Amphioxus Sp5 is a member of a conserved Specificity Protein complement and is modulated by Wnt/β-catenin signalling

SIMON C. DAILEY1,2, IRYNA KOZMIKOVA3 and ILDIKÓ M.L. SOMORJAI*1,2

1University of St Andrews, Biomedical Sciences Research Complex, North Haugh, St Andrews, UK, 2Scottish Oceans Institute, East Sands, St Andrews, UK and 3Institute of Molecular Genetics of the Czech Academy of Sciences, Prague, Czech Republic

ABSTRACT A cluster of three Specificity Protein (Sp) genes (Sp1-4, Sp5 and Sp6-9) is thought to be ancestral in both chordates and the wider Eumetazoa. Sp5 and Sp6-9 gene groups are associated with embryonic growth zones, such as tailbuds, and are both Wnt/β-catenin signalling pathway members and targets. Currently, there are conflicting reports as to the number and identity of Sp genes in the cephalochordates, the sister group to the vertebrates and urochordates. We confirm the SP complement of Branchiostoma belcheri and Branchiostoma lanceolatum, as well as their genomic arrangement, protein domain structure and residue frequency. We assay Sp5 expression in B. lanceolatum embryos, and determine its response to pharmacologically increased β-catenin signalling. Branchiostoma possesses three Sp genes, located on the same genomic scaffold. Phylogenetic and domain structure analyses are consistent with their identification as SP1-4, SP5 and SP6-9, although SP1-4 contains a novel glutamine-rich N-terminal region. SP5 is expressed in axial mesoderm and neurectoderm, and marks the cerebral vesicle and presumptive pharynx. Early exposure to increased β-catenin caused ubiquitous SP5 expression in late gastrula, while later treatment at gastrula stages reduced SP5 expression in the posterior growth zone during axis elongation. Amphioxus possess a typical invertebrate eumetazoan SP complement, and SP5 expression in embryos is well conserved with vertebrate homologues. Its expression in the tailbud, a posterior growth zone, is consistent with expression seen in other bilaterians. Branchiostoma SP5 shows a dynamic response to Wnt/β-catenin signalling.

KEY WORDS: Specificity protein, Brachyury, Branchiostoma, Wnt, Sp5

Introduction

The Specificity protein (SP) genes are important developmentally expressed transcription factors, found across almost all metazoan species with the exception of ctenophores (Presnell et al., 2015). Although the SP complement can vary significantly among species, its members can be categorised into three distinct clades: SP1-4, SP5 and SP6-9 (Schaeper et al., 2010). Expression of Sp5 and Sp6-9 genes is typically associated with embryonic regions of high proliferation and growth, such as tail or limb buds. For example, vertebrate Sp5 marks the primitive streak of mice in early development, and later the tailbud for the entire period of axis elongation (Treichel et al., 2001). Similarly, zebrafish Sp5 is expressed in both the tailbud and somites (Tallafuñ et al., 2001). Expression domains for Sp6, Sp8 and Sp9 all include the apical-ectodermal-ridge (AER) of developing limb buds (Nakamura et al., 2004; Kawakami et al., 2004). Vertebrate Sp7 is perhaps atypical, as it specifically marks osteoblasts (Nakashima et al., 2002), but for the remaining vertebrate Sp5 and Sp6-9 genes, an association with embryonic growth zones is a defining feature of these two groups. This is also true outside of the vertebrates. For example, insect Sp6-9 genes are frequently expressed in the brain, ventral

Abbreviations used in this paper: 2R, two rounds of vertebrate whole genome duplication; Azk, 1-Azakenpaullone; bp, base pairs; DiMes, di-mesencephalic primordium; HyPTs, hypothalmo-pituitary-thalamic primordium; Kb, kilobase; KLF, krüppel-like Factor; Mb, megabase; SP, Specificity Protein; RhSp, rhombencephalo-spinal primordium; PFA, paraformaldehyde; WMISH, whole mount in situ hybridisation.

Supplementary Material (six figures and one table) for this paper is available at: http://dx.doi.org/10.1387/ijdb.170205is

Submitted: 19 August, 2017; Accepted: 19 September, 2017
nerve cord and the limb primordial/buds of both antennae and legs (Schaeper et al., 2010), and Sp8 is able to regulate allometric growth of limbs in Tribolium castaneum and Oncopeltus fasciatus (Schaeper et al., 2009; Beermann et al., 2004). Similarly, insect Sp5 is expressed in the imaginal discs of Drosophila, as well as the somites and posterior growth zone of T. castaneum (Schaeper et al., 2010). Given the functional similarity of these genes both within and outside the vertebrate lineage, it would be reasonable to hypothesise that the utilisation of Sp5 and Sp6-9 in embryonic growth zones is a conserved bilaterian trait.

