Available online at www.sciencedirect.com

ScienceDirect

Genomics 89 (2007) 721–730

GENOMICS

www.elsevier.com/locate/ygeno

Diversity of olfactomedin proteins in the sea urchin

Brian J. Hillier [1], Gary W. Moy, Victor D. Vacquier *

Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093-0202, USA

Received 31 January 2007; accepted 28 February 2007
Available online 17 April 2007

Abstract

Olfactomedin (OLF) domain proteins maintain extracellular protein-protein interactions in diverse phyla. Only one OLF family member, amassin-1, has been described from the sea urchin *Strongylocentrotus purpuratus,* a basal invertebrate deuterostome. Amassin-1 mediates intercellular adhesion of coelomocytes (immunocytes). Here we describe the protein structural features of four additional OLF proteins, the total for the genome being five. Phylogenetically, four of these proteins (the amassins) form a subgroup among previously identified OLF proteins. The fifth OLF protein is within the colmedin subfamily and contains a type II transmembrane domain, collagen repeats, and an OLF domain. Sea urchin OLF proteins represent an intermediate diversification between protostomes and vertebrates. Transcripts of all five OLF family members are in coelomocytes and adult radial nerve tissue. Transcripts for some OLF proteins increase during late larval stages. Transcript levels for amassin-1 increase 1,000,000-fold, coinciding with formation of the adult urchin rudiment within the larval body.

Keywords: Amassin; Colmedin; Coelomocytes; Sea urchin genome; Cell adhesion; Gene expression

Previous work has identified the extracellular protein, amassin, as an essential factor in the rapid and dramatic cell adhesion phenomenon (clotting) that occurs between coelomocytes in the sea urchin *Strongylocentrotus purpuratus* [1]. Amassin contains an olfactomedin domain (OLF) in the C-terminal half of the protein, and it is through this domain that amassin is capable of binding the coelomocyte surface. Amassin belongs to a family of proteins that all contain an OLF domain. OLF family members occur in diverse animal phyla, including nematodes, arthropods, echinoderms, and chordates. The number of OLF proteins continues to grow. Recently described members with neurological roles include UNC-122, which functions by organizing the neuromuscular junction [2], and gliomedin, which promotes the formation of the nodes of Ranvier [3]. These OLF family proteins bind their protein targets through the OLF domain. Mutations in the OLF domain of human myocilin are associated with inherited forms of glaucoma [4–8]. Myocilin binds its target, another OLF protein named optimedin, through the OLF domain [9]. These data from disease states to developmental neurogenic events underscore the importance of understanding the functions of OLF domain-containing proteins.

Relatively few OLF family members have been identified from nonchordate invertebrates: 1 from *Drosophila melanogaster,* 2 from *Caenorhabditis elegans,* and 1, amassin, from *S. purpuratus* [1,2]. The situation is quite different in vertebrates, in which OLF domain proteins are much more common. For example, in humans there are at least 13 OLF domain proteins. This disparity between animal groups in the number of OLF family members makes it seem that vertebrates were the primary OLF innovators. Using the publicly available assemblies of the sea urchin genome at the Baylor College of Medicine Human Genome Sequencing Center (http://www.hgsc.bcm.tmc.edu/projects/seaurchin/) [10], we have identified 4 additional OLF family members, bringing the total to 5 for this sea urchin species. The results of this paper illuminate another line of OLF domain evolution specific to the sea urchin.

As stated above, the original sea urchin OLF family member is amassin [1]. In the present work, we have renamed it amassin-1 to avoid confusion with the three newly identified members

* Corresponding author. Fax: +1 858 534 7317.
E-mail address: vvacquier@ucsd.edu (V.D. Vacquier).
[1] Present address: Catalyst Biosciences, South San Francisco, CA 94080, USA.

0888-7543/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2007.02.009

that share high similarity and are named amassin-2, amassin-3, and amassin-4. The fifth and last newly identified OLF family member is named colmedin, due to its similarity to proteins of the OLF colmedin subfamily [2,3,11–13]. Here we describe the protein features of these new OLF family members. Additionally, we have quantified transcripts for each of the five OLF family members through the complete course of development, from the egg to embryo to larva to metamorphosis. The results are unusual and display a dramatic regulation of OLF transcription, implying important, as yet unknown roles for these proteins in development.

Results

Structural features of the sea urchin OLF family

An assembly of the sea urchin genome provided large genomic contigs to search for OLF domain-containing proteins. Starting from predicted coding regions of similarity to OLF domains and amassin's OLF domain, we isolated the entire coding regions of four additional OLF proteins (Fig. 1). This brings the total to five OLF proteins in this sea urchin species. The newly identified proteins are named amassin-2, amassin-3, amassin-4, and colmedin. Their NCBI accession numbers are, respectively: DQ250734, DQ250735, DQ250736, and DQ250737.

Members of the OLF family in the sea urchin share several features (Fig. 1 and Table 1), most notably an OLF domain located at the C-terminal end. Differentiating the amassins (amassin-1, -2, -3, and -4) from colmedin is the N-terminal half. This is exemplified by their overall sequence identity to amassin-1: 29% amassin-2, 28% amassin-3, 27% amassin-4, and 12% colmedin.

All amassins contain coiled coils as the next structural motif N-terminal to the OLF domain. This region functions to homodimerize amassin-1 and is stabilized by a disulfide bond near the C-terminal end of the coiled coils (B.J. Hillier and V.D. Vacquier, submitted for publication). Every amassin also contains a cysteine at a similar position and in the proper position within the heptad repeat of the coiled coils to reside at the dimer interface [14]. Colmedin, on the other hand, contains a longer region encoding collagen repeats (Gxx) in place of the coiled coils of the amassins (Fig. 1 and Table 1). Colmedin's three regions of collagen repeats encode a total of 249 amino acids, or 83 repeats of Gxx. Surrounding the first repeat region, and preceding the second, are sets of two closely spaced cysteine residues. Since collagen is typically a triple helix, each of these dual cysteine patches could form a stabilizing ring around a colmedin trimer. Although the coiled coils of the amassins and the collagen repeats of colmedin are very different in sequence, both types of domains may function to multimerize OLF protein monomers in a parallel orientation with stabilization by disulfide bonds.

