Available online at www.sciencedirect.com

SciVerse ScienceDirect

Genomics Proteomics Bioinformatics 10 (2012) 197–207

GENOMICS PROTEOMICS & BIOINFORMATICS

www.elsevier.com/locate/gpb

Original Research

Putative Chitin Synthases from *Branchiostoma floridae* Show Extracellular Matrix-related Domains and Mosaic Structures

Gea Guerriero *

Fungal Genetics and Genomics Unit, Department of Applied Genetics and Cell Biology, University of Natural Resources and Life Science Vienna, University and Research Center Campus Tulln-Technopol, Konrad Lorenz Strasse 24, A-3430 Tulln, Austria

Received 21 January 2012; revised 27 March 2012; accepted 23 April 2012
Available online 31 July 2012

Abstract

The transition from unicellular to multicellular life forms requires the development of a specialized structural component, the extracellular matrix (ECM). In Metazoans, there are two main supportive systems, which are based on chitin and collagen/hyaluronan, respectively. Chitin is the major constituent of fungal cell walls and arthropod exoskeleton. However, presence of chitin/chitooligosaccharides has been reported in lower chordates and during specific stages of vertebrate development. In this study, the occurrence of chitin synthases (CHSs) was investigated with a bioinformatics approach in the cephalochordate *Branchiostoma floridae*, in which the presence of chitin was initially reported in the skeletal rods of the pharyngeal gill basket. Twelve genes coding for proteins containing conserved amino acid residues of processive glycosyltransferases from GT2 family were found and 10 of them display mosaic structures with novel domains never reported previously in a chitin synthase. In particular, the presence of a discoidin (DS) and a sterile alpha motif (SAM) domain was found in nine identified proteins. Sequence analyses and homology modelling suggest that these domains might interact with the extracellular matrix and mediate protein–protein interaction. The multi-domain putative chitin synthases from *B. floridae* constitute an emblematic example of the explosion of domain innovation and shuffling which predate Metazoans.

Keywords: *Branchiostoma floridae*; Chitin synthase; Discoidin domain; Extracellular matrix

Introduction

Natural carbohydrate polymers are the basic constituents of the earth biomass and are used as major structural elements by plants, fungi and Metazoans. Cellulose and chitin are the major structural components of plant/fungal cell walls and of arthropod exoskeletons, while the vertebrate counterpart hyaluronan (HA) is one of the major polysaccharide components of animal collagenous extracellular matrix (ECM) [1,2]. The cellulosic/chitinous cell walls and the HA-based ECM, although different from a chemical point of view, are structures which share a supportive role in the organisms they belong to. Cell walls provide strength and protection against mechanical stresses [3] and the ECM acts as "glue", by bridging together cells and providing spatial organization to tissues [4].

The enzymes responsible for the biosynthesis of chitin, cellulose and HA are integral plasma membrane-bound proteins belonging to the glycosyltransferase (GT) Family 2 (GT2 Family) [5], which share the typical amino acid signature, QXXRW [6], of processive β-GTs that are possibly required for holding the growing glycan chain in the active site [7]. Studies have revealed several similarities in the biochemical pathways leading to chitin and HA biosynthesis [8,9]: for instance when the mouse hylauronan synthase gene (*Has2*) is introduced in *Drosophila melanogaster*, HA is successfully produced, indicating that chitin-producing organisms are capable of synthesizing HA, probably by using the endogenous chitin biosynthetic machinery [10]. A further example of the similarities in the biosynthesis of chitin and HA comes from the endoderm-specific protein

* Corresponding author.
E-mail: gea.guerriero@boku.ac.at (Guerriero G).

http://dx.doi.org/10.1016/j.gpb.2012.07.003

DG42 of *Xenopus laevis*: this protein is homologous to Rhizobium NodC and fungal chitin synthase proteins (hereafter named CHSs) and can synthesize both chitooligosaccharides [8] and HA [9]. Taken together, these findings show that the presence of chitin and HA is not mutually exclusive, since the two polymers can co-exist in the same organism. As explanation for the co-occurrence of these two biopolymers in vertebrates, it was proposed that the chitooligosaccharides could serve as primers for HA biosynthesis [8].

Further, the natural occurrence of chitin in lower chordates and the epidermal cuticle and scales of fish has been evidenced in the literature [11,12] and this raises the question about the functional role of chitin in HA-producing organisms.

The basal chordate amphioxus is considered a model organism for developmental and evolutionary studies and a reference to understand the origin and evolution of vertebrates. In amphioxus, chitin was initially reported in the skeletal rods supporting the pharyngeal gill basket [13], which is a particularly interesting structure due to its evolutionary link with the vertebrate skeleton [14–16]. However, some later studies suggested that chitin was absent and instead collagen, together with 15 nm unidentified thick filaments, were present [17,18]. In addition, other studies reported the occurrence of elastin-like molecules in the matrix of the skeletal rods [19] and more recently fibrillar collagen was found in the cells surrounding the gill skeleton [14].

To determine whether putative *Chs* genes are present in amphioxus, the genome of *Branchiostoma floridae* was mined [20]. Twelve genes were found and 10 possess novel domains, most frequently discoidin (DS, cd00057) and sterile alpha motif (SAM, superfamily cl15755) domains. Sequence analyses and homology modelling of these domains suggest that they might be involved in ECM interaction and homotypic/heterotypic associations. The possible role(s) of these domains are discussed in the present paper.

Results and discussion

Putative CHSs in amphioxus

Mining of *B. floridae* genome [20] was performed by BLASTp analysis using CHS2 from yeast (*Saccharomyces cerevisiae*) as query and revealed the presence of 12 genes with deduced protein sequence homology to yeast CHS2. The identified genes encode plasma membrane bound proteins consisting of 478–2727 amino acid residues and 4–17 predicted transmembrane helices (**Table 1**).

BLAST analyses and motif searches revealed that novel domains are present in 10 of the identified putative CHSs from *B. floridae* (**Figure 1A**). These domains show high *E*-values with sequences of proteins involved in carbohydrate binding, ECM interaction and homotypic/heterotypic interactions (Table 1). Particularly noteworthy is the occurrence of the DS and SAM domains in 9 of the identified putative CHSs (Table 1).

The evolution of tissue differentiation in the chordate lineage is strictly associated with the implementation of an ancestral "adhesome", *i.e.*, a network of proteins and glycans mediating cell adhesion and interaction [21]. The complexity of the vertebrate ECM was achieved through expansions of gene families, introduction of novel domains and domain shuffling [21]. Domain shuffling is one of the elements favouring the wave of domain innovation which is the prelude to the development of the complex metazoan ECM [21]. It is therefore possible that the presence of novel ECM-related domains in amphioxus CHSs is connected with domain shuffling events, which ultimately leads to mosaic GT2s.

The occurrence of unconventional domains associated to CHS, such as cytochrome b5-like domain, myosin motor head domain and microtubule interacting and trafficking (MIT) domain, has been documented previously [22–25]. However, the occurrence of ECM-related domains in putative CHSs has not been reported so far.

All of the identified proteins contain the pentapeptide QRRRW (Figure S1), which is a signature of all known *N*-acetylglucosaminyltransferases with processive activity [26] and also present in most processive GT2 members (in the form QXXRW) [6,7]. Detailed sequence analysis also demonstrates the presence of the conserved amino acids GCFSVYR and GEDR in 11 of the identified genes (**Figure 1B**, shaded in gray), which define region 4 and 5 in insect CHSs, respectively [27].

