Fell Muir Review: Collagen fibril formation in vitro and in vivo

Karl E. Kadler
Faculty of Biology, Medicine and Health, Wellcome Trust Centre for Cell-Matrix Research, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK

SUMMARY

It is a great honour to be awarded the Fell Muir Prize for 2016 by the British Society of Matrix Biology. As recipient of the prize, I am taking the opportunity to write a minireview on collagen fibrillogenesis, which has been the focus of my research for 33 years. This is the process by which triple helical collagen molecules assemble into centimetre-long fibrils in the extracellular matrix of animals. The fibrils appeared a billion years ago at the dawn of multicellular animal life as the primary scaffold for tissue morphogenesis. The fibrils occur in exquisite three-dimensional architectures that match the physical demands of tissues, for example orthogonal lattices in cornea, basket weaves in skin and blood vessels, and parallel bundles in tendon, ligament and nerves. The question of how collagen fibrils are formed was posed at the end of the nineteenth century. Since then, we have learned about the structure of DNA and the peptide bond, understood how plants capture the sun’s energy, cloned animals, discovered antibiotics and found ways of editing our genome in the pursuit of new cures for diseases. However, how cells generate tissues from collagen fibrils remains one of the big unsolved mysteries in biology. In this review, I will give a personal account of the topic and highlight some of the approaches that my research group are taking to find new insights.

Keywords
collagen, electron microscopy, fibril, fibripositor, procollagen, tendon

A brief introduction to collagen

There are several excellent reviews on the collagen family and collagen structure [two such examples are Mienaltowski and Birk (2014) and Bella (2016)], and therefore, only a brief account will be given here.

Collagens are a large family of proteins that have three left-handed polyproline II-like helices wound into a right-handed supercoiled triple helix. The chains have a repeating Gly-X-Y triplet in which glycine is located at every third residue position and X and Y are frequently occupied by the imino acids proline and hydroxyproline [see Bella et al. 1994; Brodsky & Ramshaw 1997; Brodsky & Persikov 2005; reviewed by Bella (2016)]. The first 20 or so collagens were identified in animal tissues at the protein level and were assigned Roman numerals [reviewed by Myllyharju and Kivirikko (2004)]. However, with the advent of genome sequencing it became apparent that many more collagens exist. We now know that there are 28 distinct collagens in vertebrates [Huxley-Jones et al. 2007; reviewed by Kadler et al. (2007) and Mienaltowski and Birk (2014)], almost 200 in Caenorhabditis elegans [reviewed by Johnstone (2000)], and further collagens in marine invertebrates (Trotter & Koob 1989; Thurmond & Trotter 1994; Exposito et al. 2010), bacteria [see Ghosh et al. (2012) and references therein] and viruses (e.g. see Rasmussen et al. 2003; Legendre et al. 2011). It has become clear that the triple helix is an important motif that is not restricted to collagens (Brodsky & Shah 1995) but which occurs in a wide range of proteins including asymmetric acetylcholinesterase (Johnson et al. 1977), macrophage scavenging receptors (Kodama et al. 1990), complement component C1q (Reid & Day 1990), ectodysplasin (Ezer et al. 1999), and the mannose-binding lectin, collectins and ficolins in the lectin pathway (Garred et al. 2016) that are involved in mediating host-pathogen interactions (Berisio & Vitagliano 2012).
The polypeptide chains in collagens are termed α-chains. Because there are numerous collagen genes and their protein products trimerize in a specific combination to produce a collagen ‘type’, a nomenclature has evolved to specify a particular α-chain based on the collagen type in which it is found. The nomenclature involves the α symbol followed by an Arabic number followed by a Roman numeral, in brackets (parentheses). The α symbol, Arabic number and Roman numeral are read together to indicate the gene that encodes that particular α-chain. Thus, α1(I) and α2(I) denote that these chains are found in type I collagen and are encoded by the genes COL1A1 and COL1A2.

Collagens can be homotrimers and heterotrimers. Moreover, some collagens of the same type can be homotrimeric or heterotrimeric, e.g. type I collagen can exist as a homotrimer of three α1(I) chains (i.e. [α1(I)]3) or a heterotrimer of two α1(I) chains and a single α2(I) chain (i.e. [α1(I)]2, α2(I)). Furthermore, heterotrimeric collagens can have three different α-chains (e.g. α1(IX), α2(IX) and α3(IX)) that are encoded by three different genes (i.e. COL9A1, COL9A2 and COL9A3 respectively), and some collagen types contain specific combinations of a family of six chains [e.g. α1(IV), α2(IV) and α3(IV), α4(IV), α5(IV) and α6(IV)] (see Hudson et al. (2003) for a review). There is chain selection specificity that of the 45 different collagen α-chains in vertebrates, only 28 different types occur (Table 1). For fibrillar collagen (Figure 1), the chain selection mechanism resides in the non-collagenous sequences at the C-terminal end of each pro-α-chain (Lees et al. 1997; Bourhis et al. 2012). The chain selection mechanism in other collagens is less well understood.