Interactions with Wnt/β-catenin signalling appear to be crucial for the roles of Sp5 and Sp6-9 in these embryological growth zones. For example, Sp8 is an effector of Wnt/β-catenin signalling in vertebrate limb outgrowth (Kawakami et al., 2005). For example, Sp5 and Sp5-like are direct targets of Wnt8/β-catenin signalling during the patterning of the neurectoderm and mesoderm (Weidinger et al., 2005), and Sp5 is upregulated in the mouse telencephalon following Wnt/β-catenin activation (Fujimura et al., 2007). Sp5-like is also a target of Wnt8 and Wnt3a in the tailbud, and is required for somitic mesoderm and notochord formation (Thorpe et al., 2005). There are numerous examples, therefore, of Sp5 and Sp6-9 acting as downstream targets of Wnt/β-catenin signalling in vertebrate embryological growth zones. Furthermore, Sp5 and Sp8 are also critical components of the canonical Wnt signalling pathway itself: both proteins are components of the TCF/LEF recruitment complex for some Wnt target genes, including Brachyury (Kennedy et al., 2016). This relationship to Brachyury is important during development, as whilst knocking out mouse Sp5 alone causes no significant embryological defects, it does significantly increase the penetrance of Brachyury (T) knockouts. These mutants develop malformed thoracic and lumbar vertebrae, and develop significantly fewer vertebrae in the tail (Harrison et al., 2000). This demonstrates the involvement of Sp5 in facilitating Wnt/β-catenin and Brachyury-mediated proliferation in the vertebrate tail. The distinction between Sp5 and Sp6-9 genes, despite their considerable similarities during development, arise from the prediction of a three-gene Sp cluster in the eumetazoan ancestor (Schaeper et al., 2010). This cluster, still present in the cnidarian Nematostella vectensis (original data from Putnam et al., 2007), discussed in Schaeper et al., (2010), contains Sp1-4, Sp5 and Sp6-9 homologues. Although the Sp complement of extant metazoan species ranges between 1 to 13 genes (Presnell et al., 2015), phylogenetic analysis supports the existence of these three clades (Presnell et al., 2015; Schaeper et al., 2010; Pei & Grishin 2015), which take their names from the human SP complement. The vertebrate Sp genes are noted for their linkage to the four vertebrate Hox clusters, leading to the conclusion that an ancestral three-gene Sp cluster was located on the same chromosome as the vertebrate Hox cluster, immediately before the two rounds of whole genome duplication (2R) at the base of the vertebrate lineage (Schaeper et al., 2010).

In order to support proposed models of Sp gene evolution, a number of studies have performed domain and amino acid frequency analyses on Sp proteins. As a gene family, they are defined by

---

**Fig. 1. Deuterostome Specificity Protein (SP) phylogeny.**

Three distinct clades of deuterostome SP proteins were found: SP1-4, SP5 and SP6-9, which take their names from human SP proteins. The cnidarian, cephalochordate, echinoderm and hemichordate lineages shown here each possess a single SP protein per clade. Midpoint-rooted RAxML tree, labelled with percentage RAxML bootstraps (above node) and MrBayes posterior probabilities (below node) for nodes with 70% or higher support from both algorithms. Trees used the Dayhoff model of protein evolution with 4 gamma rate categories. Tree branches are coloured according to their RAxML bootstrap support (green low, blue high). Cephalochordate sequences are shown in bold.
their possession of a Btd domain (Bouwman & Philipsen 2002; Wimmer et al., 1993), which is positioned shortly before a triplet of C2H2 zinc finger domains (interspaced by conserved linker regions), which all together are located close to the protein’s C-terminus (Iuchi 2001). This triplet of C2H2 domains, including their conserved linker regions, is also present in Krüppel-like factors (KLFs) and Krox/Egrs (Iuchi 2001). Outside of these conserved domains, however, SP proteins can vary substantially, both within and among species. As such, a number of previous studies have analysed the residue profiles of regions outside of the C2H2 and Btd domains to support models of Sp gene evolution (Schaeper et al., 2010; Bouwman & Philipsen 2002; Wimmer et al., 1993).

Given the considerable amount of data available for characterising SP proteins, and the convincing model for the existence of a three-gene Sp cluster prior to the vertebrate 2R, there is surprising inconsistency surrounding the SP complement of invertebrate chordates. Specifically, the cephalochordates, sister group to the vertebrate and urochordate clade, have been reported to possess either Spα, Spβ and Sp5 (Shimeld 2008), Sp1-4a, Sp1-4b, Sp5 and Sp6-9 (Presnell et al., 2015) or Sp1-4, Sp5, Sp6-9 (Pei & Grishin 2015) based on analyses in Branchiostoma floridae. As a chordate lineage, Branchiostoma is typically considered to be evolutionarily conservative, and so lineage-specific changes to their SP complement would warrant further study. Cephalochordates are also a model for chordate tail elongation, employing many of the same developmental signalling strategies with vertebrates, including the expression of Wnt ligands and Brachyury in the tailbud (Holland 2002). We might therefore expect Sp5 and Sp6-9 to interact with Wnt signalling in the amphioxus tailbud.

Previous research on Sp genes has used Branchiostoma floridae as the sole cephalochordate representative (Shimeld 2008; Presnell et al., 2015; Pei & Grishin 2015). Here, we confirm the SP complement of the B. lanceolatum and B. belcheri species, through both phylogenetic and domain structure analysis. We also demonstrate the expression of Sp5 in the B. lanceolatum tailbud, and document the responses of this Sp5 and Brachyury2 expression to pharmacological manipulation of β-catenin levels. Our data suggest that both species possess a typical invertebrate chordate SP complement, with single copies of Sp1-4, Sp5 and Sp6-9. Furthermore, in both species Sp1-4 contains a novel, Glutamine-rich expansion region between the C2H2 triplet and C-terminus. Sp5 in B. lanceolatum neurulae and early larvae marks several tissues, including the somites, neural tube and tailbud. We demonstrate here that pharmacologically increasing Wnt/β-catenin signalling in early developmental stages caused the ectopic expression of Sp5. In contrast, treatment at gastrula stage led to the subsequent loss of Sp5 expression in the posterior growth zone of neurulae and larvae.

