Proc. Indian Acad. Sci. (Chem. Sci.), Vol. 111, No. 1, February 1999, pp. 51–56

Collagen, a common thread in extracellular matrix evolution

ROBERT GARRONE
Institute of Biology and Chemistry of Proteins, CNRS UPR 412, University Lyon 1, 7, Passage du Vercors, 69367 Lyon Cedex, France
e-mail: r.garrone@ibcp.fr

Abstract. Among multicellular animals, collagen is probably one of the most constant extracellular proteins. Collagen fibrils are encountered in sponges. The organization of their genes suggests that they have not varied to a great extent during evolution. However, they are not present in all animals. Basement membrane collagen is certainly the only ubiquitous collagen.

The comparisons of collagen in lower animals suggest similarities with several collagen types of vertebrates. It is interesting to try to imagine the fundamental problems for which solutions based on collagen have been adopted. The first one is *adhesion.* Animals with poorly packed cells need a cement to be attached to the bottom of the sea. In sponges, this function has been fulfilled by microfibrils of spongin, a short-chain collagen. Later on, other organisms, such as mussels, have used a chimeric elastin-collagen protein to obtain both elasticity and strong attachment. The second important problem has been mechanical support, that is a *skeleton.* Here again, spongin has been involved in sponges. The best is the solution combining vertebrate collagen fibrils and calcium phosphate. A third vital problem has been *protection.* It is achieved by different kinds of collagens in invertebrates and lower vertebrates. It could be suggested that in vertebrates, some of the non-fibrillar collagen types are involved in tissue micro- or macro-specialization.

Keywords. Collagen; evolution; extracellular matrix.

1. Introduction

The transition from unicellular to multicellular animals has required several critical changes. The first obligatory requirement for a group of individual cells to become multicellular is a mechanism of cohesion between the units. It can be obtained simply by using an intercellular cement. Actually, several procaryotes and colonial protozoa came up this way. In such an organization, however, cell coordination, though possible, remains rudimentary. Another possibility is the establishment of cell junctions forming a layer of cells with increased coordination. This has been used clearly by some colonial protophytes which nevertheless remained unicellular organisms, and by several invertebrate larvae, which is obviously a transitory state. Most likely, the successful solution has been to combine the two systems, i.e. to connect a peripheral layer of cells, via stable junctions, and to embed some internal cells within a common matrix. This distinguishes the cells directly in contact with the environment from the cells in contact only with the intercellular matrix. In fact, the secondary specialization of a subpopulation of these cells, that is the covering layer or the internal cells, is to form the diplobastic animals. In sponges, for example, the covering layer invaginates into the cell body to make canals and chambers. The chambers are lined with highly specialized flagellate cells, constituting one fundamental layer, the endoderm. The other covering cells and the

internal cells can be replaced by each other and belong to the same category, and constitute the second fundamental layer, the ectomesenchyme. The sponges are firmly attached to a substratum with the secreted cement and the internal cells are contained within a jelly. The cement and the jelly are of the same collagenous nature. Thus, these simple organisms have invented actually *"what hold us together"* [1]. From sponges to humans, collagen is always present, in all organisms, without any exception [2].

Since variations on collagen structure and assembly are obvious in various animals, it is important to set the definitions used here for this protein.

