

SCIENTIFIC REPORTS



OPEN

Collagen cross-linking: insights on the evolution of metazoan extracellular matrix

Fernando Rodriguez-Pascual¹ & David Anthony Slatter²

Received: 31 May 2016

Accepted: 28 October 2016

Published: 23 November 2016

Collagens constitute a large family of extracellular matrix (ECM) proteins that play a fundamental role in supporting the structure of various tissues in multicellular animals. The mechanical strength of fibrillar collagens is highly dependent on the formation of covalent cross-links between individual fibrils, a process initiated by the enzymatic action of members of the lysyl oxidase (LOX) family. Fibrillar collagens are present in a wide variety of animals, therefore often being associated with metazoan evolution, where the emergence of an ancestral collagen chain has been proposed to lead to the formation of different clades. While LOX-generated collagen cross-linking metabolites have been detected in different metazoan families, there is limited information about when and how collagen acquired this particular modification. By analyzing telopeptide and helical sequences, we identified highly conserved, potential cross-linking sites throughout the metazoan tree of life. Based on this analysis, we propose that they have importantly contributed to the formation and further expansion of fibrillar collagens.

Collagens constitute a large family of extracellular matrix (ECM) proteins that play a fundamental role in providing the structural integrity and biomechanical properties of different tissues¹. In vertebrates, 28 types of collagens have been described (I–XXVIII), which are divided into several families, the most important being fibrillar collagens (I–III, V, XI, XXIV, and XXVII) and basement membrane-forming collagen IV².

Fibrillar collagens form homotrimeric (three identical α -chains) or heterotrimeric (two or three distinct polypeptide chains) molecules. Each α -chain consists of a major uninterrupted triple helical or collagenous domain (characterized by a repetition of G–X–X' triplets, where G is glycine, X is commonly proline and X' is commonly hydroxyproline) flanked by N- and C-terminal non-collagenous domains, the N- and C-propeptides.

The biosynthesis of collagen is a highly complicated process involving numerous steps, including chain association and folding, secretion, procollagen processing and cross-linking (see Supplementary Figure 1 for a graphical representation)^{3,4}. As exemplified for human type I collagen, a heterotrimeric molecule composed of two $\alpha 1$ and one $\alpha 2$ chains, after synthesis on the ribosome and their import into the rough endoplasmic reticulum, collagen chains are subjected to a series of post-translational modifications resulting in the assembly of procollagen chains. These include hydroxylation of specific proline residues, catalyzed by prolyl-4- and prolyl-3-hydroxylases (P4H and P3H enzymes); hydroxylation of specific lysine residues by lysyl hydroxylases (encoded in human by PLOD1–3 genes); N- and O-linked glycosylation, disulphide bonding and prolyl cis-trans isomerization⁵. Association of the three α -chains occurs through a process governed by the C-terminus, and the formation of the triple helix is propagated towards the N-terminal end in a zipper-like fashion to form the procollagen molecule⁶. This precursor molecule is then transported to the Golgi network where it is packaged into specialized secretory vesicles prior to export into the extracellular medium. Formation of fibrils from procollagen chains requires their proteolytic processing. The N- and C-propeptides are cleaved off by metalloproteinases belonging to the ADAMTS (a disintegrin and metalloproteinase with thrombospondin motifs) and BMP1 (bone morphogenetic protein 1)/Tolloid-like families, respectively, yielding the tropocollagen molecule, which retains a short portion of the propeptides termed telopeptides. Then, lysine or hydroxylysine residues within these non-collagenous domains are oxidatively deaminated by LOX, yielding the corresponding aldehydes which constitute the initiation products for the cross-linking formation (see Supplementary Figure 2 for details). Within hours of helix formation, these telopeptide aldehydes spontaneously react with helical lysines or hydroxylysines

¹Centro de Biología Molecular "Severo Ochoa" Consejo Superior de Investigaciones Científicas (C.S.I.C.)/Universidad Autónoma de Madrid (Madrid), Madrid, Spain. ²School of Medicine, University of Cardiff, Cardiff, United Kingdom. Correspondence and requests for materials should be addressed to F.R.-P. (email: frodriguez@cbm.csic.es)

to form immature cross-links, which further react among them and with remaining lysine or hydroxylysine residues over months/years to form permanent cross-links^{5,7–9}. The formation of these permanent or mature cross-links is fundamental as they determine the topology of adjacent molecules and contribute to the stiffness of the collagen fibril, where variations in the usage of lysine or hydroxylysine in both telopeptide and helix sites modulate the mechanical properties of the collagen matrix. In fact, defects in PLOD2, the lysyl hydroxylase isoform that specifically acts on lysine residues in collagen telopeptides, are responsible for Brück syndrome, a heritable disorder in the osteogenesis imperfecta spectrum, characterized by severe reduction or elimination of the telopeptide hydroxylysine-derived cross-links, resulting in bone fragility¹⁰. In addition to lysyl hydroxylases, collagen-associated proteins such as the small leucine-rich protein (SLRP), fibromodulin, have been shown to influence collagen cross-linking, as recently evidenced in fibromodulin-deficient mice^{11,12}.