### Fibrillar collagens

The 28 collagen types that occur in vertebrates can be classified according to domain structure, function and supramolecular assembly [for a review, see Mienaltowski and Birk (2014)]. The most abundant are the fibrillar collagens that form the basis of the fibrils in bony, cartilaginous, fibrous and tubular structures and will be the focus of the remainder of this review. The fibril-forming collagens are types I, II, III, V, XI, XXIV and XXVII. They have uninterrupted triple helices of approximately 300 nm in length and have globular domains (propeptides) at each terminus of each α-chain. Types XXIV and XXVII were identified by genome sequencing and were added to this group on the basis of protein domain structure (Koch et al. 2003) and the presence of type XXVII collagen in thin fibrils (Plumb et al. 2007).

Collagen fibrils are complex macromolecular assemblies that comprise different fibrillar collagen types (Hansen & Bruckner 2003). The fibrils are either ‘predominately type I collagen’ or ‘predominately type II collagen’. Predominately type I collagen fibrils occur in bony, tubular and fibrous tissues, whereas cartilaginous tissues contain predominately type II collagen fibrils. Collagen fibrils range in length from a few microns to centimetres (Craig et al. 1989) and therefore have molecular weights in the tera Dalton range [based on calculations described by Chapman (1989)]. The fibrils provide attachment sites for a broad range of macromolecules including fibronectin, proteoglycans and cell surface receptors such as integrins, discoidin domain-containing receptors and mannose receptors (Di Lullo et al. 2002; Joki- nen et al. 2004; Sweeney et al. 2008; Orgel et al. 2011). Furthermore, the fibrils vary in diameter depending on species, tissue and stage of development (Parry et al. 1978; Craig et al. 1989) and in response to injury and repair (Pingel et al. 2014). Collagen fibrils are arranged in exquisite

