Belg. J. Zool. — Volume 122 (1992) — issue 1 — pages 17-22 — Brussels 1992

THE COLLAGEN FAMILY OF PROTEINS : TWO DISTINCT LINES OF EVOLUTION (1)

by

ROBERT GARRONE and JEAN-YVES EXPOSITO
Institut de Biologie et Chimie des Protéines, CNRS UPR 412
Université Claude Bernard
43 Bd 11 novembre 1918
F-69622 Villeurbanne (France)

SUMMARY

Fourteen well-defined collagen types are known in vertebrates and a few of their counterparts have been described in invertebrates. A single, ancestral collagen type might have been expected in the most primitive multicellular animals, the sponges. These organisms actually contain collagen in two forms, cross-striated fibrils and a large assembly of thin filaments constituting the spongin skeleton. A molecular biology study has revealed that these two forms correspond to two distinct genetic collagen types. Several short-chain collagen sequences have been deduced from their cDNA. They bear homologies with nematode cuticular collagens, with basement membrane collagens and with the type XIII collagen of vertebrates. The corresponding genes are expressed in cells secreting the spongin. A fibrillar collagen has also been characterized. Its structure at the C-terminal non-helical domain and its gene organization allowed a comparison with the vertebrate type XI fibrillar collagen, considered as an axial core contained in vertebrate larger fibrils.

The multicellular animals are not only characterized by their cells, numerous, differentiated and coordinated, but also by their unique intercellular medium, composed of several macromolecular species. This extracellular matrix is involved in the mechanical integration of the different cells and tissues, and, in addition, modulates the cell differentiation and behaviour (TRELSTAD, 1984). Among the various components of the extracellular matrix, collagen is certainly the most universal protein (BAIRATI and GARRONE, 1985).

Its skeletal function has been obviously recognized as collagen is the main organic component of bone and constitutes the framework of cartilage. Collagen is also involved in specialized structures such as fish scales (FRANCILLON, *et al.*, 1990)

(1) Presented at the Second Belgian Congress of Zoology, Diepenbeek, Belgium, 15-16 November 1991.

and in several skeletal devices in invertebrates (BAIRATI and GARRONE, 1985). In the most primitive multicellular animals, the sponges, collagen can form a reticulate organic skeleton, for example the skeleton making the natural bath sponges. In this case, the coarse fibers are composed of the association of thin unit filaments of spongin (about 10 nm in diameter), collagenous in nature. In other species, the deposit of such filaments is used to link together siliceous needles, the spicules, building then a skeleton combining silica and collagen. In addition, these filaments are used to stick the sponges to rocks, shells, or Petri dishes in the laboratory (Pl. I). In all species of sponges, the extracellular matrix contains banded fibrils, 20 nm in diameter (Pl. I), scattered or associated in bundles, which are also composed of collagen (GARRONE, 1978). Similar fibrils are found in the cnidarians, and fibrils with a more complex banding pattern are current constituents of extracellular matrices of almost all other animals.