Results

The Sp gene complement of Branchiostoma

Three SP-like proteins were identified in the predicted protein sets of both the Branchiostoma lanceolatum and B. belcheri genomes. These each contained the characteristic buttonhead (Btd) box and C2H2 zinc finger triplets (Supplementary Fig. S2). In order to confirm the identity of these three proteins, the C2H2 triplets from these proteins were aligned against the same domain set of metazoan SP and KLF proteins (Supplementary File 1), as well as human Wilms Tumour 1 (W1) as an outgroup. Phylogenetic analysis using both RAxML and MrBayes (Supplementary Fig. 3) supported their positioning among a well-supported clade of the SP proteins. However, the RAxML analysis was not able to provide strong support values for numerous well supported clades in the MrBayes analysis, including the internal topology of the SP clade. We therefore took this subset of SP proteins and aligned both the Btd domain, C2H2 triplet and short intervening sequence (Supplementary File 2). From this analysis (Fig. 1), both RAxML and MrBayes produced strong support values for the existence of three clades, which correspond to the previously observed grouping of eumetazoan SP proteins into SP1-4, SP5 and SP6-9 clades. All invertebrates examined, including the two Branchiostoma species included here, possessed a single representative of each clade.

Outside of the Btd and C2H2-triplet domains, the SP proteins differ significantly. Several previous studies have characterised these differences, identifying both the presence of an additional region, the SP box, which is inconsistently conserved among SP genes, as well as regions enriched for specific residues (Schaeper et al., 2010; Bouwman & Philipsen 2002; Wimmer et al., 1993). The same analysis was performed here in order to further support the phylogenetic gene assignments (Fig. 2). Residue frequency analysis of the regions between the N-terminus and the Btd box of each protein indicated that SP1-4 is enriched for Q and T residues, SP5 for P, S and A residues, and SP6-9 for S residues. Despite C2H2 triplets usually being observed near the C-terminus of SP and KLF proteins (Iuchi 2001), in SP1-4 they are closer to the N-terminus, a finding supported across its full length by transcriptomic data in Branchiostoma lanceolatum (Dailey and Somorjai, unpublished) and EST data in B. belcheri (data from genome website (Huang et al., 2014)). This trailing region of approximately 480 residues is enriched for both Alanine and Glutamine, the latter of which comprises around 25% of the sequence. These residues are regularly spaced, with no repeat longer than (Q)4 (highlighted in Supplementary Fig. 4).

Previous reports indicate that these three Sp genes existed in clades as distant as the cnidarian Nematostella vectensis, in

**Fig. 2. Structural comparison of Branchiostoma SP proteins.** Comparison of domain presence and residue frequency in Branchiostoma SP proteins. Conserved SP-family domains are highlighted including the SP box, Btd box and C2H2 zinc finger domains. Sequences between the N-terminus and Btd box were analysed for residue frequency, and any enriched residues (where a single residue forms >10% of the total sequence) are indicated. The same analysis was performed for the region between the final C2H2 and C-terminus in SP1-4, but not for the remaining two SP proteins, as these regions were too short to characterise in this manner. Total protein lengths are to approximate scale.
genes in the combined developmental and adult transcriptome available for _B. lanceolatum_ (Oulion et al., 2012), suggesting that they may be developmentally expressed. A 692bp region of _Sp5_ was cloned from _B. lanceolatum_ adult cDNA, and used to assay _B. lanceolatum_ embryos via WMISH. In order to identify the earliest point of _Sp5_ expression, a range of developmental stages from fertilised egg to gastrula was assayed. The earliest expression detected here was in the late blastula stage, where a hemisphere of expression is seen, centred over the macromeres (Supplementary Fig. 5). Expression continues into all later stages assayed (Fig. 4), from the G4 gastrula to L1 larvae. In G4 gastrulae (Fig. 4A), staining is found in posterior ectoderm, mesendoderm and at the dorsal blastopore lip, but is notably absent from the entire anterior. In N0 neurulae, (Fig. 4B) staining occurs in the neural plate, dorsal blastopore lip and ventral endoderm. In N1 neurulae (Fig. 4C) expression is found in the anterior tailbud, patchily throughout the notochord, in the ventral endoderm lining the archenteron, and in the anterior and posterior of the neural tube. By the N2 stage (Fig. 4D), staining is also visible in the somites. This general pattern continues throughout the remaining neurula stages, although from N3 stage onwards (Fig. 4E) expression in the ventral endoderm is only found in the posterior (marking the chordoneural hinge and hindgut) and the anterior of the embryos. In N4 stage embryos (Fig. 4F), the staining in the anterior labels the cerebral vesicle and regions within the presumptive pharynx, and expression is seen throughout the entire tailbud. By L1 (Fig. 4G) expression in the notochord is also weak in all but the anterior

**Developmental expression of _Sp5_**

Thus far, there are no published descriptions of _Sp_ gene expression patterns in any invertebrate chordate species. We were, however, able to identify expression of all three _B. lanceolatum_ _Sp_ genes in the combined developmental and adult transcriptome available for _B. lanceolatum_ (Oulion et al., 2012), suggesting that they may be developmentally expressed. A 692bp region of _Sp5_ was cloned from _B. lanceolatum_ adult cDNA, and used to assay _B. lanceolatum_ embryos via WMISH. In order to identify the earliest point of _Sp5_ expression, a range of developmental stages from fertilised egg to gastrula was assayed. The earliest expression detected here was in the late blastula stage, where a hemisphere of expression is seen, centred over the macromeres (Supplementary Fig. 5). Expression continues into all later stages assayed (Fig. 4), from the G4 gastrula to L1 larvae. In G4 gastrulae (Fig. 4A), staining is found in posterior ectoderm, mesendoderm and at the dorsal blastopore lip, but is notably absent from the entire anterior. In N0 neurulae, (Fig. 4B) staining occurs in the neural plate, dorsal blastopore lip and ventral endoderm. In N1 neurulae (Fig. 4C) expression is found in the anterior tailbud, patchily throughout the notochord, in the ventral endoderm lining the archenteron, and in the anterior and posterior of the neural tube. By the N2 stage (Fig. 4D), staining is also visible in the somites. This general pattern continues throughout the remaining neurula stages, although from N3 stage onwards (Fig. 4E) expression in the ventral endoderm is only found in the posterior (marking the chordoneural hinge and hindgut) and the anterior of the embryos. In N4 stage embryos (Fig. 4F), the staining in the anterior labels the cerebral vesicle and regions within the presumptive pharynx, and expression is seen throughout the entire tailbud. By L1 (Fig. 4G) expression in the notochord is also weak in all but the anterior