Fibrillar collagens are present from sponges to humans, therefore often being associated with metazoan multicellularity and evolution¹³. Given the modular nature of the collagen molecule, different approaches have been used to study its evolutionary history, including phylogenetic or genomic analyses of the propeptides, the triple helix, or even of the intron-exon organization^{14–17}. Regardless of the approach, all authors agree on the appearance of an ancestral fibrillar collagen in the lineage leading up to the Metazoa, given the fact that the choanoflagellate *Monosiga brevicollis*, the sister group to Metazoa, encode proteins with triple helical and C-propeptide (COLFI) sequences in different polypeptides, while collagens present in sponges include both in the same protein chain^{14,18}. From this point on, two general hypotheses have been raised to explain the evolutionary origin of the three clades (A, B and C) identified in mammalian collagens. On one hand, the ancestor chain might have diverged into two clades (A and B/C) before the poriferan radiation, followed by a further division of B/C clade before the Parazoa/Eumetazoa split. In the second hypothesis, B and C clades are already present in demosponges and their emergence predated metazoan cladogenesis^{14,19}. Whatever the case, a remarkable expansion of collagen chains occurred in the ancestors of chordates, where gene duplications in the three clades led to the formation of specialized cartilage and bone collagens, a fundamental feature of vertebrate skeletons²⁰.

In a recent work, we have proposed that early collagen evolution has been governed by structural requirements intended to preserve the position of spatially sensitive cross-linking sites²¹. On the other hand, we also described the presence of LOX domains, not only in animals, but also in many other eukaryotes, as well as in bacteria and archaea, organisms that are devoid of recognizable fibrillar collagen-based ECM structure²². These observations indicate that the origin of LOX enzymes pre-dates the appearance of fibrillar collagens, and suggest that, by co-option of this enzymatic activity, ancestral collagen domains might have evolved into modern fibrillar collagens within the metazoan kingdom. To test this hypothesis, we have surveyed a variety of collagen sequences of different metazoan lineages in order to infer the evolutionary transitions leading to cross-linking sites. By analyzing telopeptide and helical sequences, we identified highly conserved, potential cross-linking sites throughout the metazoan tree of life. Based on this analysis, we propose that they have importantly contributed to the formation and further expansion of fibrillar collagens.

Results

In order to answer the question as to what extent potential sites for LOX-mediated cross-linking are present and conserved among metazoan lineages, we inspected the sequences of fibrillar collagens involved in the cross-linking reaction, namely the N- and C-telopeptides, and the corresponding C- and N-helical segments. C-propeptide sequences are the most conserved region of the fibrillar collagens, and the pattern of conservation of cross-linking sites can be analyzed by looking at the sequences downstream of the $(GPP)_n$ repeats, the motif marking the C-terminus of the triple helix²³. We searched for sequence similarities for the human $\alpha 1$ (I) or $\alpha 1$ (II) C-telopeptide cross-linking sites: QEKAH and REKGP (K being the lysine providing the ϵ -amino group), respectively, in several metazoan clade A collagens as homology within this clade had been previously reported⁸. Multiple sequence alignment shows the presence of homologous sequences among the groups analyzed, with the pattern $XKX'X''$, where X is any residue, K the lysine involved in the reaction, X' is glycine (mostly) or alanine (both small, non-polar aminoacids), and X'' is proline in most of cases, as shown in the weighted logo (Fig. 1A and B). N-propeptide sequences are the most variable within collagen families, including, in addition to the N-telopeptide, a cysteine-rich repeat, the von Willebrand factor-type C (VWC) module, a thrombospondin N-terminal-like domain (TSPN), or, as in some invertebrates α chains, a whey acidic protein (WAP) or von Willebrand factor A domain (VWA) modules, among others²⁴. In most cases, the presence of a short triple helix marks the beginning of the N-telopeptide. Sequence comparison shows that a certain degree of homology was also observed around the cross-linking site (Fig. 2A and B). As for C-telopeptide sequences, a significant number of species display the pattern $XKX'X''$, with little variations in X' and X'', such as in human, abalone and hydra. In addition to local sequence homology around the cross-linking telopeptide lysines, none of the collagens illustrated have any lysine residue between the C-terminal ends of the short N-terminal or main helices onto the cross-linking telopeptide lysines. Given a lysine occurrence of 7.2% in proteins, this has a probability of 1.2×10^{-9} (0.928^{275}) of occurring randomly across all 19 C-terminal telopeptide sequences, and a probability of 1.9×10^{-6} (0.928^{176}) for the 14 sequences between the (latterly removed) N-terminal helix and the cross-link, and is therefore a conserved feature. The regions between the C-terminal cross-link lysine onto the end of the molecule, and between the N-terminal cross-link lysine onto the main helix are also very lysine-poor, with no other lysine within five residues of the cross-linking one and the majority of sequences having no other lysine at all.

