Characterization of a fibrillar collagen gene in sponges reveals the early evolutionary appearance of two collagen gene families

(gene evolution/intron–exon structure/invertebrate collagens)

JEAN-YVES EXPOSITO AND ROBERT GARRONE*

Institute of Biology and Chemistry of Proteins, Centre National de la Recherche Scientifique, Unité Propre de Recherche 412, and Claude Bernard University, Lyons, France

Communicated by Jerome Gross, May 18, 1990 (received for review March 15, 1990)

ABSTRACT **We have characterized cDNA and genomic clones coding for a sponge collagen. The partial cDNA has an open reading frame encoding 547 amino acid residues. The conceptual translation product contains a probably incomplete triple-helical domain (307 amino acids) with one Gly-Xaa-Yaa-Zaa imperfection in the otherwise perfect Gly-Xaa-Yaa repeats and a carboxyl propeptide (240 amino acids) that includes 7 cysteine residues. Amino acid sequence comparisons indicate that this sponge collagen is homologous to vertebrate and sea urchin fibrillar collagens. Partial characterization of the corresponding gene reveals an intron–exon organization clearly related to the fibrillar collagen gene family. The exons coding for the triple-helical domain are 54 base pairs (bp) or multiples thereof, except for a 57-bp exon containing the Gly-Xaa-Yaa-Zaa coding sequence and for two unusual exons of 126 and 18 bp, respectively. This latter 18-bp exon marks the end of the triple-helical domain, contrary to the other known fibrillar collagen genes that contain exons coding for the junction between the triple-helical domain and the carboxyl propeptide. Compared to other fibrillar collagen genes, the introns are remarkably small. Hybridization to blotted RNAs established that the gene transcript is 4.9 kilobases. Together with previous results that showed the existence of a nonfibrillar collagen in the same species, these data demonstrate that at least two collagen gene families are represented in the most primitive metazoa.**

Collagens are multidomain, interactive proteins that constitute the main extracellular component of all multicellular animals (1). So far, 13 different collagen types have been described in vertebrates, most of them forming polymeric structures or being closely associated with other collagenous polymers. Collagen molecules are made of three α chains, which may or may not be identical. They are characterized by the presence of triple-helical domains, which contain Gly-Xaa-Yaa repeats, that are essential for helix formation (2). More than 20 genes encode the different α chains. With the exception of the gene coding for type X collagen (3, 4), the described collagen coding genes have a complex exon–intron organization. Among them, the genes coding for the fibril-forming collagens (types I, II, III, V, and XI) have a homologous organization (5–8) with, in particular, triple-helix coding exons related to a 54-base-pair (bp) unit. These exons are believed to have evolved from an ancestral unit of 54 bp encoding six Gly-Xaa-Yaa repeats (9). The nonfibrillar collagen genes examined so far have more variable structures even though a few 54-bp exon units have been described (10–13).

Very few data have been reported on the collagen gene organization in invertebrates. Variable exon sizes have been found in the nonfibrillar collagen genes, either in *Drosophila melanogaster* (14–16) or in *Caenorhabditis elegans* (17, 18), contrary to the multiple 54-bp motif described recently in a fibrillar collagen gene of the sea urchin *Paracentrotus lividus* (19).

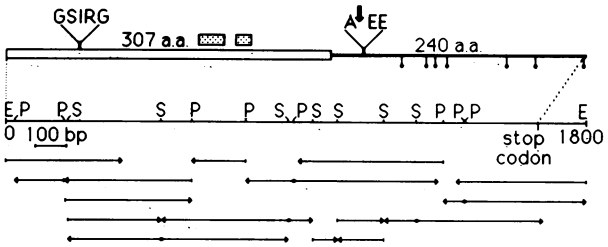

FIG. 1. Diagram showing partial restriction map and sequencing strategy of the cDNA clone C23 (*Lower*) and some characteristic features of the deduced translation product (*Upper*). Open box, helical domain; solid line, noncollagenous domain; solid circles with bars, cysteine residues; stippled boxes, triple-helical subdomains devoid of imino acids; vertical arrow, putative carboxyl-proteinase cleavage site. The Gly-Xaa-Yaa-Zaa imperfection and the number of amino acids (a.a.) in each domain are indicated. The sequencing strategy is indicated by horizontal arrows representing the directions and length of sequencing runs. E, *Eco*RI; P, *Pvu* II; S, *Sau*3A.

