Fibrillins and LTBPs (latent TGFβ (transforming growth factor β)-binding proteins) perform vital and complex roles in the extracellular matrix and are relevant to a wide range of human diseases. These proteins share a signature ‘eight cysteine’ or ‘TB (TGFβ-binding protein-like)’ domain that is found nowhere else in the human proteome, and which has been shown to mediate a variety of protein–protein interactions. These include covalent binding of the TGFβ-propeptide, and RGD-directed interactions with a repertoire of integrins. TB domains are found interspersed with long arrays of EGF (epidermal growth factor)-like domains, which occur more widely in extracellular proteins, and also mediate binding to a large number of proteins and proteoglycans. In the present paper, newly available protein sequence information from a variety of sources is reviewed and related to published findings on the structure and function of fibrillins and LTBPs. These sequences give valuable insight into the evolution of TB domain proteins and suggest that the fibrillin domain organization emerged first, over 600 million years ago, prior to the divergence of Cnidaria and Bilateria, after which it has remained remarkably unchanged. Comparison of sequence features and domain organization in such a diverse group of organisms also provides important insights into how fibrillins and LTBPs might perform their roles in the extracellular matrix.

Key words: elastic fibre, evolution, fibrillin, latent transforming growth factor β-binding protein (LTBP), microfibril, transforming growth factor β (TGFβ).

INTRODUCTION

Fibrillin/LTBP family

There are three fibrillins (fibrillin-1, -2 and -3) and four LTBPs (latent TGFβ (transforming growth factor β)-binding proteins) (LTBP-1, -2, -3 and -4) in humans, which are all disulfide-rich proteins that share a common domain architecture (Figure 1). These each include multiple TB (TGFβ-binding protein-like) domains, found in no other proteins, interspersed with cb EGF (calcium-binding epidermal growth factor)-like domains, a more widely distributed module type. Despite sharing many features, fibrillins and LTBPs can be easily distinguished as the fibrillins are longer than the LTBPs, and are composed of seven TB, 46–47 EGF/cbEGFs and two hybrid domains (Figure 1). LTBPs on the other hand have additional regions at their N-termini with no homology with any known domain type. They can also have a unique four-cysteine domain near the N-terminus. Fibrillins also have a unique C-terminal region not found in LTBPs, but the ‘unique’ N-terminal domain of fibrillin is also found in LTBP-2 and the long isoform of LTBP-1.

Roles of fibrillins/LTBPs in health and disease

The fibrillin/LTBP family of ECM (extracellular matrix) proteins perform distinct roles in connective tissue, and are important in development and tissue homoeostasis [1–3]. Fibrillins assemble into higher-order structures, known as 10–12 nm microfibrils (Figures 2A and 2B), which have been shown by rotary shadowing electron microscopy to form a beaded string with an average periodicity of ~55 nm. Prior to assembly into microfibrils, fibrillin undergoes a number of post-translational modifications, including glycosylation and cleavage of the C-terminal propeptide by furin [4–7]. Once assembled microfibrils are involved in a variety of cell–matrix interactions and developmental processes. They can be found both in association with elastin in elastic fibres, for example in the walls of elastic arteries, and also without elastin, for example in the ciliary zonules, which connect the lens of the eye to the ciliary muscles (Figure 2A). Functions of microfibrils include elastic fibre assembly and maintenance, targeting of growth factors to the ECM, via the interaction of fibrillin with BMPs (bone morphogenic proteins), GDFs (growth and differentiation factors) and LTBPs, and signalling through cell-surface integrin receptors (Figure 2B) [1–3].

LTBPs sequester the small latent complex, consisting of TGFβ and its propeptide, to the ECM (Figure 2C). LTBP-1, -3 and -4 are secreted from cells as part of the LLC (large latent complex), consisting of dimeric TGFβ non-covalently bound to its propeptide, which is itself covalently bound to LTBP. Once secreted, the LLC then associates with the ECM and maintains TGFβ in an inactive latent state. LTBP-2 is also secreted from cells, but does not bind TGFβ [8,9]. Various regions of LTBP-1 have been demonstrated to interact with the ECM [10], and the C-termini of LTBP-1, -2 and -4 have been demonstrated to interact with the N-terminus of fibrillin [11,12]. Further complexity is added by the existence of three different TGFβ proteins, which exhibit different tissue- and LTBP-binding specificities (for reviews see [13–16]). The importance of the TB protein family is illustrated by the genetic disorders that arise as a result of mutations in the fibrillin and LTBP genes. Table 1 summarizes the symptoms of a variety of genetic disorders that arise as a result of mutations in the fibrillin and LTBP genes. Table 1 summarizes the symptoms of a variety of genetic disorders that arise as a result of mutations in the fibrillin and LTBP genes.
of genetic diseases relating to TB proteins. Mutations in the fibrillin-1 gene (*FBN1*) are associated with Marfan syndrome, a disease in which patients exhibit a variety of defects in the skeletal, ocular and cardiovascular systems. These include abnormally long bones, pectus excavatum, a high arched palate and joint hypermobility. Sufferers may also develop ectopia lentis, and have a high risk of developing aortic dilatation and dissection [17]. Specific mutations in *FBN1* have also been associated with a variety of other genetic disorders, including stiff skin syndrome [18], Shprintzen–Goldberg syndrome [19] and dominant Weill–Marchesani syndrome [20].

