A Novel and Highly Conserved Collagen (proα1(XXVII)) with a Unique Expression Pattern and Unusual Molecular Characteristics Establishes a New Clade within the Vertebrate Fibrillar Collagen Family*

Received for publication, December 18, 2002, and in revised form, May 22, 2003 Published, JBC Papers in Press, May 23, 2003 DOI 10.1074/jbc.M212889200

Raymond P. Boot-Handford‡, Danny S. Tuckwell§, Darren A. Plumb‡, Claire Farrington Rock¶, and Richard Poulsom**

From the ‡Wellcome Trust Centre for Cell-Matrix Research, School of Biological Sciences, University of Manchester, Manchester M13 9PT, United Kingdom and the ¶Histopathology Unit, Cancer Research UK, 44 Lincoln’s Inn Fields, London WC2A 3PX, United Kingdom

The type XXVII collagen gene codes for a novel vertebrate fibrillar collagen that is highly conserved in man, mouse, and fish (Fugu rubripes). The proα1(XXVII) chain has a domain structure similar to that of the type B clade chains (α1(I), α3(V), α1(XI), and α2(XI)). However, compared with other vertebrate fibrillar collagens (types I, II, III, V, and XI), type XXVII collagen has unusual molecular features such as no minor helical domain, a major helical domain that is short and interrupted, and a short chain selection sequence within the NC1 domain. Proα1(XXVII) mRNA is 9 kb and expressed by chondrocytes but also by a variety of epithelial cell layers in developing tissues including stomach, lung, gonad, skin, cochlear, and tooth. By Western blotting, type XXVII antisera recognized multiple bands of 240–1000 amino acid residues. Phylogenetic analyses have revealed that the previously known vertebrate fibrillar collagens fall into two related but distinct groups or clades (3, 4). The type A clade consists of the proα1(I), proα2(I), proα1(II), proα1(III), and proα2(V), whereas the type B clade contains the remaining chains encoding types V and XI collagens (with the exception of proα3(XI)), which is derived from the COL2A1 gene. The division of the vertebrate fibrillar collagens into two clades is supported by two further observations. Firstly, the exon structures of the genes are virtually identical within a clade yet distinct between clades (5). Secondly, the members of each clade share homologous N-terminal noncollagenous domains (von Willebrand factor type C domain for type A and TSPN1 for type B clade members) with the exception of the proα2(I) chain, where the N-terminal noncollagenous domain appears to have been deleted (6).

We have recently described how the members of the two clades of vertebrate fibrillar collagens have apparently arisen early during vertebrate evolution from a single founder gene through the mechanism of “molecular incest” (7). Molecular incest is catalyzed by gene/genome duplications, events that not only characterize but are thought by many to have promoted vertebrate evolution (8). As a result, the proα chains encoding the known vertebrate fibrillar collagens (type I, II, III, V, and XI) are very closely related and interdependent in that gene products from the different clades can trimerize in complex patterns, and the resulting collagen molecules can co-polymerize in mixed fibrils. For instance, the proα1(II) chain forms a homotrimer (type II collagen) but can also heterotrimerize with the proα1(XI) and proα2(XI) chains in some forms of type XI collagen (2). Some chains of type V and type XI can heterotrimerize to form a hybrid V/XI collagen (9, 10). During fibril assembly, types III and V collagens can co-polymerize with type I collagen, and likewise, type XI co-polymerizes with type II collagen (2).

The process of vertebrate collagen evolution by molecular incest following genome duplication is perhaps most apparent in modern day fish (teleosts). Fish appear to have undergone a further round of genome duplication after their divergence, ~450 million years ago, from the lineage leading to mammals (11). As a consequence, a number of the “classical” vertebrate fibrillar collagen genes found in man and mouse are present in...
the puffer fish (Fugu rubripes) genome as duplicates that are conserved yet, at the same time, slowly diverging in sequence (7). It was while identifying fibrillar collagen genes in the puffer fish genome that we discovered a novel fibrillar collagen gene that was not only duplicated within the fish genome but also highly conserved in man and mouse. The new fibrillar collagen, pro/H9251(XXVII), has unusual molecular characteristics and a unique expression pattern compared with the classical vertebrate fibrillar collagens. The COL27A1 gene has evolved as an early offshoot of the lineage leading to the type B clade and, together with a second closely related gene, COL24A1, represents a new and third clade (type C) in the vertebrate family of fibrillar collagens.