By blasting yeast CHS2, two additional sequences were retrieved (RefSeq number XP_002602987 and XP_00260 8729). These sequences were annotated as GTs with a GT-A fold and contain a PRK12678 transcription termination factor Rho region towards the C-terminus and a DS domain, respectively. However, these sequences lack the conserved QXXRW motif and therefore, were not included in the group of 12 processive GTs coding for putative CHSs mined in the genome of amphioxus.

Interestingly, the presence of several *Chs* genes in amphioxus has recently been reported [28], suggesting that *B. floridae* might use chitin as a structural component. To check for the actual expression of the identified genes, the amphioxus EST database (http://www.amphioxus.icob.sinica.edu.tw/) [29] was mined, by blasting each of the identified genes. As result, 4 ESTs coding for peptides that possess highest *E*-values and identities with XP_002592459 and XP_002592461 were retrieved (**Table 2**). Subsequent sequence and phylogenetic analyses and homology modelling of domains were carried out on these 2 putative CHSs (hereafter named Bfl1 for XP_002592459 and Bfl2 for XP_002592461, respectively).

Phylogenetic position of B. floridae CHSs

Chitin is the dominant polysaccharide present in the cell walls of fungi and the exoskeleton of arthropods

Table 1 ***B. floridae*** **CHS Accession No., number of predicted transmembrane domains (TMDs) and deduced protein lengths**

Ref. Seq. No.	No. of TMDs	Length	Domains (*E*-values)
XP_002590818	15	2727	OTU (2.03e-21), SAM (1.45e-07), TS (1.39e-14), SRCR (1.53e-41), EGF (8.9e-13), PKD (1.59e-61)
XP_002590824	13	1752	DS (5.40e-11), OTU (4.44e-21), SAM (4.07e-08)
XP_002592457	12	1085	SSF (4.63e-45), GH18 (3.10e-20), DS (1.50e-11)
XP_002592459	7	943	DS (1.56e-27), SAM (6.28e-08)
XP_002592461	14	1437	DS (2.27e-18), SAM (2.13e-08)
XP_002592717	15	1778	DS (6.21e-38), SAM (2.40e-07)
XP_002592718	15	1627	DS (1.06e-35), SAM (3.33e-08)
XP_002592721	17	1598	SAM (6.24e-10)
XP_002592723	4	776	_
XP_002592726	5	478	_
XP_002592732	12	1023	WSC (1.62e-05)
XP_002601872	14	1337	SAM (1.09e-10)

Note: Domain abbreviations are same as shown in Figure 1A.

[27,30–35]. Chitin biosynthesis in those organisms has been thoroughly studied [30–35] and different classes of CHSs have been identified [27,35]. To analyse the relationship of CHSs from *B. floridae* with those from fungi and other Metazoans, a phylogenetic analysis was performed. The neighbor-joining tree shows a clear clustering of *B. floridae* CHSs into a Metazoan-specific clade (**Figure 2**). Intriguingly, the 2 CHSs from *B. floridae* group together with the orthologs from early branching Metazoans, namely the cnidarians *Hydra magnipapillata* (XP_002162504) and *Nematostella vectensis* (XP_001633545 and XP_001637059).

To further check the sequence similarity of putative CHSs from *B. floridae* with orthologs from other members of early-branching Metazoans, BLAST analyses of Bfl1 and Bfl2 were carried out using the biocomputational platform Compagen (http://www.compagen.org) [36]. Hits with significant *E*-values were obtained. These correspond to putative CHSs from *Acropora digitifera* (sequence ID adi_v1.04629 with *E*-value of e-108 and adi_v1.08548 with *E*-value of e-102), *Anemonia viridis* (GenBank accession FK730165 with *E*-value of 2e-28 and FK730165 with *E*-value of 4e-15) and *H. magnipapillata* (sequence ID Hma2.226142 with *E*-value of e-122 and Hma2.223960 with *E*-value of 8e-91, respectively). All of these sequences contain SAM domains towards the C-terminus (not shown). The clustering of amphioxus sequences with cnidarians orthologs is particularly interesting, considering that cnidarians occupy a key position in the subkingdom Metazoa. Cnidarians feature skeletal structures (chitinous or carbonate-based) which are spread throughout the metazoan phyla [37]. Moreover, Cnidarians are relevant to evo-devo and are examined to understand the evolution of adhesive processes (which then culminated in the anatomical complexity of bilaterians) [38], since they constitute the most primitive animals with a complex ECM.

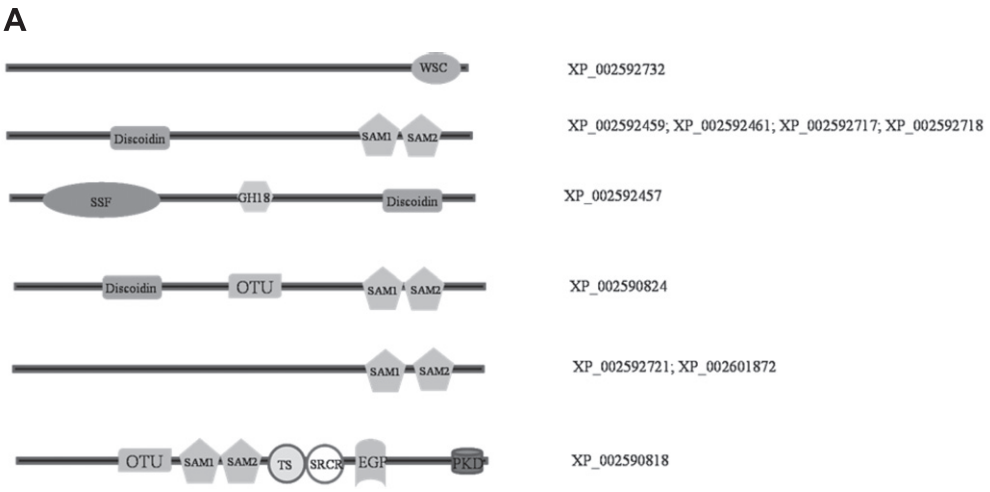
This raises the possibility as to whether the modular organization and phylogenetic position of CHSs from *B. floridae* together with the cnidarian orthologs reflect a recruitment of an early metazoan CHS to which novel domains were subsequently added.

Homology modelling of DS and SAM domains associated with amphioxus putative CHSs

The protein corresponding to Bfl1 has a computed theoretical mass of 106,463 Da and a p*I* of 6.01, while Bfl2 has a higher p*I* (8.06) and a MW of 162,908 Da. Both proteins are characterized by the occurrence of DS and SAM domains.

The DS domain, also referred to as F5/8C domain due to its occurrence at the C-terminus of blood coagulation factors 5 and 8 [39], is a motif that was originally found in the social amoeba *Dictyostelium discoideum* [40]. DS domain is present, singly or in tandem, in many prokaryotic [41] and eukaryotic proteins. For example, DS domain was found in two mammalian tyrosine kinases, discoidin domain receptor tyrosine kinase 1 and 2 (DDR1, DDR2), and shown to be responsible for binding to collagen [42]. DS domains were also found in a chitosanase [43] and β-1,3-glucanase [44], where they enhance the biological activity of the glycosyl hydrolases to which they are appended. DS domains are versatile motifs which can bind proteins, sugars, phospholipid and collagen [45]. The structure of DS domain is a jelly-roll arranged in 8 β-strands, from which loops protrude and form spikes that determine ligand specificity [45,46].