Mutations in the fibrillin-2 gene (*FBN2*) have been described that cause congenital contractural arachnodactyly or ‘Beal’s syndrome’ [21]. This has ocular and skeletal features, but lacks cardiovascular involvement. The exact reasons for the differences between Marfan syndrome and Beal’s syndrome are unclear. Fibrillin-2 is generally found in the same tissues as fibrillin-1, but does exhibit different distributions within those tissues [22]. It has been suggested that human fibrillin-2 may rely on the presence of fibrillin-1 for oligomerization into microfibrils [23], whereas fibrillin-1 oligomerization might not be dependent on fibrillin-2. However, in *fbn1*-null mice, fibrillin-2 microfibrils can still assemble, although they do demonstrate different ultrastructures when looked at by electron microscopy [24].

No pathogenic mutations have yet been definitively identified in the fibrillin-3 gene (*FBN3*). Fibrillin-3 is only present at low levels in most postnatal tissues [25], but some studies suggest its likely to be involved in polycystic ovary syndrome [26–28].

No *LTBP1* mutations have been discovered in association with any human pathology. However, recessive loss of functional mutations in *LTBP2* have been associated with ocular defects, particularly glaucoma [29–31]. Recessive mutations in *LTBP3* are believed to cause oligodontia, a condition where the teeth fail to develop properly [32]. Recessive mutations in *LTBP4* have recently been shown to cause severe developmental defects in the pulmonary, gastrointestinal, genitourinary, musculoskeletal and dermal systems [33], highlighting its crucial role in development, and consistent with its broad pattern of tissue expression [34,35].

The mechanisms of the various human pathologies involving TB domain proteins are not fully understood. However, dysregulation of TGFβ signalling is thought to play an important role in the development of defects associated with Marfan syndrome and those seen in individuals with *LTBP4* mutations. The connection between Marfan syndrome and TGFβ was initially identified by examining lung alveolar septation defects in *Fbn1* mouse models [36], and has been further supported by
Figure 2  The roles for TB domain proteins in the ECM

(A) Fibrillin microfibrils can be associated with (i) elastin to form elastic fibres in tissues such as the blood vessel wall, lung and skin, or (ii) occur independently, for example in the suspensory ligaments of the eye, periodontal ligament and mesangium of the kidney glomerulus. (B) Some of the key cell matrix interactions mediated by fibrillin microfibrils: LTBP bind microfibrils and sequester TGFβ in a latent state. A variety of BMP and GDF growth factors bind fibrillin directly and are sequestered to the ECM. Microfibrils direct the formation of elastic fibres by interacting with various proteins including fibulins. Integrins bind microfibrils, contributing to cell adhesion. Microfibrils have also been shown to interact with a variety of other proteins and ECM components (Figure 6) [1–3]. (C) The current model of LTBP organization in the ECM for maintenance of latent TGFβ. The LLC consists of dimeric TGFβ non-covalently bound to its propeptide, and the propeptide is covalently bound to LTBP via its second TB domain [13–16]. The LLC itself is sequestered to the ECM, potentially via interaction of its C-terminus with the N-terminus of fibrillin [11,12] and interactions with unknown matrix components at its N-terminus [10], which may involve transglutaminase cross-links [99].

Mouse models have also been used to examine the roles of LTBP-1 [39,40], LTBP-2 [41], LTBP-3 [42,43], LTBP-4 [44,45], and fibrillin-1 and -2 [24,36,46]. Broadly these demonstrate the involvement of LTBPs and fibrillins in a range of developmental processes, but there is some controversy over whether all of the mice in these studies represent genuine and specific knockouts. For example, deletion of exons specific to the long isoform of

the overlapping symptoms of Marfan syndrome with those seen in Loeys–Dietz patients, who have mutations in TGFβ receptor genes [37,38]. Currently there is some uncertainty in the field as to whether these TGFβ receptor mutations cause gain-of-function or loss-of-function, and further molecular experiments are required to identify the effect of these mutations on receptor activity.
Table 1 Genetic diseases associated with TB domain proteins and related pathologies

The main features of each disease are summarized, along with the gene involved and the effect on TGFβ signalling where relevant.