Preceding the coiled-coil segments of all four amassins is a short predicted β region (Fig. 1 and Table 1) [1]. This region contains three cysteines arranged in the consensus $CxCx_9C$ pattern, as found in other OLF family members [15]. This region may act to multimerize amassin-1 to states higher than dimers, but is not necessary for its biological activity of clotting coelomocytes (B.J. Hillier and V.D. Vacquier, submitted for publication). The region N-terminal to the collagen repeats of colmedin (residues 36–180) has no significant similarity to any

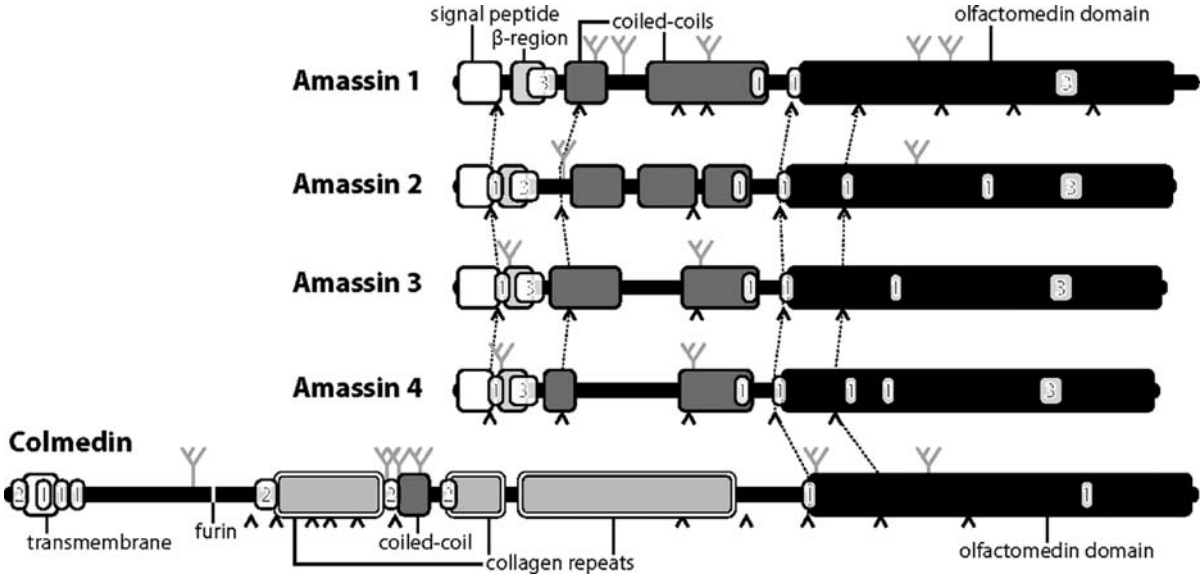

Fig. 1. All five OLF family members share a similar architecture. A schematic outlining the structural features within the OLF proteins is shown. Rounded rectangles filled with a different shades of gray denote the location of each feature and are also labeled above and below. In the amassins, from the N-terminus, there is a signal peptide, a β region, segments of coiled coils, and the OLF domain. Colmedin also contains an OLF domain at the C-terminus, but further N-terminally it contains three segments of collagen repeats, a small coiled-coil segment, a type II transmembrane helix, and a short intracellular segment. Smaller rectangles represent the locations of cysteine residues (within a proximity of 10 residues to one another), and the numeral inside indicates the quantity of cysteines contained. Exon junction locations are denoted by carets (^). Several of these junctions are well conserved within the OLF family members and are shown linked by dashed lines. Branched symbols mark the potential sites of predicted N-linked glycosylation. The furin cleavage site is labeled on the schematic for colmedin.

Table 1
Summary of protein features

Protein	Amino acids	kDa [a]	Signal peptide [b]	TM [c]	N-term. β region	Coiled coils	OLF domain	N-linked glycosylation [d]	*p*I [e]
Amassin-1	495	56.5	SKC^QE 28^29	–	39–59	75–101 130–209	230–477	5 (95, 113, 170, 310, 330)	4.4
Amassin-2	480	53.9	VSA^QC 23^24	–	29–47	78–112 123–160 167–197	221–480	2 (74, 308)	4.2
Amassin-3	474	53.6	ANA^QC 28^29	–	34–52	65–111 152–202	223–472	2 (37, 164)	4.1
Amassin-4	469	52.7	ANA^QC 23^24	–	29–47	60–80 150–198	218–467	2 (32, 159)	4.0
Colmedin	794	85.6	RARR^N [b] 142^143 [b]	13–35	–	262–282 181–250 [f] 295–331 [f] 343–484 [f]	535–790	6 (125, 255, 262, 276, 539, 614)	8.5

[a] Calculated molecular mass in kilodaltons.
[b] Shown are the amino acids in the vicinity of the cleavage site, denoted by a caret. For colmedin, the furin cleavage site is shown.
[c] Location of transmembrane domain.
[d] Number of potential N-linked glycosylation sites, the positions are in parentheses.
[e] Calculated isoelectric point.
[f] For colmedin, denotes the position of the collagen repeat regions.

protein in the NCBI database. This region is predicted by the DisEMBL server [16] to be a region of intrinsic disorder. However, one significant feature of colmedin, a furin cleavage site, RARR, is present in this region at positions 139–142 (Fig. 1 and Table 1).

At the N-terminus of all amassins is a predicted signal peptide of 23–28 residues (Fig. 1 and Table 1). Within 2 amino acids of the predicted cleavage site, all amassins contain a conserved intron–exon junction. A signal peptide can loosely be predicted for colmedin, but its confidence is low because it lacks a signal peptidase site. Instead, it is predicted to be an anchor signal, containing a type II transmembrane domain from residue 13 to 35 (Fig. 1 and Table 1). This would leave a short intracellular N-terminus of only 12 amino acids. The N-terminal region of colmedin is fairly rich in cysteine residues, containing a total of 5. There are 2 cysteines in the short intracellular segment (Cys^{9} and Cys^{11}), 1 within the transmembrane (Cys^{26}), and 2 located just extracellularly (Cys^{38} and Cys^{49}).

OLF domain characteristics

All sea urchin OLF family members contain one OLF domain, invariably at the C-terminal end. A conserved exon junction precedes the OLF domain boundary by the equivalent of 1–5 amino acids and is followed by another conserved junction ~45 amino acids downstream (Figs. 1 and 2). Identity in the OLF domain is higher than in the protein as a whole. Identity to amassin-1 is 35% for amassin-2, 34% for amassin-3, 33% for amassin-4, and 25% for colmedin. Only 40 residues in the OLF domain of ~250 amino acids are absolutely conserved between all members (Fig. 2). The two longest stretches of high conservation are from amassin-1 residue 354 to 366 (DxAVDExGLWxIY), of which 10 of 13 are identical, and from residue 451 to 466 (LxYNPRDxxLY), of which 8 of 11 are identical.

The position of a cysteine residue is absolutely conserved (amassin-1 Cys^{227}; Fig. 2), just preceding, or coincident with, the defined OLF boundary. This cysteine most likely forms an intramolecular disulfide bond with the central cysteine of a patch of three in the amassins (amassin-1 Cys^{407}) or the only other cysteine in colmedin, located at the same position. The amassins likely form an additional intramolecular disulfide between the peripheral cysteines of this patch, as is the case for amassin-1 (Cys^{403} bonded to Cys^{412}; B.J. Hillier and V.D. Vacquier, submitted for publication). Any additional cysteines beyond those already mentioned typically occur in pairs in OLF domains and so are most likely disulfide bonded. Amassin-2 and -4 contain two additional cysteines—predicted to disulfide bond—whereas amassin-3 contains only one. Amassin-3 is very similar to amassin-4, sharing 87% identical amino acids. Even so, instead of the second matching cysteine found in amassin-4, amassin-3 contains an arginine, which is also present in amassin-1 and -2. Amassin-3 may therefore contain an unbonded, free cysteine.