The DS domain is located at residues 1–149 in Bfl1 and 506–652 in Bfl2, respectively. Alignment of DS domains from Bfl1 and Bfl2 with other DS domain-containing proteins for which the structures are available indicates the presence of 8 β-strands (**Figure 3**). The predicted ligand binding residue for Bfl2 is K60 (Figure 3A, highlighted in



B

```
XP_002592723      DADVRFTPKSAMALLDLARRDPT--VGAVCGRTHPLGSGPMVWYQKFDYAVGHWYQK--- 55
XP_002592721      DADVKFTPESAIALLDLARRDPT--VGAVCGRTHPLGSGPMVWYQKFDYAVGHWYQKTAN 58
XP_002590824      DADVKFTPKSANSLLDLAQWNPD--VGAVCGRTHCLGSGPMYWLQLFDYAVGHWFQKAAN 58
XP_002590818      DADVKFTPKSANSLLDLAQWNPD--VGAVCGRTHCLGSGPMYWLQLFDYAVGHWFQKAAN 58
XP_002592717      DADVKFSPQSAKALLDIMTRDPN--VGAVCARTHPLGSGPMVWYQMFDYAVGHWFQKVAN 58
XP_002592718      DADIKFSPQSAKALMEIMSTDPN--VGAVCARTHPLGSGPLVWYQIFDYAVGHWFQKVAN 58
XP_002592461      DADVKFTPEAAKALLDITARDPA--VGAVCARTHPMGSGAVAWYQIFDYAIGHWLNKAAN 58
XP_002592459      DADVKFNPDAAKALLDITARDPA--VGAVCARTHPLGTGAVAWYQIFDYAIGHWLNKCAN 58
XP_002592457      DGDVKFKSEAVRSLLDIAVQNPS--VGAVCARTHPIGSGPVAWYQIFDYAIAHWLGKTAN 58
XP_002601872      DADVQFTPESVAALLDLASRDRK--VGAVSGRTHPSGTGPVVWYQIFDYAIGHWFQKVAE 58
XP_002592726      DGDVDFDADSIVAMLLQMLSDREEDVGAVCARTHPVGSGPLVWYQMFDYAIGHWLQKVAN 60
XP_002592732      DGDVDFDAGSIVAMLLQMLSDREGQVGAVCARTHPVGSGPFVWYQMFDYAIGHWLQKVAN 60
                  *.*: * . :  :::     :    ****..***  *:*.. * * ****:.**  *

XP_002592723      ----------------------------------------------------VSSGWRL 62
XP_002592721      SVLGSVLCCPGCFSVYRAEALREALGTYAGGVSEAREFLMKDMGEDRWLCTLMVSNGWRL 118
XP_002590824      STLGTVLCCPGCFSVYRCSAVRECVSTYATKTEVAKDFLMKDMGEDRWLCTLMVERGKRL 118
XP_002590818      STLGTVLCCPGCFSVYRCSAVRECVSTYATKTEVAKDFLMKDMGEDRWLCTLMVERGKRL 118
XP_002592717      SVLGTVLCCPGCFSVYRAKAVRDALPTYATKVTKAEEFLTKDMGEDRWFCTLLVEKGWRL 118
XP_002592718      SILGTVLCCPGCFSVYRCKAVRDTLATYASTVSKGEEFLTKDMGEDRWLCTLMVEKGWRL 118
XP_002592461      NVLGTVLCCPGCFSVYRAKAVRDGLAEYSTHVTKANEFLVKDMGEDRWFCTLLVESGWKL 118
XP_002592459      NVLGTVLCCPGCFSVYRAKAVKNTLPTYCTHVTKANEFLIKDMGEDRWLCTLM------- 111
XP_002592457      NMMGTVLCCPGCFSVYRANAVRDGLSEYSSHVKEANDFLVKDMGEDRWFCTLLIKHGWTL 118
XP_002601872      HVYGSVLCCPGCFSVFRARALR----------------QDMGEDRWLSTLLVEKGWRL 100
XP_002592726      HVLGSVLCSPGCFSVYRVEAIADVLEEYRSDVEEASDFLTKDMGEDRWFTTLLVKAGWKI 120
XP_002592732      HVLGSVLCAPGCFTVYRVEAIKKVLEEYRSDVEEASDFLTKDMGEDRWFTTLLVKAGYKI 120

XP_002592723      EYTAVSEDSTYCPEEFDEFFNQRRRWGPSTVAN 95
XP_002592721      EYTAVSEDSTYCPEEFAEFYNQRRRWGPSTVAN 151
XP_002590824      VYTSIAEDSTFVPESFDEFFNQRRRWGPSTVAN 151
XP_002590818      VYTSIAEDSTFVPESFDEFFNQRRRWGPSTVAN 151
XP_002592717      EYTAVSEDSTYCPEEFDEFFNQRRRWIPSTVAN 151
XP_002592718      EYTAVSEDSTYCPEEFDEFFNQRRRWIPSTIAN 151
XP_002592461      EYSAVSEDSTFCPETFDEFFKQRRRWLPSTVN 151
XP_002592459      --------------TFDEFFKQRRRWLPSTVAN 130
XP_002592457      EYSAMTEDFTYCPETFEELFKQRRRWLLSSLVN 151
XP_002601872      EYGAVAECKTFCPDTLEEFFKQRRRWIPSTLAN 133
XP_002592726      NYCAGAVDSTHCPEEFGEFWKQRRRWIPSTLAN 153
XP_002592732      NYCAGAVDSTHCPEEFDEFWKQRRRWIPSTLAN 153
                                : *:::*****  *::.*
```

Figure 1 CHSs from amphioxus

A. Cartoon showing the domain organization of the putative CHSs from *B. floridae*. WSC, carbohydrate binding domain (pfam91822); DS, discoidin (cd00057); SAM, sterile alpha motif (superfamily cl15755); SSF, solute symporter family (TIGR001813); GH18, glycosyl hydrolase 18 (cd02872); OTU, ovarian-tumour-like cysteine protease domain (pfam02338); TS, thrombospondin (pfam00090); SCRC, scavenger receptor (smart00202); EGF, epidermal growth factor (cd00054); PKD, polycystic kidney disease (pfam08016). **B.** Partial amino acid sequence alignment of the 12 *B. floridae* CHSs showing the conserved DxD, QXXRW motifs (underlined). Gray-shaded boxes represent region 4 and 5, according to Ref. [27].