<table>
<thead>
<tr>
<th>Disease</th>
<th>Symptoms</th>
<th>Gene involved</th>
<th>Role of TGFβ</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marfan syndrome</td>
<td>Elongation of the long bones</td>
<td>FBN1 (dominant)</td>
<td>Excessive free TGFβ seen in mouse models. Increased levels of phosphorylated Smad2 seen in Marfan syndrome aorta</td>
<td>[17,21,37,102]</td>
</tr>
<tr>
<td>Shprintzen–Goldberg syndrome</td>
<td>Craniosynostosis (early fusion of skull bones)</td>
<td>FBN1 (dominant)</td>
<td>Unknown</td>
<td>[19]</td>
</tr>
<tr>
<td>Stiff skin syndrome</td>
<td>Tight and thick skin</td>
<td>FBN1 (dominant)</td>
<td>Increase in phosphorylated Smad2 in dermis, indicative of TGFβ signalling</td>
<td>[18]</td>
</tr>
<tr>
<td>Weill–Marchesani syndrome</td>
<td>Short stature</td>
<td>FBN1 (dominant)</td>
<td>Unknown</td>
<td>[20]</td>
</tr>
<tr>
<td>LTBP2 mutation</td>
<td>Primary congenital glaucoma, or secondary glaucoma</td>
<td>LTBP2 (recessive)</td>
<td>Unknown</td>
<td>[29–31]</td>
</tr>
<tr>
<td>LTBP3 mutation</td>
<td>Oligodontia, failure of six or more teeth to develop Short stature</td>
<td>LTBP3 (recessive)</td>
<td>TGFβ signalling shown to be significantly higher when patient fibroblasts co-cultured with reporter cells</td>
<td>[33]</td>
</tr>
<tr>
<td>LTBP4 mutation</td>
<td>Severe developmental defects in the pulmonary, gastrointestinal, genitourinary, musculoskeletal and dermal systems</td>
<td>LTBP4 (recessive)</td>
<td>TGFβ signalling shown to be significantly higher when patient fibroblasts co-cultured with reporter cells</td>
<td>[33]</td>
</tr>
<tr>
<td>Loeys–Dietz syndrome</td>
<td>Cleft palate or bifid uvula</td>
<td>TGFBR1 (dominant)</td>
<td>Increased levels of phosphorylated Smad2 in aorta</td>
<td>[37,38]</td>
</tr>
<tr>
<td></td>
<td>Abnormally long limbs and fingers</td>
<td>TGFBR2 (dominant)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

LTBP-1 has been shown to be lethal to mice owing to defects in cardiac development [40], but another independent study showed that mice with deletion of the first exon of the LTBP-1 gene shared by both short and long isoforms are grossly normal [39]. Reconciling these in vivo studies will require substantial further analysis, by generation of further knockout mice or by closer analysis of those generated so far to determine the source of the inconsistencies.

Since LTBP2 and fibrillins both play an important role in maintaining TGFβ in a latent state in the ECM, understanding the structure and function of these proteins is likely to be relevant to many of the development and disease processes in which TGFβ has been implicated.

TB domain motif

The TB domain is characterized by an eight-cysteine motif that forms four disulfide bonds with a 1–3, 2–6, 4–7, 5–8 arrangement. This, together with a small hydrophobic core, stabilizes the fold. Various high-resolution structures of these domains from fibrillin-1 and LTBP-1 have been determined [47–49], either in isolation or covalently linked to cbEGF domains. In all cases they demonstrate a relatively globular fold. This is in contrast with the more elongated nature of cbEGF domains, which form calcium-stabilized rod-like structures when present as tandem repeats [50,51].

TB domains have been shown to mediate specific protein–protein interactions important for the regulatory functions associated with the LTBP/fibrillin family. These interactions include cell-surface interactions with integrins, mediated primarily by an RGD sequence exposed on the first β-hairpin in the TB4 domain of fibrillin [47,52–56]. Additionally the second TB domain of LTBP-1, -3 and -4 has been shown to bind covalently, via disulfide linkage, to the propeptide of TGFβ. This binding event has been shown by mutagenesis to be dependent on a number of specific sequence features within the TB domain, including the insertion of two amino acids (often FP) into the β sheet region prior to the seventh cysteine residue [8,9], and the presence of several acidic residues. The solution structure of this domain suggests that the two amino acid insert helps expose the 2–6 disulfide bond, making it available to react with the propeptide of TGFβ [49]. This insert and other sequence features are not found in LTBP-2, and are also absent from the fibrillin TB domains, none of which bind TGFβ. These signature residues
are sufficiently distinct from those in ‘standard’ TB domains that they can be used to predict whether or not a TB domain is likely to have a TGFβ-binding capability.

More recently, the structure of a related domain type, the hybrid domain, has been solved [57], which is also found only in fibrillin and the LTBP’s. This shows sequence and structural features of both TB and EGF domains, and has led to the suggestion that it arose from the fusion of a TB–EGF domain pair, producing an N-terminus similar to a TB domain and a C-terminus more similar to the end of a cbEGF domain, and resulting in a 1–3, 2–5, 4–6 and 7–8 disulfide arrangement [57]. The structural and sequence similarities between all three domains are compared in Figure 3.

The limited distribution of the TB domain, found only in the fibrillin/LTBP family, provides a rare opportunity to follow the evolutionary history of this group of proteins, following publication of numerous genome sequences, which offers insights into the functional specialization of each protein.

**EVOLUTION OF TB DOMAIN PROTEINS**

**Fibrillins emerge prior to the LTBP’s and exhibit a highly conserved domain architecture**

Sequence searches for TB domain homologues within available protein libraries reveal that TB domain proteins are found throughout most Eumetazoans, organisms in which the tissues are arranged into distinct germ layers, and where embryos go through a gastrulation stage. Figure 4 summarizes the key organisms in which these proteins are found and their predicted domain organizations.