EXPERIMENTAL PROCEDURES

Sequence Determination—Sequences from the collagen XXVII gene were originally identified by blast searches of the F. rubripes genome data base looking for amino acid sequence homologies with C-terminal noncollagenous domains of human fibrillar collagens. The Fugu sequences identified in this fashion were compared directly with the human to identify the collagen chain and type. In this way, F. rubripes sequences matching a previously uncharacterized fibrillar collagen conserved in both human and mouse genomes were identified and assembled by standard bioinformatics approaches including the use of available expressed sequence tags. The majority of the genomic sequences were identified directly in the human genome through expressed sequence tag clone analysis. In F. rubripes, direct translation of genomic sequence was used to identify the majority of the exons. The alignments were performed using the Clustal W (1.8) engine.

Northern and in Situ Analyses—IMAGE clones 1712826 (human) and 5289947 (mouse) were utilized for the Northern and in situ analyses. Northern blots of poly(A) RNA were obtained from BD Biosciences. cDNA inserts were 32P-labeled to a specific activity in excess of 4 x 10^8 cpm/μg DNA using an oligonucleotide labeling kit (Ready-To-Go DNA Labeling Beads; Amersham Biosciences). The Northern blots were hybridized and washed using the Church system (12), and the images were obtained either by PhosphorImager analyses or exposure to x-ray film. In situ hybridization was conducted using 35S-labeled riboprobes (13).

Antisera and Western Blotting—An antiserum was produced in rabbit to two synthetic peptides derived from the mouse variable region domain (TSPTKRSPTKPSVS and SKKFTNPTVAKSKSK) coupled by way of a C-terminal cysteine to keyhole limpet hemocyanin (Eurogentec S.A). Tissue extracts were prepared by boiling in Laemmli sample buffer containing 1% (v/v) -mercaptoethanol. 24-h conditioned medium (serum-free) from ATDC 5 cells grown under differentiation conditions for 5 days (14) was dialyzed into 50 mM ammonium bicarbonate, lyophilized, and dissolved in Laemmli sample buffer under reducing conditions. For collagenase digestion, lyophilized samples were resuspended in 50 mM Tris-HCl, pH 7.4, containing 1 M NaCl, 1% (v/v) Nonidet P-40, and digested (37 °C, 60 min) with highly purified bacterial collagenase (CLSNA, Worthington Biochemical Corp.) at a final concentration of 0.5 unit/μl in a 16-µl reaction volume. The reactions were terminated by the addition of 4 μl of Laemmli sample buffer (5× concentration). The samples were resolved on a 6% polyacrylamide gel, transferred to nitrocellulose, blocked with 5% (w/v) milk powder in TBST (20 mM Tris-

FIG. 1. Complete amino acid sequence and domain structure of type XXVII collagen; comparisons across species (A) and with NC1 domains from different fibrillar collagens (B). A, alignment of type XXVII proα chains from human (HaαXXVII), mouse (MaαXXVII), and the duplicated chains in F. rubripes (FaαXXVII and Fa2XXVII). Consensus sequence is indicated in terms of identical residues (*), conservative substitutions (:), and residues with similar properties (.). Cysteine residues (yellow) and interruptions in the triple helical domain (red) are indicated. B, alignment of the NC1 domain sequence of human type XXVII collagen with the equivalent sequences from invertebrate (hydra, cnidarian (28); abalone, mollusc (21); and sea urchin, echinoderm (27)) and the other human fibrillar collagens. Cysteine residues (yellow) and the chain selection sequence (boxed) (20) are indicated.
HCl pH 7.5, 150 mM NaCl, 0.05% (v/v) Tween 20), incubated with
primary antisera at a 1:100 dilution, and, after washing three times in
TBST, incubated with horseradish peroxidase-conjugated swine anti-
rabbit IgG (Dako) at a 1:3000 dilution. Finally, the bands were detected
using chemiluminescence (ECL; Amersham Biosciences).