To clarify the evolution of the collagen gene family, we have investigated the structure of the collagen genes in the phylum Porifera. This phylum represents the most primitive multicellular animals and contains several morphological forms of collagen (20). All the sponge species contain collagen fibrils displaying a typical banding pattern and having a small, uniform diameter of ≈20 nm. In addition, some species possess other highly variable forms of collagen aggregates, generally made up of microfibrils. The collagen in sponges is totally insoluble (20), and therefore it has been impossible to correlate these morphological differences with biochemical analyses.

We have recently cloned and sequenced a cDNA coding for a sponge collagen that possesses the structure of a short-chain collagen (21). The analysis of the corresponding gene (unpublished results) reveals variable sizes (144 and 207 bp for the two exons sequenced so far) for the exons coding for the triple-helical domain, demonstrating that this protein does not belong to the fibrillar class of collagens.

In the present study, we have characterized, in the same sponge species, cDNA and genomic clones† related to a

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*To whom reprint requests should be addressed at: Institut de Biologie et Chimie des Protéines, Bâtiment 403, Université Claude Bernard, 43 Boulevard 11 Novembre, 69622 Villeurbanne Cedex, France.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. M34640).

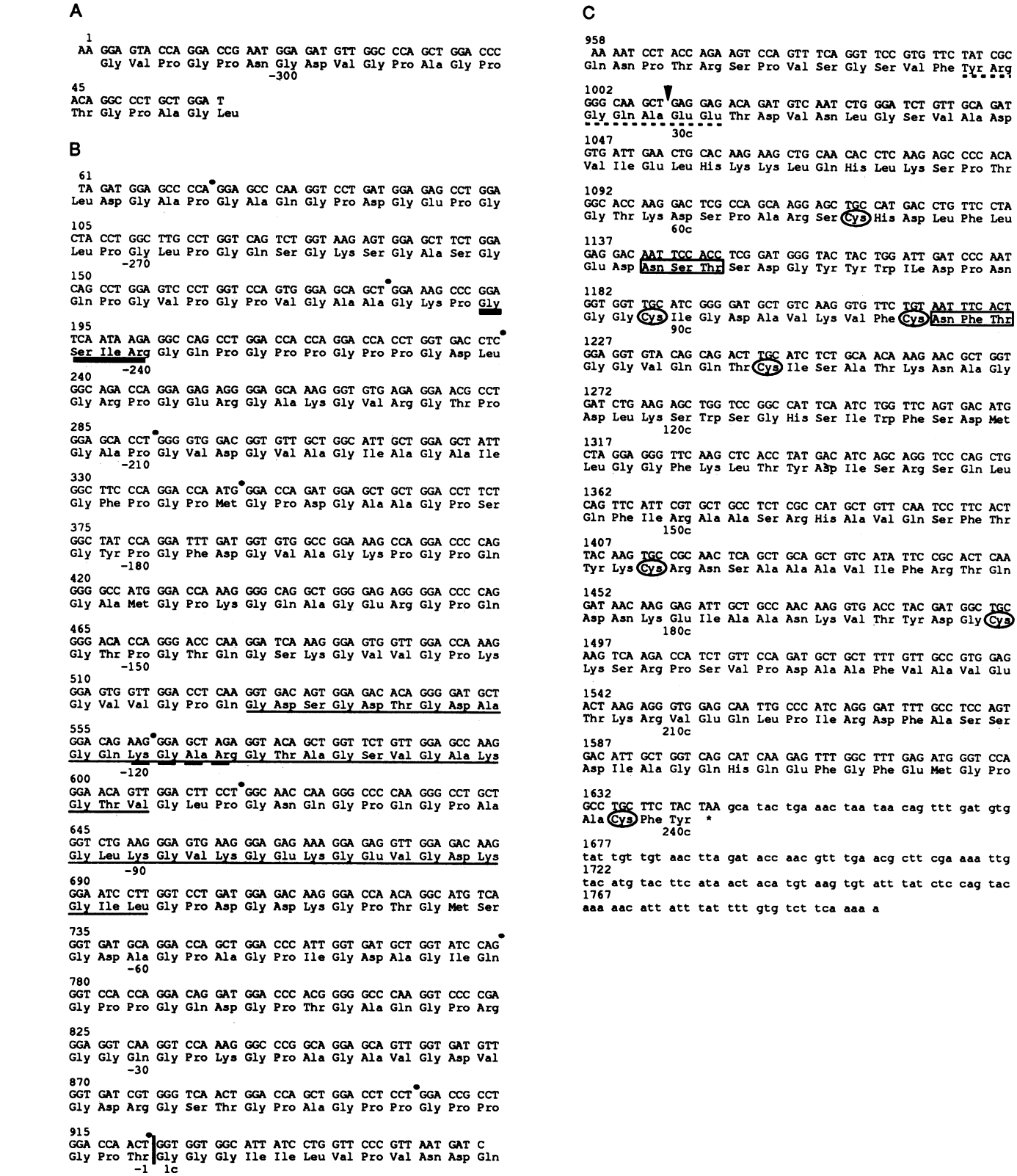

FIG. 2. Nucleotide and deduced amino acid sequences of Emf1α determined by cDNA sequencing (*A–C*) and by genomic DNA sequencing (*B*). Upper line, capital and lowercase letters indicate the coding and the 3′ noncoding sequences, respectively. Amino acid residues of the triple-helical and the C-terminal domains are numbered from −306 to −1 and from 1c to 240c, respectively. Vertical bar marks the end of the triple-helical domain and arrowhead indicates the putative carboxyl-proteinase cleaved bond contained in a characteristic sequence (underlined with a dotted line). Cysteine residues are circled, and the potential N-carbohydrate attachment sites are boxed. Thick line underlines the Gly-Xaa-Yaa-Zaa imperfection. Dashed line indicates the potential cross-linking site. Thin line shows the triple-helical regions with no imino acid residue. Solid circles in *B* demarcate the exons and asterisk in *C* indicates the termination codon.