The simplest Eumetazona group in which TB domain proteins can be found is the Cnidaria (jellyfish, anemones and hydra). TB domain proteins could not be found among the available sequences of simpler metazoans such as Porifera (sponges), Ctenophora (comb jellies) and Placozoa (trichoplax), although the lack of complete genomes for Porifera and Ctenophora groups, as well as the draft nature of the Trichoplax genome, makes this result inconclusive. However, it is clear that TB domain proteins emerged prior to the divergence of Bilateria and Cnidaria, over 600 million years ago [58], since TB domains are present in both groups. These two groups are extremely distinct from each other: Bilateria are organisms such as ourselves with bilateral symmetry, possessing a ‘front and a back’ as well as a ‘top and a bottom’, whereas Cnidaria are organisms such as jellyfish possessing only radial symmetry, giving them just a ‘top and bottom’. Both groups, however, clearly contain TB domain proteins.

From the genomes available, it can be seen that fibrillin domain organization was the first to emerge, since the sea anemone *Nematostella vectensis* TB-domain-containing protein, and many others, all show a clear fibrillin-like domain architecture when compared with human fibrillins and LTBP’s (Figure 4). In fact, the extent of conservation in domain architecture across these fibrillin-like proteins is quite remarkable. For example, the sea anemone domain sequence very closely resembles human fibrillin-1, with the exception of two cbEGF domains being replaced by EGF domains in the anemone. The only significant divergence from the signature ~47 EGF/cbEGF, seven TB and two hybrid domain organization of the fibrillins is seen in the sea urchin *Strongylocentrotus purpuratus* with the inclusion of an extra TB domain and four cbEGF domains near the C-terminus of its fibrillin.

The large number of TB domains in all fibrillins analysed suggests that the first TB domain might have emerged quite some time before it is first seen in Cnidarians. This is because a considerable amount of exon duplication and shuffling would have been required in order to reach the stable fibrillin domain organization which is seen consistently in the genomes available [59].

Diversification of the TB domain proteins, eventually giving rise to the LTBP domain organization, is not seen until the deuterostomes, a group of organisms defined by the fact that the anus forms before the mouth during gastrulation. The earliest LTBP-like protein is first observed in the sea urchin *S. purpuratus*. In LTBP’s the general domain architecture is still quite well-conserved, with the exception of an unusual LTBP-like protein in sea urchin. Some variation emerges between the different LTBP types, for example in the number of cbEGF domains in the central array and in the length of the cysteine-free linker after the last TB domain. These alterations have probably evolved in response to specific role requirements for the different LTBP types.

**Absence from Drosophila melanogaster and Caenorhabditis elegans**

A surprising feature of the evolution of TB proteins as a whole is their absence from both the model organisms *D. melanogaster* and *C. elegans*, which have been extensively studied and their genomes fully sequenced. Further analysis reveals TB proteins to be absent from predicted protein and EST (expressed sequence tag) libraries of various other related *Drosophila* and nematode species. With little information on the precise role of fibrillin microfibrils outside of vertebrates, it is difficult to propose a hypothesis to explain why these organisms have lost fibrillin when it is so well-conserved elsewhere. However, the fact that both of these organisms have been demonstrated to have rapidly evolving genomes [60,61], suggests they have undergone many changes to their biochemistry in adapting to their respective niches.

**INSIGHTS INTO FIBRILLIN ASSEMBLY INTO MICROFIBRILS**

**Microfibril organization is conserved from jellyfish to humans**

Fibrillin assembles in the ECM to form higher-order structures known as microfibrils (Figure 2). Microfibrils have been demonstrated to be present in Cnidaria (specifically in the jellyfish *Podocoryne carnea* [62]), where rotary shadowing electron microscopy shows that they form beaded string structures, similar to those seen in humans, with a periodicity of 50–60 nm. Immunohistochemistry has shown that these jellyfish fibrillins form a variety of fibrillar structures dependent upon their life-cycle stage [62].

Since microfibrils were present prior to the divergence of Cnidaria and Bilateria, sequence features of fibrillins essential for microfibril formation and function should be highly conserved. Indeed the close similarity of fibrillin domain organization between species suggests that the function of fibrillin in microfibrils is highly dependent on the precise tertiary structure and spacing of its domains. This may relate to precise structural constraints required for fibrillin’s supramolecular assembly into microfibrils and/or essential structural properties that it endows them with.

**Accessory proteins MAGP-1 (microfibril-associated glycoprotein-1) and fibronectin are not present in lower Metazoans, suggesting that fibrillin is not always dependent on these proteins for assembly**