**Fig. 3. Genomic arrangement of the _Branchiostoma_ Sp genes.** Diagram showing the genomic arrangement of _sp_ genes in _B. lanceolatum_ and _B. belcheri_. All three _B. lanceolatum_ genes are found on the same scaffold, albeit with significant distances between genes, especially in the case of _Sp1-4_. In _B. belcheri_, _Sp1-4_ is located on a separate scaffold to the other two _Sp_ genes. Arrows depict gene loci and their orientations, and between these are genomic distances in kilobase pairs (kb). Scaffold numbers are given for the respective genome projects.

**Fig. 4. Expression of _Sp5_ in _Branchiostoma lanceolatum_ embryos.** Developing _B. lanceolatum_ embryos were assayed for expression of _Sp5_ between G4 stage and L1 stage by WMISH. All embryos are orientated with the anterior to the left, except for the third panels of (A,B), which show blastoporal views with the dorsal surface to the top. Asterisks mark the blastopore, and arrowheads mark the tailbud. In G4 stage gastrulae (A), staining is found in most of the ectoderm and the mesendoderm, with the exception of the anterior, but by N0 stage neurulae (B) ectodermal expression is found only in the neural plate. In N1-4 stage neurulae (C-F), expression in neural tissues and ventral endoderm is found in the anterior and in the posterior, but not in between. Expression is seen throughout the somites, albeit with inconsistent penetrance, and throughout the length of the notochord. In late N4 neurulae and early L1 larvae (F-G), expression strongly marks the tailbud and posterior-most somitic, notochord and neural tissues as well as posterior ventral endoderm. Anteriorly, _Sp5_ is expressed in the cerebral vesicle and presumptive pharynx. Abbreviations used: _cnh_, chordoneural hinge; _cv_, cerebral vesicle; _ect_, ectoderm; _en_, endoderm; _mesen_, mesendoderm; _mo_, mouth; _nt_, notochord; _np_, neural plate; _nt_, neural tube; _pha_, pharynx; _somi_, somites.
and posterior, and somite expression is relatively weak with variable penetrance. Throughout all neurula stages and L1 larval stages, a strong domain of expression is seen overlapping the tailbud, as well as the neural and mesodermal tissues immediately anterior to this structure, and ventral posterior endoderm.

**Modulation of β-catenin levels regulates Sp5 expression**

The identification of Sp5 in the *B. lanceolatum* tailbud matches observations of this gene in the vertebrate tailbud. As reviewed in the introduction, vertebrate Sp5 acts as a component of the Wnt signalling pathway for certain Wnt/β-catenin downstream targets, such as Brachyury/T, and appears to be strongly linked to its ability to drive proliferation in the vertebrate tailbud. Amphioxus also utilises a homologous tailbud structure, which progressively extends axial tissues of the embryo during axis elongation. Interestingly, several of the Sp5 expression domains observed here, including the tailbud, overlap with previously described expression patterns for numerous Wnt ligands, as well as both Brachyury genes in *Branchiostoma* (Holland et al., 1995; Holland 2002; Bertrand et al., 2017). It is possible, therefore, that the Sp5 and Brachyury genes of amphioxus also interact with each other, or with Wnt/β-catenin signalling, in the tailbud. To test this, *B. lanceolatum* embryos were treated with 1-Azakenpaullone (Azk), an inhibitor of GSK3-β that acts to increase the nuclear accumulation of β-catenin (Kunick et al., 2004), mimicking increased Wnt/β-catenin signalling. Embryos were treated with 10μM Azk at the G4 stage, well before the formation of the tailbud, and then assayed using WMISH for Sp5 and Brachyury2 expression at N1 and L1 stages. At N1 stage, a characteristic horseshoe-shaped domain of Brachyury2 expression can be seen overlapping the tailbud; this domain was significantly expanded as a result of treatment (Fig. 5A). Most domains of Sp5 expression (Fig. 5B) were not obviously affected by Azk treatment. However expression overlapping the Brachyury2 horseshoe-shaped domain is strongly reduced in treated embryos. In L1 embryos, Brachyury2 expression exhibits a similar response to treatment (Fig. 5C), with the posterior expression domain overlapping the tailbud expanding significantly. Sp5 again acts conversely (Fig. 5D), with expression decreasing in this posterior expression domain.

The response of Sp5 expression to increased Wnt/β-catenin signalling was surprising, especially in comparison to the response of Brachyury2 expression. To confirm if these contrasting responses were reproducible under other treatment conditions, embryos were treated with 1-Azakenpaullone at blastula stage, earlier than the previous treatment. These were then assayed for expression of these two same genes at N1 stage (Fig. 6). At this stage, *Brachyury2* is usually absent from the ectoderm (Fig. 6A and Terazawa and Satoh 1997), with the exception of a ring of expression around the blastopore lip. 1-Azakenpaullone treatment however caused ectopic expression of *Brachyury2* throughout the entire ectoderm. A similar response to treatment was seen for Sp5 expression (Fig. 6B): Sp5 is detected throughout the anterior-most ectoderm and endoderm of treated embryos, whereas this region is consistently free of expression in control and wildtype embryos at N0 and G4 stage embryos, respectively.