fibrillar collagen. We thus demonstrate that the distinction between at least two collagen gene families is already established in primitive animals corresponding to the outset of collagen evolution.

MATERIALS AND METHODS

Freshwater sponges (*Ephydatia mülleri*) were cultivated in the laboratory from asexual buds (gemmules) collected in the field. Total RNA, poly(A)$^+$ RNA extract, and blot analyses were performed as described (21). By using cDNA synthesis and cloning kits (Amersham), based on the method of Gubler and Hoffman (22), a sponge cDNA library was constructed in the λgt10 vector using cDNA enriched in fragments >2 kilobases (kb) by gel electroelution. This library was screened using as a probe the 1.2-kb insert of the previously described EmC4 clone (21) labeled by the oligonucleotide labeling method (23). The conditions of hybridization and washing were as reported (21). *E. mülleri* genomic DNA was extracted (24), partially digested with *Sau*3A to obtain fragments with an average size of 15 kb, and submitted to a partial fill-in with dATP and dGTP according to a method described elsewhere (25). A genomic library was constructed by inserting the prepared *E. mülleri* genomic DNA fragments into the *Xho* I site of Lambda GEM-11 DNA (Promega), partially filled in with dTTP and dCTP. High stringencies for hybridization (50% formamide/900 mM NaCl/90 mM sodium citrate, pH 7.0; 42°C) and washing (15 mM NaCl/1.5 mM sodium citrate, pH 7.0/0.1% sodium dodecyl sulfate; 65°C) were used to screen the genomic library with ^{32}P-labeled cDNA used as a probe. Sequencing of both strands was performed in M13mp18 and/or M13mp19 vectors by the dideoxynucleotide chain-termination method (26) with a Sequenase kit (United States Biochemical). The orientation of the cloning fragments was tested by S1 nuclease digestion as described elsewhere (27). Computer analysis was performed using the program DNAid (28).