Table 2 EST sequences from different developmental stages of *B. floridae* showing amino acid sequence homology to CHSs

Clone ID	Corresponding amphioxus putative CHS	% Identity	*E*-value	Query coverage	Developmental stage
bflv038m04	XP_002592459	90	7e-81	99	36 h Larvae
	XP_002592717	43	4e-20	96	
	XP_002592718	42	2e-18	98	
bflv061b06	XP_002592459	98	2e-83	100	36 h Larvae
	XP_002592461	68	1e-54	100	
	XP_002592717	38	2e-20	94	
	XP_002592718	34	9e-19	90	
bfad030d07	XP_002592459	63	4e-57	81	Adult
	XP_002592461	99	1e-54	49	
	XP_002601872	39	8e-20	55	
	XP_002592718	41	1e-18	66	
	XP_002592717	32	7e-18	60	
bfad046i17	XP_002592717	43	3e-28	79	Adult

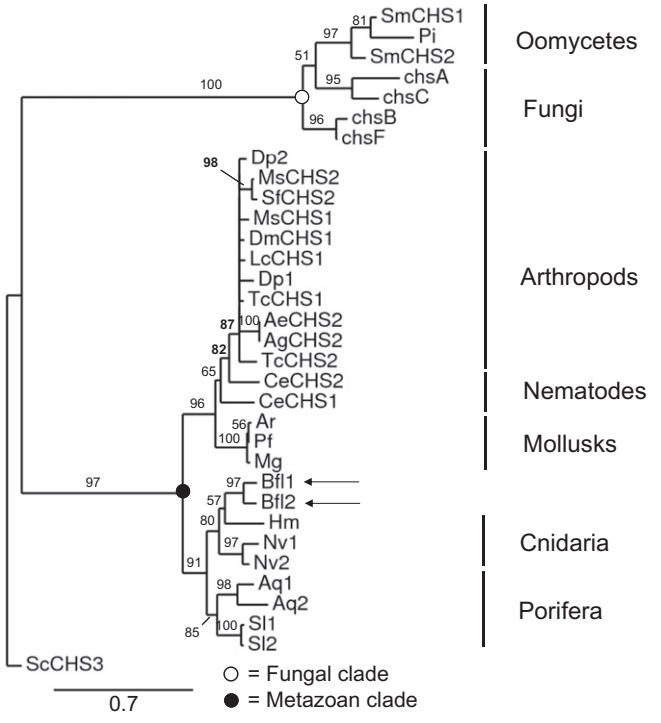


Figure 2 Neighbor-joining phylogenetic tree of *B. floridae* CHSs Phylogenetic tree of *B. floridae* XP_002592459 (Bfl1) and XP_002592461 (Bfl2) was constructed using the neighbor-joining method. The GenBank accession numbers used to build the tree are the same as those listed in Figure S1. CHS3 from *S. cerevisiae* was used as outgroup to root the tree. Bootstrap = 1000. Numbers refer to % of branch support values. The scale bar indicates the number of amino acid substitutions per site. Open circle represents the fungal clade and the filled circle represents the Metazoan clade.

yellow), which is located between the third and fourth β-strands and it is predicted to bind galactose and fucose as heterogens (Figure 3B). The predicted ligand binding residues for Bfl1 are K46, E49 and Q51 in loop 4 and N139 and A140 before the eighth β-strand (Figure 3A, highlighted in yellow), with *N*-Acetylglucosamine (GlcNAc), galactose, *N*-Acetylgalactosamine (GalNAc), mannose and fucose as heterogens (Figure 3B). The predicted binding affinity of the Bfl1 DS domain for the hexosamines GlcNAc and GalNAc (Figure 3B) is quite interesting, since it might indicate a binding of the DS domain to the GalNAc and GlcNAc present in glycosaminoglycans (GAGs). Moreover, the predicted binding of the Bfl1 DS domain to galactose, fucose and mannose and of Bfl2 to galactose and mannose (Figure 3B) might indicate affinity to the carbohydrate moieties of collagen. Indeed, several studies have shown that after synthesis, collagen is extensively modified through *N*- and *O*-linked carbohydrates [42,47,48]. Monosaccharide analysis of mouse tail collagen using fluorophore-assisted carbohydrate electrophoresis technology, for instance, has shown the presence of high amounts of galactose and smaller amounts of fucose and mannose [49].

The predicted 3D structures of Bfl1 and Bfl2 DS domains are jelly-rolls (Figure 3B). The estimated precision value (*i.e.*, the percentage of times a match with a given *E*-value was found to be a true homology) for the Bfl1 DS domain was 100%. DS domain of Bfl1 showed 38% identity with the B domain of human blood coagulation factor 8 (PDB 3CDZ) and 28% with the DS domain of DDR2 (PDB 2Z4F). Similarly, the DS domain of Bfl2 showed 100% confidence, 28% identity with the B1 domain of neuropilin1 (PDB 2QQI) and 22% identity with the DS domain of DDR2.

According to the topological models of the putative amphioxus CHSs, the DS domains are predicted to be located extracellularly, together with the conserved pentapeptide QRRRW (not shown). If the DS domains are extracellular, this would agree with their potential role in binding GAG hexosamines and/or the carbohydrate moieties of collagen in the ECM, but would imply the presence of a mechanism to translocate the substrate UDP-GlcNAc. Topological predictions with other GT2 members favour the orientation of the catalytic domain towards the cytoplasmic side, since the cytoplasmic pool of UDP-GlcNAc

