The Proteinase PAPP-A has Deep Evolutionary Roots Outside of the IGF System

Caroline M. N. Kjeldsen [1], Claus Oxvig [1,*]

[1]Department of Molecular Biology and Genetics, Aarhus University, Aarhus DK-8000 C, Denmark

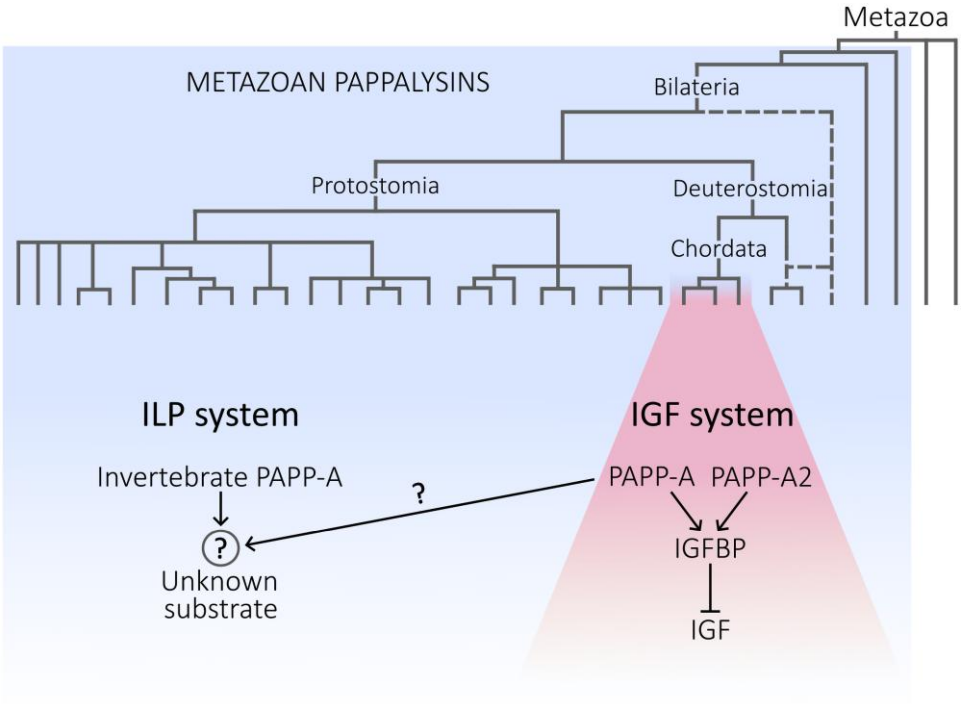
*Corresponding author: E-mail: co@mbg.au.dk.

Accepted: February 28, 2025

Abstract

The animal pappalysin metalloproteinases, PAPP-A and PAPP-A2, are highly specific regulatory enzymes of the insulin-like growth factor (IGF) system. Cleavage of their only known substrates, a subset of IGF binding proteins (IGFBPs), releases bioactive IGFI and IGFII, thus promoting IGF signaling. Stanniocalcin-1 and -2 (STC1 and STC2) are potent pappalysin inhibitors, completing the STC-PAPP-A-IGFBP-IGF axis. Utilizing homology searches and phylogenetic analyses, we examined the occurrence of pappalysins in the animal kingdom and their functional conservation. This revealed the extensive presence of pappalysins across metazoans, as well as the presence of 3 pappalysins: PAPP-A, PAPP-A2, and a third group of invertebrate pappalysins, which we name invertebrate PAPP-A (invPAPP-A). We show that PAPP-A and PAPP-A2 arose by duplication during early vertebrate evolution. Despite significant evolutionary distance, the domain architecture of the metazoan pappalysins is completely conserved, and several functional domains and motifs are highly conserved across all pappalysins. However, invPAPP-A exists outside the context of IGFBPs, suggesting that the animal pappalysins may have substrates beyond the IGFBPs for PAPP-A and PAPP-A2 that remain to be discovered. Since PAPP-A is an emerging drug target, it is important to understand potential involvement in regulatory systems other than the IGF system, which might be affected upon targeting of PAPP-A.