We also searched N- and C-terminal helix sequences for homology in the cross-linking sites. While the major triple helix is frequently interspersed with lysine residues, particularly in the third position of the collagen triplet, a recognizable and highly conserved pattern is observed at both N- and C-terminal helical cross-linking sites: $YKGY''Y'''$ (Fig. 3A and B), where Y is any residue, K the lysine involved in the condensation reaction, the third position is invariably glycine as essential part of the collagen triplet repeats, and Y'' and Y''' are histidine and arginine, respectively, in most of the cases.

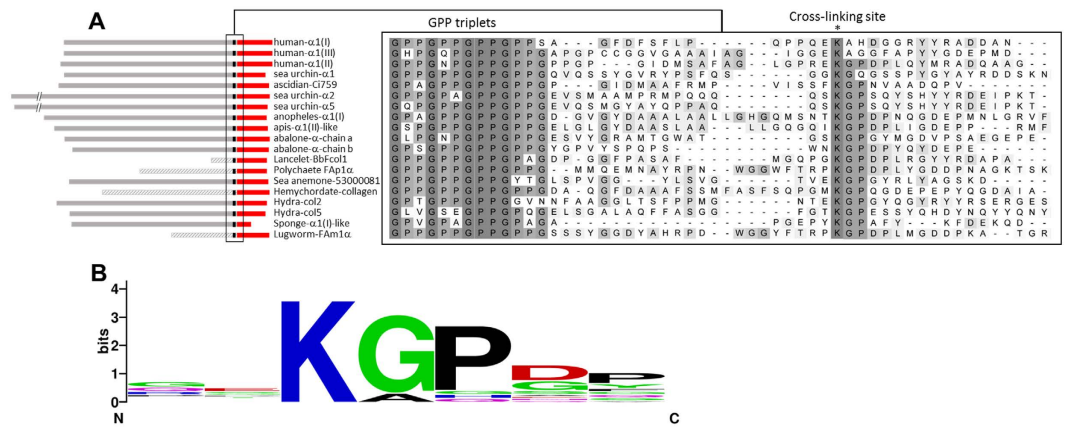


Figure 1. Multiple sequence alignment of potential C-telopeptide cross-linking sites among several metazoan collagens. (A) A group of 19 clade A (or A-like) collagens representing different metazoan lineages were aligned using the ClustalW algorithm and potential cross-linking sites within the C-telopeptide sequences were manually checked in the alignment⁴⁷. Left panel shows the location of the aligned segments within the analyzed chains, with the fibrillar collagen C-terminal domain (COLFI, pfam01410) marked in red. Right part shows the alignment, which displays the degree of identity on a gray scale. The position of the GPP triplets, the known motif marking the C-terminal end of the helix region, and the potential cross-linking sites are indicated above the alignment. (B) Sequence logo of the position weight matrix of the putative cross-linking sites analyzed in the C-telopeptides collagen chains obtained using WebLogo⁴⁸.

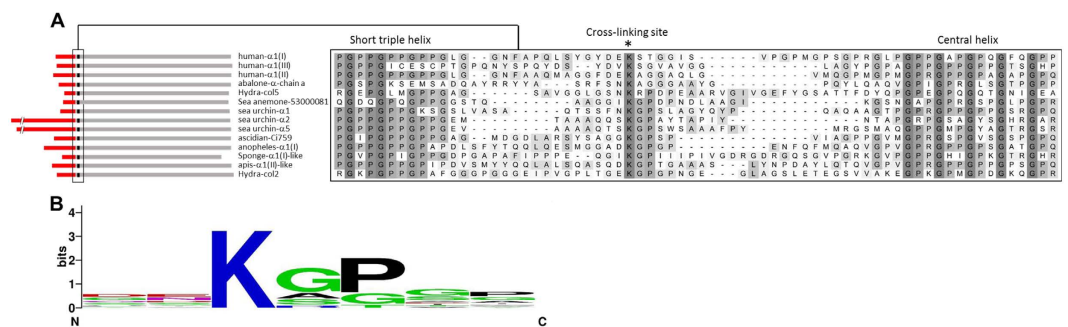


Figure 2. Sequence alignment of N-telopeptide cross-linking sites. (A) Metazoan collagen chains used in Fig. 3 were aligned at the N-telopeptide sequences to search for potentially conserved cross-linking sites. Left panel shows the location of the aligned segments, with the N-propeptides marked in red. Right part shows the alignment, which is delimited by the position of the short triple helix and the beginning of the central helical segment. Sequences omitted from the C-propeptide analysis were either incomplete (lancelet, polychaete, lugworm, hemychordate) or lack of significant homology (abalone- α -chain b). (B) Sequence logo of the position weight matrix of the putative cross-linking sites analyzed in the N-telopeptides collagen chains.

