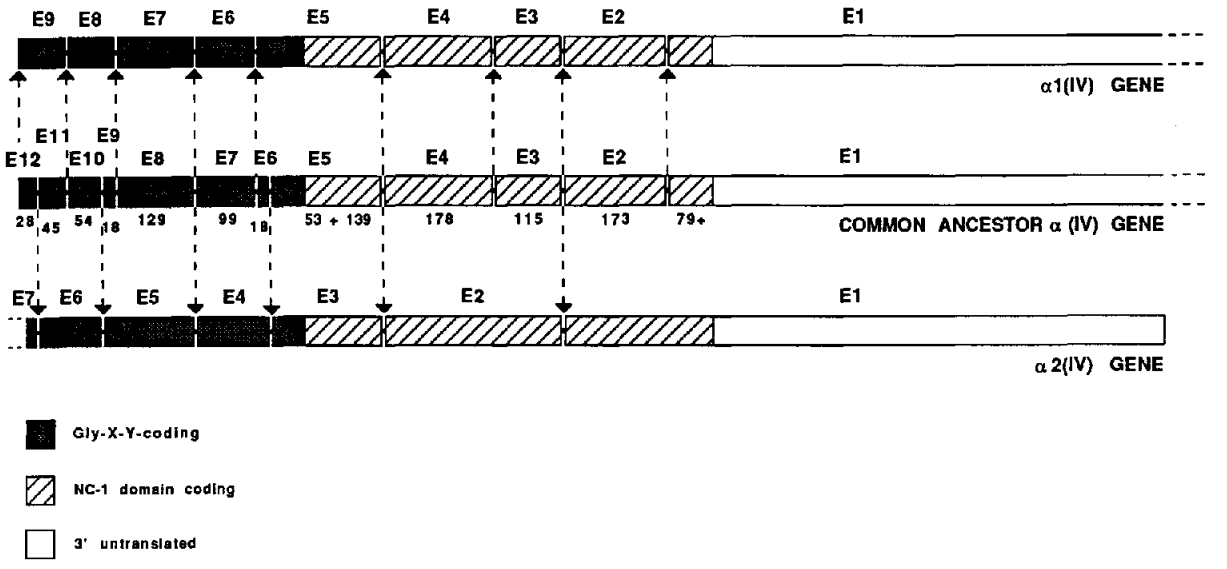

Fig.5. Potential evolutionary process of the 3'-end of the human α_1(IV) and α_2(IV) collagen gene after the duplication of a common ancestor gene. According to the hypothesis of Gilbert and others [36,37], differential loss of introns from a multi-intron ancestor gene can explain cases where intron positions are not always conserved between related genes. This hypothesis suggests that the part of the duplicated ancestor gene examined here had at least 11 introns. Only three of those, possibly functionally important, introns have been conserved whereas some of the other introns were lost from the α_1(IV) gene and others from the α_2(IV) gene.

if at all [44]. This finding cannot be generalized, however, as an 11 bp long exon containing a part of the sequence coding for the signal peptide cleavage site has been identified in the chicken proα_2(I) collagen gene [45]. The variability in exon sizes and also the locations of introns in the parts of the α_1(IV) and α_2(IV) genes that code for the Gly-X-Y repeats could also, at least partially, have developed through recombinational arrangements after an ancient common ancestor was duplicated. The presence of conserved sequence elements such as those coding for the Gly-X-Y triplets should favour such recombination [46]. The generation of 45 and 99 bp exons in the group 1 collagen genes may have occurred through unequal crossing-over between two 54 bp exons and the formation of 108 and 162 bp exons by recombination between the gene and a cDNA intermediate [19,46]. The occurrence of this kind of rearrangements cannot be excluded for the type IV collagen genes. The exons in the hypothetical common ancestor α(IV) collagen gene depicted in fig.5 contain 45, 54 and 99 bp long exons but it is difficult to envisage how the hypothesis of recombination of 54 bp primordial unit exons can be applied to explain the development of the variable sizes of exons in the type IV collagen genes. Rather, it appears that such recombination processes have mainly involved unequal crossing-over of single Gly-X-Y coding units of 9 bp in some cases with imperfections that have led to the formation of the interruptions typical for type IV collagen, and whose presence is tolerated in this collagen type.