Graphical abstract



Key words: pappalysin phylogeny, IGF, proteinase, IGFBP, PAPP-A, PAPP-A2.

Significance

The pappalysin metalloproteinases, PAPP-A and PAPP-A2, are known regulatory enzymes in the IGF system, cleaving IGFBPs with high specificity to release bioactive IGF. However, their metazoan distribution and phylogeny, along with the possibility of substrates existing outside the IGF system, remains unexplored. Pappalysins were identified nearly throughout the animal kingdom, with PAPP-A and PAPP-A2 being specific to vertebrates. Invertebrate PAPP-A (invPAPP-A) lies outside the IGF system yet shares a globally conserved sequence, highly conserved motifs, and all domains with PAPP-A/PAPP-A2. These findings suggest that metazoan pappalysins have additional, unknown substrates that are not IGFBPs, expanding their potential regulatory roles beyond the IGF system.

Introduction

The large multidomain proteinase, pregnancy-associated plasma protein-A (PAPP-A, pappalysin-1), was discovered in the plasma of pregnant women where it is abundantly secreted from the syncytiotrophoblast of the human placenta. PAPP-A was later found to be synthesized ubiquitously in multiple other cell types and tissues independent of pregnancy. Discovery of its proteolytic activity connected PAPP-A functionally to the insulin-like growth factor (IGF) system (Conover and Oxvig 2023).

The IGF system is well characterized in placental mammals and teleost fish, where it is involved in growth, metabolism, cell proliferation, reproduction, and aging (Junnila et al. 2013; Ipsa et al. 2019; Miller et al. 2022). The IGF growth factors, IGFI and IGFII (collectively referred to as IGF), are synthesized locally in an autocrine or paracrine manner, or in an endocrine manner by the liver in response to growth hormone (Fig. 1). The IGFs bind the tyrosine kinase IGF receptor (IGFIR), as well as the structurally similar insulin receptor (IR), albeit with reduced affinity (Frasca et al. 1999). Insulin (INS) is also capable of binding IGFIR, but with markedly reduced affinity compared to the IGFs (Belfiore et al. 2009). Activation of IGFIR leads to increased proliferation, protein synthesis, decreased