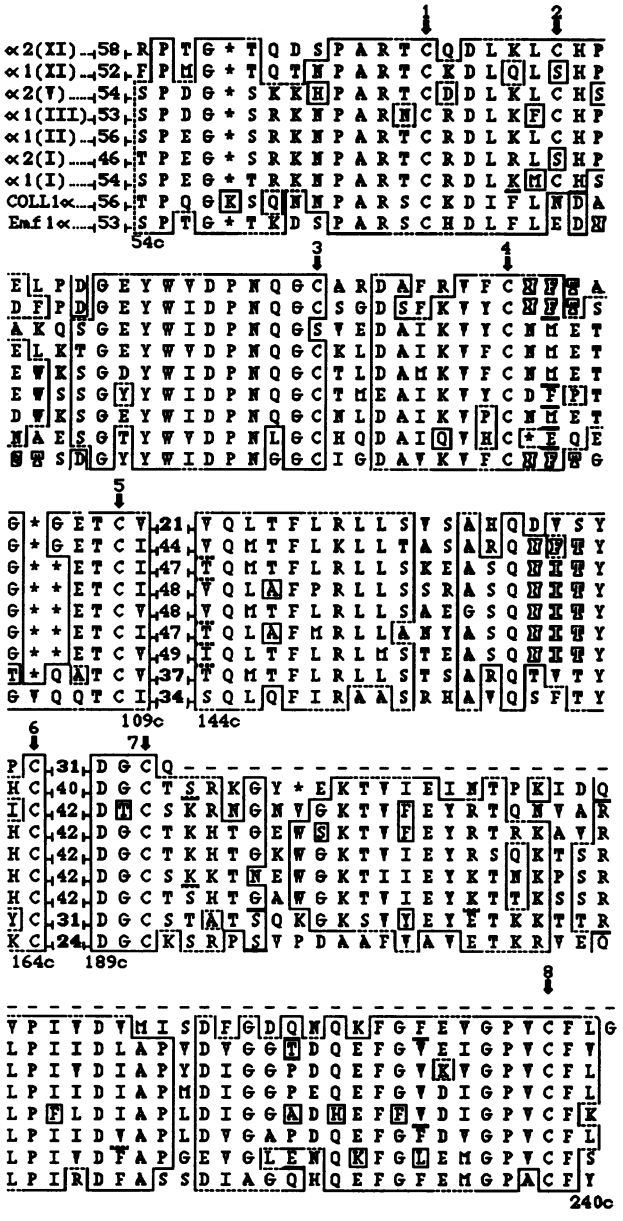

FIG. 3. Alignment of portions of the carboxyl propeptide of Emf1α, sea urchin COLL1α (19), and human fibrillar collagens (7, 8, 30–35). Computer alignments were done with the sequence analysis program DNAid (28). The less conserved regions are only represented by their number in amino acids. Identical residues are boxed, whereas chemically similar amino acids (ILMV, DNEQ, KHR, FWY, ST, and AG) are indicated by dashed boxes. Numbered arrows indicate cysteine residues and asterisks represent gaps inserted to give the best alignment. Dashes indicate the nonsequenced region of α2(XI).

RESULTS

Cloning of *E. mülleri* Collagen cDNAs. A previous study led to the isolation of a sponge cDNA clone, EmC4, that encodes a nonfibrillar collagen (21). The 1.2-kb insert cDNA of EmC4 was used as a probe to screen a sponge cDNA library at low and high stringencies. A clone (C23) positive only at low stringency was isolated. Its 1.8-kb insert was digested with the restriction endonuclease *Sau*96I, which recognizes the site GGNCC encoding the sequence Gly-Pro frequently observed in the triple-helical domain of collagens (9, 29). The fragments obtained, which were multiples of 9 bp (data not shown), suggested that this clone contained a collagen coding sequence. The partial restriction map, the sequencing strategy, and the structure of the deduced translation product (designated Emf1α) are presented in Fig. 1. This clone contains an open reading frame coding for 547 amino acids, terminated by a stop codon (TAA) with 151 nucleotides downstream (Figs. 1 and 2). The conceptual translation product can be divided into two domains, an uninterrupted collagenous domain of 307 amino acids with one Gly-Xaa-Yaa-Zaa imperfection and a noncollagenous, C-terminal domain of 240 amino acids with 7 cysteine residues. By Northern blot analysis, it was shown that the C23 insert hybridized to a 4.9-kb mRNA (data not shown).

The noncollagenous, C-terminal domain of Emf1α was compared by computer alignment to similar domains of human and sea urchin fibrillar collagen chains (Fig. 3) (7, 8, 19, 30–35). The homology of Emf1α with fibrillar collagens is evident and a conservation of the general organization of this region is observed. As shown in Table 1, comparison of sequences corresponding to highly conserved domains of the C-terminal region showed that Emf1α is more similar to α1(XI) and α2(XI) collagen chains than to any other chain, including the sea urchin collagen. This similarity is further emphasized by distinct features. As in α1(XI), cysteine number 2 is missing. As in α2(XI), the polypeptide chain is shorter between cysteines 5, 6, and 7, and the classical N-linked carbohydrate attachment site is absent. However, a putative N-glycosylation site (Asn-Phe-Thr) following cysteine 4 is observed as in the α1(XI) and α2(XI) chains.