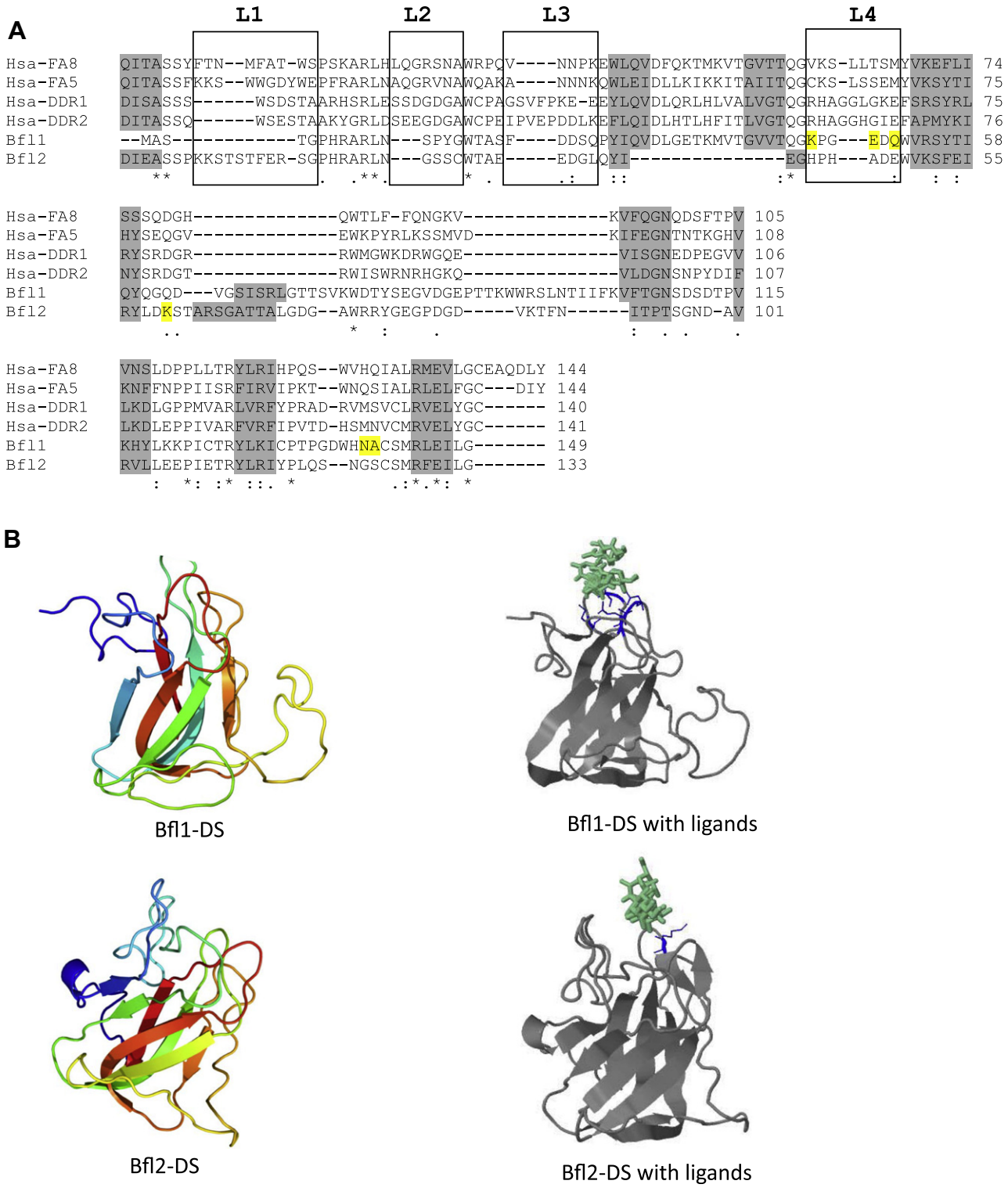


Figure 3 DS domains from *B. floridae* CHSs

A. Domains alignment. Alignment of the DS domain of XP_002592459 (Bfl1) and XP_002592461 (Bfl2) with those of human coagulation factors (Hsa-FA) 5 (CAI23065) and 8 (NP_063916), DDR1 (CAQ09766) and DDR2 (CAE45946). Shaded in grey are the β-strands, the boxed sequences refer to loops 1 to 4 (L1-L4) according to Ref. [70] and highlighted in yellow are the amino acid residues predicted to be involved in ligand binding. **B.** Homology models. Homology models of DS domains and the heterogens, for Bfl1 and Bfl2. Homology models of the DS domains of Bfl1 and Bfl2 (blue is the N-terminus, red the C-terminus) are shown on the left. The same structures shown on the right indicate the amino acid residues (indicated in blue) involved in binding to the heterogens (indicated in light green). For the DS domain of Bfl1 the heterogens are GlcNAc, GalNAc, galactose, fucose and mannose and for the DS domain of Bfl2 the heterogens are galactose and fucose.

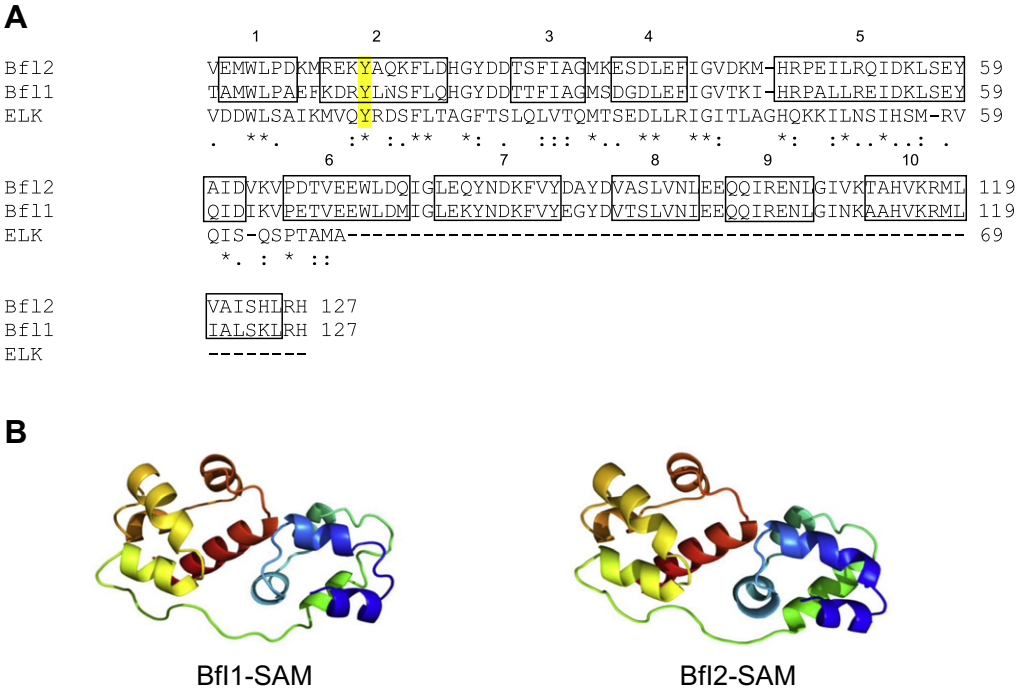


Figure 4 SAM domains from *B. floridae* CHSs

A. Domains alignment. Alignment of the SAM domains of XP_002592461 (Bfl2) and XP_002592459 (Bfl1) with that of ELK (P54762). Alpha helices are boxed and numbered from 1 to 10. Highlighted in yellow are the Tyr residues predicted to be phosphorylated. **B.** Homology models. Homology models of the SAM domains of Bfl1 and Bfl2 (blue is the N-terminus, red the C-terminus).

could easily access the active site [27]. If the DS domains are located intracellularly, this might indicate a potential role in protein-lipid interactions, similar to the DS domains of blood coagulation factors [39], which could assist trafficking and targeting of the enzyme. The presence of domains associated with GT2 members which may mediate trafficking and/or targeting of the enzyme has been already described in the literature. In this respect, it is noteworthy to mention the recent study on the pleckstrin homology domain (PH) of another member of the GT2 family, namely cellulose synthase 2 (SmCesA2), from the oomycete *Saprolegnia monoica*. This domain binds phosphoinositides, F-actin and microtubules *in vitro*, co-localizes with F-actin *in vivo*, and might therefore play a role in trafficking and/or targeting of the enzyme [50].

The SAM is a domain characterized by a bundle of 5 alpha helices [51]. Two SAM domains are found towards the C-terminus of Bfl1 (aa 616–742) and Bfl2 (aa 1097–1223) (**Figure 4**). Topology model predicts the motifs to be located towards the cytoplasm (not shown). The cytoplasmic localization of the SAM domains present in putative CHSs from *B. floridae* agrees with a role in protein–protein interaction. Additionally, SAM domains interact with SH2-containing proteins via a phosphorylated Tyr and trigger signal transduction [52]. In this study, two Tyr residues in the SAM domains of putative amphioxus CHSs are predicted to be phosphorylated (not shown) and the first one corresponds to the Tyr928 of ELK (Accession No. P54762) (Figure 4A), whose phosphorylation is required for interaction with Gbr10 [53]. ELK belongs to the Eph-related tyrosine kinase family, which includes many members possessing SAM domains [52].The predicted 3D structures of *B. floridae* SAMs show, as expected, 2 bundles of 5 alpha helices (Figure 4B). The SAM domains of Bfl1 and Bfl2 show 35% and 32% identity, respectively, with human caskin-1 (PDB 3SEN), both with 100% confidence. Caskin-1, a scaffolding protein organizing the active zones of neural synapses, has been recently shown to be capable of forming a new helical polymer via its SAM domains [54]. The SAM domain can function as a protein interaction platform, due to the ability to form homo- or hetero-oligomers with other SAM domains [52]. Moreover, a binding affinity of this domain for lipids [55], as well as RNA [56], has also been reported.