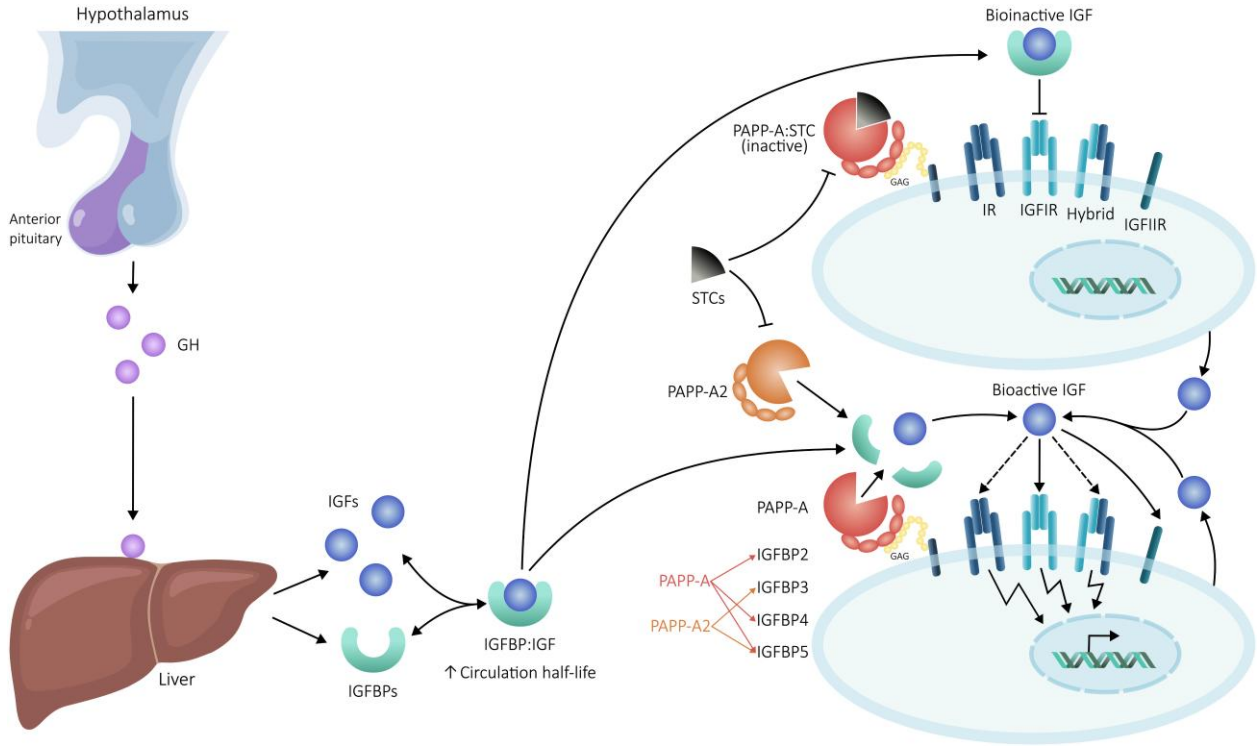


Fig. 1. Simplified model of the STC-PAPP-A-IGFBP-IGF axis according to studies in mammals (Oxvig and Conover 2023). The vast majority of IGFI and IGFII are present in IGFBP:IGF complexes, which increase the half-life of IGF in circulation. PAPP-A or PAPP-A2 is able to specifically cleave different IGFBPs and thus release bound IGF (lower right). The free bioactive IGF is able to bind IGFIR (or IR and IR/IGFIR hybrid receptors). IGFII also binds IGFIIR, which leads to its internalization and degradation. The 2 stanniocalcins (STC1 and STC2) are able to proteolytically inhibit PAPP-A and PAPP-A2. The inhibited enzymes cannot release IGF from IGFBP:IGF complexes (upper right). In the circulation, IGFI and IGFII are derived from the liver, with IGFI synthesis being dependent on pituitary GH. Both IGFI and IGFII are also synthesized locally in tissues in an autocrine and paracrine fashion.

apoptosis, and differentiation (Nagao et al. 2021). Both INS and the IGFs bind the hybrid IGFIR/IR receptor, though the physiological significance of this in normal physiology is unclear (Belfiore et al. 2009). Additionally, IGFII binds the non-signaling IGFIIR that internalizes and degrades bound IGFII (Brown et al. 2008).

In the circulation, <1% of IGF is in a bioactive state (Juul 2003). The vast majority is inactive, bound to IGF-binding proteins (IGFBPs), which increases the half-live of the IGFs from 10 to 12 min to 30 to 90 min and further up to 12 to 15 h if in a ternary complex with acid-labile subunit (only IGFBP-3 and IGFBP-5) (Guler et al. 1989; Twigg and Baxter 1998). Six homologous IGFBPs (IGFBP-1 through -6) exist in humans which bind IGFI and -II with variable but higher affinity than IGFIR, thus rendering the large majority of IGF in a bound, inactive state. The IGFBPs (28 to 35 kDa) consists of an N-terminal and C-terminal domain connected by a flexible central linker domain (CLD). IGF can be released from the IGFBPs through proteolytic cleavage of the IGFBP CLD (Forbes et al. 2012) by PAPP-A and its paralogue PAPP-A2 (pappalysin-2) (Oxvig 2015). In mammals, PAPP-A has been shown to cleave IGFBP-2 (Monget et al. 2003), IGFBP-4 (Lawrence et al. 1999), and IGFBP-5 (Laursen et al. 2001), while PAPP-A2 cleaves IGFBP-3 and IGFBP-5 (Overgaard et al. 2001).