Diversity of the OLF family

A phylogenetic tree was constructed for OLF family proteins from a dozen animal species by aligning their OLF domains (Fig. 3). Major groupings correspond to those obtained in a recent work by Zeng et al. [17]. The subgroups, labeled I–VII, represent these dominant members: (I) noelins and optimedin, (II) latrophilins, (III) myocilins, (IV) photomedins, (V) olfactomedin and tiarin, (VI) colmedins, and (VII) olfactomedins-like. However, previous phylogenetic work was inconsistent in assigning amassin-1 to any one group. In one instance it was assigned to subgroup VII, but in another tree, it was placed in subgroup V [17]. Here, we show that the four amassins identified from the sea urchin are more similar to themselves than to any other group. Thus, the sea urchin amassins form an eighth distinctive OLF subgroup.

By far, the most diverse subgroup is the colmedins (named by conjoining collagen and olfactomedin) [2], subgroup VI

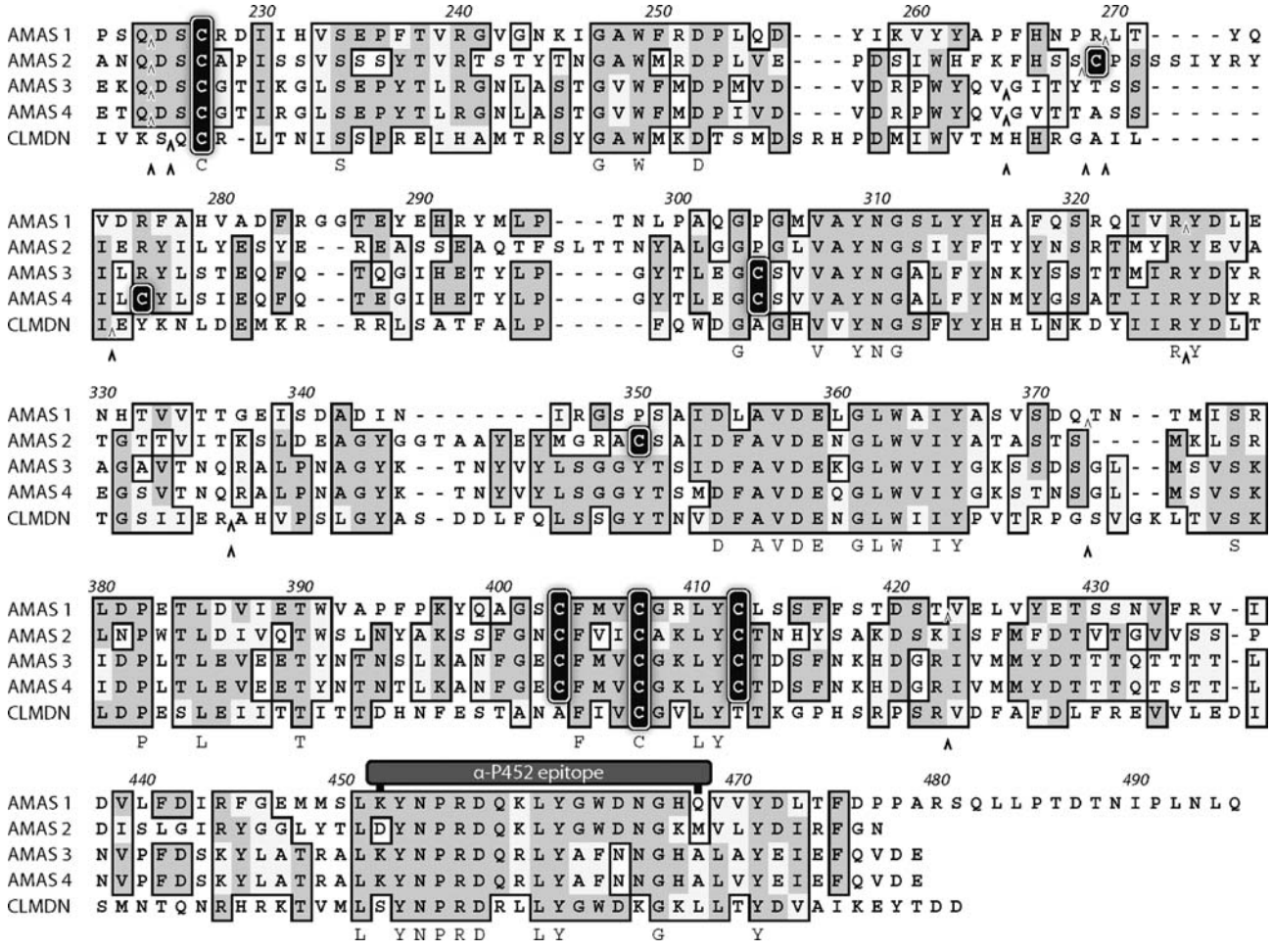

Fig. 2. Conservation in the OLF domain. Shown is a multiple sequence alignment of the OLF domains from all five family members. AMAS 1, amassin-1; AMAS 2, amassin-2; AMAS 3 amassin-3; AMAS 4, amassin-4; CLMDN, colmedin. Locations of cysteines are highlighted by black rectangles. Amino acid identities are dark shaded, similarities are light shaded, and boxes enclose both. The 40 amino acids that are absolutely conserved are shown below the colmedin sequence. The numbering is for the amassin-1 protein sequence. Carets (^) mark the locations of exon junctions. The anti-P452 peptide epitope (amassin-1 residues 452–468) is indicated.

(Fig. 3). To date, colmedins are represented by members from four animal phyla: *Nematoda, Arthropoda, Chordata,* and now *Echinodermata.* Sea urchin colmedin shares the characteristic presence of a short intracellular N-terminal portion, followed by a transmembrane domain and collagen repeats. This is the only subgroup representing both protostomes and deuterostomes; all other subgroups have only deuterostome members. As such, it seems that this is the most primitive subgroup. Further diversification of OLF members has occurred only within the deuterostomes. With the exception of the colmedins, the only nonchordate deuterostome known to contain an OLF domain protein is the sea urchin. The sea urchin is intermediate to the diversity of the vertebrates, being represented in two of the eight phylogenetic subgroups. Human sequences are present in seven of the eight OLF subgroups.

OLF presence in coelomocytes

A highly conserved region among all sea urchin OLF family members (amassin-1 residues 452–468) was used as a peptide epitope to raise a polyclonal antibody (anti-P452). Amassin-2 contains 14 of 17 identical residues in this region, amassin-3 12 of 17, amassin-4 12 of 14, and colmedin 11 of 17 (Fig. 2). Because of the high identity, the antibody would be expected to react with all members of the sea urchin OLF family.