In light of the now contrasting responses of Sp5 to GSK3-β inhibition with 1-Azakenpaullone, a third treatment was performed with an alternate GSK3-β inhibitor (CHIR99021) from 8-cell stage to G4 gastrula stage (Supplementary Fig. 6A). This treatment again resulted in ubiquitous Sp5 expression (Supplementary Fig. 6B), in contrast to control embryos where no expression is detected in the anterior of the embryo. Quantitative RT-PCR of these embryos showed a marked increase in Sp5 expression following treatment (Supplementary Fig. 6C), indicating that GSK3-β inhibition increases Sp5 expression under these conditions.

**Discussion**

**Amphioxus possesses a typical invertebrate chordate SP complement**

The current hypothesis is that of a eumetazoan ancestor possessing a clustered set of three SP proteins: SP1-4, SP5 and SP6-9 (see introduction). Central to these evolutionary models is the assumption that such a cluster existed in the vertebrate ancestor, and by extension the chordate ancestor, with subsequent expansions in various vertebrate lineages (Schaeper et al., 2010). However, genomic data from *Branchiostoma floridae* resulted in conflicting conclusions surrounding the number and identity of these genes.

![Fig. 5. 1-Azakenpaullone treatment expands the Brachyury2 expression domain overlapping the tailbud, but reduces that of Sp5.](image)
in the basal chordate lineage, the cephalochordates. We therefore analysed genomic datasets from two additional species, *B. belcheri* and *B. lanceolatum*, in order to resolve this confusion. Our results indicate the presence of three *Branchiostoma* SP-like proteins, which grouped strongly along with other eumetazoan SP proteins in a KLF/SP phylogenetic analysis. Further support for this assignation was found when considering domain presence, as all three possessed the Btd and C2H2 domain triplets that are defining of this protein family (Iuchi 2001). It appears therefore that the typical *Branchiostoma* condition is the possession of three *Sp*-like genes.

Phylogenetic analysis of deuterostome SP proteins provided initial support for the identity of the three *Branchiostoma* SP proteins as SP1-4, SP5 and SP6-9. To confirm this finding, we examined the residue frequency of these proteins outside of the highly conserved Btd and C2H2 domains. In the regions between the N-terminus and Btd domain (the ‘pre-Btd region’), we find that SP1-4 is enriched for Q and T, which is consistent with vertebrate, cnidarian and arthropod states (Schaeper et al., 2010). The same is true of our observation of P and A residues for *Branchiostoma* SP5 and SP6-9: although variable, they are frequently observed in these two groups of genes among the same set of species (Schaeper et al., 2010). Together, these findings indicate that *Branchiostoma* possess a typical eumetazoan three *Sp*-gene complement.

Our domain structure analysis of *Branchiostoma* SP proteins also identified the inconsistently conserved SP box in *Branchiostoma* SP1-4. This domain is present in all three SP genes in humans (with only SP6 and SP9 lacking this domain) (Schaeper et al., 2010), and thus the absence of this domain in both *Branchiostoma* SP5 and SP6-9 must represent lineage specific losses. The function of this domain has never been experimentally proven. However, the SP box possesses an endoproteolytic cleavage site, located near the N-terminal region in SP1, which is necessary for proteasome-dependent degradation *in vitro* (Su et al., 1999), and as such it is considered likely to be involved in SP turnover (Bouwman & Philipse 2002). Given that SP boxes are lost in some vertebrate SP proteins (Schaeper et al., 2010), their independent loss in several amphioxus SP proteins is not unprecedented.

A second, more surprising observation was the discovery of a novel expansion in *Branchiostoma* SP1-4. Typically, the C2H2 triplet of SP and KLF proteins is found close to the C-terminus of these proteins, which is not the case for amphioxus SP1-4. When comparing the length of this C2H2 ‘trailing region’ across the entire pan-metazoan KLF/SP dataset of Presnell et al. (2015), in almost all cases these are less than 100 residues in length, compared to almost 500 in *Branchiostoma* SP1-4. Unusually, 25% of the residues in this region are Glutamine. Whilst it is not uncommon in other contexts to see poly-Glutamine repeats occur through DNA replication slippage, in this instance there are no repeats longer than (Q)_n. This region of low-complexity overlaps with the assembly artefacts seen in the *B. lanceolatum* Sp1-4 loci, where 100% identical introns and exons were repeated multiple times. Low-complexity regions such as there are relatively difficult for assembly software to reconstruct, providing a potential explanation for these artefacts. We also speculate, therefore, that the additional "SP1-4b" protein (which contains only the region from the Btd box to final C2H2 domain) from Presnell et al. (2015) may be the product of a similar computational error. The purpose of the Glutamine-rich trailing region of *Branchiostoma* SP1-4 remains unclear. However, the Glutamine rich pre-Btd region of SP1 allows the cross-linking of SP1 monomers, permitting both proximally- and distally-bound SP1 molecules to synergistically act upon a single target gene (Courey et al., 1989). Given that amphioxus SP1-4 still retains a Glutamine rich pre-Btd region, it is possible that this trailing region facilitates novel interactions in *Branchiostoma*.