The proteolytic activity of PAPP-A and PAPP-A2 can be inhibited by the 2 paralogs stanniocalcin-1 and -2 (STC1 and STC2) (Oxvig and Conover 2023), adding another layer of complexity to the regulation of the IGF system. STC1 inhibits through high-affinity binding ($K_i = 68$ pM) (Kløverpris et al. 2015), while STC2 forms an irreversible covalent complex mediated by a specific disulfide bond between STC2 (Jepsen et al. 2015) and PAPP-A (Kobberø et al. 2022). In mice, overexpression of STC1 or STC2 results in a dwarf phenotype with a size reduction of 30% to 50% (Varghese et al. 2002) or 45% (Gagliardi et al. 2005), respectively, while STC2 knockout mice are 10% to 15% larger (Chang et al. 2008). Together, these components form the STC-PAPP-A-IGFBP-IGF axis (Fig. 1), that has been shown to be physiologically important in several systems (Oxvig and Conover 2023).

PAPP-A and PAPP-A2 are part of the pappalysin family of the metzincin metalloproteinase superfamily (Guevara et al. 2020). Human PAPP-A (hPAPP-A) consists of 1,546 amino acids and forms a covalent homodimer of ~400 kDa. The mature form of human PAPP-A2 (hPAPP-A2), which is

45% similar to hPAPP-A (Boldt et al. 2001; Overgaard et al. 2001), consists of 1,557 amino acids and has been reported as both a 220 kDa monomer (Sridar et al. 2023) and 400 kDa homodimer (Weyer et al. 2007). PAPP-A and PAPP-A2 share the same domain composition (Fig. 2a). It begins with a β-sandwich-forming laminin-G-like (LG) domain (E82-P324 of hPAPP-A) that has no reported functional role but is believed to be part of a rigid body in the hPAPP-A structure (Kobberø et al. 2022). The LG domain is succeeded by the catalytic domain (CD, P325-K671), the overall fold of which is highly similar to that of other more distant metzincins (Kobberø et al. 2022). The CD contains a metzincin-characteristic so-called extended zinc-binding consensus sequence (HExxHxxGxxH), harboring 3 zinc-coordinating histidine residues and a glutamate catalytic base (Bode et al. 1993). The zinc-binding motif is followed by a calcium-binding gap sequence and a canonical methionine of the Met-turn (supplementary fig. S1a, Supplementary Material online), which is essential for the structural integrity of the zinc-binding site geometry (Oberholzer et al. 2009). This active site environment of the pappalysins is highly similar to the archaeal ulilysin and the bacterial mirolysin, which are classified as part of the lower (unicellular) pappalysins, containing the CD, but none of the other domains of PAPP-A/PAPP-A2 (Gomis-Rüth 2009; Guevara et al. 2020).

The central (M) region of hPAPP-A (P672-D1214) contains 6 domains (M1-M6, as defined by Kobberø et al. 2022) binding a total of 4 calcium ions. Briefly, M1 (P672-F701 + L884-G943), M2 (F702-P883, looping out from M1), and M5 (F1019-F1181) fold into large β-sandwich structures, where M2 and M5 form a substrate binding groove together with the CD (Judge et al. 2022) (supplementary fig. S2, Supplementary Material online). M2 and M5 are separated by the smaller coils and strands of M3 (S944-G995) and M4 (D996-G1018). M6 (D1182-D1214) contains the dimerization cysteine of hPAPP-A (C1210-C1210) and forms intersubunit contacts together with M5 (Kobberø et al. 2022). The M region is followed by 5 short consensus repeat (SCR/CCP/Sushi) domains (SCR1-5, C1215-C1554), where SCR1 and SCR2 of hPAPP-A anchor the string of SCR domains through *in trans* interaction with M5 and the LG domain, respectively (Kobberø et al. 2022) (Fig. 2b). SCR3-5 do not interact with other hPAPP-A domains and show pronounced flexibility (Judge et al. 2022; Kobberø et al. 2022) (Fig. 2b). Positively charged residues of SCR3 and SCR4 of hPAPP-A, but not hPAPP-A2, mediate cell surface adhesion through glycosaminoglycan (GAG) binding (Laursen et al. 2002; Weyer et al. 2004), providing proximity to IGFIR (Laursen et al. 2002). The SCR region is followed by a short C-terminal domain (C1554-G1627), the C domain. Lastly, each subunit contains 3 calcium-binding Lin12-Notch repeat (LNR1-3) modules, otherwise only known from the family of Notch receptors. In contrast to the Notch receptors, where the LNR modules are in tandem, LNR1-2 (V411-C440, R441-C473) are located in the CD, while LNR3 (V1555-C1583) is located more than 1000 residues away in the C domain (Boldt et al. 2004) (Fig. 2a and b). All 3 LNR modules of hPAPP-A are essential for the cleavage of IGFBP-4, but not for cleavage of IGFBP-5 (Boldt et al. 2004). The importance of the LNR modules in hPAPP-A2 is not understood.