The experimental data presented here provide further evidence that type IV collagen genes belong to a distinct class of collagen genes that have followed a different evolutionary pathway from that of the genes for interstitial collagens of group 1. The genes for group 1 collagens studied thus far have a highly conserved pattern of exon sizes (see [17–21]). All the genes have four exons coding for the carboxyl-terminal propeptides and six exons coding for the noncollagenous amino-terminal end of the α chains. Furthermore, the size distribution of exons coding for the triple-helical domain is identical with one exception [19]. This highly conserved distribution of exon sizes is very different from the diverged size pattern of exons in the type

IV collagen genes. This fact raises the question as to why such an evolutionary drift has been tolerated in the type IV collagen genes but not in the genes for group I collagens. One explanation for this could lie in the fact that the formation of collagen fibrils requires the lateral alignment of collagen molecules of an optimal length and this would not allow a drift in gene structure leading to α chains of variable length [19]. In contrast, an exactly equal length of collagen molecules may not be a prerequisite for the formation of type IV collagen networks, so that there has not been a severe functional constraint to maintain exact lengths of the discontinuous collagenous domain. Still, the fact that there is a conservation of sites of interruptions in the triple helix [5] indicates that the evolutionary changes of exons have not happened completely at random. On the other hand, there has clearly been evolutionary pressure to maintain the primary structure of the NC-1 domain of the type IV collagen α chains from *Drosophila* to man, probably because a certain structure of this domain is required for the proper assembly of the molecules as well as intermolecular crosslinking.

ACKNOWLEDGEMENTS

We thank Benoit De Crombrugghe and Raija Soininen for valuable comments. This work was supported in part by a grant from the National Research Council for Natural Sciences, Academy of Finland.

REFERENCES

[1] Miller, E.J. (1985) Ann. NY Acad. Sci. 460, 1–13.
[2] Timpl, R. and Dziadek, M. (1986) Int. Rev. Exp. Pathol. 29, 1–112.
[3] 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.
[4] Brinker, J.M., Gudas, L.J., Loidl, H.R., Wang, S.-Y., Rosenbloom, J., Kefalides, N.A. and Myers, J.C. (1985) Proc. Natl. Acad. Sci. USA 82, 3649–3653.
[5] Hostika, S.L., Kurkinen, M. and Tryggvason, K. (1987) FEBS Lett. 216, 281–286.
[6] 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.
[7] Schwarz-Magdolen, U., Oberbäumer, I. and Kühn, K. (1986) FEBS Lett. 208, 203–207.
[8] Kurkinen, M., Condon, M.R., Blumberg, B., Barlow, D., Quinones, S., Saus, J. and Pihlajaniemi, T. (1987) J. Biol. Chem. 262, 8496–8499.
[9] Babel, W. and Glanville, R.W. (1984) Eur. J. Biochem. 143, 545–556.
[10] Glanville, R.W., Qian, R., Siebold, B., Risteli, J. and Kühn, K. (1985) Eur. J. Biochem. 152, 213–219.
[11] Nath, P., Laurent, M., Horn, E., Sobel, M.E., Zon, G. and Vogeli, G. (1986) Gene 43, 301–304.
[12] Schuppan, D., Glanville, R.W. and Timpl, R. (1982) Eur. J. Biochem. 123, 505–512.
[13] Schwarz, U., Schuppan, D., Oberbäumer, I., Glanville, R.W., Deutzmann, R., Timpl, R. and Kühn, K. (1986) Eur. J. Biochem. 157, 49–56.
[14] Vogeli, G., Horn, E., Carter, J. and Kaytes, P.S. (1986) FEBS Lett. 206, 29–32.
[15] Blumberg, B., MacKrell, A.J., Olson, P.F., Kurkinen, M., Monson, J.M., Natzle, J.E. and Fessler, J.H. (1987) J. Biol. Chem. 262, 5947.
[16] Cecchini, J.-P., Knibiehler, B., Mirre, C. and LeParco, Y. (1987) Eur. J. Biochem. 165, 587–593.
[17] Boedtker, H., Finer, M. and Aho, S. (1985) Ann. NY Acad. Sci. 460, 85–116.
[18] Ramirez, F., Bernard, M., Chu, M.-L., Dickson, L., Sangiorgi, F., Weil, D., De Wet, W., Junien, C. and Sobel, M. (1985) Ann. NY Acad. Sci. 460, 117–129.
[19] De Crombrugghe, B., Schmidt, A., Liau, G., Setoyama, C., Mudryj, M., Yamada, Y. and McKeon, C. (1985) Ann. NY Acad. Sci. 460, 154–162.
[20] Upholt, W.B., Strom, C.M. and Sandell, L.J. (1985) Ann. NY Acad. Sci. 460, 130–140.
[21] Weil, D., Bernard, M., Gargano, S. and Ramirez, R. (1987) Nucleic Acids Res. 15, 181–198.
[22] Soininen, R., Tikka, L., Chow, L., Pihlajaniemi, T., Kurkinen, M., Prockop, D.J., Boyd, C.D. and Tryggvason, K. (1986) Proc. Natl. Acad. Sci. USA 83, 1568–1572.
[23] Soininen, R., Chow, L., Kurkinen, M., Tryggvason, K. and Prockop, D.J. (1986) EMBO J. 5, 2821–2823.
[24] Sakurai, Y., Sullivan, M. and Yamada, Y. (1986) J. Biol. Chem. 261, 6654–6657.
[25] Kurkinen, M., Bernard, M.P., Barlow, D.P. and Chow, L.T. (1985) Nature 317, 177–179.