2. Some definitions

2.1 Collagen conformation, collagen module or domain, collagen molecule

The *collagen conformation* is well known since the pioneering work of Ramachandran [3]. The repeat of the amino acid triplet [Gly–Xaa–Yaa], in which the position Xaa and Yaa are frequently occupied by proline and hydroxyproline respectively, sets a left-handed polypeptide helix, the collagen α chain. Intracellularly, three of these chains assemble according to a recognition mechanism, not totally understood, and form a right-handed coiled-coil triple helix [4]. Recent data obtained by X-ray diffraction [5] and NMR analysis [6] have confirmed the model and revealed the importance of imino acids (and above all hydroxyproline formed during the post-translational modifications of the chains) in the interactions with water molecules [6,7]. This hydration is critical in inter-chain bonding and lateral assembly of triple helices. Such triple helical, *collagen module or domain*, interrupted or not, is encountered within each of the nineteen collagen types so far discovered in vertebrates [8]. However, it can be found also in a large variety of proteins which are not named collagen [9,10]. The definition of a *collagen molecule* then implies a function in the organization of the extracellular matrix. In upper vertebrates, the extracellular matrix is inside the body and underlies or surrounds most of the cells (except circulating blood cells, neurons of the central nervous system and myelinated peripheral nerve cells). However, in invertebrates and some lower vertebrates, the extracellular matrix can lie over epithelial cells and constitutes a cuticle surrounding the animal body. This definition eliminates the circulating collectins, some cell surface receptors and some forms of acetylcholine-esterases. It includes the lung surfactant protein which, however, does not belong to the extracellular matrix.

2.2 Fundamental versus specialized collagens

Among the nineteen collagen types [8], which are the fundamental ones and which are the secondary or specialized ones? The answer to this question is uncertain as the collagen types in invertebrates have not been investigated as thoroughly as in vertebrates. However, fibril-forming collagens and basement membrane collagens appear widely distributed [11]. Due to this general occurrence they have to be considered fundamental. Collagen fibrils are observed in almost all multicellular animals indeed [2,11]. Sponges, sea anemones and jelly fish contain thin fibrils displaying a simplified banding pattern of 67 nm. cDNA and genomic cloning in sponges [12,13], biochemical studies in sea anemones [14], have demonstrated that these fibrillar collagens are very close to the group of vertebrate fibrillar collagens V and XI. As it seems now established that collagen V and XI have evolved differently than other fibrillar collagens [15,16], it is consistent to suggest that they are the most primitive fibrillar collagens. However, collagen fibrils have never

been observed in such organisms as *Caenorhabditis elegans*, *Drosophila melanogaster* and a few other groups such as the rotifers [2]. It is tempting to speculate that in these animals there are few individual wandering cells other than cells involved in defence mechanisms (for example the haemocytes in *Drosophila*). So, as a rule collagen fibrils are absolutely necessary in migrating cells: this is the case in most multicellular organisms, even (and may be above all) in the most primitive as sponges.

Basement membranes are present everywhere, except in most sponges. In fact, as explained above, sponge tissues are poorly differentiated. The only sponge group where basement membrane have been demonstrated [17] is considered to be an evolved branch due to different features (for example these sponges possess spermatozoa with acrosome, unlike the others). Therefore, basement membranes probably appeared with the establishment of stable differentiated tissues. From these evolved sponges to humans, this is always the case.

2.3 Which ancestor(s) for collagen?

The characterization of the extracellular matrix in sponges soon revealed a polymorphism of collagen. Besides collagen fibrils [12] (and type IV collagen in certain peculiar sponges [17]), a short-chain collagen was identified [18,19]. It makes large aggregates, giving rise to macroscopic structures, the best-known being the skeleton of the bath sponges [2]. This material, originally identified by Gross [20] and named "spongin B", has features similar to a variety of non-fibrillar collagens [18,19]. For example:

- the C-terminal, non-collagenous domain of individual chains appears duplicated as in the NC1 (non-triple helical) domain of type IV collagen; eight of the nine cysteine residues of this sponge collagen molecule domain are in a similar position as eight of the twelve cysteine residues of the NC1 domain of type IV collagen;
- the high number of genes involved in coding the spongin molecule and the general organization of this molecule is comparable to some of the cuticular collagens in nematodes, specially considering the length and the three cysteine residues of one of the non-collagenous domains;
- the length of the two collagenous domains of the spongin molecule (66 – following a sequence including two cysteine residues – and 171 residues) is strikingly similar to the length (66 – following a sequence including two cysteine residues – and 172 residues) of two collagenous domain of type XIII collagen; moreover, these two collagens can have an N-terminal transmembrane domain (Lu, unpublished results);
- the mixture of non-collagenous and collagenous domains obviously brings this sponge collagen and the fibril associated collagen with interrupted triple helices (FACITs) collagens near again.