### Table 1 Collagen types and their chain compositions

<table>
<thead>
<tr>
<th>Type</th>
<th>Gene</th>
<th>α-chain</th>
<th>Molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>COL1A1</td>
<td>α1(I)</td>
<td>[α1(I)]3</td>
</tr>
<tr>
<td></td>
<td>COL1A2</td>
<td>α2(I)</td>
<td>[α1(I)]2, α2(I)</td>
</tr>
<tr>
<td>II</td>
<td>COL2A1</td>
<td>α1(II)</td>
<td>[α1(II)]3</td>
</tr>
<tr>
<td></td>
<td>COL2A2</td>
<td>α2(II)</td>
<td>[α1(II)]2, α2(II)</td>
</tr>
<tr>
<td>III</td>
<td>COL3A1</td>
<td>α1(III)</td>
<td>[α1(III)]3</td>
</tr>
<tr>
<td>IV</td>
<td>COL4A1</td>
<td>α1(IV)</td>
<td>[α1(IV)]3, α2(IV)</td>
</tr>
<tr>
<td></td>
<td>COL4A2</td>
<td>α2(IV)</td>
<td>[α3(IV), α4(IV), α5(IV)]</td>
</tr>
<tr>
<td></td>
<td>COL4A3</td>
<td>α3(IV)</td>
<td>α5(IV), α5(IV), α6(IV)</td>
</tr>
<tr>
<td></td>
<td>COL4A4</td>
<td>α4(IV)</td>
<td>[α5(IV)]</td>
</tr>
<tr>
<td></td>
<td>COL4A5</td>
<td>α5(IV)</td>
<td>[α5(IV)]</td>
</tr>
<tr>
<td></td>
<td>COL4A6</td>
<td>α6(IV)</td>
<td>[α6(IV)]</td>
</tr>
<tr>
<td>V</td>
<td>COL5A1</td>
<td>α1(V)</td>
<td>[α1(V)]2, α2(V)</td>
</tr>
<tr>
<td></td>
<td>COL5A2</td>
<td>α2(V)</td>
<td>[α1(V)], α2(V), α3(V)]3</td>
</tr>
<tr>
<td></td>
<td>COL5A3</td>
<td>α3(V)</td>
<td>[α3(V)]3</td>
</tr>
<tr>
<td>VI</td>
<td>COL6A1</td>
<td>α1(VI)</td>
<td>α1(VI), α2(VI) and any of α3(VI), α4(VI), α5(VI) and α6(VI) (Maass et al. 2016)</td>
</tr>
<tr>
<td></td>
<td>COL6A2</td>
<td>α2(VI)</td>
<td>[α3(VI)]3</td>
</tr>
<tr>
<td></td>
<td>COL6A3</td>
<td>α3(VI)</td>
<td>[α4(VI)]3</td>
</tr>
<tr>
<td></td>
<td>COL6A4</td>
<td>α4(VI)</td>
<td>[α6(VI)]</td>
</tr>
<tr>
<td></td>
<td>COL6A5</td>
<td>α5(VI)</td>
<td>[α6(VI)]</td>
</tr>
<tr>
<td></td>
<td>COL6A6</td>
<td>α6(VI)</td>
<td>[α6(VI)]</td>
</tr>
<tr>
<td>VII</td>
<td>COL7A1</td>
<td>α1(VII)</td>
<td>[α1(VII)]3</td>
</tr>
<tr>
<td>VIII</td>
<td>COL8A1</td>
<td>α1(VIII)</td>
<td>[α1(VIII)]3</td>
</tr>
<tr>
<td></td>
<td>COL8A2</td>
<td>α2(VIII)</td>
<td>[α2(VIII)]3</td>
</tr>
<tr>
<td>IX</td>
<td>COL9A1</td>
<td>α1(IX)</td>
<td>α1(IX), α2(IX), α3(IX)</td>
</tr>
<tr>
<td></td>
<td>COL9A2</td>
<td>α2(IX)</td>
<td>[α3(IX)]3</td>
</tr>
<tr>
<td></td>
<td>COL9A3</td>
<td>α3(IX)</td>
<td>[α3(IX)]3</td>
</tr>
<tr>
<td>X</td>
<td>COL10A1</td>
<td>α1(X)</td>
<td>[α1(X)]3</td>
</tr>
<tr>
<td>XI</td>
<td>COL11A1</td>
<td>α1(XI)</td>
<td>[α1(XI)], α2(XI), α3(XI), α2(XI), α3(XI)*</td>
</tr>
<tr>
<td></td>
<td>COL11A2</td>
<td>α2(XI)</td>
<td>[α2(XI)]3</td>
</tr>
<tr>
<td></td>
<td>COL12A1</td>
<td>α1(XII)</td>
<td>[α1(XII)]3</td>
</tr>
<tr>
<td></td>
<td>COL13A1</td>
<td>α1(XIII)</td>
<td>[α1(XIII)]3</td>
</tr>
<tr>
<td></td>
<td>COL14A1</td>
<td>α1(XIV)</td>
<td>[α1(XIV)]3</td>
</tr>
<tr>
<td></td>
<td>COL15A1</td>
<td>α1(XV)</td>
<td>[α1(XV)]3</td>
</tr>
<tr>
<td></td>
<td>COL16A1</td>
<td>α1(XVI)</td>
<td>[α1(XVI)]3</td>
</tr>
<tr>
<td></td>
<td>COL17A1</td>
<td>α1(XVII)</td>
<td>[α1(XVII)]3</td>
</tr>
<tr>
<td></td>
<td>COL18A1</td>
<td>α1(XVIII)</td>
<td>[α1(XVIII)]3</td>
</tr>
<tr>
<td></td>
<td>COL19A1</td>
<td>α1(XIX)</td>
<td>[α1(XIX)]3</td>
</tr>
<tr>
<td></td>
<td>COL20A1</td>
<td>α1(XX)</td>
<td>[α1(XX)]3</td>
</tr>
<tr>
<td></td>
<td>COL21A1</td>
<td>α1(XXI)</td>
<td>[α1(XXI)]3</td>
</tr>
<tr>
<td></td>
<td>COL22A1</td>
<td>α1(XXII)</td>
<td>[α1(XXII)]3</td>
</tr>
<tr>
<td></td>
<td>COL23A1</td>
<td>α1(XXIII)</td>
<td>[α1(XXIII)]3</td>
</tr>
<tr>
<td></td>
<td>COL24A1</td>
<td>α1(XXIV)</td>
<td>[α1(XXIV)]3</td>
</tr>
<tr>
<td></td>
<td>COL25A1</td>
<td>α1(XXV)</td>
<td>[α1(XXV)]3</td>
</tr>
<tr>
<td></td>
<td>COL26A1</td>
<td>α1(XXVI)</td>
<td>[α1(XXVI)]3</td>
</tr>
<tr>
<td></td>
<td>COL27A1</td>
<td>α1(XXVII)</td>
<td>[α1(XXVII)]3</td>
</tr>
<tr>
<td></td>
<td>COL28A1</td>
<td>α1(XXVIII)</td>
<td>[α1(XXVIII)]3</td>
</tr>
</tbody>
</table>