Our extended analyses also give further insight into the genomic arrangement of *Branchiostoma* *Sp* genes, and evolution of the cluster. The current model of *Sp* gene evolution predicts a eumetazoan gene cluster containing a single *Sp1*-4, *Sp5* and *Sp6*-9, based primarily upon the existence of such a cluster in the cnidarian *Nematostella vectensis* (Schaeper et al., 2010). This model also points to the non-teleost vertebrate condition, where *Sp* genes are found on four separate chromosomes in an arrangement congruent with an original three *Sp*-gene cluster expanding during the 2R event (Schaesper et al., 2010). We were able to identify *Sp* gene arrangements from a number of invertebrate species that support this model (Fig. 7). These included a second cnidarian...
species, *Orbicella faveolata*, as well as the hemichordate *Saccoglossus kowalevskii*. Finally, of course, we were able to identify all three *B. lanceolatum* Sp genes on a single genomic scaffold. The location of *B. belcheri* Sp1-4 on a separate scaffold, whilst not conclusive, does not preclude it from being located on the same chromosome as the other *B. belcheri* genes. Therefore, all of the additional data collated here support the existence of Sp1-4, Sp4 and Sp6-9 existing on the same chromosome in both the deuterostome and chordate ancestor. Interestingly, however, whilst previous studies refer to this arrangement as a ‘cluster’ (Schaepert *et al.*, 2010), the intragenic distances for the species observed here (Fig. 7) are considerable. Even in the cnidarian *O. faveolata*, these distances are double (50kb) those in *N. vectensis* (25-29kb). In *B. lanceolatum*, Sp1-4 is 1.1Mb from the other Sp genes, which we feel makes it unlikely that there is meaningful “clustering” of this gene with the others.

**Amphioxus Sp5 expression is dynamic**

To date, expression of Sp genes has not been characterised in any invertebrate chordate species. The combined developmental and adult transcriptome for *Branchiostoma lanceolatum* (Oulion *et al.*, 2012) contains fragments matching all three Sp genes from this species, and so it is possible that all three are expressed during amphioxus development. Here, we cloned the *B. lanceolatum* Sp5 gene; as the only Sp not to have duplicated during the vertebrate 2R events, this allows for more direct comparison among the Metazoa, and an insight into its role in the common chordate ancestor. Between the onset of neurulation (N0) and early larva (L1), Sp5 expression marks recently specified neural, notochord and somitic tissue, which is consistent with observations in other species for roles in specifying neurectoderm and axial mesoderm (Weidinger *et al.*, 2005). As axis elongation occurs, and the more anterior and central regions of these tissues begin to differentiate, Sp5 expression in most neural and notochord tissue is lost, and somitic expression is more sporadic. As a result of this, most Sp5 expression is found in the newly formed regions of these tissues just proximal of the tailbud, as well as the tailbud itself. This is therefore consistent with Sp5’s conserved role in embryological growth zones in the Bilateria.

Several anterior domains of Sp5 expression persist during axis elongation. Expression in the presumptive pharynx is matched by that of mouse Sp5, which is expressed in the first three branchial pouches and clefts (Treichel *et al.*, 2001). Expression also remains in some anterior neural tissues such as the cerebral vesicle. In the mouse, Sp5 is reportedly expressed both in the forebrain, in regions such as the dorsal posterior telencephalon, and at the midbrain-hindbrain boundary (Treichel *et al.*, 2001; Sahara *et al.*, 2007). Classically, the cerebral vesicle in amphioxus was considered to be homologous to the vertebrate forebrain (Lacalli *et al.*, 1994). However, recent studies indicate that by early neurula stages (N2), the amphioxus neural tube can be subdivided into three regions: the hypothalamo-prethalamic primordium (HyPTh); the Di-Mesencephalic primordium (DiMes) and Rhombencephalo-Spinal primordium (RhSp) regions, which are homologous to the vertebrate midbrain, thalamus, and pretectum; and the hindbrain and spine (Albuixech-Crespo *et al.*, 2017). It would therefore be interesting to determine if the neural expression of Sp5 overlaps the DiMes and RhSp boundary in amphioxus, or rather the more anterior HyPTh, via double-labelling experiments. Despite this outstanding question, these findings indicate a high degree of conservation between amphioxus and vertebrate Sp5 expression. Amphioxus may therefore be a useful model system in which to study the role of Sp5 in contexts such as chordate posterior elongation or neural development.

**Contrasting responses of Sp5 expression to increased Wnt/β-catenin signalling**

The role of Sp5 as both a downstream effector and component of the Wnt/β-catenin pathway is well documented in vertebrates (see introduction), particularly its relationship with Brachyury/T in promoting proliferation in the tailbud. Cephalochordates also possess an embryonic tailbud structure, which expresses multiple Wnt ligands and two Brachyury genes derived from a lineage-specific
duplication event (Holland et al., 1995; Terazawa and Satoh., 1997; Holland 2002; Bertrand et al., 2017). The two amphioxus Brachyury genes are virtually identical within the coding region, and in situ hybridisation that utilises 3’ UTR sequence as a probe is unsuccessful, such that the expression pattern may be a conflation of both genes (Holland et al., 1995; Inoue et al., 2017). Regardless, we show that Sp5 expression overlaps with both the tailbud and Brachyury2 expression in B. lanceolatum, raising the question as to whether these two genes also share a relationship with Wnt/β-catenin in the amphioxus tailbud.

When B. lanceolatum embryos were treated here with 1-azakenpaullone (Azk) to increase Wnt/β-catenin signalling, Brachyury2 expression expanded significantly. This finding is generally consistent with β-catenin acting to promote Brachyury2 expression, similar to the relationship of these two factors in vertebrates. In contrast, Sp5 expression decreased in the Brachyury2+ region of the tailbud. Taken in isolation, this result could be interpreted as an inhibition of Sp5 expression, resulting directly from increased β-catenin levels. We felt however that this was unlikely, given that Sp5 is an effector of Wnt/β-catenin signalling in vertebrates, and both Wnt ligands and Sp5 are co-expressed in domains overlapping the amphioxus tailbud. In order to determine if this response was consistent, we performed additional treatments at an earlier developmental stage with two different GSK3-β inhibitors. In contrast to prolonged exposure to the drug from the gastrula stage, treatments at either the blastula or 8-cell stage resulted in upregulated expression of Sp5 by the G4 stage, as assessed by both WMISH and qPCR. Importantly, both Sp5 and Brachyury2 show similar responses, becoming ectopically expressed in the ectoderm. This finding is therefore consistent with both Sp5 and Brachyury2 acting as downstream targets of Wnt/β-catenin signalling in amphioxus, similarly to vertebrates such as mouse (Fujimura et al., 2007).