PAPP-A and PAPP-A2 have so far only been described and experimentally studied in mammals (Lawrence et al. 1999; Overgaard et al. 2001; Søe et al. 2002; Conover et al. 2011) and zebrafish (Kjaer-Sorensen et al. 2013, 2014). However, the recent growth of sequence databases has revealed the presence of PAPP-A orthologs in species other than those currently studied. Here, we investigate the presence of PAPP-A and PAPP-A2 in the animal kingdom, assess the conservation or divergence of sequence elements between species, and explore the potential existence of PAPP-A and PAPP-A2 outside of the STC-PAPP-A-IGFBP-IGF axis.

Results and Discussion

Pappalysin is Ancient in the Animal Kingdom

Putative homologs of human PAPP-A (hPAPP-A) were identified throughout the animal kingdom (Metazoa). Due to the high sequence conservation of PAPP-A, only sequence hits with ≥90% sequence coverage of mature hPAPP-A, thereby containing the full complement of PAPP-A domains, was used to identify phyla of the animal kingdom containing hPAPP-A homologs. Pappalysin was found to be ancient in the animal kingdom (Fig. 3), with identifiable hPAPP-A homologs present before the emergence of bilaterians ~540 million years ago (Marshall 2006). The earliest branching metazoan phyla where hPAPP-A orthologs could be identified, were Placozoa and Cnidaria. In both phyla, the identified homologs exhibited a global sequence identity of 33% to 39% to hPAPP-A (Fig. 3) despite the considerable evolutionary distance. hPAPP-A homologs were found in all Deuterostomia branches, and identified in several protostome phyla, covering both the Ecdyzoa and Lophotrochozoa, with a global sequence identity of 28% to 40% to hPAPP-A. Note that the apparent absence of hPAPP-A homologs in several phyla may be caused by incomplete sequencing, incomplete depositing, or the depositing of partial sequences. Additionally, the presence of hPAPP-A homologs in a phylum cannot be equated to all members of the phyla containing hPAPP-A homologs. For instance, hPAPP-A homologs are present in species of both Arthropoda (arthropods) and Nematoda (round worms), but e.g. *Drosophila melanogaster* and *Caenorhabditis elegans* have no identifiable hPAPP-A