Discussion

This study identified potential collagen cross-linking sites, which are conserved throughout the metazoan lineage. Several assumptions can be raised when these patterns of sequence conservation are analyzed across the tree of life (Fig. 4). First, these sequences were not found in the triple helix- or COLFI-containing polypeptides of the choanoflagellate *Monosiga brevicollis*. As this organism has been described not to express true fibrillar collagen, the acquisition of potential cross-linking sites is genuinely linked to the ability to produce this important constituent^{18,22,25}. This observation, together with the fact that LOX domains were shown to have a pre-metazoan origin, suggests that LOX enzymes might have been co-opted for collagen cross-linking in the lineage leading up to the Metazoa, presumably contributing to the formation of the ancestral fibrillar collagen. According to our study, N- and C-propeptide KGP sites first appeared in sponges, where fibrillar collagen has been identified to form the mesohyl, the central cavity that acts as an endoskeleton, supporting the tubular shape of sponges²⁶. Interestingly, the acquisition of these cross-linking sites is coincident with the appearance of LOX proteins containing scavenger receptor cysteine-rich (SRCR) domains, suggesting these domains were key for the co-option of LOX enzymes to the remodeling of fibrillar collagens²². In fact, with the exceptions of some hydra, abalone and sea urchin chains, N- and C-propeptide KGP sites remain invariably until the appearance of vertebrate collagens, when this triplet experiences numerous changes (KAH, KAG, KST, KSG). From a LOX perspective, these variants were associated with new LOX protein architectures, without the SRCR domains and with

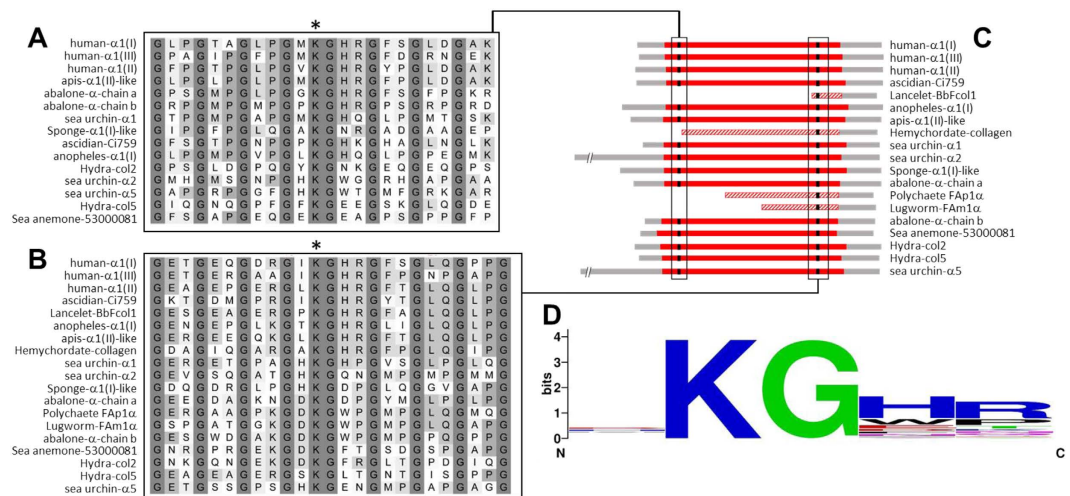


Figure 3. Sequence alignment of the helical cross-linking sites. Metazoan collagen chains used in Fig. 3 were aligned at the N- (A) and C- (B) ends of the central helical segment to search for homology at cross-linking sites. Panel C shows the location of the aligned segments, with the central helix marked in red. Available hemychordate, polychaete, lugworm and lancelet chains were incomplete at the N-terminal end. (B) Sequence logo of the position weight matrix of the putative cross-linking sites analyzed in the helical segments.