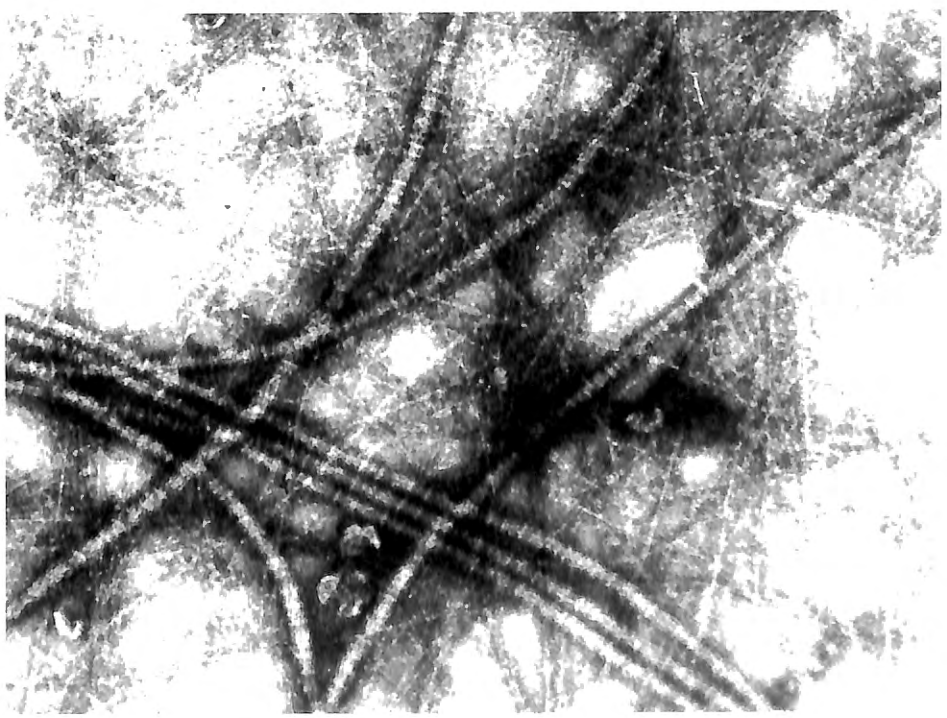

Plate 1. — Electron microscope view of the two collagen structural units present in sponges : collagen fibrils (the largest fibrils) and spongin filaments (the thinnest elements) in the background. Negative staining, X 100,000.

These collagen fibrils are formed by the regular association of collagen molecules. Each molecule itself is a triple helix formed by the intertwining of three helical polypeptides, the alpha chains. The amino acid sequences of the collagen

alpha chains are characterized by the repetition of triplets Gly-Xaa-Yaa. This conformation is stabilized by the presence of imino acids in about 30 % of the Xaa and Yaa positions (prolyl and hydroxyprolyl residues respectively). Actually, the extremities of the molecule are composed of non helical domains (often termed non-collagenous domains) which may or may not be removed during the maturation of the molecule. For some collagens, short, non helical domains can also interrupt the helical domain. Fourteen different collagen types are known in vertebrates (VAN DER REST and GARRONE, 1991). They form a variety of extracellular assemblies : banded fibrils (types I, II, III, V and XI), fibril coats (types IX, XII, perhaps XIII, and XIV) ; sheets (types IV — the basement membranes —, VIII and X), beaded filaments (type VI), short anchoring segments (type VII).

Several hypotheses have been considered to determine the most primitive of these collagen types. In order to bring a contribution to this point, we investigated the collagen of sponges using a molecular biology approach. We choose a freshwater species (*Ephydatia mülleri* LIEB) easy to use in laboratory experiments. From 3-day-old sponges, total RNA populations were isolated and translated *in vitro*. Among the translation products, four peptides were sensitive to collagenase digestion. Two of them had an estimated molecular mass of about 200 and 160 kDa respectively. They can be compared to precursors of vertebrate alpha chains of fibrillar collagens. The two other collagenase-sensitive translated peptides had a molecular mass of 81 and 48 kDa respectively and they correspond to short-chain collagens (EXPOSITO *et al.*, 1990). Poly(A)-rich RNA was then used to construc a cDNA library with λ gt 10 as vector. This library was screened first with a probe corresponding to a human collagen (type I, α1 chain). A clone, EmC4, was isolated and characterized. It coded for a short-chain, non-fibrillar collagen (EXPOSITO *et al.*, 1990). The 1.2kb insert of this clone was then used to screen the same library at high stringency. Several additional positive clones were detected, which all coded for collagenous polypeptides with a structure similar to the previous one, encoded by EmC4 (Fig. 1). These polypeptides contain two helical domains (66 and

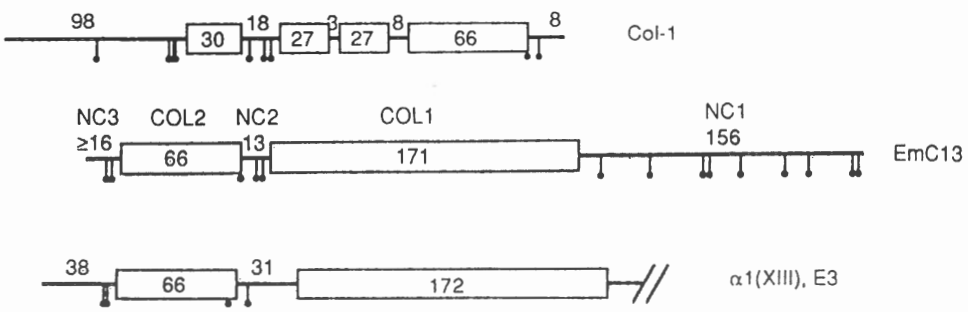

Fig. 1. — Comparison of the structure of a sponge short-chain collagen chain (EmC13), a nematode cuticular chain (Col-1) and a type XIII collagen chain (α 1 XIII, E3). The solid lines correspond to non-helical domains (NC) ; the boxes indicate the helical domains (COL). The numbers correspond to the number of amino acid residues in each domain. The position of cysteines are shown by vertical bars with a dot (modified from EXPOSITO *et al.*, 1991).