Finally, one might wonder if this short-chain collagen in sponges is the collagen ancestor or if fibrillar collagen is the real ancestor. While there is no evidence to favour one hypothesis or the other, an interesting observation is the report of collagenous sequences in bacteriophages [21], with the suggestion that this phage collagen was then passed to eukaryotes. More interesting is the fact that some phage collagen sequences code for tail fiber trimeric proteins, while others encode a trimeric protein lying on the surface of the head. Here again, it is tempting to speculate that one sequence might have been the origin of fibrillar collagen (the one coding for bacteriophage tail fibers) while the other (coding for the head surface protein) has given rise to the membrane-anchored spongin. However,

a more accurate biochemical characterization of these bacteriophage proteins would be necessary to demonstrate that they are trimeric and stable in the most likely absence of hydroxyproline. As hydroxyproline is present in unicellular algae and in plants, a further transmission and adaptation of the enzyme to this new substrate is conceivable.

Once it is assumed that collagen originates from two distinct lines, it might be interesting to question the reasons for such an extended polymorphism of collagen. A tentative answer is that the varieties of collagens originate during evolution as adaptations developed to fulfill special functions, namely to promote adhesion, build a skeleton and provide protection.

3. Collagen, an adaptable protein during evolution

3.1 Adhesion, one of the necessities of life

A recent comment from Ruoslahti [22], "stretching is good for a cell", implies that for most cells contact with an extracellular matrix is necessary for proper growth and survival. This assumption can be extended to entire organisms. Generally, the action of muscles, attached onto a cell layer or a skeleton provides the necessary forces that maintain cells and organs under stress. However, when the covering cell layer is loosely tightened, as it is in sponges, the solution is a strong linkage to a substratum (i. e. rocks or shells). The basal cell layer of sponges expresses the short-chain collagen on the side which faces the substratum. Some of the molecules remain attached to the cell plasma membrane (Lu, unpublished results), while others are shedded. The result is a dense feltwork of fine filaments making a strong attachment device functioning in a watery environment. The induction of such a secretion is easy to obtain in fresh water sponge cultures, just by dropping a glass cover-slide on a sponge developing onto a microscope glass slide. The result is a genuine "sandwich", transparent sponge, expanding between the two glass layers. This kind of adhesion is highly efficient as it is difficult to remove a sponge from its support without damaging the attached lower part of the animal. Peculiarly, intronless collagen genes have been characterized in marine sponges [23] and could encode similar sticky collagens.

Another collagenous device involved in animal adhesion is constituted by the byssal apparatus of mussels [24,25]. To adhere to solid surfaces, those animals secrete filaments containing an extracorporeal collagen molecule (forming a stiff rope) flanked at each end by elastic domains, functioning as shock absorbers, and terminated by histidine-rich sequences, able to form metal-involved cross-links.

3.2 The flesh is weak and needs a skeleton

The increasing size of the body of sponges is held by a skeleton which can be composed of calcareous or siliceous units, the spicules, a mixture of a calcareous mass and siliceous units or a mixture of siliceous units embedded in the short-chain collagens [2]. Very few sponge species have no other skeleton than bundles of collagen fibrils, and they never reach a large size. The mixture of collagen and an inorganic compound is interesting as it foreshadows the skeleton of vertebrates. Several stages can be observed, from small collagen gluing points just joining siliceous needles, to siliceous spicules completely embedded in collagen and even almost entirely collagen as in the skeleton of the bath sponges. Interestingly, this skeletal collagen is in direct continuity with the basal collagen attaching the animal to its substratum [2].