Other proteins have been implicated in fibrillin assembly by virtue of their presence within the microfibril (MAGP-1 [63–66]) or from functional data (fibronectin [67,68]). Searches for these
proteins in a range of organisms (Table 2) show them to be absent from more primitive groups, such as Cnidarians (which in the case of MAGPs is supported by other studies [69]). MAGP-1 has consistently been observed in mammalian microfibrils, both by immunohistochemistry and by proteomic analysis [63–66], suggesting that it may be required for microfibril formation. However, as Cnidaria also contain microfibrils, the absence of MAGP from these organisms implies that MAGPs are not a prerequisite for microfibril assembly. This is further supported by the observation that MAGP-1-knockout mice are still able to
Proteins were detected using the NCBI (National Center for Biotechnology Information) blast tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi) with the fibrillin-1 TB3 domain as the search query. Further proteins were also discovered by blasting genomes from specific organisms in the JGI (Joint Genome Institute) database (http://www.jgi.doe.gov/). Domains were annotated initially using the SMART web-based domain prediction tool [100], followed by manual inspection of regions of interest, and location of cysteine residues not associated with domains. It should be noted that many of these protein sequences are predicted from shotgun genome sequences of varying quality and are not backed up by cDNAs, so many may be missing exons or include other errors. These errors probably give rise to the various unrecognized cysteine-containing domains shown in orange. Where possible reprediction of proteins was carried out from genomic DNA, using the WISE2 web-based tool (http://www.ebi.ac.uk/Tools/Wise) with human fibrillin-1 as an example sequence. This gave significant improvements in the protein sequence predictions in some cases, such as in the fibrillin of N. vectensis and Ixodes scapularis. Sequence references are given in Supplementary Table S1 at http://www.BiochemJ.org/bj/433/bj4330263add.htm. The phylogenetic tree shown is based on the Tree of Life web project (http://tolweb.org/tree/).

### Figure 4  Domain organization of TB proteins in a variety of organisms and the evolutionary relationships between them

The table and diagram illustrate the domain organization of TB proteins across various organisms, highlighting the evolutionary relationships through a phylogenetic tree. The table provides a summary of TB proteins found in different groups, their evolutionary relationships, and the name of the organism they are present in. The phylogenetic tree visually represents the evolutionary connections among these proteins, with different colors and labels indicating specific domains and evolutionary branches.
assemble microfibrils and suggests that MAGPs perform some other function [70].

Fibronectin has also been implicated in the early stages of assembly, since microfibrils initially appear to co-localize with fibronectin in cell culture [68], and peptides blocking fibronectin assembly have recently been shown to prevent formation of a fibrillin network in culture [67]. The absence of fibronectin in many fibrillin-containing species (Table 2) may suggest that it acts as a chaperone or organizer in the more complex ECM observed in higher metazoans.

High conservation across Eumetazoa of fibrillin sequences implicated in mammalian microfibril assembly

Fibrillin sequence comparisons across genomes can offer insights into regions that might mediate assembly. Pulse-chase experiments in chick aortae have previously shown rapid formation of disulfide cross-linked fibrillin multimers [71]. This study also indicated that Cys204 in the hybrid 1 domain is present as a free thiol in fragments of human fibrillin-1 and -2 when expressed in cell culture. The hybrid 1 domain is unique in that it has nine cysteine residues rather than the eight seen in the hybrid 2 and TB domains. This unpaired cysteine residue is an excellent candidate for formation of intermolecular disulfide bonds and is absolutely conserved in all fibrillins from Cnidaria to humans (Figure 5A). Furthermore, a homology model from our recent study on hybrid 2 has suggested that residue 204 is present in a surface-exposed loop of hybrid 1 and does not participate in the disulfide-stabilized fold of hybrid 1 [57].

However, a recent study in mice has deleted the first hybrid domain with no serious detriment to the homozygous mice and without affecting microfibril assembly [72]. This is surprising since, in addition to the high conservation of this domain, various mutations associated with Marfan syndrome are located in this domain. These include a missense mutation resulting in a W271G substitution [73], and a G214S substitution in this region has also been associated with dominant ectopia lentis [74]. One possible explanation for the apparent lack of function of hybrid 1 in mice is that fibrillin-2 may compensate for fibrillin-1 in mice, but not in humans. Development of cell-based models of microfibril assembly should allow a greater range of fibrillin hybrid domain constructs to be studied, exploring the effect of deletion and pathogenic mutations, and would help clarify this issue.

There is currently no evidence to indicate which other cysteine residue might be the disulfide bonding ‘partner’ of Cys204. The single cysteine residue in the hybrid domain of another fibrillin molecule might be a candidate, but this would only lead to dimerization, rather than the observed oligomerization. Identification of this partner using proteomic approaches would provide an important constraint for future models of microfibril organization. A variety of models have been proposed for the exact organization of fibrillin within microfibrils. One aspect common to all of these models is a head-to-tail/ N-to-C-terminal interaction of fibrillin, proposed when fibrillin was first purified and characterized [75]. Numerous studies since have demonstrated this N-C interaction by various binding assays [23,76]. Biochemical data support an important role for the C-terminal half of fibrillin, which appears to multimerize during the early stages of assembly [77]. The unique C-terminal region was shown to contribute significantly to this multimerization. Figure 5(B) shows the absolute conservation of a CXXC motif within this region throughout all fibrillins, which is found before the C-terminal furin-cleavage site. These cysteine residues may be candidates for intermolecular disulfide formation, as CXXC motifs are also found in thioredoxins, some of which can help direct the folding of extracellular proteins by driving disulfide exchange [78].