Chitin synthases from amphioxus: fascinating GT2 members for evolutionary studies

The GT2 family of glycosyltransferases comprises proteins encoded by the *CesA* superfamily, *Chs* and *Has genes* and other transferases with different substrate specificities. The family is indeed very broad: there are more than 26,000 entries in the CaZy database (http://www.cazy.org/) [57], distributed among Archaea, viruses and eukaryotes. This finding implies that this family is quite ancient and that UDP-Glc and UDP-GlcNAc are among the first sugar nucleotide precursors found in a hypothetical common

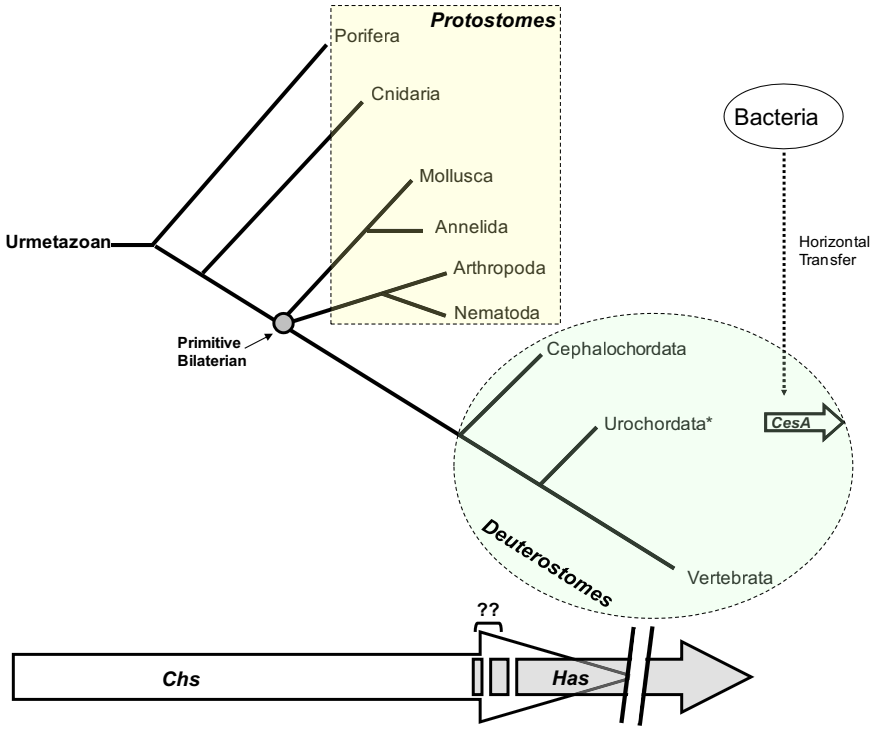


Figure 5 Overview of GT2 member distribution among Metazoans
"??" indicates the hypothesis according to which *Has* derives from *Chs*, either by addition of the β3 transferase activity [61], or by mutation [2]. The break on the arrow corresponding to *Has* indicates the absence of this gene in Urochordates. Asterisk refers to putative CHSs from Urochordata (XP_002119921 and ENSCSAVP00000016559).

ancestor [2]. In Metazoans, chitin is widespread among several members belonging to different phyla (**Figure 5**). Putative CHS sequences could be retrieved in the Metazome portal (http://www.metazome.net), also for the urochordates *Ciona intestinalis* (accession number XP_002119921, which interestingly contains an OTU-like cysteine protease domain) and *C. savigny* (ENSCSAVP00000016559). Therefore, chitin and *Chs* gene(s) were likely present in the Urmetazoan (Figure 5) [2].

Chitin is a template for biomineralization [58] and its occurrence in early Metazoans suggests that it could be considered as a basic metazoan character [59]. The CHSs from amphioxus could promote biomineralization, by providing a chitinous template and participate in the interaction with the ECM, via the novel domains here described. Bioinformatics analysis shows high homology in the structure of the DS domains from putative amphioxus CHSs with those from other proteins (DDR1 and DDR2) involved in collagen binding and ECM interaction. Particularly interesting in this respect is the annotation of XP_002590818 as a putative GT with a GT-A fold possessing thrombospondin (TS; pfam00090) and epidermal growth factor (EGF) domains (cd00054; Figure 1 and Table 1). These domains define a class of calcium-binding motifs commonly found within several glycoproteins mediating angiogenesis and connective tissue organization [60]. The presence of these domains in a CHS could point to a role in the interaction with the ECM, since *in vitro* studies have indeed revealed that TS can bind to growth factors and to structural components of the ECM, thus tuning its adhesive properties [60].

The putative CHSs from amphioxus are interesting targets for the study of the diversification of GTs involved in ECM biosynthesis. One of the major polysaccharides of vertebrate ECM is HA. In this respect, since HASs are capable of synthesizing chitooligosaccharides under specific conditions *in vitro* [8], it was hypothesized that they probably evolve from ancestral CHSs by either addition of the β3 transferase activity [61], or by mutation (Figure 5) [2]. The study of CHSs in living Metazoans preceding the vertebrate branch could provide important information to test this hypothesis and identify the occurrence of other putative mosaic GT2 members.

CHSs in amphioxus: part of a primordial skeletogenic-like program?

The presence of putative chitin synthase genes in amphioxus points towards the occurrence of chitin in this organism. The multi-domain CHSs identified here through a bioinformatics approach represent interesting targets for functional studies.

Chitin and collagen are very fascinating organic matrices which constitute templates for the process of biomineralization [62]: they nucleate and control the deposition of the mineral phase in superb natural composite structures

like the mollusc nacre [63,64] and sponge spicules [58], respectively. Given the mechanical properties and the biomineralization-promoting activity of chitin, a provocative hypothesis would link the presence of putative CHSs with a skeletogenic-like program in this basal chordate and with an evolutionary stage preceding the vertebrate lineage, characterized by the occurrence of a primordial form of endoskeleton (as for instance a pharyngeal endoskeleton). Chitin, together with other ECM components, might be involved in hardening and stiffening of particular structures in amphioxus.

Since the matrix component of the rods has yet to be unambiguously described and the processes regulating skeletogenesis in this organism are still poorly explored, a functional study of the reported CHSs and their novel domains would constitute a first step towards understanding the role of chitin in amphioxus development.

For instance, the use of CHS inhibitors during different *B. floridae* developmental stages (similarly to what has been performed on the marine bivalve mollusc *Atrina rigida* with Nikkomycin Z) [63] could provide important information about the role of chitin *in vivo*, its relationship with the ECM and its eventual involvement in skeletogenesis.

Materials and methods

Data mining and sequence analysis

The identification of putative full-length CHSs from *B. floridae* was carried out by performing BLASTp searches against the *B. floridae* filtered gene models database (http://www.genome.jgi-psf.org/pages/blast.jsf?db=Brafl1) using as query a representative CHS from yeast (CHS2, GenBank Accession No. AAA34493). Sequence alignment was performed with ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) [65].

The identification of novel domains in the CHSs was carried out with Motif Scan (http://www.myhits.isb-sib.ch/cgi-bin/motif_scan). Isoelectric point and MW determinations were carried out at http://www.web.expasy.org/compute_pi/. Phosphorylation sites prediction was carried out using the NetPhos 2.0 server (http://www.cbs.dtu.dk/services/NetPhos/) [66].