171 amino acid residues) separated by a short non-helical domain (13 to 16 amino acid residues, with 2 or 3 cysteines), and a C-terminal, non-helical domain of 156 amino acid residues with 9 cysteines (EXPOSITO *et al.*, 1991). These collagen chains had similarities with various other vertebrate and invertebrate collagen chains. The helical domains are comparable in size to the helical domains COL-1 and COL-2 of vertebrate type XIII collagen (66 and 172 amino acid residues) ; the central, short non-helical domain is similar in size and cysteine positions with interruptions in nematode cuticular collagens (EXPOSITO *et al.*, 1991) ; and the C-terminal non-helical domain can be compared to the same domain of type IV (basement membrane) collagen in vertebrates (EXPOSITO *et al.*, 1990). The genomic DNA of *E. mülleri* was digested with several restriction endonucleases and analyzed by Southern blotting. Using the EmC4 probe, 10 to 12 bands were revealed for each enzyme used, indicating that approximately 10 highly homologous genes belong to the same family. They all encode short chain collagens (EXPOSITO *et al.*, 1991). This situation is reminiscent of the nematode cuticular collagen gene family (COX *et al.*, 1984). By *in situ* hybridization methods, it was demonstrated that the cells expressing these genes were secreting the collagenous filaments (spongin) around spicules and in the basal, sticking cement (EXPOSITO *et al.*, 1991).

Screening of the cDNA library also revealed a clone, C 23, positive only at low stringency, and coding for a different collagen chain named Emf1α. The conceptual translation product comprises an uninterrupted helical domain (with one imperfection due to an additional amino acid insertion in one of the triplets) and a non-helical, C-terminal domain containing 7 cysteines (EXPOSITO and GARRONE, 1990). Computer alignement of this domain with comparable domains of vertebrate and sea urchin fibrillar collagen chains revealed homologies with vertebrate type XI collagen (EXPOSITO and GARRONE, 1990).

A genomic library was then constructed by inserting *E. mülleri* genomic DNA fragments into the *Xho* I site of λ GEM-11 DNA (Promega). This library was screened using the available sponge collagen probes. With the probe corresponding to the fibrillar collagen, the isolated gene encoded the Emf1α collagen chain (EXPOSITO and GARRONE, 1990). All the exons coding for the helical domain begin with a complete Gly codon and terminate with a complete Yaa codon. Moreover, for most of them, their size is 54 bp or a multiple of 54 bp. These two features are typical of the gene organization of fibrillar collagens (Fig. 2). Additional characters lying in the 3' end of the regions coding for the helical domain strenghten the comparison with type XI collagen (EXPOSITO and GARRONE, 1990). This comparison is highly significant if one considers that in vertebrates the type XI collagen (and also type V) has been postulated as forming a thin axial core inside large collagen fibrils (MENDLER *et al.*, 1989). Using the probes corresponding to the short-chain collagens, the genes which were cloned had not the characteristic organization of fibrillar collagen genes. The exons coding for the helical domains had variable sizes and they all began with split Gly codons (EXPOSITO *et al.*, 1991).

In conclusion, sponges contain at least two collagen types (Pl. I). One is a fibrillar collagen, presumably forming the cross-striated fibrils present in the intercellular spaces of all species. The small diameter of these fibrils (20 nm) and their

