International Journal of Biological Macromolecules xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

Review

Not only tendons: The other architecture of collagen fibrils

Mario Raspanti*, Marcella Reguzzoni, Marina Protasoni, Petra Basso

Department of Medicine & Surgery, Insubria University, Varese, Italy

ARTICLE INFO

Article history:
Received 12 July 2017
Received in revised form
29 September 2017
Accepted 6 October 2017
Available online xxx

Keywords:
Collagen
Fibrillogenesis
Extracellular matrix

ABSTRACT

For many decades the fibrillar collagens have been the subject of a remarkable body of ultrastructural research. The vast majority of the studies, however, were carried out on tendon or on tendon-derived material. For many reasons this reflects an obvious choice but at the same time it also is an unfortunate circumstance, because this flooding of tendon-related data can easily encourage the false confidence that all connective tissues are similar. The reality is quite the opposite, and a different fibrillar structure has been long time observed on collagen fibrils from different tissues, the most notable example being offered by corneal fibrils. The same architecture can be found in a number of disparate tissues and may actually be the prevalent one on a whole-body scale. Although these fibrils diverge from those of tendon in their architecture, size, D-period, composition, cross-linking and fibrillogenesis mechanism, their structure was the subject of rather sparse ultrastructural studies and even today their mere existence is often overlooked or ignored. This paper summarizes the main aspects of the structural biology of these forgotten fibrils.

Contents

1. Introduction: a collagen primer

The collagen represent one of the most ancient protein families, being present in all the Metazoa and tracing its lineage up to protein motifs in choanoflagellates [1,2]. In mammalians we recognize about thirty genetically defined types identified with Roman numbers and which, as whole, represent more than one third of all proteins of the body. The hallmark of all collagens is the repeated basic structure $(Gly\text{-}Xaa\text{-}Yaa)_n$, where the residues designated as Xaa and Yaa can be any amino acid but are frequently represented by proline. All collagens have a quaternary structure, each functional molecule being composed of three distinct polypeptide chains (the so-called alpha-chains). In addition, collagens also have a highly complex supramolecular structure where molecules interact with each other at different hierarchical levels in order to form a variety of higher-order structures, including fibrils, microfibrils and felt-like sheets.

A few collagen types, defined fibrillar collagens, have an uninterrupted sequence of about 300 Gly-Xaa-Yaa triplets, flanked by shorter globular domains. Other collagen types have more or less interrupted sequences and are classified as fibril-associated collagens with interrupted triple helices (FACITs), membrane-associated collagens with interrupted triple helices (MACITs),

* Corresponding author at: Laboratory of Human Morphology, Dept. of Medicine and Surgery, Insubria University, Via Monte Generoso 71, 21100 Varese, Italy.
E-mail address: mario.raspanti@uninsubria.it (M. Raspanti).

https://doi.org/10.1016/j.ijbiomac.2017.10.037
0141-8130/

multiple triple-helix domains and interruptions (MULTIPLEXINs), microfibrillar collagens, etc. [3].

In this context we will deal exclusively with the main fibrillar collagens, i.e. the biochemical types I, II, III, V and XI. Collagen types XXIV and XXVII have recently joined the fibrillar collagens, but their role is still less defined [4]. The types II and III are homopolymers, each of them composed of three identical chains respectively named α1(II) and α1(III), while the type I is a heteropolymer made of two identical αI(I) chains and an α2(I). Collagen type V is also a heteropolymer made of three different chains, α1(V), α2(V) and α3(V), mixed in different, tissue-dependent combinations. The type XI seems to be, from a phylogenetical viewpoint, a more recent protein: it is closely related to the type V to the point that the α2(V) chain may replace the α2(XI) chain in some tissues, and the α3(XI) chain seems to be an alternative splicing of the α1(II) gene. All the genes coding for the fibrillar collagens are highly interrupted, being made of dozens of exons, and apparently descend from a common ancestor formed by 45 or 54 nucleotides, precisely a string of five or six triplets of the form Gly-Xaa-Yaa, coding for a short 15- or 18-residue chain. This ancient gene seems to have subsequently originate by successive duplications and mutations the actual library of large, highly interrupted genes, whose exons still betray the 45- or 54-base motif.

Although the primary structure of the α-chains is known, a word of caution is necessary since all collagens undergo massive post-translational modifications so that the correspondence between nucleotides and amino acid is far from granted. In particular the hydroxylation of proline residues to 4-hydroxyproline (or, less commonly, 3-hydroxyproline) is critical to the correct folding of the α-chains and therefore to the structure and function of collagen [5]. Another post-translational modification essential for the maturation of collagen is the oxidation of some lysine and hydroxylysine residues into an aldehyde group, which subsequently react with other lysine of hydroxylysine residues to form insoluble cross-links. Some hydroxylysine residues are glycosylated by specific enzymes into galactosyl- or glucosyl-galactosyl-hydroxylysine. Other non-enzymatic glycosylation processes [6] lead to the slow formation of advanced glycation end-products (AGE).