Phylogenetic Analyses
— Phylogenetic analyses were performed essentially as described (15) using the following sequences:

- H9251 (I) human, CA11_HUMAN;
- H9251 (II) human, HUMCOL2A1Z_1;
- H9251 (III) human, CA13_HUMAN;
- H9251 (V) chick, A40695;
- H9251 (V) human, CGHU1V/CA15_HUMAN;
- H9251 (IX) chick, B34493;
- H9251 (IX) human, P20849/CA2B_HUMAN;
- H9251 (X) mouse, A53317;
- H9251 (XII) chick, S23810;
- H9251 (XII) human, BAA07368.1;
- H9251 (XVIII) human, AAC39659.1;
- H9251 (XIX) human, BAA23578.1;
- H9251 (XX) chick, AAN03650;
- H9251 (XXI) human, A56101;
- H9251 (XXIV) human, CA25_HUMAN; o3(V) mouse, Q9JL2; o4(V) rat, Q6J04; o1(IX) chick, B34493; o1(IX) human, P20849/CA1B_HUMAN; o1(IX) mouse, Q05722; o1(XI) human, AAF04725.1; o2(XI) human, P13942/CA2B_HUMAN; o2(XII) mouse, Q64739; o1(IX) chick, P13944; o1(IX) human, Q99715; o1(IX) mouse, Q60847; o1(IX) chick, S1211; o1(IX) human, CAAT24902.1; o2(IX) human, S23810; o1(IX) human, BAA07368.1; o1(IX) human, AAC39659.1; o1(IX) mouse, A56101; o2(IX) mouse, BAA23578.1; o2(IX) chick, (16); CLE-1B, AF164959; o2(IX) hu-
man, AAN03650; Alvinella pompejana fibrillar collagen, AF053538.1;
Collagen XXVII: a New Vertebrate Fibrillar Collagen

**Table I**

<table>
<thead>
<tr>
<th></th>
<th>TSPN</th>
<th>Variable</th>
<th>Collagenous</th>
<th>NCI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% identity</td>
<td>aa</td>
<td>pl</td>
<td>aa</td>
</tr>
<tr>
<td>Human proo1(XXVII)</td>
<td>100</td>
<td>154</td>
<td>11.74</td>
<td>403</td>
</tr>
<tr>
<td>Mouse proo1(XXVII)</td>
<td>81</td>
<td>156</td>
<td>11.54</td>
<td>387</td>
</tr>
<tr>
<td>Fugu proo2(XXVII)</td>
<td>55</td>
<td>150</td>
<td>9.46</td>
<td>226</td>
</tr>
<tr>
<td>Fugu proo2(XXVII)</td>
<td>47</td>
<td>153</td>
<td>10.86</td>
<td>313</td>
</tr>
<tr>
<td>Human proo1(XXVII)</td>
<td>36</td>
<td>154</td>
<td>5.16</td>
<td>261</td>
</tr>
</tbody>
</table>

**RESULTS**

Sequence Analyses and Comparisons—The aligned sequences of the human and mouse proo1(XXVII) and the duplicated *F. rubripes* (proo1 and proo2(XXVII)) chains are presented in Fig. 1A. The sequences for human and mouse were derived from a combination of expressed sequence tag analyses and direct identification of conserved amino acid sequences in translated genome sequences and are identical to those recently posted on sequence databases (accession numbers AY149237 and AY167568) and designated proo1(XXVII). The signal peptide sequences for human and mouse, which are presented for completeness, are from AY149237 and AY167568. The signal peptides for the *Fugu* genes are not included because they are not conserved and could not therefore be identified with any degree of certainty from the genomic sequences.