Phylogenetic analysis, prediction of transmembrane domains and 3D structure

Transmembrane domain prediction was performed using the online program TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/) and the neighbor-joining phylogenetic tree was built by aligning the conserved amino acid regions of 33 sequences (Figure S1) using ClustalW [65] and BIONJ [67]. The number of bootstraps was set to 1000 and the CHS3 sequence from *S. cerevisiae* (accession number NP_009579) was used to root the tree.

The predicted 3D structures were obtained using the automatic fold recognition server Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) [68]. Prediction of ligand binding residues was performed using the 3DLigandSite server (http://www.sbg.bio.ic.ac.uk/3dligandsite/about.html) [69].

Competing interests

I declare no competing interests.

Acknowledgements

I gratefully acknowledge financial support from the Autonomous Province of Bozen/Bolzano-South Tyrol (Promotion of Educational Policies, University and Research Department) and the Austrian Science Fund (FWF, Grant No. M1315). I would also like to thank Dr. Sanja Baric for fruitful discussion and the anonymous reviewers for their constructive comments.

Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.gpb.2012.07.003.

References

[1] DeAngelis PL. Hyaluronan synthases: fascinating glycosyltransferases from vertebrates, bacterial pathogens and algal viruses. Cell Mol Life Sci 1999;56:670–82.
[2] DeAngelis PL. Evolution of glycosaminoglycans and their glycosyltransferases: implications for the extracellular matrices of animals and the capsules of pathogenic bacteria. Anat Rec 2002;268:317–26.
[3] Liepman AH, Wightman R, Geshi N, Turner SR, Scheller HV. Arabidopsis – a powerful model system for plant cell wall research. Plant J 2010;61:1107–21.
[4] Teti A. Regulation of cellular functions by extracellular matrix. J Am Soc Nephrol 1992;2:S83–7.
[5] Coutinho PM, Deleury E, Davies GJ, Henrissat B. An evolving hierarchical family classification for glycosyltransferases. J Mol Biol 2003;328:307–17.
[6] Campbell JA, Davies GJ, Bulone V, Henrissat B. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. Biochem J 1997;326:929–39.
[7] Saxena IM, Brown Jr RM, Dandekar T. Structure–function characterization of cellulose synthase: relationship to other glycosyltransferases. Phytochemistry 2001;57:1135–48.
[8] Semino CE, Specht CA, Raimondi A, Robbins PW. Homologs of the Xenopus developmental gene DG42 are present in zebrafish and mouse and are involved in the synthesis of Nod-like chitin oligosaccharides during early embryogenesis. Proc Natl Acad Sci U S A 1996;93:4548–53.
[9] Meyer MF, Kreil G. Cells expressing the DG42 gene from early Xenopus embryos synthesize hyaluronan. Proc Natl Acad Sci U S A 1996;93:4543–7.
[10] Takeo S, Fujise M, Akiyama T, Habuchi H, Itano N, Matsuo T, et al. In vivo hyaluronan synthesis upon expression of the mammalian hyaluronan synthase gene in *Drosophila*. J Biol Chem 2004;279:18920–5.
[11] Wagner GP, Lo J, Laine R, Almeder M. Chitin in the epidermal cuticle of a vertebrate. Experientia 1993;49:317–9.