Later, combinations of collagen and calcium carbonate have been observed in corals, sea anemones, sea urchins[26]. But the collagen component (fibrils or non-fibrillar collagen) is just surrounded by calcium carbonate and is not impregnated as it is with calcium phosphate in bone and dentin. Curiously, collagen-based skeletons have not been successful during evolution, except in vertebrates.

3.3 Protection with collagen, either soft or hard but suitable

The peculiar stiffness of the collagen triple helix and its resistance to proteolysis make it a good candidate for protective devices. Actually, this function of collagen has been used in a few but very well suited ways. The first use has little to do with the extracellular matrix and has been described in cnidarians, specially in *Hydra* and in reef-building corals. In addition to various forms of interstitial collagens, cnidarians (sea anemones and jelly-fish) contain a peculiar collagen molecule forming the inner layer of nematocysts capsule[27], a component of a stinging cell. The molecule is composed of three polypeptides each containing a very short repeat of collagen triplets (14 to 16) flanked by stretches of proline residues (14 to 23 in the N-terminal domain, 6 to 9 in the C-terminal domain). As the contents of the nematocyst capsule are under pressure (15 Mpa) the necessary tensile strength of the capsule must be nearly that of steel[28].

It seems that the highly toxic substances in sponges and the stinging organelles in cnidarians constitute efficient protection systems needing no additional device. Entire groups of animals have developed calcareous shells (molluscs, sea-urchins) or chitin exoskeletons (the arthropods). Few invertebrates are devoid of a covering protective, as, for example, flat worms, which in turns are able to secrete active proteolytic enzymes. Two interesting solutions have been developed in round worms. In nematodes, a collagenous cuticle surrounds the animal and is made of a family of collagenous proteins[29]. They are comparable to the spongin secreted outside the body of sponges and used for attachment. In annelids, which are larger than most of the nematodes, the cuticle is totally different and made by large fibers, arranged in a plywood-like manner. The fibers are composed of uninterrupted collagen molecules[30] which are the longest known, about 2400 nm, 8 times the length of a vertebrate fibrillar collagen molecule!

The protective function of collagen is also encountered within more evolved organisms, for example in cartilageneous fishes. For example, the egg-case of the dogfish[31] contains a collagenous protein heavily crosslinked by tannic agents.

4. Human collagens: The achievement?

In vertebrates, there is of course a mechanical function of collagen, namely the adhesion with the collagen types IV (attachment of epithelial and endothelial layers, muscle cells), VII (attachment of the basal lamina to the dermis), XVII (attachment of keratinocytes to the basal lamina), and probably with type XIII collagen due to its transmembrane domain. The skeletal function is obviously fulfilled through collagen types I, II, IX, XI, X, and is completed by the attachment of muscles and the transmission of mechanical forces via the tendons. Protection is efficient in the dermis with the fibrillar collagens I, III, V etc. However, one function which appears outstanding, even if it was probably present in invertebrates, is the direct effect of collagen on cells via cell-matrix interactions. It seems that the micro-heterogeneity of collagen expression during development modulates organogenesis. Considering the large non-collagenous domains of several collagen types (i. e. The FACITS, or type VI) the consequences of cell interactions with such domains

should have potent effects. Moreover, some collagen molecules participate in the "economy of means" of living material, that is to say that a given molecule can have several different functions, according to the part of the molecule considered. The most recent example is type XVIII collagen, the C-terminal part of which forms endostatin, an inhibitor of angiogenesis[32].

Thus, it can be assumed that in the beginning collagen was probably a structural protein only. With evolution, it became diversified and specialized. Adhesive function has become discrete and is restricted to cells and tissues; skeletal role is clearly established and characterizes the vertebrates; protective effect, although developed in some animals (i.e. the bovines, giving rise to the leather industry) remains internal and located mainly in the dermis and in some internal envelopes. The new and most relevant function is that of a messenger molecule, acting directly on cell behaviour. Now, the next challenge for collagen research is to determine which is (or are) the active domain(s) of collagen and to elucidate the mechanisms of signal transduction.