Taking all of these results together, the response of Sp5 to ectopically increased β-catenin levels appears to be variable, for which there are a number of explanations. In the simplest scenario, it may simply be that these responses are specific to certain tissues, such as Brachyury2+ tailbud cells. Alternatively, the direction of response may be dependent on the developmental stages at which the embryos were treated. However, we also note that the gastrula-stage treatment window was significantly longer, and that this was the only instance in which GSK3-β inhibition resulted in decreased Sp5 expression. This could suggest that there is some compensatory regulatory mechanism that acts to combat ectopic increases in β-catenin, which is only detected after prolonged exposure to treatment. To resolve this question, qPCR time courses for Sp5 expression following both shorter and longer treatments would be informative, and it would be important to determine if these responses differ between early and later development. It may also be possible to ectopically express Sp5 mRNA in B. lanceolatum embryos. This would bypass any hypothetical transcriptional negative feedback loop, and we may expect to find a similar (or indeed additive) tailbud phenotype to that seen following pharmacological Wnt/β-catenin activation.

Conclusions

The protein structure, phylogenetic signal and genomic position of the three Branchiostoma Sp-like genes found in B. lanceolatum and B. belcheri are consistent with this genus possessing a three-gene Sp cluster, as is typical for invertebrate eumetazoans. During embryonic development in B. lanceolatum, Sp5 is specifically expressed in a number of tissues of the embryo, including the cerebral vesicle and tailbud, which is broadly consistent with its roles in other bilaterians. We demonstrate that expression of Sp5 is modulated by pharmacological increases in Wnt/β-catenin signalling, although its response varies depending upon treatment conditions. We hypothesise that the decrease of Sp5 expression seen only in long treatments is indicative of a negative-feedback loop in the Wnt/β-catenin-Sp5 signalling pathway.

Materials and Methods

Identification of Sp-like genes

Human SP proteins were BLAST searched against the predicted protein set of the genome projects for Branchiostoma belcheri (Huang et al., 2014) and B. lanceolatum (Unpublished data, kindly made available by the B. lanceolatum genome consortium). Matches were then reciprocally BLAST searched against the NCBI nr protein database, and apparent SP-like proteins for each species were taken for further analysis. The locus for each gene was confirmed using IBLASTn searches of the predicted protein against the genome scaffold set. Two apparent assembly artefacts are present in the scaffold over one gene in B. lanceolatum (shown in Supplementary Figure 1), which lead to the repetition of several exons and intronic regions with 100% identity, flanked by patches of unidentified nucleotides (Ns). These repeat-artefacts were however not present in the predicted protein.

In order to confirm the identity of these three Branchiostoma SP-like proteins, two sets of alignments and trees were generated. The first contained the C2H2 domains of a selection of metazoan SP and KLF proteins, as well as the human Wilms Tumour 1 (Wt1) as an outgroup. The subset of these proteins identified as genuine SP proteins in this analysis were then used to create an alignment of Btd and C2H2 domains in order to confirm internal SP clades. Sequences for Saccoglossus kowalevskii, Lingula anatina and Orbicella faveolata were retrieved through BLASTp searches of the NCBI nr database, and the remaining sequences (and their associated identifications) were taken from Presnell et al. (2015) (all sequences and accessions provided in Supplementary File 1). Both alignments were created using MAFFT on default settings, and then analysed using model_generator (Keane et al., 2006) to determine the most suitable model of protein evolution for subsequent phylogenetic trees (given in the appropriate figure legends). The trees shown were created using RAxML (Stamatakis 2014), with additional support values from MrBayes (Ronquist et al., 2012), and all trees were generated on the CIPRES Science Gateway (Miller et al., 2010). The Btd and C2H2 domain alignments of Branchiostoma SP proteins were also used to identify the sequence between the N-terminus and the Btd box (referred to here as the ‘pre-Btd region’) and the sequence following the final C2H2 domain (referred to here as the ‘trailing region’). These were then analysed for residue frequency.

Embryo collection and treatments

Adult Branchiostoma lanceolatum with mature gonads were collected from Argelès-sur Mer (France), and heat-induced to spawn, as per the protocol detailed in (Fuentes et al., 2007). Embryos used in treatment experiments were kept in a known volume of seawater, allowing the addition of stock solutions of either 1-Azakenpaullone (Azk) or CHIR99021 (both prepared in DMSO) to be added at the specified treatment time. Final concentrations for treatments were 25μM and 10μM Azk when treating at blastula and gastrula stages, respectively, and 10μM CHIR99021 at cleavage (8 cell) stages. Control embryos were treated with an equal volume of DMSO. Once embryos had developed to the desired stage, they were fixed with 4% PFA in MOPS salts (0.1M MOPS, 2mM MgSO4, 1mM EGTA, and 0.5M NaCl) and stored in 70% EtOH. Embryo staging was performed according to Hirakow & Kajita (1994) and Hirakow & Kajita (1991), with modification as per (Zhang et al., 2013).
Cloning and probe template preparation

RNA was extracted from a whole unripe *B. lanceolatum* adult using Trisrene (Bioline) and phenol chloroform extraction; cDNA was prepared using the Tetro cDNA Synthesis Kit (Bioline). A partial length gene fragment for Sp5 was obtained through PCR (primer sequences: forward: 5’ GTCTCGCTCTTTTCTTCGACG 3’, reverse: 5’ TTGTGAATGATCCGGCCCAAG 3’). This fragment was ligated into pGEM-T Easy (Promega) and then transformed into XL10-Gold competent *E. coli* using standard protocols. Plasmids were harvested using pGOLD (VWR) and sequenced to confirm successful ligation, after which the insert sequence was submitted to GenBank (accession: MF162314.1). For Brachury2, the pBluescript II SK+ plasmid originating from (Somorjai et al., 2008) (accession: EU685284.1) was used.