Gene Organization. A genomic library from *E. mülleri* was screened by using the 1.8-kb insert of C23 as a probe. One positive clone (G238) was characterized. Its 3′ end lies in an exon coding for the beginning of the noncollagenous C-terminal domain. The sequenced fragment covers the exons coding for almost the entire collagenous sequence of the cDNA clone. The corresponding gene was designated *COLF1*. Nine complete and two partial exons were sequenced. They all begin with a complete Gly codon and terminate with a Yaa codon. Six exons have a size of 54 bp

Table 1. Percentages of sequence identities between the nonhelical C-terminal domain of the sponge Emflα chain and similar domains of other fibrillar collagen chains

	COLL1α	α1(I)	α2(I)	α1(II)	α1(III)	α2(V)	α1(XI)	α2(XI)
Emflα	47.5	51.9	51.9	50.6	51.9	50.6	53.8	56.2

Results were obtained with the computer program DNAid (28). The highest homology was obtained with the α1(XI) and α2(XI) chains. The comparison was limited to the sequences presented in Fig. 3 (amino acids 54c–109c, 144c–164c, and 189c–192c).

or multiples of 54 bp. One exon of 57 bp contains the coding sequence for the imperfection Gly-Xaa-Yaa-Zaa. Two exons of 126 and 18 bp, respectively, are different from exons described so far for fibrillar collagens and are localized at the 3′ end of the region coding for the triple-helical domain (Fig. 4). The 18-bp exon codes for the last two Gly-Xaa-Yaa repeats of this domain. The sequences of the donor and acceptor splice junction sites are in agreement with the consensus sequences observed in eukaryotes (36) (data not shown).

DISCUSSION

Our data demonstrate that Emflα, the protein encoded by the cDNA clone C23, is homologous to vertebrate fibrillar collagens. This homology is supported by the high conservation of the primary structure of the carboxyl propeptide and by the presence of predominantly 54-bp unit exons in the triple-helix coding region of the corresponding gene.

The complete insolubility of sponge collagens (20) has precluded any detailed structural study at the protein level. Our present data suggest that Emflα participates in a heterotrimeric collagen molecule. Indeed the carboxyl propeptide of Emflα contains seven cysteines, as observed for vertebrate collagen chains participating only in heterotrimeric assemblies (7, 19). Furthermore, our previous results of cell-free translation experiments (21) have suggested that at least two collagen chains of a size similar to that of vertebrate fibrillar collagens are present in sponges.

No sequence identical to known carboxyl-proteinase cleavage sites has been identified in Emflα. Nevertheless, a possible site of cleavage could be the Ala-Glu bond in the sequence Tyr-Arg-Gly-Gln-Ala-Glu-Glu (Fig. 2), which is comparable to known sequences containing a carboxyl-proteinase cleavage site (presence of Tyr and Arg close to the cleavage site, Ala at the P1 position, and two acidic residues at P1′ and P2′ positions). Such a propeptide removal would leave a C-terminal telopeptide of 29 amino acids, beginning by a triplet Gly-Gly-Gly as observed in the α2(I) chain (30) and following a triple helix stabilized by a high imino acid content in the last few triplets as already described for other collagen chains (37). This telopeptide does not contain any lysine residue that could participate in covalent cross-linking, a situation already encountered in the α2(I) chain (30).

As the vertebrate collagen chain α1(I) and α2(V) (31, 35), Emflα contains, in the triple-helical region, long stretches of amino acids without proline residues. One stretch of 27 amino acids (−131 to −105) is located ≈100 amino acids from the carboxyl end of the triple helix and contains a sequence (Lys-Gly-Ala-Arg) reminiscent of (but not identical to) the vertebrate consensus sequence for lysyl oxidase attachment (38). This site could be involved in cross-link formation. Another stretch of 18 amino acids (−92 to −75) contains 4 lysine residues in position Yaa. This is frequently observed in Emflα as well as in nonfibrillar sponge collagen (21) and in vertebrate type IV collagen (39, 40). Most of these lysine residues could be hydroxylated and subsequently glycosylated since sponge collagens are highly glycosylated (20, 41). The presence of an imperfection (Gly-Xaa-Yaa-Zaa) in the triple-helical domain is exceptional for fibrillar collagen. So far, it has been described only in human and mouse α2(IV) chain (42, 43). It should be noted that in Emflα, two Gly-Pro-Pro stabilizing triplets occur near the imperfection.