[12] Zaku SG, Emmanuel SA, Aguzue OC, Thomas SA. Extraction and characterization of chitin; a functional biopolymer obtained from scales of common carp fish (*Cyprinus carpio* L.): a lesser known source. Afr J Food Sci 2011;5:478–83.
[13] Sannasi A, Hermann HR. Chitin in the cephalochordata *Branchisotoma floridae*. Experientia 1970;26:351–2.
[14] Rychel AL, Swalla BJ. Development and evolution of chordate cartilage. J Exp Zool (Mol Dev Evol) 2007;308B:325–35.
[15] Hecht J, Stricker S, Wiecha U, Stiege A, Panopoulou G, Podsiadlowski L, et al. Evolution of a core gene network for skeletogenesis in chordates. PLoS Genet 2008;4:e1000025.
[16] Kaneto S, Wada H. Regeneration of amphioxus oral cirri and its skeletal rods: implications for the origin of the vertebrate skeleton. J Exp Zool (Mol Dev Evol) 2011;316:409–17.
[17] Azariah J. Studies on the cephalochordates of the Madras coast. 16. A theory of structural stabilization. Acta Histochem 1974;50:62–74.
[18] Rähr H. Ultrastructure of gill bars of *Branchiostoma lanceolatum* with special reference to gill skeleton and blood vessels (Cephalochordata). Zoomorphology 1982;99:167–80.
[19] Wright GM, Keeley FW, Robson P. The unusual cartilagenous tissues in jawless craniates cephalochordates and invertebrates. Cell Tissue Res 2001;304:165–74.
[20] Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, et al. The amphioxus genome and the evolution of the chordate karyotype. Nature 2008;453:1064–71.
[21] Özbek S, Balasubramanian PG, Chiquet-Ehrismann R, Tucker RP, Adams JC. The evolution of extracellular matrix. Mol Biol Cell 2010;21:4300–5.
[22] Choquer M, Boccara M, Gonçalves IR, Soulié MC, Vidal-Cros A. Survey of the *Botrytis cinerea* chitin synthase multigenic family through the analysis of six euascomycetes genomes. Eur J Biochem 2004;271:2153–64.
[23] Takeshita N, Yamashita S, Ohta A, Horiuchi H. Aspergillus nidulans class V and VI chitin synthases CsmA and CsmB each with a myosin motor-like domain perform compensatory functions that are essential for hyphal tip growth. Mol Microbiol 2006;59:1380–94.
[24] Weiss IM, Schönitzer V, Eichner N, Sumper M. The chitin synthase involved in marine bivalve mollusk shell formation contains a myosin domain. FEBS Lett 2006;580:1846–52.
[25] Guerriero G, Avino M, Zhou Q, Fugelstad J, Clergeot PH, Bulone V. Chitin synthases from Saprolegnia are involved in tip growth and represent a potential target for anti-oomycete drugs. PLoS Pathog 2010;6:e1001070.
[26] Ruiz-Herrera J, González-Prieto JM, Ruiz-Medrano R. Evolution and phylogenetic relationships of chitin synthases from yeasts and fungi. FEMS Yeast Res 2002;4:247–56.
[27] Merzendorfer H. Insect chitin synthases: a review. J Comp Physiol 2006;B176:1–15.
[28] Huang S, Wang X, Yan Q, Guo L, Yuan S, Huang G, et al. The evolution and regulation of the mucosal immune complexity in the basal chordate amphioxus. J Immunol 2011;186:2042–55.
[29] Yu JK, Wang MC, Shin-I T, Kohara Y, Holland LZ, Satoh N, et al. A cDNA resource for the cephalochordate amphioxus *Branchiostoma floridae*. Dev Genes Evol 2008;218:723–7.
[30] Lenardon MD, Whitton RK, Munro CA, Marshall D, Gow NA. Individual chitin synthase enzymes synthesize microfibrils of differing structure at specific locations in the *Candida albicans* cell wall. Mol Microbiol 2007;66:1164–73.
[31] Tsuizaki M, Takeshita N, Ohta A, Horiuchi H. Myosin motor-like domain of the class VI chitin synthase CsmB is essential to its functions in *Aspergillus nidulans*. Biosci Biotechnol Biochem 2009;73:1163–7.
[32] Lenardon MD, Milne SA, Mora-Montes HM, Kaffarnik FA, Peck SC, Brown AJ, et al. Phosphorylation regulates polarisation of chitin synthesis in *Candida albicans*. J Cell Sci 2010;123:2199–206.
[33] Zimoch L, Merzendorfer H. Immunolocalization of chitin synthase in the tobacco hornworm. Cell Tissue Res 2002;308:287–97.
[34] Zimoch L, Hogenkamp DG, Kramer KJ, Muthukrishnan S, Merzendorfer H. Regulation of chitin synthesis in the larval midgut of Manduca sexta. Insect Biochem Mol Biol 2005;35:515–27.
[35] Horiuchi H. Functional diversity of chitin synthases of *Aspergillus nidulans* in hyphal growth conidiophore development and septum formation. Med Mycol 2009;47:S47–52.
[36] Hemmrich G, Bosh TC. Compagen, a comparative genomics platform for early branching metazoan animals reveals early origins of genes regulating stem-cell differentiation. Bioessays 2008;30:1010–8.
[37] Boero F, Schierwater B, Piraino S. Cnidarian milestones in metazoan evolution. Integr Comp Biol 2007;47:693–700.
[38] Magie CR, Martindale CQ. Cell-cell adhesion in the cnidaria: insights into the evolution of tissue morphogenesis. Biol Bull 2008;214:218–32.
[39] Foster PA, Fulcher CA, Houghten RA, Zimmerman TS. Synthetic factor VIII peptides with amino acid sequences contained within the C2 domain of factor VIII inhibit factor VIII binding to phosphatidylserine. Blood 1990;75:1999–2004.
[40] Poole S, Firtel RA, Lamar E, Rowekamp W. Sequence and expression of the discoidin 1 family in *Dictyostelium discoideum*. J Mol Biol 1981;153:273–89.
[41] Baumgartner S, Hofmann K, Chiquet-Ehrismann R, Bucher P. The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction. Protein Sci 1998;7:1626–31.
[42] Vogel W, Gish GD, Alves F, Pawson T. The discoidin domain receptor tyrosine kinases are activated by collagen. Mol Cell 1997;1:13–23.
[43] Kimoto H, Akamatsu M, Fujii Y, Tatsumi H, Kusaoke H, Taketo A. Discoidin domain of chitosanase is required for binding to the fungal cell wall. J Mol Microbiol Biotechnol 2010;18:14–23.
[44] Cheng YM, Hsieh FC, Meng M. Functional analysis of conserved aromatic amino acids in the discoidin domain of *Paenibacillus* beta-1,3-glucanase. Microb Cell Fact 2009;8:62.
[45] Kiedzierska A, Smietana K, Czepczynska H, Otlewski J. Structural similarities and functional diversity of eukaryotic discoidin-like domains. Biochim Biophys Acta 2007;1774:1069–78.
[46] Ichikawa O, Osawa M, Nishida N, Goshima N, Nomura N, Shimada I. Structural basis of the collagen-binding mode of discoidin domain receptor 2. EMBO J 2007;26:4168–76.
[47] Kivirikko KI, Myllylä R. Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. Methods Enzymol 1982;82:245–304.
[48] Gelse K, Pöschl E, Aigner T. Collagens–structure, function, and biosynthesis. Adv Drug Deliv Rev 2003;55:1531–46.
[49] Higgins E, Friedman Y. A method for monitoring the glycosylation of recombinant glycoproteins from conditioned medium using fluorophore assisted carbohydrate electrophoresis. Anal Biochem 1995;228:221–5.
[50] Fugelstad J, Brown C, Hukasova E, Sundqvist G, Lindqvist A, Bulone V. Functional characterization of the pleckstrin homology domain of a cellulose synthase from the oomycete *Saprolegnia monoica*. Biochem Biophys Res Commun 2012;417:1248–53.
[51] Smalla M, Schmieder P, Kelly M, Ter Laak A, Krause G, Ball L, et al. Solution structure of the receptor tyrosine kinase EphB2 SAM domain and identification of two distinct homotypic interaction sites. Protein Sci 1999;8:1954–61.
[52] Schultz J, Ponting CP, Hofmann K, Bork P. SAM as a protein interaction domain involved in developmental regulation. Protein Sci 1997;6:249–53.
[53] Stein E, Cerretti DP, Daniel TO. Ligand activation of ELK receptor tyrosine kinase promotes its association with Grb10 and Grb2 in vascular endothelial cells. J Biol Chem 1996;271:23588–93.
[54] Stafford RL, Hinde E, Knight MJ, Pennella MA, Ear J, Digman MA, et al. Tandem SAM domain structure of human caskin1: a presynaptic self-assembling scaffold for CASK. Structure 2011;19:1826–36.

[55] Barrera FN, Poveda JA, González-Ros JM, Neira JL. Binding of the C-terminal sterile alpha motif (SAM) domain of human p73 to lipid membranes. J Biol Chem 2003;278:46878–85.
[56] Kim CA, Bowie JU. SAM domains: uniform structure diversity of function. Trends Biochem Sci 2003;28:625–8.
[57] Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res 2009;37:D233–8.
[58] Ehrlich H, Worch H. Collagen: a huge matrix in glass sponge flexible spicules of the meter-long *Hyalonema sieboldi*. In: Bäuerlein E, editor. Handbook of biomineralization: biological aspects and structure formation. Weinheim Germany: Wiley-VCH Verlag GmbH; 2007. p. 23–41.
[59] Muzzarelli RAA. Chitin nanostructures in living organisms. In: Gupta NS, editor. Chitin: formation and diagenesis. Topics in geobiology. Springer; 2011. p. 1–34.
[60] Adams JC, Bentley AA, Kvansakul M, Hatherley D, Hohenester E. Extracellular matrix retention of thrombospondin 1 is controlled by its conserved C-terminal region. J Cell Sci 2008;121:784–95.
[61] Lee JY, Spicer AP. Hyaluronan: a multifunctional megaDalton stealth molecule. Curr Opin Cell Biol 2000;12:581–6.
[62] Ehrlich H. Chitin and collagen as universal and alternative templates in biomineralization. Int Geol Rev 2010;52:661–99.
[63] Schönitzer V, Weiss IM. The structure of mollusc larval shells formed in the presence of the chitin synthase inhibitor Nikkomycin Z. BMC Struct Biol 2007;7:71.
[64] Cartwright JH, Checa AG. The dynamics of nacre self-assembly. J R Soc Interface 2007;4:491–504.
[65] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. ClustalW and ClustalX version 2. Bioinformatics 2007;23:2947–8.
[66] Blom N, Gammeltoft S, Brunak S. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol 1999;294:1351–62.
[67] Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, et al. Phylogenyfr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 2008;36:W465–9.
[68] Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 2009;4:363–71.
[69] Wass MN, Kelley LA, Sternberg MJ. 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Res 2010;38:W469–73.
[70] Leitinger B. Molecular analysis of collagen binding by the human discoidin domain receptors DDR1 and DDR2 Identification of collagen binding sites in DDR2. J Biol Chem 2003;278:16761–9.