The type XXVII proo chains in man, mouse, and fish (in which the gene has been duplicated and the chains therefore designated a1 and a2(XXVII)), are highly conserved (Fig. 1A) and have unmodified molecular masses of 185 kDa (human and mouse) and 165–170 kDa (*Fugu* chains). The N terminus of type XXVII collagen proo chains has a similar domain structure to that exhibited by the type B clade vertebrate fibrillar collagens including a TSPN domain of ~150 residues and a variable domain ranging in size from 403 residues in man to 226 residues in the *Fugu proo1(XXVII)* chain (Fig. 1A and Table I). The TSPN is highly conserved between man, mouse, and fish (Table I). Although the variable domain is relatively well conserved between man and mouse (51% identity), no conservation is apparent when comparing the *Fugu* sequences either with each other or with those derived from the mammalian species (Fig. 1A). However, the chemical nature of the variable domain is conserved in that the mammalian and both fish sequences have high pl values (9.46–11.74), and this is in marked contrast to the low pl exhibited by the equivalent domain of the related but distinct proo1(XXIV) chain (Table I). In collagen XXVII, the variable domain fuses directly with the major triple helical domain (Fig. 1A), and these proo chains therefore lack the telopeptide and minor triple helical domains found in all previously characterized vertebrate fibrillar collagen chains. In addition, the highly conserved collagen XXVII triple helical domain (994–997 residues; Table I) is both shorter than the equivalent domain in the previously described vertebrate fibrillar collagens (1014–1020 residues) and contains two conserved interruptions in the Gly-X-Y repeat (Fig. 1A). The *Fugu* proo1 chain contains a further two (nonconserved) interruptions in the collagenous domain (Fig. 1A). The NCI domain or C-propeptide is joined to the collagenous domain by a short relatively variant sequence of 15–33 residues, depending on the species. The remaining ~200 residues of the NCI domain are highly conserved (Fig. 1A and Table I) with in excess of 50% identity between man, mouse, and *Fugu* chains. The NCI domain of proo1(XXVII) is also highly conserved compared with the C-propeptide sequences of fibrillar collagens from across the animal kingdom and contains all 8 of the characteristic cysteine residues (Fig. 1B). It is particularly noteworthy, however, that unlike the proo chains of classical vertebrate fibrillar collagens (types I, II, III, V, and XI), proo1(XXVII) contains a short chain selection sequence, a feature previously only found in fibrillar collagens of invertebrates (Fig. 1B and Ref. 7).

Gene Expression—Northern analysis of poly(A) RNA extracted from fetal human tissues were hybridized to 32P-labeled cDNA encoding the NCI domain of proo1(XXVII). The mRNA encoding type XXVII collagen is ~9 kb. The mRNA for collagen XXVII is relatively highly expressed in fetal liver and pancreas with lower levels being found in heart, brain, skeletal muscle, and kidney. Type III collagen mRNA gives an entirely different profile (Fig. 2).

To examine the expression of proo1(XXVII) mRNA in more detail, a series of in situ analyses was performed. In fetal human tissues, collagen XXVII expression is particularly strong in chondrocytes in the cartilaginous anlagen of the developing skeleton and in tracheal cartilage (Fig. 3, A–F). The
gene is also expressed in the epidermal region of fetal skin (Fig. 3, G–I), by the developing mucosal layer in fetal stomach (Fig. 3, J–L), by bronchial epithelium in developing lung (Fig. 3, M–O), and in epidermal-derived cells in hair follicles but not by the dermal papilla at the base of the follicle (Fig. 3, P–R). In fetal mouse tissues, the pattern of expression of collagen XXVII is similar to that seen in human with strong expression in cartilaginous anlagen of skeletal elements and epithelial cell layers in developing lung and stomach (Fig. 4, D–H). Collagen XXVII expression is evident in cochlear epithelium as well as the cartilaginous structures associated with inner ear development (Fig. 4, A–C). Sagittal sections through the head (Fig. 4, J and K) reveal strong expression in both nasal cartilage and ameloblasts, the epithelial cells responsible for synthesizing tooth enamel.

**Western Analyses**—Antiserum to two synthetic peptides containing sequences derived from the variable domain of mouse procollagen XXVII were generated in rabbits. Western blots of SDS extracts of newborn mouse skin and tail revealed a fairly complex pattern of polypeptides with major bands at 240, 150, and 110 kDa (Fig. 5). This pattern presumably arises from a combination of covalent cross-linking and degradation. Cells in culture frequently secrete procollagens and partially processed collagens into the medium presenting a simpler pattern to interpret than that found in tissues. We therefore Western blotted the conditioned medium from a chondrogenic cell line
FIG. 4. **Expression of type XXVII collagen in fetal mouse tissues.** Darkfield images of $^{35}$S-labeled antisense riboprobes coding for type XXVII collagen (B, E, H, and K) or β actin (C, F, and I) hybridized to sections through embryos at embryonic days 14.5 (A–I) and 19 (J and K). A–C represent enlargements of the square-boxed areas in D–F, and likewise, G–I represent enlargements of the rectangular-boxed areas. The arrows in A and B indicate the cochlear epithelium. G, I, lung; g, gonad; s, stomach. J, nc, nasal cavity; it, incisor tooth; mt, molar tooth. Brightfield images (A, D, G, and J) are also presented.
Collagen XXVII: a New Vertebrate Fibrillar Collagen