[26] Monson, J.M., Natzle, J.E., Friedman, I. and McCarthy, B.I. (1982) Proc. Natl. Acad. Sci. USA 79, 1761–1765.
[27] Boyd, C.D., Weliky, K., Toth-Fejel, S., Deak, S.B., Christiano, A.M., Mackenzie, J.W., Sandell, L.J., Tryggvason, K. and Magenis, E. (1986) Hum. Genet. 74, 121–125.
[28] Griffin, C.A., Emanuel, B.S., Hansen, J.R., Cavenee, W.K. and Myers, J.C. (1987) Proc. Natl. Acad. Sci. USA 84, 512–516.
[29] Vasios, G., Nishimura, I., Konomi, H., Van der Rest, M., Ninomiya, Y. and Olsen, B.R. (1987) submitted.
[30] Ninomiya, Y., Gordon, M., Van der Rest, M., Schmid, T., Linsenmayer, T. and Olsen, B.R. (1986) J. Biol. Chem. 261, 5041–5050.
[31] Maniatis, T., Fritsch, E. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[32] Messing, J. (1983) Methods Enzymol. 101, 20–78.
[33] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[34] Holmquist, R., Jukes, T.H., Moise, H., Goodman, M. and Moore, G.W. (1976) J. Mol. Biol. 105, 39–74.
[35] Gilbert, W. (1979) in: Eucaryotic Gene Regulation, ICN-UCLA Symposia on Molecular and Cellular Biology (Axel, R. et al. eds) p.1, New York Academic Press.
[36] Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, I. (1980) Cell 20, 555–566.
[37] Kondel, D.A., Tilgham, S.M. and Leider, P. (1978) Cell 15, 1125–1132.
[38] Bernard, O., Hozum, N. and Tonagawa, S. (1987) Cell 15, 1133–1142.
[39] Inana, G., Piatigorsky, J., Nosman, B., Slingsby, C. and Blundell, T. (1983) Nature 302, 310–314.
[40] Steinmetz, M., More, K.W., Frelinger, J.G., Sher, B.T., Sher, F.W., Boyse, E.A. and Hood, L. (1981) Cell 25, 683–692.
[41] Eiferman, F.A., Young, P.R., Scott, R.W. and Tilgham, S.M. (1981) Nature 294, 713–718.
[42] Craik, C.S., Ruffer, W.J. and Fletterick, R. (1983) Science 220, 1125–1129.
[43] Crabtree, G.R., Comeau, C.M., Fowlkes, D.M., Fornace, A.J., Malley, J.D. and Kant, J.A. (1985) J. Mol. Biol. 185, 1–19.
[44] Furdon, P.J. and Kole, R. (1986) Proc. Natl. Acad. Sci. USA 83, 927–931.
[45] Aho, S., Tate, V. and Boedtker, H. (1984) Nucleic Acids Res. 12, 6117–6125.
[46] Yamada, Y., Liau, G., Mudryj, M., Obici, S. and De Crombrugghe, B. (1984) Nature 310, 333–337.