The central $(Gly\text{-}X\text{-}Y)_n$ domain of the collagen α-chains folds into a tight, left-handed helix with an average axial residue-to-residue spacing of about 0.286 nm and an angular separation of 108°, values which can vary slightly according to the size and shape of the different amino acids [7]. This coiling is largely due to steric repulsion between proline residues in the X position and 4-hydroxyprolines in Y-position, and is therefore highly dependent on the post-translational hydroxylation of proline. The peptide bonds form the backbone of the helix, leaving the side chains of amino acids exposed on the outside [8]. Since there are about three residues per turn and since every third residue of this domain is a glycine residue, on the surface of the helix appears a row of (almost) superimposed glycine residues. Since the glycine has no side chain (represented by a single hydrogen atom), this glycine row ultimately appears as a 'clean' line slowly spiraling along the molecule in a slightly right-handed helix with a pitch of approx. 8.58 nm.

This glycine line makes possible the aggregation of three chains into a right-handed triple helix, with all the glycine residues buried inside the structure and the side chains of all other residues exposed on the outside, where they are responsible of the intermolecular interactions. It is worth mentioning the reciprocal stabilizing effect of left-handed threads wound into a right-handed rope.

Although the collagen family consists of approximately 30 distinct types, the term "collagen" without other specifications is often used in reference to collagen type I, which is also the most extensively studied. Here the formation of the triple helix is directed by the carboxy-terminal globular domain (C-propeptide) and proceeds along the C → N direction [9]. Critically, this depends on the lack of steric hindrances in the glycine rows along the central axis. A point mutation can be tolerated in the X and Y position, but the substitution of a glycine in the third position with any other residue stops the formation of the triple helix and results in defective molecules. The clinical outcome is some form of Osteogenesis Imperfecta.

Normally the process leads to the formation of a long, somewhat flexible rod-like molecule 1.5 nm wide and over 300 nm long, topped at both ends by globular domains. In this form, also defined procollagen, the molecule is complete.

This is the starting point for the formation of supramolecular aggregates.

2. The formation of fibrils

In the extracellular space (or maybe intracellularly [10]) the terminal globular domains are cleaved by specific enzymes leaving only two small non-helical fragments (the telopeptides), which fold compactly along the helical portion. In this form the collagen molecule is essentially insoluble and begins to aggregate even at nanomolar concentrations. At this stage the intermolecular interactions are inherently labile, i.e. hydrophilic-hydrophobic or electrostatic; a careful analysis of the amino acid sequence reveals the superposition of several patterns and motifs coexisting along the molecule [11] that may be enough to direct an ordered layout in supramolecular aggregates [12]. In collagen type I the distribution of some amino acids (Lysine, Glutamine and Arginine) shows a periodicity of 18 residues, which reflects the length of the exons, while the distribution of polar and of hydrophobic residues shows a periodic repeat of 234 residues. The presence of other molecules (ATP, polyanions, cations, other collagen types etc.) can modify the balance of these labile interactions and thus promote one aggregation form rather than another, even without taking part of the final structure. This way a whole range of supramolecular aggregates, finite or periodic, symmetrical or asymmetrical, can appear [13–17].

The longitudinal distribution of polar and hydrophobic residues normally directs a lateral intermolecular interaction with an axial stagger of approximately 234 amino acid residues, i.e. about 67 nm, or an integer multiple of this measure. Since this measure is not an integral divider of the molecule length, after four repeats a short interval remains between the end of each molecule and the beginning of the next one [18]. This interaction pattern directly translates into a periodic structure with a period of 67 nm (the D-period), divided into an 'overlap' zone and a 'gap' zone, whose 5:4 mass ratio has subsequently found extensive experimental confirmation.

Other fibrillar collagens can co-precipitate to form mixed fibrils, but not all combinations are allowed. Collagens I, III and V are reciprocally miscible and form mixed fibrils present in all fibrous connective tissues, where type I seems always to be preeminent while types III and V appear in widely variable amounts in different tissues. The types II and XI are not miscible with other fibrillar collagens but co-form mixed fibrils typical of cartilage, together with collagen type IX and perhaps other FACITs.

Cartilage fibrils have been shown to have an epitaxial structure with an axial core including four collagen XI and ten collagen II molecules [19], surrounded by type II molecules and eventually by surface-bound type IX. Considering the similitude between type V and type XI it seems not implausible that type V collagen have a similar nucleating role in the formation of mixed fibrils with types I and/or type III [20–25]. Undoubtedly type V and I coform mixed fibrils *in vitro* whose diameter is somewhat proportional to the type I/type V ratio [20,26,27], and it has been suggested that type V has some nucleating/initiating role [24,25], but the same can be said of type III [28]. It has also been reported that fibrillogenesis *in vivo* is