Indeed, in coelomocyte lysates analyzed by Western blotting, there are multiple reactive bands when probed with anti-P452 (Fig. 4). In contrast, when only the cell-free coelomic fluid is analyzed, only a single reactive band running at ~80 kDa is evident. This band is determined to be amassin-1 due to reactivity with its specific polyclonal antibody at an identical relative mobility (M_r). Washed coelomocytes are lacking in amassin-1. This is evidenced by faint reactivity at 80 kDa when probed with highly specific anti-amassin-1 and the lack of a prominent band at that mobility when probed with anti-P452. However, amassin-1 is abundant in lysates produced from clotted coelomocytes, as seen with both antibodies.

By inference, the other reacting bands present in washed coelomocytes can be ascribed to the remaining OLF family members. Colmedin has a predicted molecular weight of 85.6 kDa (Table 1). This is in close agreement with the reacting

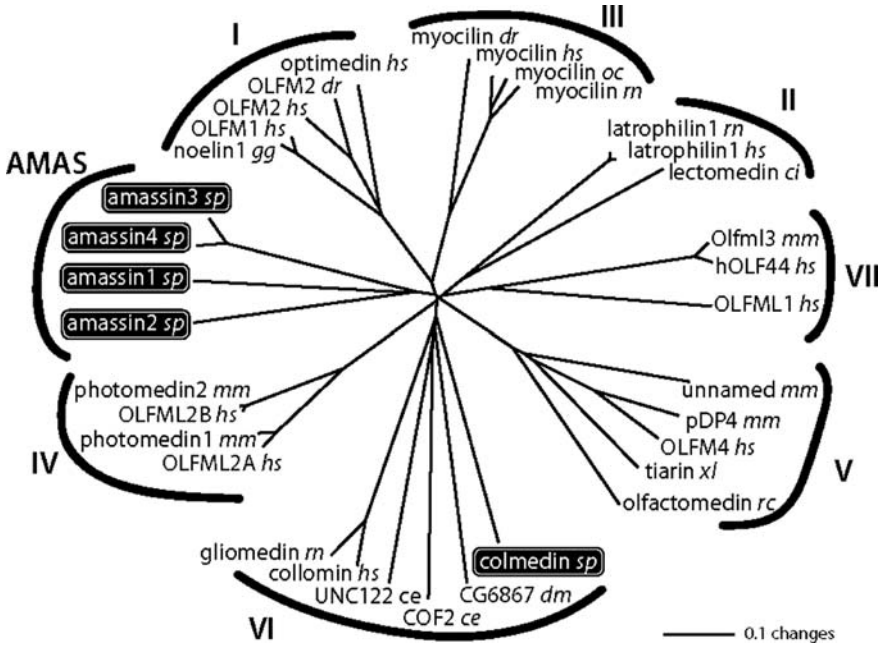

Fig. 3. Amassins-1, -2, -3, and -4 form their own subgroup of OLF proteins. An unrooted phylogenetic tree made from an alignment of the OLF domains is shown. Seven previously described [17] subgroups are apparent (labeled I–VII), with one additional subgroup containing the four amassins from sea urchin. Sea urchin colmedin is found in subgroup VI. Following the protein name, species names are abbreviated as follows: sp, *S. purpuratus;* hs, *Homo sapiens;* ce, *C. elegans;* dm, *D. melanogaster;* rn, *Rattus norvegicus;* ci, *Ciona intestinalis;* mm, *Mus musculus;* gg, *Gallus gallus;* dr, *Danio rerio;* oc, *Oryctolagus cuniculus;* xl, *Xenopus laevis;* rc, *Rana catesbeiana.* NCBI accession numbers of all proteins used in this tree can be found in the supplemental data (Table S2).

band migrating at ~90 kDa. Collagen repeat-containing proteins are typically present in a triple helix and can be cross-linked through lysine residues by a lysyl-oxidase-type enzyme [18–20]. A covalently cross-linked colmedin trimer would have a calculated molecular weight of 257 kDa, similar to the reactive band migrating at ~240 kDa in washed coelomocytes. Indeed, a protein similar to human lysyl-oxidase-like 2 was cloned from coelomocytes during this study (NCBI Accession No. DQ250738), leaving the occurrence of a covalently linked trimer possible. Interestingly, when coelomocytes are induced to clot, the presumed colmedin bands are no longer present. Instead, there is a band at ~190 kDa. This may be due to cleavage at the furin site (RARR)—removing the first 142 residues of each subunit in the covalently linked trimer—which would reduce its calculated molecular mass to 208 kDa. Monomeric colmedin would decrease to a calculated molecular mass of 69.4 kDa and be obscured by the other intense bands in that region.

There remain two discernible bands, one at ~60 kDa and the other at ~70 kDa (Fig. 4). For certain they are not amassin-1, as they are unreactive with its specific antibody, antiamassin-1. The protein sequences of amassin-3 and -4 are very similar and may migrate together as one band on gels. For these remaining amassins, there is not sufficient experimental evidence at this time to assign them to a reactive band on a Western blot.

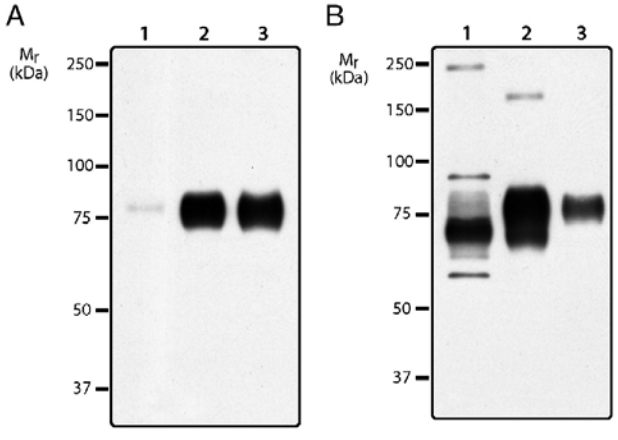

Fig. 4. OLF family proteins are found in coelomocytes. Lanes were loaded identically in both (A) and (B). Washed coelomocytes (lane 1), clotted coelomocytes (lane 2), and cell-free coelomic fluid (lane 3) were analyzed by immunoblots probed with (A) anti-amassin-1 or (B) anti-P452. Molecular mass standards, shown on the left, are in kDa.