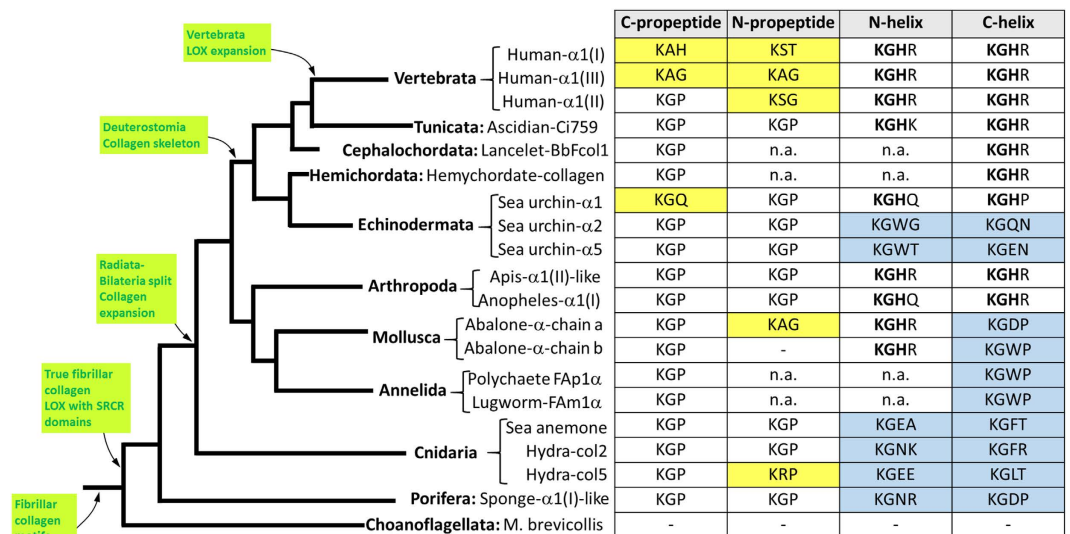


Figure 4. Evolution of collagen cross-linking sites across the tree of life. The figure combines the sequence information on the collagen cross-linking sites identified in the study and a cladogram representing a consensus view of the eukaryotic tree of life from the common ancestor of Metazoa and choanoflagellates. Yellow- and blue-colored table cells represent cross-linking sites in propeptides and helix, respectively, which are different from the consensus KGP (propeptide) and KGH (helix). Green boxes represent milestones in the evolution of fibrillar collagen and LOX enzymes. Blank cells indicate lack of homology, n.a.: sequences at those sites not available.

propeptide and proline-rich regions, presumably contributing to the expansion of vertebrate collagens. Our study also shows different patterns of conservation in propeptide sites with respect to those in helical segments. This can be explained by the fact that helical lysines or hydroxylysines are not directly modified by LOX, but rather they are acceptor sites for the attack of telopeptide aldehydes, without participation of LOX enzymes. This circumstance likely determined distinct evolutionary events compared to those having occurred in the propeptides. With the exception of some sea urchin chains, our analysis shows a transition from sequences displaying only KG as conserved pattern, to those with KGH, with some organisms having both, one in N-helix and another in C-helix, such as the abalone chains. Considering that KG doublet is very frequently found in the helical segments, the acquisition of the specific KGH pattern might have contributed to fix the length of the helix and thereby its orientation within the supramolecular structure of the fiber²¹. Interestingly, this evolutionary transition predated the Radiata-Bilateria split, when a significant expansion of collagen chains has been proposed to occur¹⁴. Therefore, a

model for the acquisition and evolution of cross-linking sites in fibrillar collagen is presented here that suggests, on one hand, that sequences becoming the target for LOX-catalyzed oxidation appeared soon in the metazoan lineage, and were invariably conserved until vertebrate LOX and collagen expansion. On the other one, helical cross-linking KGH sites might have fueled bilaterian evolution by strengthening and fixing the length and the structural orientation of collagen chains.

Collagen triple helix repeat-containing proteins have been also identified in bacteria and other non-metazoan forms of life, for which the formation of fibrils and higher order structures has not been described^{27–30}. These collagen-like sequences, proposed to be horizontally transferred from metazoan, are always flanked by non-collagenous domains with structural motifs of surface or spore-associated proteins, which lack of a recognizable signature for the action of LOX enzymes, making highly improbable the formation of initiation products, and hence, the subsequent steps to the generation of covalent cross-links. Therefore, the identification of a highly conserved pattern of potential cross-linking sites in the C- and N-terminal domains is strongly associated to metazoan collagen chains, presumably having contributed to the formation and expansion of fibrillar collagens.

Solid biochemical data to support these assumptions is missing and there is only limited and very fragmented information about the existence of collagen cross-links in several organisms. For example, LOX-generated cross-links have been isolated from a sponge (*Haliciona oculata*), a sea urchin (*Strongylocentrotus droebachensis*), sea cucumbers (*Sclerodactyla briarius* and *Holothuria forskali*), and some cephalopods (*Loligo vulgaris* and *Sepia officinalis*) as well as from several annelids and molluscs^{31–34}. Indirect evidences also suggest the cross-linking of collagen in other organisms. In the cnidarian *Hydra vulgaris*, which encodes six type I-like chains (Hcol1 to Hcol6), lathyric agents, well-known inhibitors of collagen cross-linking, have been shown to impair head regeneration, an evidence indicating a role for this collagen modification in *Hydra* morphogenesis^{35–37}. Putative cross-linking sites were also identified in the marine polychaetous annelid *Alvinella pompejana*, for whose structural material a high thermal and chemical stability has been described, probably reflecting a high degree of cross-linking³⁸. Nevertheless, it should be taken into account that other enzymes different than LOX, such as transglutaminases, have been also reported to cross-link collagens and other matrix proteins, and therefore, LOX-mediated remodeling might not be the exclusive route to a high stability³⁹.