References

1. Prockop D J 1998 *Matrix Biol.* **16** 519
2. Garrone R 1978 *Phylogenesis of connective tissue* (Basel: Karger)
3. Ramachandran G N and Kartha G 1954 *Nature (London)* **174** 269
4. Rich A and Crick F H C 1961 *J. Mol. Biol.* **3** 483
5. Bella J, Eaton M, Brodsky B and Berman H M 1994 *Science* **266** 75
6. Naito A, Tuzi S and Saitô H 1994 *Eur. J. Biochem.* **224** 729
7. Kramer R Z, Vitagliano L, Bella J, Berisio R, Mazarella L, Brodsky B, Zagari A and Berman H M 1998 *J. Mol. Biol.* **280** 623
8. van der Rest M and Garrone R 1991 *FASEB J.* **5** 2814
9. Brodsky B and Shah N K 1995 *FASEB J.* **9** 1537
10. Brodsky B and Ramshaw J A M 1997 *Matrix Biol.* **15** 545
11. Garrone R 1998 In *Molecular evolution: towards the origin of metazoa* (ed.) W E G Müller (Berlin: Springer) pp. 119–139
12. Exposito J Y and Garrone R 1990 *Proc. Natl. Acad. Sci. USA* **87** 6669
13. Exposito J Y, van der Rest M and Garrone R 1993 *J. Mol. Evol.* **37** 254
14. Tillet E, Franc J M, Franc S and Garrone R 1996 *Comp. Biochem. Physiol.* **113** 239
15. Takahara K, Hoffman G G and Greenspan D S 1995 *Genomics* **29** 588
16. Vuoristo M M, Pihlajamaa T, Vandenberg P, Prockop D J and Ala-Kokko L 1995 *J. Biol. Chem.* **270** 22873
17. Boute N, Exposito J Y Boury-Esnault N, Vacelet J, Noro N, Miyazaki K, Yoshizato K and Garrone R 1996 *Biol. Cell* **88** 37
18. Exposito J Y, Ouazana R and Garrone R 1990 *Eur. J. Biochem.* **190** 401
19. Exposito J Y, Le Guellec D, Lu Q and Garrone R 1991 *J. Biol. Chem.* **266** 21923
20. Gross J, Sokal Z and Rougvie M 1956 *J. Histochem. Cytochem.* **4** 227
21. Smith M C M, Burns N, Sayers J R, Sorrell J A, Casjens S R and Hendrix R W 1998 *Science* **279** 1834
22. Ruoslahti E 1997 *Science* **276** 1345
23. Aho S, Turakainen H, Onnela M L and Boedtker H 1993 *Proc. Natl. Acad. Sci. USA* **90** 7288
24. Coyne K J, Qin X X and Waite J H 1997 *Science* **277** 1830
25. Waite J H, Qin X X and Coyne K J 1998 *Matrix Biol.* **17** 93
26. Lethias C, Exposito J Y and Garrone R 1997 *Eur. J. Biochem.* **245** 434
27. Kurz E M, Holstein T W, Petri B M, Engel J and David C N 1991 *J. Cell Biol.* **115** 1159
28. Holstein T W, Benoit M, Herder G V, Wanner G, David C N and Gaub H E 1994 *Science* **265** 402
29. Kramer J M *FASEB J.* **8** 329
30. Gaill F, Mann K, Wiedemann H, Engel J and Timpl R 1995 *J. Mol. Biol.* **246** 284
31. Luong T T, Boutillon M M, Garrone R and Knight D P 1998 *Biochem. Biophys. Res. Commun.* **250** 657
32. Sasaki T, Fukai N, Mann K, Göhring W, Olsen B R and Timpl R 1998 *EMBO J.* **17** 4249