Regulation of OLF family members throughout larval stages

The regulation of OLF family transcript levels at 11 developmental stages was investigated by real-time quantitative PCR (RTQ-PCR). Samples were taken at the stages unfertilized egg, 2-day-old late gastrula, 4-day-old late prism, 6-day-old early pluteus, 10-day-old pluteus, 15-day-old 4-arm stage, 20-day-old 6-arm stage, 25-day-old 8-arm stage, 30-day-old late larvae, and 35-day-old metamorphs and from an entire small adult. Previous work on sea urchins has shown that the ubiquitin transcript level remains constant through early development [21,22]. We found that the copy number of ubiquitin transcripts remains fairly constant throughout development at ~10^4 to 10^5 copies per 4.2 ng total RNA, until reaching the adult, which contains 1.9×10^3 copies (Fig. 5). The copy number of ubiquitin

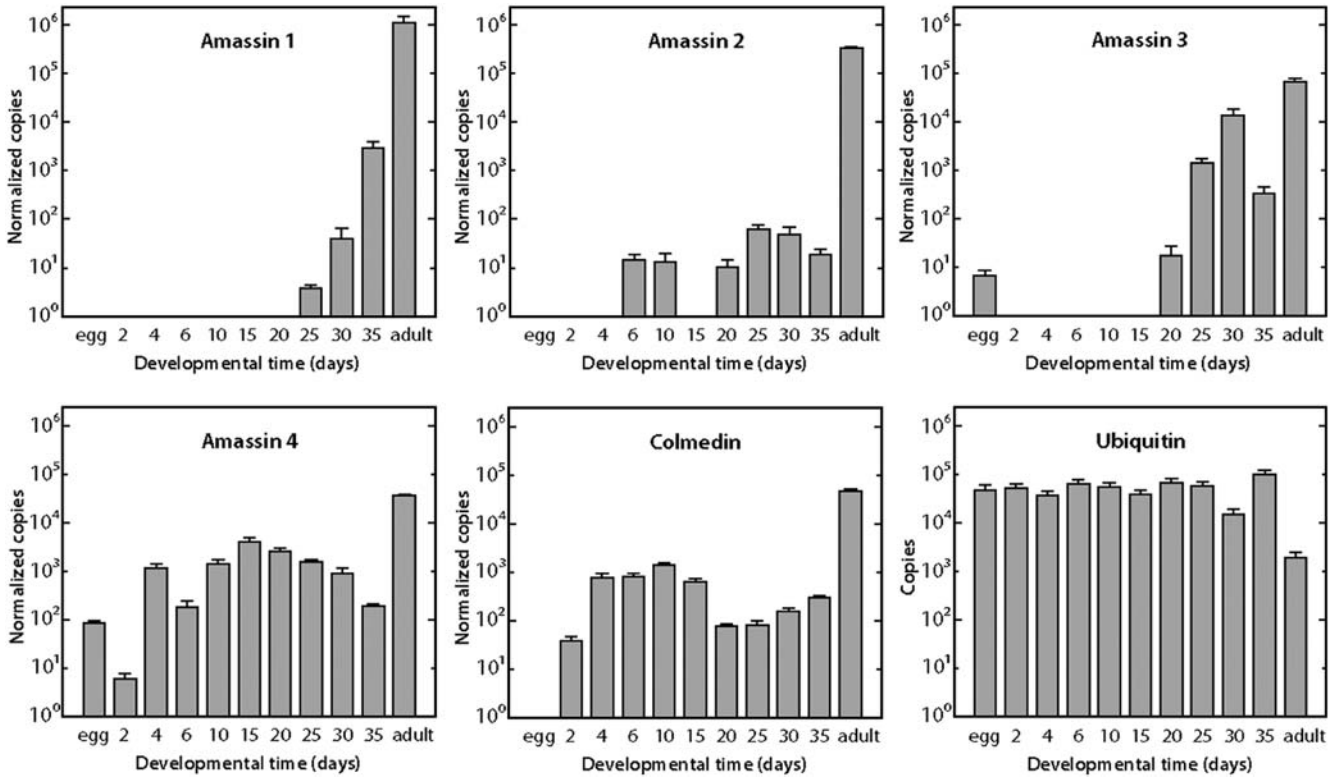

Fig. 5. Regulation of OLF transcript levels throughout all stages of development. Transcript copy numbers (per 4.2 ng total RNA) at each of 11 developmental stages for all five OLF family members are shown. Obtained by RTQ-PCR, results are the averages of three replicates, standard error is shown. Copy numbers were calculated from a standard curve generated in the same run and were normalized to the level of ubiquitin transcript. The transcripts of all six genes were analyzed from the same pools of cDNA.

transcripts was used for normalization of OLF family transcript levels.

The amassin transcripts are developmentally regulated. A dramatic regulation is observed for amassin-1 expression (Fig. 5). No transcripts are detectable until the 25-day-old larvae—a stage at which the developing rudiment is visible. By the adult, its levels increase 1,000,000-fold to 1.1×10^6 copies. Amassin-2 is also highly enriched in the adult, rising to 3.3×10^5 copies. But amassin-2 is also found at very low levels in earlier stages of development (from 10 to 62 copies), with undetectable levels until 6 days, and cannot be detected at 15 days. Amassin-3 is primarily expressed in late development and in the adult, with a slight detectable expression in the egg (6.9 copies). Although amassin-4 is very similar to amassin-3 in sequence, regulation of its gene is quite different. Unique among all OLF members, amassin-4 is detectable in all stages of development. In general, its levels increase through development until 15 days, with a gradual loss to metamorphosis, and again a high level in the adult.

The colmedin transcript is also regulated throughout development (Fig. 5). Although absent in the unfertilized egg, it is detectable at all other stages. Its levels peak at 10 days, followed by a reduction at 20 days, and then an increase through metamorphosis.

Proteins were also analyzed from the same developmental samples by Western analysis. Interestingly, the amassin-1 protein remained undetectable throughout development, rendering no reactivity with the specific anti-amassin-1 IgG (data not shown). With the anti-P452 antibody, which should recognize all OLF family members, several bands are visible (Fig. 6). The ~240-kDa band is present in all stages of development, with lower intensity in the egg and 35-day metamorphosis samples. This band is hypothesized to react with a cross-linked trimer of colmedin, described previously. A dramatic increase in the presence of the ~60-kDa band occurs beginning at 25 days and continues through metamorphosis. Upon overexposure, faint, possibly doublet bands are also visible at ~70 kDa in most developmental stages. It is undetectable in the egg and 10-day samples.

Discussion

Common structure of sea urchin OLF proteins

There are five OLF domain-containing proteins in this sea urchin species, four amassins and one colmedin. Among the four amassins, the overall structural elements are remarkably conserved. Beginning from the N-terminus, all amassins have a signal peptide that would direct their secretion to the extracellular space, as is the case for amassin-1, which is found in the coelomic fluid. Having no doubt arisen by gene duplications, the four amassins have a common exon junction close to the C-terminal end of the signal peptide and another farther downstream following the next shared protein feature