Figure 5. Western blot of mouse tissue extracts and conditioned ATDC5 cell culture medium probed with an antisera to type XXVII collagen. Duplicate tracks were probed with either the antiserum (I) or the preimmune serum (PI). Samples of conditioned medium were digested with highly purified bacterial collagenase as described under "Experimental Procedures." Conditioned medium (M), conditioned medium plus collagenase (M+C), and the collagenase enzyme alone (C) were used.

The antisera specifically recognized a doublet at ~150 kDa and a third band at 140 kDa (Fig. 5). These bands are collagenous, as demonstrated by their susceptibility to highly purified bacterial collagenase digestion (Fig. 5, lane M+C). Unfortunately, this antiserum was not of use for immunohistochemical analyses.

Phylogenetic Analyses—Distance and parsimony analyses of the NC1 (C-propeptide) domains of collagens pro1(XXVII), pro1(XXIV), and other fibrillar collagens showed that pro1(XXVII) and pro1(XXIV) form a separate and distinct clade (Type C) of fibrillar collagens, supported by very high bootstrap values (Fig. 6, A and B, respectively). Significantly, pro1(XXVII) and pro1(XXIV) can be distinguished from the two previously described clades within the family of vertebrate fibrillar collagens (Types A and B; Fig. 6, A and B). Invertebrate NC1 sequences were included in these analyses to root the trees. Both distance and parsimony analyses indicated that the collagen XXVII and XXIV NC1 sequences are more closely related to vertebrate fibrillar collagens of the type B rather than type A clade.

Distance and parsimony analyses of the N-terminal TSPN domains of pro1(XXVII), pro1(XXIV), and other collagens again showed that pro1(XXVII) and pro1(XXIV) form a distinct clade (Type C; Fig. 6, C and D, respectively) strongly supported by bootstrap values and clearly separate from both the fibrillar type B clade and the FACIT collagen clade. The trees were rooted using collagen XVIII and its Caenorhabditis elegans homologue CLE-1B, because this is the only known TSPN domain collagen with an invertebrate homologue. The distance analysis (Fig. 6C) suggests that the TSPN domains of the FACIT and type B clade fibrillar collagens are more closely related to each other than to the TSPN domains of collagens XXVII and XXIV. However, the parsimony analyses (Fig. 6D) suggest that the TSPN domains of the FACIT and type B clade fibrillar collagens are more equally related to each other.

Structure of the XXVII Gene—The human COL27A1 gene is ~156 kb in length and located on chromosome 9q32. The mouse gene is ~118 kb and is located on chromosome 4c1. The F. rubripes genes are considerably shorter (~a1 gene, 37 kb, located on scaffold 2117, partial gene sequence reported under accession number SINFRUG00000038690; ~a2 gene, 41 kb, located on scaffold 1015, partial gene sequence reported under accession number SINFRUG00000034716).