From a structural point of view, sequence requirements for LOX activity have been studied *in vitro* using synthetic oligopeptides⁴⁰. By providing polyglycine peptide substrates to purified LOX in enzymatic activity assays, these authors investigated the influence of sequences found in the N- or C-telopeptides of mammalian $\alpha 1(I)$ collagen chains. While the C-telopeptide sequence –QEK– was moderately favored when compared to the corresponding peptide with the lysine residue alone, the N-telopeptide sequence –DEK– was not oxidized at all. Although this study did not analyze the contribution of residues C-terminal to the reactive lysine, found to be highly conserved in our work, these observations reflect a higher degree of complexity in the recognition of collagen chains by LOX than that brought about by short peptides in solution. To this respect, LOX enzymes have been shown to be more active with collagen forming fibrils than with solubilized forms as a substrate, and more importantly, their binding capacity highly influenced by collagen-binding proteins such as small leucine-rich proteins (SRLP)^{11,12,41}. Closely linked to this, LOX enzymes are also responsible for the cross-linking of tropoelastin monomers during the process of formation of elastic fibers, and sequence recognition within the tropoelastin chain very much differs from that observed in collagens⁴². Here too, the specificity of LOX on elastin substrates is complex and highly regulated by structural constraints and ancillary proteins, such as fibulins and fibrillins^{43–45}. Interestingly, from an evolutionary perspective, it is much more evident that LOX were co-opted for a new function as elastin appeared much later in evolution compared to collagens and other matrix proteins⁴⁶.

As a conclusion, this article has attempted to provide a general view of the process of collagen cross-linking by LOX enzymes, as well as to address the question about how this post-translational modification might have been incorporated into the collection of collagen biochemical reactions during evolution. Based on the identification of putative LOX-mediated cross-linking sites throughout the metazoan lineage, we propose that pre-existing LOX enzymes might have been co-opted for collagen cross-linking at the dawn of the Metazoa, thereby promoting the intermolecular association of ancestral collagen chains and the subsequent impact on their mechanical properties. Further refinement of N- and C- telopeptide cross-linking sites in vertebrates was associated with the appearance of new LOX genes likely contributing to collagen expansion, while helical sites might have evolved into the highly conserved KGH sequence to support the structural properties of the collagen chain, confirming the previous assumption that cross-linking has played a fundamental role in the evolution of the collagen fibril²¹. It should be nevertheless taken into account that this *in silico* analysis has only considered sequences in the clade A collagens, deliberately excluding those belonging to clade B, such as collagen XI (fibril-forming) or IX (fibril-associated), described also to be the subject of LOX-mediated cross-linking³. During the next years, further molecular and evolutionary genomic analyses will permit to get more insight about the contribution of this chemical modification to the evolution of the collagen fibrils.

Methods

Collagen sequences used in the analysis were:

Human $\alpha 1(I)$ [Genbank: AAB94054].

Human $\alpha 1(III)$ [AGL34959].

Human $\alpha 1(II)$ [NP_001835].

Sea urchin- $\alpha 1$ collagen precursor *Strongylocentrotus purpuratus* [NP_999674].

Ascidian-Ci759 fibrillar collagen α chain *Ciona intestinalis* genome: jgi-psf.org/ciona4/ciona4.home.html [ciona4]150759[ci0100150759].

Sea urchin $\alpha 2$ collagen *Strongylocentrotus purpuratus* [NP_999675].

Sea urchin $\alpha 5$ collagen *Paracentrotus lividus* [CAE53096].

Anopheles- α 1 (1) *Anopheles sinensis* [KFB49448].
 Apis- α 1(II)-like *Apis mellifera* [XP_393523].
 Abalone- α -chain a *Haliotis discus* [BAA75669].
 Abalone α -chain b *Haliotis discus* [BAA75668].
 Lancelet-BbFcol1 fibrillar collagen *Branchiostoma belcheri* [BAD97679].
 Polychaete FAp1a fibrillar collagen chain *Alvinella pompejana* [AAC35289].
 Sea anemone-53000081 collagen α -chain *Nematostella vectensis* genome: jgi.doe.gov/Nemve1/Nemve1.home.html [Nemve1|204742|fgenesh1_pg.scaffold_53000081].
 Hemichordate fibrillar collagen *Saccoglossus kowalevskii* [ABB83364].
 Hydra-col2 *Hydra vulgaris* [ABG80449].
 Hydra-col5 *Hydra vulgaris* [ABG80451].
 Sponge- α 1(I)-like collagen *Amphimedon queenslandica* [XP_003388783].
 Lugworm-FAm1 α fibrillar collagen *Arenicola marina* [AAC47545].
Monosiga brevicollis [XP_001748906] and [XP_001749460] with triple helix, [XP_001744748], [XP_001747370] and [XP_001744747] with COLFI domains.

Sequence alignments are available by the corresponding author upon request.