*The α3(XI) chain is encoded by the COL2A1 gene.
three-dimensional architectures in vivo including parallel bundles in tendon and ligament, orthogonal lattices in cornea, concentric weaves in bone and blood vessels and basket weaves in skin. How the fibrils assemble, how length and diameter are regulated, how molecules attach to fibril surfaces and how the multiscale organization is achieved are questions for which answers are either sketchy or not available. The extreme size and compositional heterogeneity of collagen fibrils mean that they are extremely difficult to study by conventional molecular, genetic and biochemical approaches. Collagen molecules are also hydroxylated at specific prolyl residues (by prolyl hydroxylases), lysyl residues (by lysyl hydroxylases or PLODs, procollagen-lysine 5-dioxygenases) and are non-reducibly cross-linked [by lysyl oxidase (LOX) enzymes; Eyre et al. 1984; Gistelinck et al. 2016], which adds to the difficulty of studying the fibrils. Collagen fibrils in tendon (Heinemeier et al. 2013) and cartilage (Heinemeier et al. 2016) are extremely long lived with estimates exceeding hundreds of years. Therefore, the collagen in the fibrils is particularly prone to modification by advanced glycation end products (Verzijl et al. 2000; Thorpe et al. 2010). Thus, the two major experimental approaches used in the study of collagen fibril assembly have been electron microscopy of tissues to describe the organization of the fibrils in vivo (explained below) and reconstitution of fibrils in vitro using collagen extracted from tissues (explained below) or recombinant collagens (e.g. Fertala et al. 1996).

Collagen fibril assembly in vitro

Studies by Gross (Gross & Kirk 1958), Wood & Keech (Wood & Keech 1960), Hodge & Petruska (Hodge 1989), Silver (Silver & Trelstad 1980) and Chapman (Bard & Chapman 1968), to name a few, showed that exposure of animal tissues (typically skin and tendon) to weak acidic solutions (typically acetic acid) or neutral salt buffers yielded a solution of collagen molecules that when neutralized and warmed to approximately 30°C, produced elongated fibrils that had the same alternating light and dark transmission electron microscope banding appearance as fibrils occurring in vivo (Holmes & Chapman 1979) (Figure 2). The characteristic banding pattern of the fibrils arises from D-staggering of triple helical collagen molecules that are 4.4 × D in length (where D is 67 nm, to a close approximation). The electron-dense stain used at neutral pH penetrates more readily into regions of least protein packing (the ‘gaps’) between the N- and C-termini of collagen molecules that are aligned head-to-tail along the long axis of the fibril. The fact that fibrils with D-periodic banding could be formed in vitro from purified collagen showed that all the information required to form a collagen fibril was contained within the amino acid sequence and triple helical structure of the collagen molecule (Hułmes et al. 1973).

Subsequent studies showed that collagen fibrils from embryonic tendon (which are predominantly type I collagen) exist in two isoforms: unipolar and bipolar (Holmes et al. 1994; Figure 3). Unipolar fibrils have all collagen molecules in the fibril oriented in one direction, which gives the fibril a carboxyl and an amino tip. Bipolar fibrils (more precisely, N, N-bipolar fibrils) have two amino-terminal ends and a molecular polarity switch region (or transition zone) in which the orientation of collagen molecules switches, for example, from N-to-C to C-to-N (Holmes et al. 1994). The switch in orientation occurs over an 8 D-period range in chick tendon collagen fibrils (Holmes et al. 1994). Notably, sea cucumbers (Actinopyga echinites) lack unipolar fibrils and all their bipolar fibrils have the molecular switch region located precisely midway from each fibril tip; also, the switch varies in extent from 14 to 41 D-periods in invertebrate fibrils (Trotter et al. 1998, 2000). Earlier studies had shown that collagen fibrils formed by cleavage of procollagen to collagen (explained below) grow from pointed tips (i.e. the pointed ends of fibrils) and the collagen molecules were oriented in one direction along the long axis of the fibril (Kadler et al. 1990). Moreover, the C-tip of a unipolar

![Figure 1 Schematic diagram of the chain composition of the fibril-forming collagens.](image-url)
fibril is required for end-to-end fusion of either two unipolar fibrils (to generate a new N, N-bipolar fibril) or to one end of an N, N-bipolar to generate a longer N, N-bipolar fibril (Graham et al. 2000; Kadler et al. 2000). Notably, C, C-bipolar collagen fibrils have not been described. Presumably, the structure of C-tips exposes binding sites to promote carboxyl-to-amino fusion of fibril tips.

Two schools of thought developed about how collagen molecules assemble into fibrils: (i) precipitation from a solution of ‘bulk’ collagen by liquid crystalline ordering of molecules (e.g. see Martin et al. 2000) or (ii) ‘nucleation and propagation’ in which a finite number of collagen molecules form a nucleus that then grows in length and diameter to become the mature fibril (Gross et al. 1954). This latter mechanism is analogous to the formation of inorganic crystals. The existence of fibrils of different lengths supports the notion that the fibrils grow in size (which supports the nucleation and propagation model) but collagen and procollagen molecules can form a liquidlike structure when packed in high concentration (which supports the liquid crystalline model). In reality, these two hypotheses might not be mutually exclusive; work by Hulmes and Bruns showed that procollagen molecules can align in zero-D register in secretory vacuoles of fibroblasts (analogous to liquid crystalline packing) (Bruns et al. 1979; Hulmes et al. 1983), which might increase the rate of conversion of procollagen to collagen to nucleate fibrils. Therefore, it is possible that elements of both assembly mechanisms exist in vivo.
A system for generating collagen fibrils in vitro starting with procollagen