### EVOLUTION OF FIBRILLIN-INTERACTING PROTEINS

**Perlecan and BMPs are present in simple metazoans**

Fibrillin-1 has been demonstrated to bind a large number of other ECM proteins (Figure 6) [12,54,67,79–87]. Analysis of the evolutionary history of these proteins offers insight into the acquisition of specific microfibril functions. Table 2 summarizes a representative set of species in which different fibrillin-interacting proteins can and cannot be found. It can be seen that many fibrillin-interacting proteins are not present outside of vertebrates, but those that are offer interesting insights into the evolution of microfibril functions.
Figure 5  Alignments of selected regions of functional interest in fibrillins and LTBPs
The heparin-sulfate proteoglycan perlecan, which has been shown to bind to mammalian fibrillin and co-localize with fibrillin in basement membrane zones [86], can be found throughout Metazoa, and is also found in the simple placazoan trichoplax, which only has three cell layers and lacks a basement membrane. If perlecan still binds fibrillin in simple organisms, such as the Cnidarians, it has the potential to be one of the oldest matrix interactions established by microfibrils and may have an important, but as yet undefined, role to play in microfibril biology.

Various other proteoglycans, such as versican and aggrecan, have been demonstrated to interact with fibrillin in the same region as perlecan [87]. However, these are not found outside Chordates. Decorin is another proteoglycan that has been demonstrated to form a ternary complex with fibrillin and MAGP-1 [85], but like MAGP-1 it also appears to be limited to vertebrates.

Various TGFβ family growth factors have also been shown to interact directly with fibrillin (although not TGFβ itself). In Table 2, BMP7 was selected, as its interaction has been best characterized [79,80]. BMP7-like growth factors can be found throughout Metazoa, including Cnidarians, making it possible that this interaction may have evolved prior to the LTBP–TGFβ association. Interestingly C. elegans lacks a BMP7-like protein (where the closest homologue has more sequence similarity to BMP5), and C. elegans also lacks fibrillin, but this could be coincidental. Unfortunately, unlike the LTBP–TGFβ interaction, the exact amino acids involved in fibrillin binding to BMP7 have not yet been defined, so the evolutionary history of this interaction cannot be further investigated by sequence alignment.

Emergence of the LTBP–TGFβ complex

Table 2 shows that both the first LTBP and the first TGFβ-like proteins emerge in the sea urchin S. purpuratus. The sea urchin LTBP has some striking differences in domain architecture to the other LTBPs, which cannot be explained just by errors in gene prediction (Figure 4). Its N-terminal region does have some similarities with typical LTBPs, starting with a fibrillin-like N-terminus and EGF, separated from the hybrid–EGF–TB domains by a unique linker, but it lacks the unique four cysteine domain seen in most typical LTBPs. There are also only four EGF domains between the first and second TB domains, and then a large stretch of EGF and cbEGF domains after the second TB domain, and another TB domain at the C-terminus. This does not resemble any known domain organization; however, the second TB domain of this protein does contain the necessary sequence characteristics to bind the propeptide of TGFβ, so it is described here as ‘LTBP-like’.

As previously mentioned, the amino acid requirements for TGFβ propeptide binding have been determined by mutagenesis studies [8,9]. Of particular importance is a two amino acid insertion prior to the seventh cysteine residue, not seen in any other TB domains. Figure 5(D) shows that this insertion is absolutely conserved in the second TB domain of LTBP-like proteins from various species and also in the unusual sea urchin LTBP. Also highlighted are the various acidic residues shown to be involved in binding, which are also highly conserved (numbered according to [9]).

Human LTBP-2, -3 and -4 are also shown below the main alignment. Human LTBP-2 does not have the two amino acid insert, consistent with its inability to bind TGFβ. Interestingly in our searches we only found LTBP-2-like proteins lacking this insert in mammals. Despite its aforementioned role in glaucoma [31], its precise function remains unclear, although it may be involved in elastogenesis through interaction with fibrillin-5 [88].

The close association of LTBPs and TGFβ, and the fact that they both appear at the same time in sea urchin, might suggest their functions have co-evolved. However the sea squirt Ciona intestinalis possesses a TGFβ-like growth factor, but no LTBP. This may be because TGFβ can function independently of LTBP in this organism or LTBP has been missed by current predictions from its genome.

LTBP–fibrillin interaction

In order to sequester latent TGFβ to the ECM, LTBPs must interact with various insoluble matrix components. In humans, the only interaction characterized so far is the binding of the C-terminus of LTBP-1, -2 and -4 to the N-terminus of fibrillin. No specific LTBP residues have yet been implicated as being involved in the interaction, but in LTBP-1 and -4 the binding site has been localized to a C-terminal region [11,12]. Figure 5(E) shows an alignment of part of this region from LTBP-1 across species and demonstrates strikingly strong conservation of the third EGF domain. This conservation may be due to the importance of the domain for fibrillin binding, as most residues are also conserved in LTBP-2 and -4, but not in LTBP-3, which does not bind fibrillin. This EGF domain is conserved in various fish species, but comparison of this region with the sea urchin LTBP-like protein is not possible, as it does not have the equivalent domains at its C-terminus, or any others with a similar protein sequence. It is interesting to speculate from these observations that the sea urchin LTBP might not bind fibrillin, and that this function is only conserved in fish and higher organisms. This can be tested experimentally using smaller domain fragments of the LTBP C-terminus in binding assays with the fibrillin N-terminus.