References

- Fratzl, P. Collagen: Structure and Mechanics, an Introduction. In: *Collagen* (ed. Fratzl, P.). Springer, US (2008).
- Kadler, K. E., Baldock, C., Bella, J. & Boot-Handford, R. P. Collagens at a glance. *J Cell Sci* **120**, 1955–1958 (2007).
- Hulmes, D. J. S. Collagen Diversity, Synthesis and Assembly. In: *Collagen* (ed. Fratzl, P.). Springer, US (2008).
- Kadler, K. E., Holmes, D. F., Trotter, J. A. & Chapman, J. A. Collagen fibril formation. *Biochem J* **316** (Pt 1), 1–11 (1996).
- Yamauchi, M. & Sricholpech, M. Lysine post-translational modifications of collagen. *Essays In Biochemistry* **52**, 113–133 (2012).
- Brodsky, B. & Persikov, A. V. Molecular Structure of the Collagen Triple Helix. In: *Advances in Protein Chemistry: Fibrous Proteins: Coiled-Coils, Collagen and Elastomers* (ed. Parry, D. A. & Squire, J. M.). Academic Press (2005).
- Eyre, D. R., Paz, M. A. & Gallop, P. M. Cross-Linking in Collagen and Elastin. *Annual Review of Biochemistry* **53**, 717–748 (1984).
- Eyre, D. R. & Wu, J. J. Collagen cross-links. In: *Topics in Current Chemistry: Collagen* (ed. Brinckmann, J., Notbohm, H. & Müller, P. K.) Springer Berlin Heidelberg (2005).
- Yamauchi, M. & Shiiba, M. Lysine hydroxylation and cross-linking of collagen. *Methods Mol Biol* **446**, 95–108 (2008).
- van der Slot, A. J. *et al.* Identification of PLOD2 as Telopeptide Lysyl Hydroxylase, an Important Enzyme in Fibrosis. *Journal of Biological Chemistry* **278**, 40967–40972 (2003).
- Kalamajski, S. *et al.* Increased C-telopeptide Cross-linking of Tendon Type I Collagen in Fibromodulin-deficient Mice. *Journal of Biological Chemistry* **289**, 18873–18879 (2014).
- Kalamajski, S., Bihan, D., Bonna, A., Rubin, K. & Fardale, R. W. Fibromodulin Interacts with Collagen Cross-linking Sites and Activates Lysyl Oxidase. *Journal of Biological Chemistry* **291**, 7951–7960 (2016).
- Ozbek, S., Balasubramanian, P. G., Chiquet-Ehrismann, R., Tucker, R. P. & Adams, J. C. The evolution of extracellular matrix. *Mol Biol Cell* **21**, 4300–4305 (2010).
- Exposito, J.-Y., Valcourt, U., Cluzel, C. & Lethias, C. The Fibrillar Collagen Family. *Int J Mol Sci* **11**, 407–426 (2010).
- Ivanova, V. P. & Krivchenko, A. I. A current viewpoint on structure and evolution of collagens. I. Fibrillar collagens. *J Evol Biochem Phys* **48**, 127–139 (2012).
- Boot-Handford, R. P. & Tuckwell, D. S. Fibrillar collagen: the key to vertebrate evolution? A tale of molecular incest. *Bioessays* **25**, 142–151 (2003).
- Huxley-Jones, J., Robertson, D. L. & Boot-Handford, R. P. On the origins of the extracellular matrix in vertebrates. *Matrix Biol* **26**, 2–11 (2007).
- King, N. *et al.* The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783–788 (2008).
- Exposito, J.-Y. *et al.* Demosponge and Sea Anemone Fibrillar Collagen Diversity Reveals the Early Emergence of A/C Clades and the Maintenance of the Modular Structure of Type V/XI Collagens from Sponge to Human. *Journal of Biological Chemistry* **283**, 28226–28235 (2008).
- Wada, H., Okuyama, M., Satoh, N. & Zhang, S. Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. *Evolution & Development* **8**, 370–377 (2006).
- Slatter, D. A. & Fardale, R. W. Structural constraints on the evolution of the collagen fibril: convergence on a 1014-residue COL domain. *Open Biology* **5** (2015).
- Grau-Bové, X., Ruiz-Trillo, I. & Rodriguez-Pascual, F. Origin and evolution of lysyl oxidases. *Scientific Reports* **5**, 10568 (2015).
- Eyre, D. R., Weis, M., Hudson, D. M., Wu, J.-J. & Kim, L. A Novel 3-Hydroxyproline (3Hyp)-rich Motif Marks the Triple-helical C Terminus of Tendon Type I Collagen. *Journal of Biological Chemistry* **286**, 7732–7736 (2011).
- Exposito, J.-Y. & Lethias, C. Invertebrate and Vertebrate Collagens. In: *Evolution of Extracellular Matrix* (ed. Keeley F. W., Mecham R. P.). Springer Berlin Heidelberg (2013).
- Williams, F., Tew, H. A., Paul, C. E. & Adams, J. C. The predicted secretomes of *Monosiga brevicollis* and *Capsaspora owczarzaki*, close unicellular relatives of metazoans, reveal new insights into the evolution of the metazoan extracellular matrix. *Matrix Biol* **37**, 60–68 (2014).
- Heinemann, S. *et al.* Ultrastructural studies on the collagen of the marine sponge *Chondrosia reniformis* Nardo. *Biomacromolecules* **8**, 3452–3457 (2007).
- Yu, Z., An, B., Ramshaw, J. A. M. & Brodsky, B. Bacterial collagen-like proteins that form triple-helical structures. *Journal of Structural Biology* **186**, 451–461 (2014).
- Xu, C., Yu, Z., Inouye, M., Brodsky, B. & Mirochnitchenko, O. Expanding the Family of Collagen Proteins: Recombinant Bacterial Collagens of Varying Composition Form Triple-Helices of Similar Stability. *Biomacromolecules* **11**, 348–356 (2010).
- Rasmussen, M., Jacobsson, M. & Björck, L. Genome-based Identification and Analysis of Collagen-related Structural Motifs in Bacterial and Viral Proteins. *Journal of Biological Chemistry* **278**, 32313–32316 (2003).
- Xu, Y., Keene, D. R., Bujnicki, J. M., Höök, M. & Lukomski, S. Streptococcal Scl1 and Scl2 Proteins Form Collagen-like Triple Helices. *Journal of Biological Chemistry* **277**, 27312–27318 (2002).
- Eyre, D. R. & Glimcher, M. J. Comparative biochemistry of collagen crosslinks: Reducible bonds in invertebrate collagens. *Biochimica et Biophysica Acta (BBA) - Protein Structure* **243**, 525–529 (1971).
- Van Ness, K. P., Koob, T. J. & Eyre, D. R. Collagen cross-linking: distribution of hydroxyproline cross-links among invertebrate phyla and tissues. *Comparative biochemistry and physiology B, Comparative biochemistry* **91**, 531–534 (1988).
- Shadwick, R. Crosslinking and Chemical Characterisation of Cephalopod Collagens. In: *Biology of Invertebrate and Lower Vertebrate Collagens* (ed. Bairati, A., Garrone, R.), Springer, US (1985).