In 1984, I joined Darwin Prockop’s laboratory at UMDNJ, Piscataway, NJ, USA, to develop a system of studying collagen fibril formation by cleavage of procollagen with its physiological convertases, the procollagen N- and C-proteinases (Figure 4). Procollagen had previously been shown to be the biosynthetic precursor of collagen (Bellamy & Bornstein 1971) and there had been initial success in purifying the N- and C-proteinases that convert procollagen to collagen (Njieha et al. 1982; Tuderman & Prockop 1982). With the collaboration of Yoshio Hojima who purified the procollagen N- and C-proteinases from chick tendon (Hojima et al. 1985, 1989), we developed a method of purifying type I procollagen and cleaving it with N-proteinase to generate pCcollagen, and then cleaving the repurified pCcollagen with the C-proteinase in a bicarbonate buffer. pCcollagen is a cleavage intermediate of procollagen that retains the C-propeptide but lacks the N-propeptide. This system allowed us to study collagen fibril formation in the absence of lysyl oxidase and cross-link precursors (Eyre et al. 2008). The presence of cross-link precursors in extracted collagen can affect collagen fibril formation in vivo (Herchenhan et al. 2015). Using this new system of forming fibrils by cleavage of procollagen, we defined the thermodynamic parameters of the assembly process (Kadler et al. 1987), the temperature dependence of collagen fibril assembly (Kadler et al. 1988), and showed that the fibrils form as a nucleus that grows at its pointed tips (Kadler et al. 1990). These observations indicated that collagen fibrils (in the absence of lysyl oxidase-derived cross-links) exhibit a critical concentration of assembly, analogous to the self-formation of inorganic crystals. Our ability to purify procollagen from cells paved the way to study how mutations in collagen genes that cause osteogenesis imperfecta affect procollagen structure and fibril assembly. These studies showed that mutations in type I collagen genes can produce procollagen molecules that are ‘kinked’ (Vogel et al. 1988), slow the rate of conversion of procollagen to collagen (Lightfoot et al. 1992), lead to the formation of abnormal collagen fibrils (Kadler et al. 1991) and impair the ability of collagen fibrils to be mineralized during the formation of bone (Culbert et al. 1995). These studies led to a better understanding of how mutations in collagen genes can change the structure and processing of collagen molecules and how the resultant collagen fibrils are poorer scaffolds for mineralization, as occurs in osteogenesis imperfecta in vivo (Culbert et al. 1996). In parallel studies, we also showed that the tips are the sites of diameter regulation (Holmes et al. 1998), that fibrils formed at low C-proteinase/pCcollagen ratios bore the closest resemblance to fibrils in vivo (Holmes et al. 1996) and that the tips of fibrils are paraboloidal in shape (Holmes et al. 1992).

Collagen fibril formation in vivo

Although collagen molecules can spontaneously self-assemble into fibrils in vitro, additional factors must exist in vivo to explain the exquisite three-dimensional supramolecular organization of fibrils, as well as the regulation of diameter, length and composition, that depend on tissue, stage of development, state of tissue ageing and repair, and which vary in disease. The in vivo regulation of collagen fibril formation has been studied for over a century, and although enormous progress has been made, the cellular mechanisms of fibril assembly and organization in vivo remain elusive.

Some of the earliest reports on the existence of collagen fibrils date back to the end of the 19th century and beginning of the 20th century. For example, Mallory described a ‘fibrillar substance’ produced by connective tissue cells (i.e. fibroblasts) (Mallory 1903). Studies of collagen fibrils continued during the 1920s and 1930s during which time several groups attempted to develop methods to observe the assembly of the fibrils in vivo. A breakthrough came in 1940 when Mary Stearns published her first observations of fibroblasts secreting and assembling collagen fibres (Stearns 1940). Her paper is a ‘must-read’ for students of collagen fibril formation; the 46 hand-drawn plates are exquisite. Stearns used the camera lucida to visualize and draw details of cytoplasmic connections between cells, striations within cells, ‘vacuoles de secretion’ and fibres growing at the cell surface. In so doing, she produced the first evidence that fibroblasts are instrumental in assembling collagen fibrils in tissues. Almost 40 years later, Trelstad and Hayashi used transmission electron microscopy (TEM) to show that collagen fibrils occurred in invaginations of the plasma.
membrane of embryonic fibroblasts (Trelstad & Hayashi 1979). A decade later this observation was extended using high-voltage TEM to study collagen fibrillogenesis in cornea as well as embryonic chick tendon (Birk & Trelstad 1984, 1985, 1986; Trelstad & Birk 1985). In 2006, we used serial section TEM and immunoEM of embryonic tendon to describe a variety of structures at the plasma membrane that contained collagen fibrils, and which we collectively called ‘fibripositors’ (Canty et al. 2004). Collectively, these studies demonstrate the exquisite control the cell exerts over the self-assembly of collagen fibrils to generate tissues with highly organized collagen matrices.