Transcriptional evidence and Whole Mount in situ Hybridisation (WMISH)

*B. lanceolatum* SP proteins were used to tBLASTn search publicly available transcript databases for this species to identify their expression. Matches of over 90% identity (for the full length of the transcript) were then confirmed by reciprocally BLASTX searching them against the predicted protein set of their respective species, before being counted (accessions given in Supplementary Table 1).

Sp5 and Brachury2 containing plasmids were used as PCR templates for standard M13F (5’ GTAAACGACGGCCAGT 3’) and M13R (5’ AAGGTGATCGGGAACAAGTG 3’); Polr2l_Rev: 5’ GGTCCACATGTGAG 3’). Following primers were used: Rpl32_Fwd: 5’ GGCTTCAAGAAATTCCTC (VWR) and used as templates for the production of DIG-labelled RNA probes, using T7 RNA polymerase with standard protocols. Probes were purified using Roche miniQuick spin columns or sodium precipitation with sodium acetate (3M, pH 5.2) and ethanol.

WMISH was performed as previously described (Dailey et al., 2016), with minor modifications. Specifically, embryos were permeabilised using proteinase K at 1.25U/ml, for empirically-tested periods based upon embry stage and enzyme batch, and post-fixed for 60 minutes in PFA. The pre-hybridisation wash, hybridisation wash and initial post-hybridisation washes were performed at either 60 or 62°C, dependent on experiment. The chromogenic reaction was performed using NBT/BCIP (Roche), and embryos were then post-fixed in PFA for 60 minutes. Four WMISH experiments were performed for embryonic Sp5 expression, and three for Sp5 and Brachury2 expression in treated embryos, with a minimum of 3 embryos per stage. Embryos were mounted in 90% glycerol/10% PBS, and photographed under a Leitz DMRB microscope (Leica Microsystems) with Normarski optics. Photographs were taken with the Retiga 2000R camera and the QCapture software suite (Qimaging), and processed using Fiji (Schindelin et al., 2012) and Adobe Fireworks CS5.

Quantitative RT-PCR (qRT-PCR)

Embryos were treated with DMSO (control) or 10µM CHIR99021 at the 8-cell stage and harvested for RNA at the G4 gastrula stage. cDNA was prepared using a VILO cDNA synthesis kit (Invitrogen). qRT-PCR reactions were run in a LightCycler® 480 Instrument (Roche) using LightCycler® 480 DNA SYBR Green I Master (Roche). Embryos were mounted in 90% glycerol/10% PBS, and photographed under a Leitz DMRB microscope (Leica Microsystems) with Normarski optics. Photographs were taken with the Retiga 2000R camera and the QCapture software suite (Qimaging), and processed using Fiji (Schindelin et al., 2012) and Adobe Fireworks CS5.

References


Conflict of interest

The authors declare that there is no conflict of interest.

Author contributions

SCD performed experiments with contributions from IMLS, and analysed genomic data. IK performed and analysed CHIR99021 treatment experiments and qPCR. IMLS conceived the study and contributed reagents. SCD and IMLS wrote the paper. All authors approved the final manuscript.

Funding and Acknowledgements

SCD was funded by a MASTS (Marine Alliance for Science & Technology Scotland) PhD Studentship. Work in the IMLS lab is in part funded by the European Union Horizon 2020 research and innovation programme under grant agreement numbers 654428 (“CORBEL”) and 730984 (“ASSEMBLE+”), and the Wellcome Trust ISSF grant number 204821/Z/16/Z. I.K. was supported by grant 15-21285J from the Czech Science Foundation. This work was supported by the Branchiostoma lanceolatum genome consortium, which provided access to the Branchiostoma lanceolatum genome sequence.


Further Related Reading, published previously in the *Int. J. Dev. Biol.*

**The Wnt signaling mediator tcf1 is required for expression of foxd3 during Xenopus gastrulation**
Sylvie Janssens, Olaf Van Den Broek, Ian R. Davenport, Robbert C. Akkers, Fei Liu, Gert Jan C. Veenstra, Stefan Hoppler, Kris Vleminckx and Olivier Destrée
Int. J. Dev. Biol. (2013) 57: 49-54

**Wnt signaling in planarians: new answers to old questions**
Maria Almuedo-Castillo, Miquel Sureda-Gómez and Teresa Adell

**From the American to the European amphioxus: towards experimental Evo-Devo at the origin of chordates**
Jordi Garcia-Fernández, Senda Jiménez-Delgado, Juan Pascual-Anaya, Ignacio Maeso, Manuel Irimia, Carolina Minguillón, Élia Benito-Gutiérrez, Josep Gardenyes, Stéphanie Bertrand and Salvatore D’Aniello
https://doi.org/10.1387/ijdb.072436jg

**Wnt signaling in hydroid development: ectopic heads and giant buds induced by GSK-3beta inhibitors**
Werner Müller, Uri Frank, Regina Teo, Ofer Mokady, Christina Guette and Günter Plickert

**Formation and maintenance of the organizer among the vertebrates.**
K Joubin and C D Stern

*5 yr ISI Impact Factor (2013) = 2.879*