The gene coding for Emflα sponge collagen chain, *COLF1*, has an organization that is related to fibrillar collagen genes (9) but with some unique features. Among the prevalent 54-bp unit exons, an unusually large 216-bp exon is present. The triple-helical domain is terminated by an 18-bp exon, the shortest exon ever described coding for a triple-helical sequence. Contrary to other known fibrillar collagen genes, the junction between the triple-helical domain and the carboxyl propeptide is not encoded by a single junction exon. Finally, the introns are short compared to other fibrillar collagen genes. Almost all the characterized exons of this sponge collagen gene are multiples of 18 bp and begin with a complete glycine

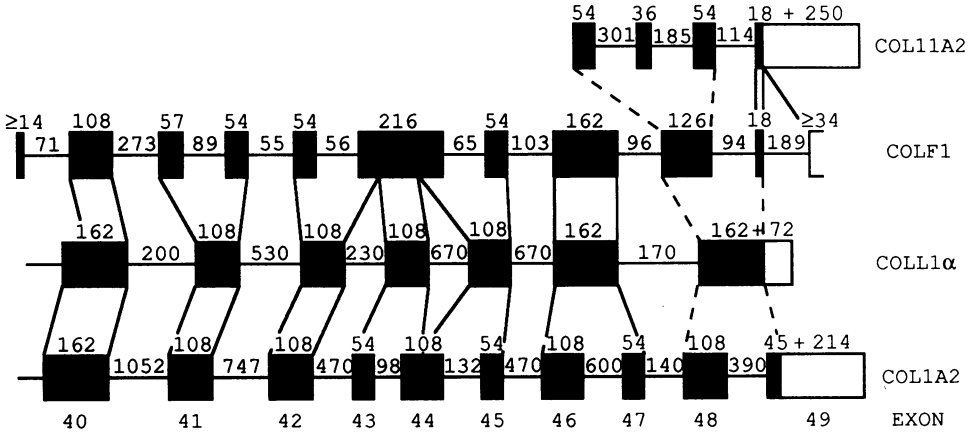

FIG. 4. Comparison of the exon–intron organization of the *COLF1* gene with the equivalent regions of human proα2(XI) and proα2(I) genes, *COL11A2* and *COL1A2*, and the sea urchin *COLL1α* gene. Exons are represented by solid boxes for the triple-helical domain coding exons, and by open boxes for the carboxyl propeptide coding exons. Introns are indicated by lines. Numbers indicate the size in bp of exons and introns, without a scale for the latter. The numbering of the human proα2(I) exons is indicated at the bottom. Two approximations have been accepted in constructing this diagram. One is due to the Gly-Xaa-Yaa-Zaa imperfection (a 57- plus a 54-bp exon corresponding to a 108-bp exon), and the second comes from the heterogeneous sizes of the exons coding for the end of the triple-helical domains (6). The latter heterogeneities are represented by dashed lines.

codon. The 57-bp exon has very likely arisen from a 3-base insertion in a 54-bp exon. It is worth noting that other exons that are multiples of 18 bp (36 and 72 bp) have been described in *COL9A2*, *COL11A2*, *COL12A1*, and *COL13A1* genes (8, 10–13). It is generally admitted that the 54-bp unit is itself the result of amplification events (9) of a primordial unit such as a 9-mer GGN CCN CCN oligonucleotide, coding for the triplet Gly-Pro-Pro (44, 45). Interestingly, the 18-bp sponge exon encodes the sequence Gly-Pro-Pro-Gly-Pro-Thr.

Although a more thorough treatment of the sequences would be needed to establish definite relationships between Emf1α and vertebrate fibrillar collagens, the comparison with type XI collagen is highly suggestive. A similarity is also observed at the end of the triple-helical coding region. Within the junction exon of *COL11A2*, there are just 18 bp coding the last two Gly-Xaa-Yaa repeats, while in *COLF1* these residues are encoded by a separate 18-bp exon. These similarities between two collagens so distant in evolution should have some significance. The postulated role of an axial template (46) for type XI collagen might reflect some ancestral structural function. The fibrillar collagen of sponges could then constitute what in higher animals will be the backbone of a collagen fibril.