The exon organization of the genes in man, mouse, and fish is virtually identical. The first two exons in man and mouse code for the signal peptide. The third exon encodes the remainder of the N-terminal noncollagenous sequence (TSPN and variable domains plus the first 4 Gly-X-Y repeats of the collagenous domain in man, mouse, and Fugu a2 and 5 Gly-X-Y repeats in Fugu a1). Because of the lack of sequence conservation in the variable domains, the size of the very large exon 3 differs considerably between species (human, 1775 bp; mouse, 1736 bp; Fugu a1, 1245 bp; Fugu a2, 1497 bp). The size of exons encoding the collagenous domain in the type XXVII gene follows a similar but distinct pattern (based on 54- and 45-bp motifs) to that exhibited by the genes encoding the type A and B clades of vertebrate fibrillar collagens (Fig. 7). Exons 4–36 of the collagen XXVII gene coding for the N-terminal region of the collagenous domain appear to have undergone no "exon fusion" events, whereas several such events appear to have occurred independently in each of the progenitor genes for the type A and B clades. In contrast, the more C-terminal parts of the collagenous domains in all three clades (equivalent to exons 38–56 in the type XXVII gene) appear to have been subjected to numerous separate exon rearrangement events (Fig. 7). Comparisons of exon size/pattern across all three clades of vertebrate fibrillar collagen suggests that in the type XXVII genes, one of the most 5-prime 54 bp exons has been deleted in comparison with the genes of other clades (Fig. 7), accounting for the shorter collagenous domain. The conserved interruptions in the collagenous domains of collagen XXVII genes are due to base substitutions in exon 6 (54 bp) and a 3-bp insertion within exon 42 (57 bp; Fig. 7). The NC1 domain is encoded by exons 56 (18 bp of Gly-X-Y and 9 bp of noncollagenous sequence) and exons 57–61 (66, 169, 110, 219, and 147 bp including the stop codon, respectively).

Discussion

Fibrillar collagens are among the most abundant proteins in the vertebrate body. Until very recently, there was thought to be only five types of vertebrate fibrillar collagens (I, II, III, V, and XI) encoded by nine distinct genes. We report here the initial characterization of a new member of the vertebrate fibrillar collagen family, collagen XXVII, that is highly conserved in man, mouse, and puffer fish. When examining the amino acid alignments presented in Fig. 1A, it is must be remembered that the puffer fish proa1 and proa2(XXVII) sequences have been evolving in the same organism but as separate genes, because of a teleost-specific gene duplication occurring soon after the fish lineage diverged from that leading to mammals 450 million years ago (11, 21). In contrast, the human and mouse sequences have been evolving in separate organisms for the relatively short period of 40 million years (21). Not unexpectedly therefore, the human and mouse amino acid sequences are very similar. However, when comparing the two Fugu sequences with each other and with the mammalian sequences, it is intriguing to note the different types of evolutionary constraint exerted on the separate domains of the type XXVII collagen proa chains. The sequence of the TSPN domain is clearly conserved and presumably plays an important role in the biology of the tissues in which the type XXVII gene is expressed. In stark contrast, neither the length nor the precise sequence of the adjacent variable domain seems crucial in that there is no sequence conservation either between the two Fugu sequences or between Fugu and mammalian sequences (Fig. 1A).

Nevertheless, the chemical characteristics of this domain of the collagen XXVII gene are conserved in terms of the high pl exhibited across species (Table I). It is interesting to note that sequence variation produced within the variable domain of some proa chains of type B clade collagens by alternative splicing (see Refs. 22 and refs therein) cannot occur in type
XXVII collagen because the entire TSPN and variable domain are encoded by a single large exon. The level of sequence conservation for the triple helical domains (including the interruptions shared by the fish and mammalian genes) and for the NC1 domains is high, suggesting that there is little tolerance for sequence variation in these regions of the type XXVII molecule. It is only in the context of one of the duplicated genes in Fugu (proa1(XXVII)) that any significant sequence deviation in the helical domain has been tolerated in the form of two additional minor interruptions in the collagenous domain (Fig. 6).

Fig. 6. Phylogenetic analysis of NC1 (A and B) and TSPN (C and D) domains of types XXVII, XXIV, and other collagens. A and C, distance trees. Distance matrices were calculated for the conserved regions of the NC1 (A) and TSPN (C) domain alignments using PROTDIST, and the trees were derived using FITCH. For bootstrapping, 100 replicates were generated using SEQBOOT, distance matrices were calculated using PROTDIST, trees were derived using NEIGHBOUR, and the bootstrap consensus values were calculated using CONSENSE. The bootstrap values are given to the left of the nodes; only values of greater than 50 are shown. The scale bar length corresponds to the expected number of substitutions per site. B and D, parsimony trees. The most parsimonious trees were inferred for the conserved regions of the NC1 (B) and TSPN (D) domain alignments using PROTPARS. For the NC1 domain, seven trees were identified of 914.0 and 915.0 steps. For the TSPN domain, two were identified of 1222.0 and 1221.0 steps. The consensus topologies are shown. For bootstrapping, 100 replicates were generated using SEQBOOT, most parsimonious trees were inferred using PROTPARS, and the bootstrap consensus values were calculated using CONSENSE. Bootstrap values are given to the left of the nodes; only values of greater than 50 are shown.
A). The collagen XXVII gene therefore appears similar to all of the other vertebrate fibrillar collagen genes in terms of its resistance to sequence change, and one would predict that mutations of conserved residues are likely to be the cause of specific heritable conditions in man. However, there are no obvious candidate diseases based on the chromosomal location of the gene.