Fibripositors

In 1989, I returned to the UK as a Wellcome Trust Senior Research Fellow in Basic Biomedical Science and joined Michael Grant’s Department of Medical Biochemistry. During the next 10 years, we extended our knowledge of how mutations in collagen genes affect procollagen structure and fibril formation. In collaboration with Peter Byers and Gillian Wallis, these studies focussed on the Ehlers–Danlos syndrome (type VII) that is caused by mutations in COL1A1 and COL1A2 genes that encode the chains of type I procollagen. PhD students Rod Watson, Samantha Lightfoot and Ainsley Culbert, and a postdoc David Holmes, joined my laboratory, and together we showed how mutations in COL1A1 and COL1A2 that cause EDS VII disrupt the structure of procollagen, slow the cleavage of procollagen by N-proteinase and lead to the ‘cauliflower’ appearance of collagen fibrils in affected individuals (Wallis et al. 1992; Watson et al. 1992, 1998; Holmes et al. 1993; Culbert et al. 1996). We also studied the function of the CUB domains in bone morphogenetic protein-1, which is a potent procollagen C-proteinase (Garrigue-Antar et al. 2001, 2002, 2004; Hartigan et al. 2003; Petropoulou et al. 2005; Canty et al. 2006a,b). Here, CUB is an evolutionary conserved protein domain named after its discovery in complement components (C1r/C1s), the sea urchin protein Uegf and BMP-1 [for a review, see Bork and Beckmann (1993)].

However, it was during a staff meeting in 2002 that I heard good advice that scientists should change their experimental approach every 10 years. Up until this time, I had used cells as a factory for procollagen production and had overlooked the importance of the cell in fibril assembly. A new postdoc in the laboratory, Elizabeth Canty, took up the challenge of taking our laboratory into new, in vivo, directions. We were inspired by the work of Hayashi, Trelstad and Birk and decided to ask questions about how cells regulate fibril assembly and fibril number. With the assistance of David Holmes, Roger Meadows, Tobias Starborg and Yin-hui Lu in the laboratory, Liz Canty embarked on studying collagen fibril formation in embryonic chick tendon using serial section electron microscopy and 3D reconstruction. Our first paper, in 2004, showed 3D reconstructions from 50 × 100 nm serial sections of embryonic chick tendon, cut perpendicular to the tissue long axis. These were the deepest and most detailed 3D reconstructions at the time and showed fingerlike projections of the plasma membrane containing thin collagen fibrils (Figure 5). The 3D reconstructions showed that the projections were part of an invagination of the plasma membrane and that the fibril within the invagination and the projection were co-aligned to the long axis of the tendon (Canty et al. 2004) (Figure 6). We called these structures ‘fibripositors’ (a portmanteau of ‘fibril’ and ‘depositors’). We also showed that fibripositors are actin-dependent structures (Canty et al. 2006a,b) that projected into intercellular channels stabilized by cadherin-11 containing junctions (Richardson et al. 2007).

Figure 5 Transmission electron microscopy of embryonic tendon. Embryonic tendon contains bundles of collagen fibrils between adjacent fibroblasts. The image shows profiles of fibripositors. Image obtained by Yin-hui Lu.

Figure 6 A fibripositor at the plasma membrane of an embryonic fibroblast. Transmission electron microscope image of a collagen fibril contained within a fibripositor at the surface of an embryonic mouse tail-tendon fibroblast. Image obtained by Yin-hui Lu.
Serial block face-scanning electron microscopy

The fact that fibripositors are too thin to be seen by light microscopy and that no marker has been identified that can aid in their visualization by fluorescence light microscopy has been a severe hurdle to studies of fibripositor structure, function and formation. Also, the effort and time involved in producing serial sections for electron microscopy is a significant hurdle to further progress; sections can be lost or distorted during processing, and the process requires exceptional skills in ultrathin sectioning and handling. A major breakthrough came with the commercialization of serial block face-scanning electron microscopy (SBF-SEM) (Denk & Horstmann 2004). Here, images of a block face are recorded using a scanning electron microscope prior to the removal of a section by an in-microscope ultramicrotome. The ability to produce serial images without manual sectioning opened up new opportunities to explore fibripositor function. After optimization of sample preparation and staining, image acquisition and data analysis, Toby Starborg, Nick Kalson and Yinhui Lu showed that we could use SBF-SEM as a semi-high-throughput system to examine fibripositor structure and function at the cell–matrix interface (Starborg et al. 2013) (Figure 7 and Movie S1). With this new approach, we were able to show that fibripositors are the site of fibril assembly in tendon and that non-muscle myosin II is required for fibril transport and formation (Kalson et al. 2013). We also showed that fibripositor-like structures called keratopodia exist in corneal keratocytes (Young et al. 2014). SBF-SEM also gave us the opportunity to explore how collagen fibril formation contributes to tendon development. In a tour de force of SBF-SEM, Nick Kalson, Yinhui Lu and Susan Taylor outlined a new hypothesis for tendon development in which the number of collagen fibrils is determined by embryonic tendon fibroblasts, and that the growth in lateral size of the tendon is driven by matrix expansion caused by the increase in girth and length of collagen fibrils (Kalson et al. 2015). SBF-SEM studies have also revealed a new function for membrane type I-matrix metalloproteinase (MT1-MMP or MMP14) in being essential for tendon development (Taylor et al. 2015). Taylor et al. showed that release of collagen fibrils from fibripositors at birth requires MT1-MMP and that the process does not rely on the cleavage of collagen at the ¼¾vertebrate collagenase cleavage site in the molecule.