The existence in sponges of at least one fibrillar collagen gene, as demonstrated in the present work, and one nonfibrillar collagen gene (ref. 21; unpublished results) clearly establishes that the divergence between two collagen gene families (or the convergence of two unrelated genes) has occurred very early during evolution. It is clear that the fibrillar collagen gene family has evolved relatively little. Contrarily, the vertebrate nonfibrillar collagen genes appear to form several distinct families (47). The nonfibrillar collagen encoded in sponges (21) shares features with basement membrane collagen (types IV) and with nematode cuticular collagens and might thus reflect two lines of evolution. One line might have been the "exocollagens," such as collagen attaching sponges to their substratum, exoskeletons of Cnidarians, cuticles of worms, secreted collagens of mussels, and the egg case of Selacians. The other line of evolution might concern an internalization of such collagens, leading to the differentiation of basement membrane collagen.

We wish to thank Dr. R. Ouazana for his help during the course of this work, Prof. B. Olsen for constructive comments, and Prof. M. van der Rest for many helpful discussions and suggestions during the preparation of the manuscript.

1. Gross, J. (1985) in *Biology of Invertebrate and Lower Vertebrate Collagens*, eds. Bairati, A. & Garrone, R. (Plenum, New York), pp. 1–28.
2. Miller, E. J. & Gay, S. (1987) *Methods Enzymol.* **144,** 3–41.
3. Ninomiya, Y., Gordon, M., van der Rest, M., Schmid, T., Linsenmayer, T. & Olsen, B. R. (1986) *J. Biol. Chem.* **261,** 5041–5050.
4. Lu Valle, P., Ninomiya, Y., Rosenblum, N. D. & Olsen, B. R. (1988) *J. Biol. Chem.* **263,** 18378–18385.
5. De Crombrugghe, B., Schmidt, A., Liau, G., Setotama, C., Mudryj, M., Yamada, Y. & McKeon, C. (1985) *Ann. N.Y. Acad. Sci.* **460,** 154–162.
6. Ramirez, F., Bernard, M., Chu, M. L., Dickson, L., Sangiorgi, F., Weil, D., de Wet, W., Junien, C. & Sobel, M. E. (1985) *Ann. N.Y. Acad. Sci.* **460,** 117–124.
7. Bernard, M., Yoshioka, H., Rodriguez, E., van der Rest, M., Kimura, T., Ninomiya, Y., Olsen, B. R. & Ramirez, F. (1988) *J. Biol. Chem.* **263,** 17159–17166.
8. Kimura, T., Cheah, K. S. E., Chan, S. D. H., Lui, V. C. H., Mattei, M. G., van der Rest, M., Ono, K., Solomon, E., Ninomiya, Y. & Olsen, B. R. (1989) *J. Biol. Chem.* **264,** 13910–13916.
9. Yamada, Y., Avvedimento, V. E., Mudryj, M., Ohkubo, H., Vogeli, G., Irani, M., Pastan, I. & de Crombrugghe, B. (1980) *Cell* **22,** 887–892.
10. Lozano, G., Ninomiya, Y., Thompson, H. & Olsen, B. R. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4050–4054.
11. Olsen, B. R. (1989) *Connect. Tissue Res.* **23,** 115–121.
12. Gordon, M. K., Gerecke, D. R., Dublet, B., van der Rest, M. & Olsen, B. R. (1989) *J. Biol. Chem.* **264,** 19772–19778.
13. Tikka, L., Pihlajaniemi, T., Henttu, P., Prockop, D. J. & Tryggvason, K. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7491–7495.
14. Blumberg, B., MacKrell, A. J., Olson, P. F., Kurkinen, M., Monson, J. M., Natzle, J. E. & Fessler, J. H. (1987) *J. Biol. Chem.* **262,** 5947–5950.
15. Blumberg, B., MacKrell, A. J. & Fessler, J. H. (1988) *J. Biol. Chem.* **263,** 18328–18337.