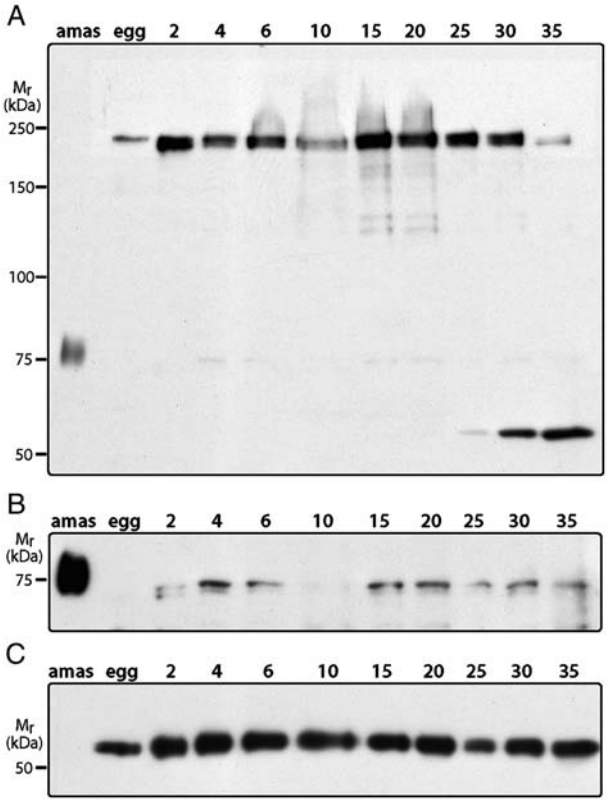

Fig. 6. OLF family proteins are present in development. Protein lysates from each developmental stage were resolved by SDS–PAGE on a 6% gel and immunoblotted with anti-P452 that should react with all OLF proteins. (A) The 240-kDa band is present throughout development, while the 60-kDa band is present only beginning at day 25. (B) In an overexposure of the blot shown in (A), a reactive band becomes visible at 70 kDa (possibly a doublet) in the majority of stages except for egg and 10 days. (C) Immunoblotting with anti-α-tubulin controls for equal protein load. The lane labeled amas contained 10 ng of purified amassin-1.

(Fig. 1). This next feature is a short predicted β region that partially encompasses three cysteines. This region is believed to be involved in forming higher multimeric states in amassin-1. Farther C-terminal, there are segments of coiled coils that end with a cysteine in the proper location to form a dimerizing disulfide. The OLF domains comprise the C-terminal half of all four amassins (Fig. 1).

With the exception of the latrophilins, a similar domain arrangement can be found in all OLF proteins. These similarities include the presence of signal peptides that can be found in every subgroup aside from the colmedins, subgroup VI. The conservatively spaced cysteines, assigned to be part of the N-terminal β region, are found in a smaller subset, which excludes the colmedins (subgroup VI) and the olfactomedin-like proteins (subgroup VII). But common to every OLF group is a helical multimerizing motif located N-terminal to the OLF domain. In most OLF proteins these are dimeric coiled-coil regions, while in the colmedins they are trimeric collagen repeat regions. This would result in a common structure of multiple OLF domains being displayed together as part of one larger molecule. Because it is the OLF domain that is the functional portion of these proteins that bind their target proteins via protein–protein interaction, the result of closely spaced multiple OLF domains would be an increase in avidity. This could be the reason for the marked conservation of structure of OLF proteins.

The OLF domain itself displays further similarities among all five sea urchin OLF family members. Exon junctions are well conserved, supporting their evolution by gene duplication followed by functional specialization (Fig. 1). Cysteine positions, probably involved in intradomain disulfides, are also conserved.

OLF presence and function in coelomocyte clotting

Overall identity between the OLF domains was fairly low, ranging from 25 to 35% identity to amassin-1, with only two stretches of high percentages of identity among all proteins. One of these stretches was utilized to produce an antibody that recognizes all OLF family members. Multiple reactive bands resulted from sea urchin protein preparations probed with this antibody and logical inferences were made in assigning probable OLF proteins to specific bands. Of note is the assignment of the bands at 90 and 240 kDa in coelomocyte preparations. The calculated molecular masses of 85.6 kDa for monomeric colmedin and 257 kDa for cross-linked trimeric colmedin led to the assignment of these two bands as the colmedins. Attempts to identify the antibody-reacting bands by mass spectrometry were unsuccessful.

There are three facts supporting the hypothesis that colmedin is a likely candidate to be the binding partner of the coelomocyte clot mediator, amassin-1. First, heterotypic OLF–OLF interactions have previously been demonstrated for other proteins, for example, myocilin and optimedin [9]. Second, like myocilin and optimedin, the isoelectric points of amassin-1 and colmedin are separated by at least four units (Table 1), which could be an indicator of complementarily charged interaction surfaces. Third, colmedin contains a transmembrane domain, providing a membrane anchor and extending its OLF domains away from the cell surface, where it could interact with OLF domains of amassin-1 and bind coelomocytes together in clots. Additionally, the colmedin-reactive band could be truncated, possibly at the furin cleavage site, following clot formation (Fig. 4). Coelomocytes remain tightly bound during this stage, and secondary cell-adhesion events may be occurring. Interestingly, all the OLF family proteins identified in this work are expressed in the coelomocytes, suggesting that they might be involved with the robust innate immune system of this animal [23].

OLF family gene regulation during development

Because many OLF family proteins are developmentally regulated, especially in neuronal tissue, the appearance of their transcripts during development was investigated. The sea urchin under investigation, *S. purpuratus,* goes through indirect larval development. The free-living larva bears no resemblance to the developing adult urchin body.

Amassin-1 gene expression displays a rather unusual developmental profile. It is not detectable until well into larval development, first appearing at 25 days when the adult-forming

rudiment becomes visible. To our knowledge, this is the latest developmental gene activation known for this species. Beginning at 25 days, amassin-1 expression increases dramatically, climbing to a 1,000,000-fold increase in the young adult. The mechanism of its transcriptional regulation would be important to study. Perhaps its expression is regulated in introns or 3′ to the gene, such as in the OLF family member UNC-122 [2]. The amassin-1 protein is undetectable at all stages until the adult, where it is found in the coelomic fluid and expressed by coelomocytes. These observations indicate that amassin-1 is an adult-functioning protein and most likely plays no developmental role until metamorphosis to the adult.

The transcriptions of the other amassin genes are also highly regulated. The lowest expressing amassin, amassin-2, is present in small amounts in the latter half of development and then is plentiful in the adult. Amassin-3 displays a marked increase beginning at 20 days, again coinciding with rudiment formation. An OLF family protein that matches this profile is detectable; its presence begins 5 days later at 25 days and it migrates at 60 kDa on gels. This band may correspond to amassin-3. The final amassin, amassin-4, displays a more constant level of expression and is found throughout development. The OLF family member running at 70 kDa may correspond to this protein, as it is faintly detectable in most stages and its sequence corresponds to that molecular mass. Regulation of colmedin follows a cyclic profile, with an early peak at 10 days and then a later peak in the adult. The 240-kDa band has been assigned to this protein, and it is found throughout development. Functionally, colmedin may follow that of other related proteins, UNC-122 and gliomedin, with neurogenic roles important during development [2,3]. It is important to note that the same pools of larval cDNA were used in the analysis of all genes in this work.

In summary, these data show that the genes coding for OLF family members in sea urchin are highly transcriptionally regulated during larval development, which most probably is a reflection of their as yet unknown functional importance. Neurological roles for the OLF family members in the sea urchin should be seriously considered, as transcripts for all members can be easily amplified from cDNA prepared from adult radial nerve tissue (data not shown), although no other tissues were analyzed. Further research should reveal interesting functions for sea urchin OLF proteins.

Materials and methods

Immunoblots and antibodies

Immunoblots were performed as described previously with the anti-amassin-1-specific antibody [1]. Anti-amassin-1 was used at a 1:25,000 dilution. The monoclonal anti-α-tubulin (Sigma T5168; St. Louis, MO, USA) was used in immunoblots at a dilution of 1:20,000.