The fibrillin N-terminus (aligned in Figure 5A) has been shown to interact with LTBPs in the region around its first hybrid domain. Specifically, it has been shown that one residue in particular,
Asn164 in EGF3, has a significant affect on the affinity of the interaction with LTBP-2 and -4 if substituted with serine [12]. This substitution has also been associated with dominant ectopia lentis and Marfan syndrome [89]. Table 5(A) shows that this residue is also not conserved outside vertebrates and sea urchin, consistent with the distribution of LTBP's.

**RGD-mediated integrin interactions**

A fibrillin–protein interaction which has been well studied is the interaction of the TB4 domain with integrins [52–54], specifically αβ3, αβ6 and α5β1, which are important for cell adhesion. One absolute requirement for these fibrillin–integrin interactions is the presence of an RGD motif in the first loop region of TB4, which may be implicated in the pathology of stiff skin syndrome [18]. Comparison of fibrillin sequences in Figure 5(C) shows that this RGD sequence is not conserved outside of vertebrates. Figure 4 also shows that various fibrillins have no RGD motifs in their sequence at all. Collectively, this suggests that the role of fibrillin in cell adhesion via RGD-dependent integrins is limited to vertebrates, even though integrins themselves are found throughout metazoa [90].

**Emergence of elastic fibres and interacting proteins**

Sequences for elastin, the major component of elastic fibres, are only found in vertebrates (Table 2). This is supported by previous protein-based studies, which show fibres with an elastin-like amino acid composition are only present in vertebrates and also absent from the cyclostomes (hagfish and lamprey) [91]. Some other proteins outside this group have been compared with elastin such as ‘lamprin’ from lamprey, and more distantly *Bombyx mori* chorion B and spider dragline silk, as they all contain the GGLG repeats characteristic of elastin sequences [92]; however, no interaction of microfibrils with these proteins has ever been proposed. The absence of elastin from more primitive species, despite the presence of fibrillins, is consistent with a role for microfibrils independent of the elastic fibre.

Various studies have shown that fibulin-2, -4 and -5 play an important role in elastic fibre formation [93–95]. These fibrils interact with the N-terminus of fibrillin in the same region implicated in LTBP binding [12]. This interaction also appears to involve the Asn164 residue, which is limited to vertebrates and sea urchin. Table 2, supported by a recent study of fibulin evolution [96], shows that fibulin-like proteins exist throughout the Eumetazoa, even in organisms that lack elastin, suggesting other, elastic fibre-independent, roles for these proteins. This is supported by the fact that fibulins have been shown to play roles in a variety of microfibril-independent cell signalling and developmental processes in various organisms, including the fibrillin-less *C. elegans* [97,98]. Table 2 records the presence of general fibulin-like sequences, and does not look specifically for fibulin-2, -4 and -5-like proteins, as the molecular features that define them, and give rise to their interaction with fibrillin, are currently unknown. Interestingly, the main diversification of the fibulins can be seen in fish at the same time as the emergence of elastic fibres [96], perhaps allowing new members of the fibulin family to evolve their specialized roles in elastogenesis. Fibrillin diversification coinciding with the emergence of elastic fibres, together with the poor conservation of the fibulin binding Asn164 residue, suggests that the fibulin–fibrillin interaction might be restricted to vertebrates, along with other elastic fibre components.

It is also interesting to note from Figure 6 that many proteins in addition to the LTBP's and fibulins have been implicated in binding the same region of the fibrillin N-terminus. Competition for binding to fibrillin has so far only been demonstrated between the LTBP's and the fibulins [12]; if competition at this site is to constitute a novel regulatory mechanism for the proteins involved, quantification of these interactions in the context of the established microfibril, rather than isolated fibrillin fragments, will be important for furthering our understanding.

**CONCLUSIONS**

A summary of the evolution of the function of TB domain proteins is shown in Figure 7, based on the genomic and functional data reviewed here. Fibrillin microfibrils appear to initially play a structural role in the ECM, potentially binding BMP7-like growth factors. Divergence of the LTBP's and TGFβ then allowed the separate evolution of the LTBP–TGFβ apparatus, creating a complex extracellular sensor, which is now critical to the regulation of a wide variety of human development and disease processes. Subsequent to the divergence of LTBP's, the fibrillins also acquired a vital role in directing elastic fibre assembly and now perform this parallel structural function in a variety of tissues.

The release of future genome studies should cast more light on the initial emergence of TB domain proteins. Analysis of the conservation of sequence of these proteins and their interacting partners should aid experimental studies and help to decipher their complex roles in the ECM.

The TB domain proteins still pose many exciting challenges in the understanding of their function; various models of microfibril
organization need to be reconciled and the arrangement and dynamics of microfibril accessory proteins in different tissues determined. A better understanding of the structure and functional mechanisms of the fibrillin–LTBP–TGFβ apparatus should cast light on pathologies associated with its dysregulation and may have future therapeutic implications.

ACKNOWLEDGEMENT

We would like to thank Dr Chris Ponting for helpful discussions and critical comments.

REFERENCES


276

I. Robertson, S. Jensen and P. Handford

The Authors Journal compilation © 2011 Biochemical Society

Received 23 August 2010/14 September 2010; accepted 17 September 2010
Published on the Internet 22 December 2010, doi:10.1042/BJ20101320
SUPPLEMENTARY ONLINE DATA
TB domain proteins: evolutionary insights into the multifaceted roles of fibrillins and LTBPs

Ian ROBERTSON, Sacha JENSEN and Penny HANDFORD
Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K.