16. Cecchini, J. P., Kniebiehler, B., Mirre, C. & Le Parco, Y. (1987) *Eur. J. Biochem.* **165,** 587–593.
17. Cox, G. N., Fields, C., Kramer, J. M., Rosenzweig, B. & Hirsh, D. (1989) *Gene* **76,** 331–344.
18. Guo, X. & Kramer, J. M. (1989) *J. Biol. Chem.* **264,** 17574–17582.
19. D'Alessio, M., Ramirez, F., Suzuki, H. R., Solursh, M. & Gambino, R. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 9303–9307.
20. Garrone, R. (1978) *Phylogenesis of Connective Tissue. Morphological Aspects and Biosynthesis of Sponge Intercellular Matrix,* (Karger, Basel).
21. Exposito, J. Y., Ouazana, R. & Garrone, R. (1990) *Eur. J. Biochem.*, in press.
22. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25,** 263–269.
23. Feinberg, A. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
24. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, New York).
25. Zabarovsky, E. R. & Allikmets, R. (1986) *Gene* **42,** 119–123.
26. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
27. Aegerter, E. & Trachsel, H. (1987) *Nucleic Acids Res.* **15,** 372.
28. Dardel, F. & Bensoussan, P. (1988) *Cabios* **4,** 483–486.
29. Gordon, M. K., Gerecke, D. R. & Olsen, B. R. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6040–6044.
30. Bernard, M. P., Chu, M. L., Myers, J. C., Ramirez, F., Eikenberry, E. F. & Prockop, D. J. (1983) *Biochemistry* **22,** 5213–5223.
31. Bernard, M. P., Myers, J. C., Chu, M. L., Ramirez, F., Eikenberry, E. F. & Prockop, D. J. (1983) *Biochemistry* **22,** 1139–1145.
32. De Wet, W., Bernard, M., Benson-Chanda, V., Chu, M. L., Dickson, L., Weil, D. & Ramirez, F. (1987) *J. Biol. Chem.* **262,** 16032–16036.
33. Sangiorgi, F. O., Benson-Chanda, V., de Wet, W. J., Sobel, M. E., Tsipouras, P. & Ramirez, F. (1985) *Nucleic Acids Res.* **13,** 2207–2225.
34. Chu, M. L., Weil, D., de Wet, W., Bernard, M., Sippola, M. & Ramirez, F. (1985) *J. Biol. Chem.* **260,** 4357–4363.
35. Weil, D., Bernard, M., Gargano, S. & Ramirez, F. (1987) *Nucleic Acids Res.* **15,** 181–198.
36. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55,** 1119–1150.
37. Yamada, Y., Kühn, K. & De Crombrugghe, B. (1983) *Nucleic Acids Res.* **11,** 2733–2744.
38. Kühn, K. (1987) in *Structure and Function of Collagen Types*, eds. Mayne, R. & Burgeson, R. E. (Academic, New York), pp. 1–42.
39. Brazel, D., Oberbäumer, I., Dieringer, H., Babel, W., Glanville, R. W., Deutzmann, R. & Kühn, K. (1987) *Eur. J. Biochem.* **168,** 529–536.
40. Schwarz, U., Schuppan, D., Oberbäumer, I., Glanville, R. W., Deutzmann, R., Timpl, R. & Kühn, K. (1986) *Eur. J. Biochem.* **157,** 49–56.
41. Junqua, S., Robert, L., Garrone, R., Pavans de Ceccatty, M. & Vacelet, J. (1974) *Connect. Tissue Res.* **2,** 193–203.
42. Hostikka, S. L. & Tryggvason, K. (1988) *J. Biol. Chem.* **263,** 19488–19493.
43. Saus, J., Quinones, S., MacKrell, A., Blumberg, B., Muthukumaran, G., Pihlajaniemi, T. & Kurkinen, M. (1989) *J. Biol. Chem.* **264,** 6318–6324.
44. Ycas, M. (1972) *J. Mol. Evol.* **2,** 17–27.
45. Benveniste-Schrode, K., Doering, J. L., Hauck, W. W., Kendra, K. L. & Drexler, B. K. (1985) *J. Mol. Evol.* **22,** 209–219.
46. Mendler, M., Eich-Bender, S. G., Vaughan, L., Winterhalter, K. H. & Bruckner, P. (1989) *J. Cell Biol.* **108,** 131–137.
47. Fields, C. (1988) *J. Mol. Evol.* **28,** 55–63.