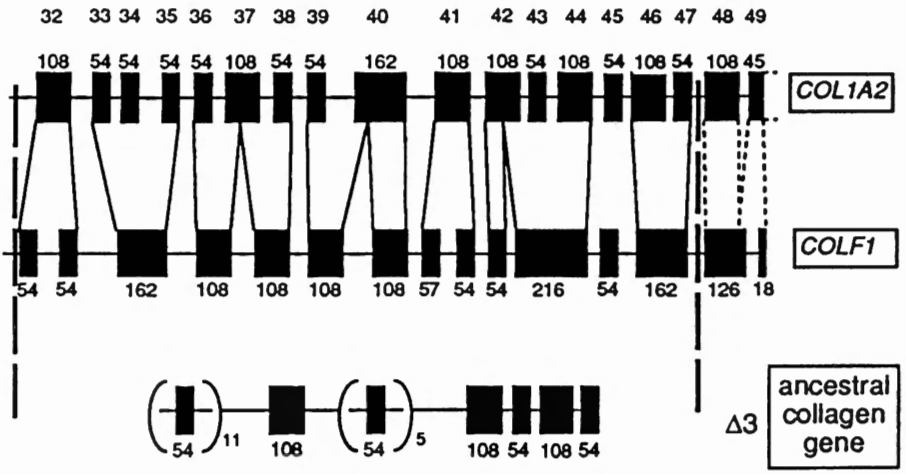

Fig. 2. — Comparison the organization of vertebrate and sponge fibrillar collagen genes for the region coding for the C-terminal end of the helical domain. The black boxes indicate the exons. *COL1A2* : gene coding for the chain α 2 of the type I collagen ; *COLF1* : sponge fibrillar collagen gene. The upper line indicates the vertebrate exon numbers. The numbers above (verbrate) and below (sponge) the exons indicate their size in base pairs. In the bottom of the figure is represented a tentative reconstruction of an hypothetical ancestral collagen gene.

very simplified banding pattern make them rudimentary collagen fibrils. It is reasonable to think that their counterpart in vertebrates (and probably in more organized invertebrates) are the fibrillar collagen types V and XI supposed to constitute the internal core of large collagen fibrils. The second sponge collagen type might be present as a basal cement in most species and as a skeletal component in certain groups. It is clearly the main constituant of the thin spongin filaments. This collagen has in fact a surface localization, even if it is secondarily internalized to make the organic skeleton. It is then not surprizing if it has features reminiscent of nematode cuticular collagens. As it is also deposited by cell layers, the comparison with basement membrane collagens is explainable. This collagen might then be the percursor of (1) external, secreted collagens such as some skeletal collagen in cnidarians, cuticular collagens in nematodes and annelids, byssus threads in molluscs, egg case in selachians, and (2) internal basement membrane collagens (BAIRATI and GARRONE, 1985). These two collagen types have probably envolved in completely separate ways.

REFERENCES

BAIRATI, A. and GARRONE (1985) — *Biology of invertebrate and lower vertebrate collagens.* Plenum Press, New York, 583 pp.

COX, G.N., J.M. KRAMER and D. HIRSH (1984) — Number and organization of collagen genes in *Caenorhabditis elegans. Mol. Cell. Biol.*, **4** : 2389-2395

EXPOSITO, J.Y. and R. GARRONE (1990) — Characterization of a fibrillar collagen gene in sponges reveals the early evolutionary appearance of two collagen gene families. *Proc. Natl. Acad. Sci. USA*, **87** : 6669-6673.

EXPOSITO, J.Y., R. OUAZANA and R. GARRONE (1990) — Cloning and sequencing of a Porifera partial cDNA coding for a short-chain collagen. *Eur. J. Biochem.*, **190** : 401-406.

EXPOSITO, J.Y., D. LE GUELLEC, Q. LU and R. GARRONE (1991) — Short chain collagens in sponges are encoded by a family of closely related genes. *J. Biol. Chem.*, **266** : 21923-21928.

FRANCILLON-VIEILLOT, H., V. DE BUFFRENIL, J. CASTANET, J. GERAUDIE, F.J. MEUNIER-,J.Y. SIRE, L. ZYLBERBERG and A. DE RICQLES (1990) — Microstructure and mineralization of vetrebrate skeletal tissues. In : *Skeletal biomineralization : patterns, processes and evolutionary trends*. vol. I (edit. J.G. CARTER) Van Nostrand Reinhold, New York, 471-530.

GARRONE, R. (1978) — *Phylogenesis of connective tissue. Morphological aspects and biosynthesis of sponge intercellular matrix*. S. Karger, Basel, 250 pp.

MENDLER, S.G. EICH-BENDER, L. VAUGHAN, K.H. WINTERHALTER and P. BRUCKNER (1989) — Cartilage contains mixed fibrils of collagen type II, IX and XI. *J. Cell Biol.*, **108** : 191-197.

TRELSTAD, R.L. (1984) — *The role of extracellular matrix in development*. Alan Liss, New York, 643 pp.

VAN DER REST, M. and R. GARRONE (1991) — Collagen family of proteins. *FASEB J.*, **6** : 2814-2823.