Negative regulation of collagen fibril formation during intracellular protein trafficking

Canty et al. also made the observation that procollagen can be cleaved to collagen prior to secretion by tendon fibroblasts in vivo. Evidence that procollagen can be cleaved to collagen within the cell without forming fibrils demonstrates active negative control of the self-assembly properties of collagen fibrillogenesis in vivo (Humphries et al. 2008). These observations are in contrast to what happens in conventional cell culture, in which procollagen is readily purified.

Figure 7 Serial block face-scanning electron microscopy for studies of the cell–matrix interface. Three images from the downloadable Movie S1 generated by serial block face-scanning electron microscopy. The coloured circles show fibripositors. Numbers refer to the image sequence. Images recorded by Tobias Starborg and Nicholas Kalson.
from the cell culture medium. Presumably the environment of the cell and matrix influence the trafficking of procollagen. A halfway house between in vivo and in vitro is the use of 3D cell culture systems; Kapacee et al. showed that fibroblasts incubated in fibrin gels under linear tension replace the fibrin with collagen fibrils that are aligned parallel to the lines of stress and exhibit features of embryonic fibroblasts in vivo, including fibripositors (Kapacee et al. 2008, 2010; Bayer et al. 2010; Kalson et al. 2010, 2011). This approach facilitates studies of the role of cells, in a near-physiological environment with tissue-derived mechanical forces, in assembling collagen fibrils.

**Regulators of collagen fibril assembly in vivo**

The fact that collagen fibrils are comprised of different collagens, that they occur in different numbers and with different diameters and packing densities in different tissues, that the supramolecular organization of fibrils is different in different tissues and that collagen molecules provide interaction sites for receptors and a wide range of extracellular matrix molecules suggests that there are multiple steps in the assembly and organization of fibrils, for example decorin (Daniel et al. 2002), osteoglycin (Tasheva et al. 2008, 2010; Bayer et al. 2010), biglycan (Heegaard et al. 2007) [for a review, see Kalamajski and Oldberg (2010)]; enzymes required for post-translational modification of collagen α-chains, for example prolyl 4-hydroxylase (Mussini et al. 1967), lysyl hydroxylases (Takaluoma et al. 2007) and lysyl oxidases (Maki et al. 2002); proteins involved in transporting collagens through the secretory pathway, for example HSP47 (Satoh et al. 1996), sedlin (Venditti et al. 2012) and TANGO1 (Saito et al. 2009; Wilson et al. 2011); and proteinases involved in collagen turnover, for example MMP14 (Taylor et al. 2015). Loss of the collagen network in cartilage occurs in end-stage osteoarthritis (Ehrlich et al. 1977). Conversely ectopic or excessive accumulation of collagen occurs in fibrosis, which can be stimulated by TGF-β (Roberts et al. 1986), and can affect any organ often resulting in death. Thus, collagen fibrillogenesis is a precisely regulated process in which the mechanisms that maintain the appropriate number, size and organization of collagen fibrils in adult tissues appear to be sensitive to a wide range of genetic mutations and environmental stimuli.