An 18-mer peptide was synthesized (Bio-Synthesis, Lewisville, TX, USA) corresponding to amassin-1 residues 452–468 with the addition of one cysteine residue at the N-terminus (amine-CKYNPRDQKLYGWDNGHQ-amide). The peptide was coupled to maleimide-activated mariculture keyhole limpet hemocyanin (Pierce Chemical Co., Rockford, IL, USA) and used to raise rabbit antiserum (Strategic Biosolutions, Newark, DE, USA). The peptide was also linked to a solid support (SulfoLink Coupling Gel; Pierce Chemical Co.) following the manufacturer's instructions. This peptide-linked gel was used to affinity purify the antiserum. The antibodies (anti-P452) were concentrated to 0.5 mg/ml with the addition of 10 mg/ml BSA. Anti-P452 was used at a dilution of 1:250 for Western blots.

Cloning full-length cDNAs

The publicly available assemblies of the *S. purpuratus* genome at the Baylor College of Medicine Human Genome Sequencing Center (http://www.hgsc.bcm.tmc.edu/blast/) contigs dated November 23, 2004, were searched using tBLASTn. Queries were submitted in an exhaustive search using most well-characterized OLF domains currently recognized in the Pfam database as search models. The genomic search results were manually scored for confidence by the presence of appropriately spaced conserved sequence features typically present in an OLF domain. Few absolutely conserved stretches of residues actually exist between all OLF domains, but the tripeptide GLW present midway in the domain at amassin residues 361–363 and the highly conserved cysteine (Cys^{407}) located ~44 residues later were most useful. Four unique OLF domain-containing genomic sequences (other than amassin) were identified after the positive contigs were tabulated and redundancies removed.

PCR primers that would amplify hypothetical exon regions were designed and synthesized. Total RNA was prepared from a whole adult animal [1] and from total coelomocytes with the Trizol reagent (Invitrogen, Carlsbad, CA, USA). First-strand cDNA was synthesized by priming with either random decamers or oligo$(dT)_{12–18}$ (Invitrogen) and extension with SuperScript III (Invitrogen) following the manufacturer's protocol. PCR products were cloned by TOPO-TA (Invitrogen) and plasmids sequenced.

Extension of the clones farther terminal (5′ and 3′) was accomplished by a combination of two techniques. First, contigs were translated in silico in all possible reading frames and examined for the presence of structural features typically found in OLF family members, namely coiled coils or collagen repeats. Once these regions were found, additional primers were designed to amplify a larger region of cDNA. Second, complete coding regions were identified using 5′ and 3′ rapid amplification of cDNA ends (Ambion, Austin, TX, USA). Additional primers that would anneal in the newly identified untranslated regions were used in combination with a proofreading polymerase (Platinum *Pfx* polymerase; Invitrogen) to verify complete sequences. Using the complete cDNA sequences as queries for a BLAST search against the genomic database, other contigs that individually spanned one or more exons were identified. Exon boundaries were identified by manual comparison of the cDNA and genomic contig sequences with conformation to the typical AG–exon–GT rule [24,25], the results of which are tabulated in the supplemental data (Table S1]).

Primary sequence analysis

The locations of the OLF domain boundaries were obtained from the Pfam database [26]. Coiled-coil predictions were made with the Coils server [14] using the MTIDK matrix, with no weighting, and a window size of 21. Continuous regions with a probability score >0.5 were considered likely. The presence and location of signal peptides were predicted with SignalP [27] and a transmembrane domain identified with the TMHMM Server v2 [28]. The short N-terminal predicted β region was identified by alignment with the amassin sequence. The furin cleavage site and location of predicted N-linked glycosylation sites were found with assistance from the MacVector (Accelrys) computer program.

Phylogenetic analysis

A multiple sequence alignment was constructed from the OLF domains of 34 proteins with the default settings of the ClustalW function of MacVector. NCBI accession numbers of the proteins used can be found in the supplementary materials (Table S2). The highly conserved cysteine located just N-terminal to the currently defined OLF domain boundary (positions synonymous to amassin Cys^{227}) and five additional residues beyond were included. An unrooted distance-based neighbor-joining tree was created from this alignment with PAUP* [29].

Coelomocyte preparations

Washed clot-inhibited coelomocytes were prepared as previously described [1]. Washed cells (0.5 ml at 2×10^6 cells/ml) were pelleted by centrifugation at 500 g for 4 min. Cells were lysed by resuspension into 100 μl 2× Laemmli sample buffer and boiled 5 min, and 10 μl/lane was resolved by SDS–PAGE on an 8% gel. Clotted cell samples from the same batch of coelomocytes were prepared similarly. To the washed cells were added 250 ng purified amassin and 10 mM $CaCl_2$ with rotation end-over-end for 5 min to induce complete clotting. Further processing of the clotted samples was identical to that of the washed-alone samples.

Culture of larvae and preparation of their biomolecules

Sea urchin larvae were reared essentially as described [30]. Larvae were incubated at 14°C, with no added agitation. The medium, 0.45-μm-filtered natural seawater, was replaced every 2 days by gentle capture of larvae on a 20-μm mesh of Nytex and resuspension in fresh seawater. Separate cultures of the unicellular algae, *Rhodomonas,* were maintained and added to the medium after every seawater change.

Biomolecules were prepared at various stages of their developmental growth by taking samples of the larvae after a seawater change, but before diluting the culture and adding algae. Sufficient larvae were pelleted and washed with filtered seawater to give a pellet of ~100 μl. Protein lysates from each larval stage were prepared by resuspension in 200 μl 2× Laemmli buffer with the aid of a micropestle and boiling for 5 min. The protein lysates were clarified by centrifugation for 5 min at 14,000 g and stored at −20°C until Western analysis. From identical samples, total RNA from each stage was prepared. RNA was solubilized by resuspension in 1 ml of Trizol reagent (Invitrogen), and total RNA was purified according to the standard procedure. Synthesis of cDNA was performed with 1 μg of total RNA using the SuperScript III polymerase and oligo$(dT)_{12-18}$ primers by standard procedures (Invitrogen), resulting in a final volume of 24 μl.

RTQ-PCR

Primers were designed to amplify the specific genes as follows (5′ to 3′): *AMAS1,* CAACTGCGTGAAGAACTG and AGACGGAGGCATAGATTGC (622-bp product); *AMAS2,* GAGGACCTGGCACAATGCGA and CATTCTCATCCACAGCAAAG (544-bp product); *AMAS3,* CGATCAATATATCCTTGACC and GTAAGATAGACGAAGTATA (356-bp product); *AMAS4,* TGATCAATTTATCATCGTCA and ATAGGATAGACGAAGCTGT (356-bp product); *CLMDN,* GTTTGGACCGAAAGGAATC and GGATGATAGTTGGAACAAGTCG (655-bp product); *UBIQ,* CGAGTATTTGCCAGATGAACCC and ATTGGATTTTTTGCCCCTGC (233-bp product; NCBI Accession No. M61772). These primer pairs were used to PCR amplify cDNA prepared from whole adult animal. The products were gel purified (Qiagen, Valencia, CA, USA), and concentrations were determined with the PicoGreen reagent (Invitrogen). Standard template quantities, expressed as number of copies, were prepared from these products by serial dilution. RTQ-PCR was performed on a MX3000P instrument with the Brilliant SYBR Green reagent (Stratagene, La Jolla, CA, USA) in a 96-well format. A passive reference dye (monitored on the ROX channel) was included in all samples. Template cDNA from each unknown (staged larvae) was added at 0.1 μl/reaction. A standard curve, run on the same plate as unknowns, was generated by determining the threshold cycle of each dilution of standard. Copy numbers of unknowns, obtained in triplicate and analyzed individually, were then determined from this curve. Variation between stages was corrected by normalization to the determined copy number of ubiquitin. The product specificity of each RTQ-PCR was assayed by a dissociation (melting) curve analysis and by agarose gel electrophoresis.