See the following page for Supplementary Table S1.

1 To whom correspondence should be addressed (email penny.handford@bioch.ox.ac.uk).

© The Authors Journal compilation © 2011 Biochemical Society
Table S1  Accession numbers used in the construction of Figure 4 of the main text

<table>
<thead>
<tr>
<th>Organism</th>
<th>Database used</th>
<th>Accession no. of protein sequence used (order as in Figure 4 of main text)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Homo sapiens</strong></td>
<td>NCBI</td>
<td>P35555.2  NP_001990.2  NP_115823.3  NP_996826.1  NP_000419.1  NP_001123616.1  NP_001036009.1</td>
</tr>
<tr>
<td><strong>Mus musculus</strong></td>
<td>NCBI</td>
<td>NP_032019  NP_034311  NP_064303  NP_039617  NP_032546  NP_001107021</td>
</tr>
<tr>
<td><strong>Gallus gallus</strong></td>
<td>NCBI</td>
<td>XP_41815.2  XP_424715.2  XP_41813.2  XP_419510  XP_426444</td>
</tr>
<tr>
<td><strong>Xenopus tropicalis</strong></td>
<td>JGI</td>
<td>C_scaffold_221000019  C_scaffold_7000008  Ext_fgenesh1_p.g.C_5550008  C_scaffold_98000019  gw1.409.46.1  gw1.146.1.1</td>
</tr>
<tr>
<td><strong>Danio rerio</strong></td>
<td>NCBI</td>
<td>NP_001129262.1  ADI29314.1+XP_68500.4 (overlapping sequences combined)  XP_691109.4  WISE2 re-prediction based on  XP_00265326.1  XP_001342567.2  XP_001921093.1  XP_691374.3  XP_690670.3  XP_690560.3</td>
</tr>
<tr>
<td><strong>Ciona intestinalis</strong></td>
<td>NCBI</td>
<td>XP_002119988.1</td>
</tr>
<tr>
<td><strong>Strongylocentrotus purpuratus</strong></td>
<td>NCBI</td>
<td>WISE2 re-prediction of:  XP_787529.2  XP_784696.2</td>
</tr>
<tr>
<td><strong>Tribrulus castaneum</strong></td>
<td>NCBI</td>
<td>XP_974344.2</td>
</tr>
<tr>
<td><strong>Apis mellifera</strong></td>
<td>NCBI</td>
<td>XP_001122812.1</td>
</tr>
<tr>
<td><strong>Acrystophorus pisum</strong></td>
<td>NCBI</td>
<td>XP_001945971.1</td>
</tr>
<tr>
<td><strong>Pediculus humanus corporis</strong></td>
<td>NCBI</td>
<td>XP_002428778.1</td>
</tr>
<tr>
<td><strong>Daphnia pulex</strong></td>
<td>JGI</td>
<td>240066_SNAP_00008124</td>
</tr>
<tr>
<td><strong>Ixodes scapularis</strong></td>
<td>NCBI</td>
<td>WISE2 re-prediction based on XP_002412357.1</td>
</tr>
<tr>
<td><strong>Lottia gigantea</strong></td>
<td>JGI</td>
<td>C_sca_28000110 + e_gw1.28.29.1</td>
</tr>
<tr>
<td><strong>Capitella teleta</strong></td>
<td>JGI</td>
<td>estExt_fgenesh1_p.g.C_1980013</td>
</tr>
<tr>
<td><strong>Hydra magnipapillata</strong></td>
<td>NCBI</td>
<td>XP_002167158.1</td>
</tr>
<tr>
<td><strong>Podocoryne carnea</strong></td>
<td>JGI</td>
<td>AAA91336</td>
</tr>
<tr>
<td><strong>Nematostella vectensis</strong></td>
<td>JGI</td>
<td>WISE2 reprediction based on fgenesh1_p.g.scaffold_462100001_e_gw.113.12.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(b) Other proteins used in alignments</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anolis LTBP-1</td>
<td>Ensembl</td>
</tr>
<tr>
<td>Equus caballus LTBP-1</td>
<td>Ensembl</td>
</tr>
<tr>
<td>Canis familiaris LTBP-1</td>
<td>Ensembl</td>
</tr>
<tr>
<td>Monodelphis domestica LTBP-1</td>
<td>Ensembl</td>
</tr>
<tr>
<td>Ornithorhynchus anatinus LTBP-1</td>
<td>Ensembl</td>
</tr>
<tr>
<td>Gasterosteus aculeatus LTBP-1</td>
<td>Ensembl</td>
</tr>
<tr>
<td>Oryctes latipes (medaka fish) LTBP-1</td>
<td>Ensembl</td>
</tr>
<tr>
<td>Xenopus laevis LTBP-1</td>
<td>NCBI</td>
</tr>
<tr>
<td>Taeniopygia guttata LTBP-1</td>
<td>NCBI</td>
</tr>
</tbody>
</table>

Received 23 August 2010/14 September 2010; accepted 17 September 2010
Published on the Internet 22 December 2010, doi:10.1042/BJ20101320