**Table 2** Diseases caused by mutations in genes encoding fibrillar collagens

<table>
<thead>
<tr>
<th>Collagen type</th>
<th>Gene</th>
<th>OMIM</th>
<th>Disease</th>
<th>Mouse models</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>COL1A1</td>
<td>120150</td>
<td>Osteogenesis imperfecta (OI); Ehlers–Danlos syndrome type VII</td>
<td>Mov13 (Bonadio et al. 1990); Col1a1(Jrt/+) OI/EDS mouse (Chen et al. 2014)</td>
</tr>
<tr>
<td></td>
<td>COL1A2</td>
<td>120160</td>
<td>Osteogenesis imperfecta (OI); Ehlers–Danlos syndrome type VII</td>
<td>OIM (Chipman et al. 1993); Col1a2(+/G610C) OI (Amish) mouse (Daley et al. 2010)</td>
</tr>
<tr>
<td></td>
<td>COL2A1</td>
<td>120140</td>
<td>Stuckler syndrome; achondrogenesis; familial avascular necrosis of the femoral head; Legg–Calve–Perthes disease</td>
<td>Garofalo et al. (1991), Vandenberge et al. (1991), Li et al. (1995a,b), Gäser et al. (2002), Donahue et al. (2003)</td>
</tr>
<tr>
<td>III</td>
<td>COL3A1</td>
<td>120180</td>
<td>Ehlers–Danlos syndrome type IV; intracranial berry aneurysm</td>
<td>Liu et al. (1997); Tsk2 mouse (Long et al. 2015)</td>
</tr>
<tr>
<td>V</td>
<td>COL5A1</td>
<td>120215</td>
<td>Nail patella syndrome; Ehlers–Danlos syndrome classic type</td>
<td>Wenstrup et al. (2004)</td>
</tr>
<tr>
<td></td>
<td>COL5A2</td>
<td>120190</td>
<td>Ehlers–Danlos syndrome type I or type II</td>
<td>Andrikopoulos et al. (1995)</td>
</tr>
<tr>
<td></td>
<td>COL5A3</td>
<td>120216</td>
<td>Ehlers–Danlos syndrome type II</td>
<td>Huang et al. (2011)</td>
</tr>
<tr>
<td>XI</td>
<td>COL11A1</td>
<td>120280</td>
<td>Stickler syndrome; otospondyloepiphyseal dysplasia (OSMED); Marshall syndrome</td>
<td>Cho/cho mouse (Li et al. 1995a,b)</td>
</tr>
<tr>
<td></td>
<td>COL11A2</td>
<td>120290</td>
<td>Stickler syndrome; otospondyloepiphyseal dysplasia</td>
<td>McGuirt et al. (1999), Li et al. (2001)</td>
</tr>
<tr>
<td>XXIV</td>
<td>COL2A1</td>
<td>610025</td>
<td>Steel syndrome (Gonzaga-Jauregui et al. 2015)</td>
<td></td>
</tr>
<tr>
<td>XXVII</td>
<td>COL27A1</td>
<td>608461</td>
<td>Steel syndrome</td>
<td>Plumb et al. (2011)</td>
</tr>
</tbody>
</table>
the patterning of the matrix. Perhaps novel insights into organogenesis will come from a better understanding of the interplay between cell positioning, cell–cell communication, cell–matrix interactions, cell polarity, the role of the secretory pathway in directing matrix assembly, and mechanical forces.

A further exciting area of research is matrix homeostasis; it will be fascinating to learn how changes in this process lead to diseases such as osteoarthritis, tendinopathies, fibrosis and cell migration through the matrix. The realization that the bulk of the collagen in tendon and cartilage is synthesized during adolescence and remains unchanged during the lifetime of a person raises intriguing questions about how the collagen network is maintained during life despite countless cycles of mechanical loading. Advances in genome editing and super-resolution light microscopy are all likely to be brought to bear on this question. These approaches are expected to lead to a better understanding of how matrix homeostasis goes wrong in diseases such as fibrosis, where ectopic and excessive deposition of collagen fibrils can cause death. Recent discoveries show that matrix-rich tissues are peripheral circadian clock tissues and that defects in the rhythm in these tissues lead to pathologies such as calcific tendinopathy (Yeung et al. 2014), osteoarthritis (Guo et al. 2015; Dudek et al. 2016a,b) and intervertebral disc disease (Dudek et al. 2016a,b). Thus, the mechanical environment of the cell, the role of the matrix in modulating cell behaviour and peripheral circadian clocks are all likely to contribute to matrix homeostasis.

Acknowledgements

I am indebted to The Wellcome Trust for continuous support of my research for over 30 years. Numerous students, postdoctoral fellows and technical staff have made valuable contributions to this research. However, a few people need a special mention: Laure Garrigue-Antar, Samantha Lightfoot and Rod Watson performed the studies on osteogenesis imperfecta, the Ehlers–Danlos syndrome and bone morphogenetic protein-1 that established my laboratory at the University of Manchester. In recent years, my gratitude goes out to Nick Kalson, Zohre Kapacee, Elizabeth Laird (nee Canty), Yinhu Li, Susan Taylor (nee Richardson), Toby Starborg and Chloë Yeung who led the work on 3D electron microscopy, fibripositors and tendonlike constructs. Finally, very special thanks go to David Holmes, who I have known since my PhD days, and who joined my group in the 1990s. David is an enormously talented electron microscopist who has made outstanding contributions to our research.

Conflict of Interest

There is no conflict of interest to declare.

References


International Journal of Experimental Pathology
Dudek M., Gossan N., Yang N.  
Daley E., Streeten E.A., Sorkin J.D.  


International Journal of Experimental Pathology


Collagen fibril formation


Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Movie S1. Step-through movie of consecutive images of embryonic mouse tendon generated by serial block face-scanning electron microscopy.