The phylogenetic analyses together with data relating to exon structure (Figs. 6 and 7, respectively) demonstrate that type XXVII collagen represents (together with type XXIV collagen) a new clade or subgroup (type C clade) within the vertebrate fibrillar collagen family. A model representing the most likely way in which the type C clade evolved, based on separate phylogenetic analyses of the NC1 and NC2 sequences (Fig. 6) combined with recent new insights into the evolution of fibrillar collagens in vertebrates (7), is presented in Fig. 8. The type C clade appears to have arisen as a result of a duplication of the common ancestor shared with the type B clade fibrillar collagen genes.

Although the model of type XXVII collagen evolution (Fig. 8) fits well with the phylogenetic data, the model does not provide an explanation for why type XXVII collagen differs so markedly in molecular characteristics compared with its close relatives, the classical fibrillar collagens (A and B clade). For example type XXVII collagen has a short chain selection sequence in its NC1 domain in comparison with the A/B clade collagens. The collagen XXVII major triple helical domain contains two interruptions, whereas the A/B clade α chains have uninterrupted helical domains. Furthermore, XXVII triple helical domain is shorter (994–997 amino acid residues depending on species; Table I) compared with the classical vertebrate fibrillar collagens (helical domains in all nine chains of the A/B clade fall in the range of 1011–1020 residues). Lastly, in collagen XXVII, the N-terminal end of the major helix, defined by the 54 bp-repeat exon structure encoding this region (Fig. 7), fuses directly with the NC2 domain and therefore lacks the telopeptide/telopeptide-like sequences and minor triple-helical domains found in all the A/B clade vertebrate fibrillar collagen chains (see Ref. 5 for exon structures through these domains). These unusual molecular features of the collagen XXVII polypeptide have previously only been found in invertebrate fibrillar collagens (7). However, in type XXVII, these features appear to have evolved within the vertebrate lineage after the type C clade diverged from the type B clade (Fig. 8). We have recently proposed a mechanism for how vertebrate fibrillar collagens evolved termed molecular incest that accounts for why the A/B clade polypeptides share such similar molecular characteristics in terms of the length of chain selection sequence and the length and nature of triple helical domain. Briefly, these classical vertebrate fibrillar collagen chains share the listed characteristics because various chains from both clades co-trimerize to form specific collagen types and different collagen types then co-polymerize to form mixed fibrils. Functionally, the chains encoded by the A and B clade genes are therefore interdependent and subject to a common constraining evolutionary pressure. The type XXVII collagen gene has not been subjected to the common evolutionary constraint experienced by the A/B clade genes, raising the possibility that collagen XXVII has evolved functions that are not dependent upon direct interactions (i.e. co-trimerization and

**Fig. 7.** Sizes of exons encoding the collagenous domains of proα1(XXVII) (type C clade), proα1(I) (type A clade), and proα1(V) (type B clade) collagens. The numbers of the type XXVII exons are given for reference purposes. Asterisks indicate the number of bases coding for Gly-X-Y repeats present in the 5' and 3' junction exons that code for both collagenous and noncollagenous sequences. Exons that appear to be of conserved size are lightly shaded. Conserved interruptions in collagenous domain of proα1(XXVII) are coded for by exons shaded darkly.
co-polymerization) with polypeptides/collagens encoded by the type A and B clade genes.