Acknowledgments

We thank Russell F. Doolittle, Jonathan Flowers, and Jay McCarren for helpful discussions; Eddie Kisfaludy for assistance collecting adult urchins; Elisa Maldonado and Michael Latz for their help and lab space to rear larvae; and Robert Burke for adult radial nerve RNA. This work was supported by the Graduate Department at Scripps Institution of Oceanography.

Appendix A. Supplementary data

Supplementary data for this article may be found on ScienceDirect, at doi:10.1016/j.ygeno.2007.02.009.

References

[1] B.J. Hillier, V.D. Vacquier, Amassin, an olfactomedin protein, mediates the massive intercellular adhesion of sea urchin coelomocytes, J. Cell Biol. 160 (2003) 597–604.
[2] P.M. Loria, J. Hodgkin, O. Hobert, A conserved postsynaptic transmembrane protein affecting neuromuscular signaling in Caenorhabditis elegans, J. Neurosci. 24 (2004) 2191–2201.
[3] Y. Eshed, et al., Gliomedin mediates Schwann cell–axon interaction and the molecular assembly of the nodes of Ranvier, Neuron 47 (2005) 215–229.
[4] M.P. Fautsch, C.K. Bahler, D.J. Jewison, D.H. Johnson, Recombinant TIGR/MYOC increases outflow resistance in the human anterior segment, Invest. Ophthalmol. Visual Sci. 41 (2000) 4163–4168.
[5] K.M. Hardy, E.A. Hoffman, P. Gonzalez, B.S. McKay, W.D. Stamer, Extracellular trafficking of myocilin in human trabecular meshwork cells, J. Biol. Chem. 280 (2005) 28917–28926.
[6] T.D. Nguyen, et al., Gene structure and properties of TIGR, an olfactomedin-related glycoprotein cloned from glucocorticoid-induced trabecular meshwork cells, J. Biol. Chem. 273 (1998) 6341–6350.
[7] K. Ray, A. Mukhopadhyay, M. Acharya, Recent advances in molecular genetics of glaucoma, Mol. Cell. Biochem. 253 (2003) 223–231.
[8] E.M. Stone, et al., Identification of a gene that causes primary open angle glaucoma, Science 275 (1997) 668–670.
[9] M. Torrado, R. Trivedi, R. Zinovieva, I. Karavanova, S.I. Tomarev, Optimedin: a novel olfactomedin-related protein that interacts with myocilin, Hum. Mol. Genet. 11 (2002) 1291–1301.
[10] E. Sodergren, et al., The genome of the sea urchin Strongylocentrotus purpuratus, Science 314 (2006) 941–952.
[11] Q. Cao, et al., Expression of an olfactomedin-related gene in rat hair follicular papilla cells, J. Invest. Dermatol. 125 (2005) 24–33.
[12] C.W. Franzke, P. Bruckner, L. Bruckner-Tuderman, Collagenous transmembrane proteins: recent insights into biology and pathology, J. Biol. Chem. 280 (2005) 4005–4008.
[13] C.R. Graveel, S.R. Harkins-Perry, L.G. Acevedo, P.J. Farnham, Identification and characterization of CRG-L2, a new marker for liver tumor development, Oncogene 22 (2003) 1730–1736.
[14] A. Lupas, M. Van Dyke, J. Stock, Predicting coiled coils from protein sequences, Science 252 (1991) 1162–1164.
[15] Y. Furutani, et al., Identification and characterization of photomedins: novel olfactomedin-domain-containing proteins with chondroitin sulphate-E-binding activity, Biochem. J. 389 (2005) 675–684.
[16] R. Linding, et al., Protein disorder prediction: implications for structural proteomics, Structure 11 (2003) 1453–1459.
[17] L.C. Zeng, Z.G. Han, W.J. Ma, Elucidation of subfamily segregation and intramolecular coevolution of the olfactomedin-like proteins by comprehensive phylogenetic analysis and gene expression pattern assessment, FEBS Lett. 579 (2005) 5443–5453.
[18] C. Jourdan-Le Saux, et al., The LOXL2 gene encodes a new lysyl oxidase-like protein and is expressed at high levels in reproductive tissues, J. Biol. Chem. 274 (1999) 12939–12944.
[19] J.M. Maki, H. Tikkanen, K.I. Kivirikko, Cloning and characterization of a fifth human lysyl oxidase isoenzyme: the third member of the lysyl oxidase-related subfamily with four scavenger receptor cysteine-rich domains, Matrix Biol. 20 (2001) 493–496.

[20] J. Molnar, et al., Structural and functional diversity of lysyl oxidase and the LOX-like proteins, Biochim. Biophys. Acta 1647 (2003) 220–224.
[21] M. Nemer, E. Rondinelli, D. Infante, A.A. Infante, Polyubiquitin RNA characteristics and conditional induction in sea urchin embryos, Dev. Biol. 145 (1991) 255–265.
[22] A. Ransick, J.P. Rast, T. Minokawa, C. Calestani, E.H. Davidson, New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization, Dev. Biol. 246 (2002) 132–147.
[23] J.P. Rast, L.C. Smith, M. Loza-Coll, T. Hibino, G.W. Litman, Genomic insights into the immune system of the sea urchin, Science 314 (2006) 952–956.
[24] S.M. Mount, A catalogue of splice junction sequences, Nucleic. Acids Res. 10 (1982) 459–472.
[25] M.B. Shapiro, P. Senapathy, RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression, Nucleic. Acids Res. 15 (1987) 7155–7174.
[26] A. Bateman, et al., The Pfam protein families database, Nucleic. Acids Res. 30 (2002) 276–280.
[27] H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, Protein Eng. 10 (1997) 1–6.
[28] E.L. Sonnhammer, G. von Heijne, A. Krogh, A hidden Markov model for predicting transmembrane helices in protein sequences, Proc. Int. Conf. Intell. Syst. Mol. Biol. 6 (1998) 175–182.
[29] D.L. Swofford, PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods), Sinauer, Sunderland, MA, 1998.
[30] G.A. Wray, C. Kitazawa, B. Miner, Culture of echinoderm larvae through metamorphosis, Methods Cell Biol. 74 (2004) 75–86.