34. Bailey, A. The Collagen of the Echinodermata. In: *Biology of Invertebrate and Lower Vertebrate Collagens* (ed. Bairati, A., Garrone, R.), Springer, US (1985).
35. Barzansky, B. & Lenhoff, H. M. On the chemical composition and developmental role of the mesoglea of hydra. *Integrative and Comparative Biology* **14**, 575–581 (1974).
36. Deutzmann, R. *et al.* Molecular, biochemical and functional analysis of a novel and developmentally important fibrillar collagen (Hcol-I) in hydra. *Development* **127**, 4669–4680 (2000).
37. Zhang, X. *et al.* The Collagens of Hydra Provide Insight into the Evolution of Metazoan Extracellular Matrices. *Journal of Biological Chemistry* **282**, 6792–6802 (2007).
38. Pradillon, F. & Gaill, F. Adaptation to deep-sea hydrothermal vents: some molecular and developmental aspects. *J Mar Sci Technol* **15**, 37–53 (2007).
39. Collighan, R. J. & Griffin, M. Transglutaminase 2 cross-linking of matrix proteins: biological significance and medical applications. *Amino Acids* **36**, 659–670 (2009).
40. Nagan, N. & Kagan, H. M. Modulation of lysyl oxidase activity toward peptidyl lysine by vicinal dicarboxylic amino acid residues. Implications for collagen cross-linking. *Journal of Biological Chemistry* **269**, 22366–22371 (1994).
41. Siegel, R. C. Biosynthesis of collagen crosslinks: increased activity of purified lysyl oxidase with reconstituted collagen fibrils. *Proceedings of the National Academy of Sciences of the United States of America* **71**, 4826–4830 (1974).
42. Kozel, B. A., Mecham, R. P. & Rosenbloom, J. Elastin. In: *The Extracellular Matrix: an Overview* (ed. Mecham, R. P.), Springer-Verlag (2011).
43. Horiguchi, M. *et al.* Fibulin-4 conducts proper elastogenesis via interaction with cross-linking enzyme lysyl oxidase. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 19029–19034 (2009).
44. Liu, X. *et al.* Elastic fiber homeostasis requires lysyl oxidase-like 1 protein. *Nat Genet* **36**, 178–182 (2004).
45. Thomassin, L. *et al.* The Pro-regions of Lysyl Oxidase and Lysyl Oxidase-like 1 Are Required for Deposition onto Elastic Fibers. *Journal of Biological Chemistry* **280**, 42848–42855 (2005).
46. Keeley, F. W. The evolution of elastin. In: *Evolution of Extracellular Matrix* (ed. Keeley, F. W., Mecham, R. P.), Springer, Berlin Heidelberg (2013).
47. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* **22**, 4673–4680 (1994).
48. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Research* **14**, 1188–1190 (2004).

Acknowledgements

We thank Iñaki Ruiz-Trillo (C.S.I.C.-Universitat Pompeu Fabra, Barcelona, Spain) for critical reading the manuscript and valuable suggestions.

Author Contributions

Both authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. F.R.-P. and D.A.S. Analysis and interpretation of data: F.R.-P. Drafting of the manuscript. D.A.S. Critical revision of the manuscript. Both authors gave final approval for publication.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Rodriguez-Pascual, F. and Slatter, D. A. Collagen cross-linking: insights on the evolution of metazoan extracellular matrix. *Sci. Rep.* **6**, 37374; doi: 10.1038/srep37374 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016