The unique expression pattern exhibited by collagen XXVII lends weight to the prediction that type XXVII collagen may perform functions that are independent of the collagen types encoded by the type A and B clade genes (i.e. types I, II, III, V, and XI). Although type XXVII collagen is expressed by chondrocytes in cartilage as is found for collagens II and XI, it is also expressed by a variety of epithelial cell types (Figs. 3 and 4). Because the classical vertebrate fibrillar collagens are not expressed by these epithelial layers, the function of collagen XXVII at these locations cannot be dependent upon the co-trimerization or co-polymerization with collagen chains of the A/B clade.

Antisera were raised to synthetic peptides derived from the variable domain that would remain part of the mature \( \alpha \) chain if proteolytic processing of the N terminus of type XXVII collagen follows the same pattern as that of the homologous type B clade \( \alpha \) chains (23). The detection of type XXVII collagen by Western blotting indicates firstly that this gene, the amino acid sequence of which was predicted largely on the basis of genome analysis, does encode an expressed protein. Secondly, the tissue analysis suggests the protein may be part of the co-trimerization or co-polymerization with collagens chains of the A/B clade.

Antisera were raised to synthetic peptides derived from the variable domain. If proteolytic processing of the N terminus of type XXVII collagen follows the same pattern as that of the homologous type B clade \( \alpha \) chains (23). The detection of type XXVII collagen by Western blotting indicates firstly that this gene, the amino acid sequence of which was predicted largely on the basis of genome analysis, does encode an expressed protein. Secondly, the tissue analysis suggests the protein may be part of the co-trimerization or co-polymerization with collagen chains of the A/B clade.

Fig. 8. A model for the evolution of the three clades of vertebrate fibrillar collagens. The founder vertebrate fibrillar collagen gene acquired an elongated chain selection sequence (7). Gene/genome duplication events led to the evolution of the founder genes for the type A, B, and C clades. During this early phase of molecular evolution (indicated by the red dashed lines), rearrangements between the exons encoding the collagenous domains and swapping of the exons encoding the von Willebrand factor type C domain for those encoding a TSPN domain occurred. Once the essential features of each clade had evolved, exon rearrangements ceased. Subsequent gene duplication events produced the multiple members of each clade with exon patterns that are common within a clade but distinct between clades (see Fig. 7).

Types XXVII and XXIV collagen are very closely related fibrillar collagen genes, and it is not altogether clear whether these two genes should have been given separate numbers or alternatively, designated as separate \( \alpha \) chains of the same type of collagen. On the basis of the highly conserved nature of the triple helical domains including length and position of interruptions and the similarities in the NC1 domains where both have a short chain selection sequence, it would appear entirely possible that the two types of pro\( \alpha \) chain can heterotrimerize. However, preliminary analyses suggest that the pro\( \alpha \)1(XXVII) and pro\( \alpha \)1(XXIV) chains are not co-ordinately expressed and appear to have distinct patterns of expression in tissues. On balance, therefore, it seems likely that the type XXVII and XXIV collagen genes are named correctly in that they appear most likely to form distinct collagen types. It will be very interesting to determine the function of type XXVII collagen, given its unusual structure and expression pattern, and work on a mouse knockout model is in progress.

\[ \text{M. Gordon and R. P. Boot-Handford, unpublished observations.} \]
Acknowledgments—The MRC Tissue Bank provided access to the fixed human tissues used in this study. Toby Hunt is thanked for expert technical assistance. Marion Gordon, James Pace, and Peter Byers are thanked for sharing data on types XXIV and XXVII collagen prior to publication.

Note Added in Proof—While this paper was under review, the paper by Pace et al. (25) on collagen XXVII was published.

REFERENCES
18. Felsenstein, J. (1993) Phylogenetic Inference Package (PHYLIP), version 3.5c, Department of Genetics, University of Washington, Seattle, WA
A Novel and Highly Conserved Collagen (proc1(XXVII)) with a Unique Expression Pattern and Unusual Molecular Characteristics Establishes a New Clade within the Vertebrate Fibrillar Collagen Family

Raymond P. Boot-Handford, Danny S. Tuckwell, Darren A. Plumb, Claire Farrington

Rock and Richard Poulson

doi: 10.1074/jbc.M212889200 originally published online May 23, 2003

Access the most updated version of this article at doi: 10.1074/jbc.M212889200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 21 references, 8 of which can be accessed free at
http://www.jbc.org/content/278/33/31067.full.html#ref-list-1