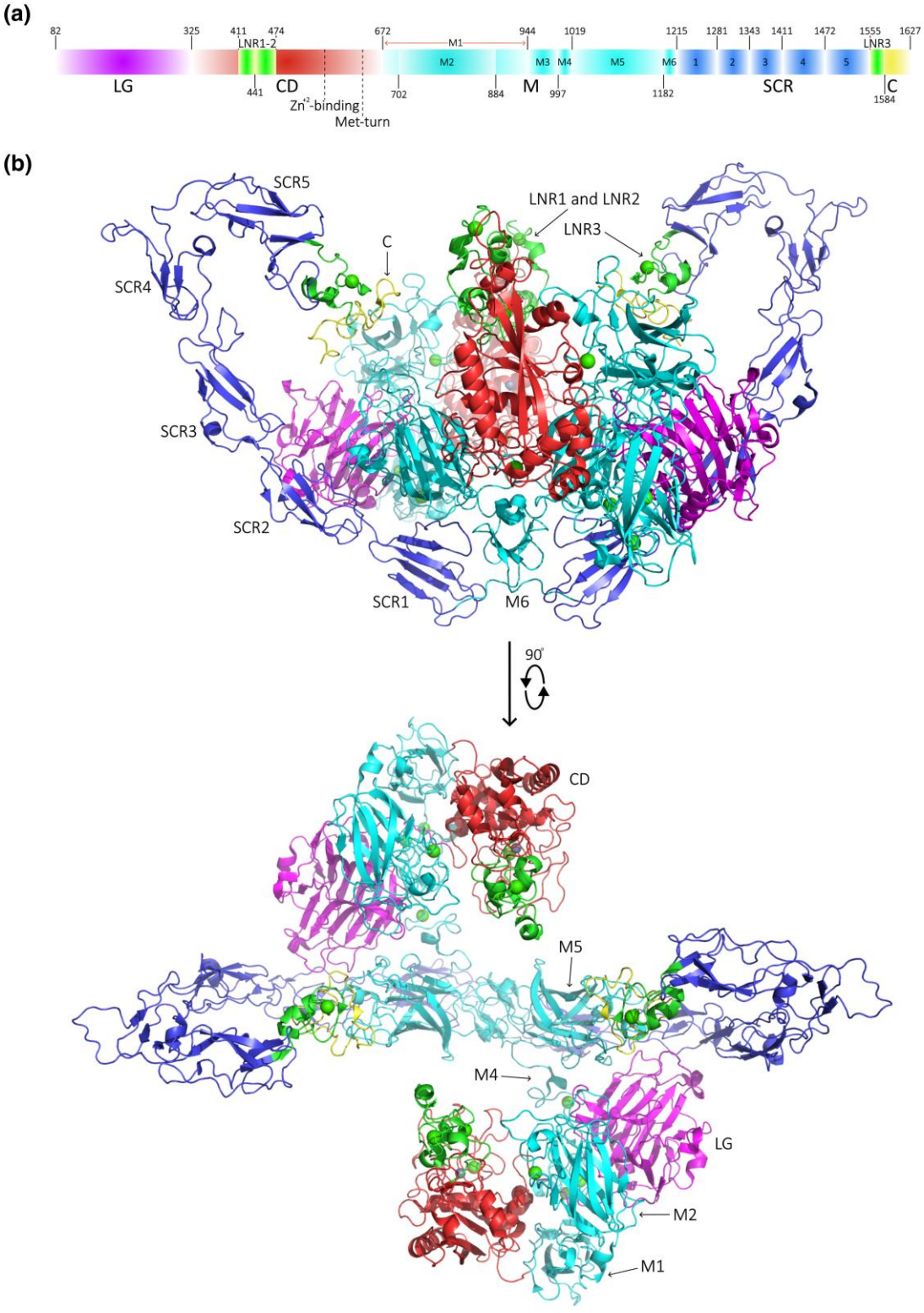


Fig. 2. PAPP-A domain overview. a) Modular structure of the 1546-residue mature hPAPP-A subunit with the positions of domains indicated (LG, laminin G-like domain; CD, catalytic domain; M, middle region; SCR, short consensus repeats; C, C domain). Note that Lin12-Notch repeat (LNR) modules 1 and 2 (LNR1 and LNR2) are inserted into the CD, while LNR3 is present as part of the C domain. b) Cartoon representation of hPAPP-A homodimer in 2 orientations, as indicated, with domains colored and labeled according to (a). The illustrated structure is based on the structure of the hPAPP-A:hSTC2 complex (PDB 8A7E) (Kobberø et al. 2022) with the hSTC2 dimer removed.

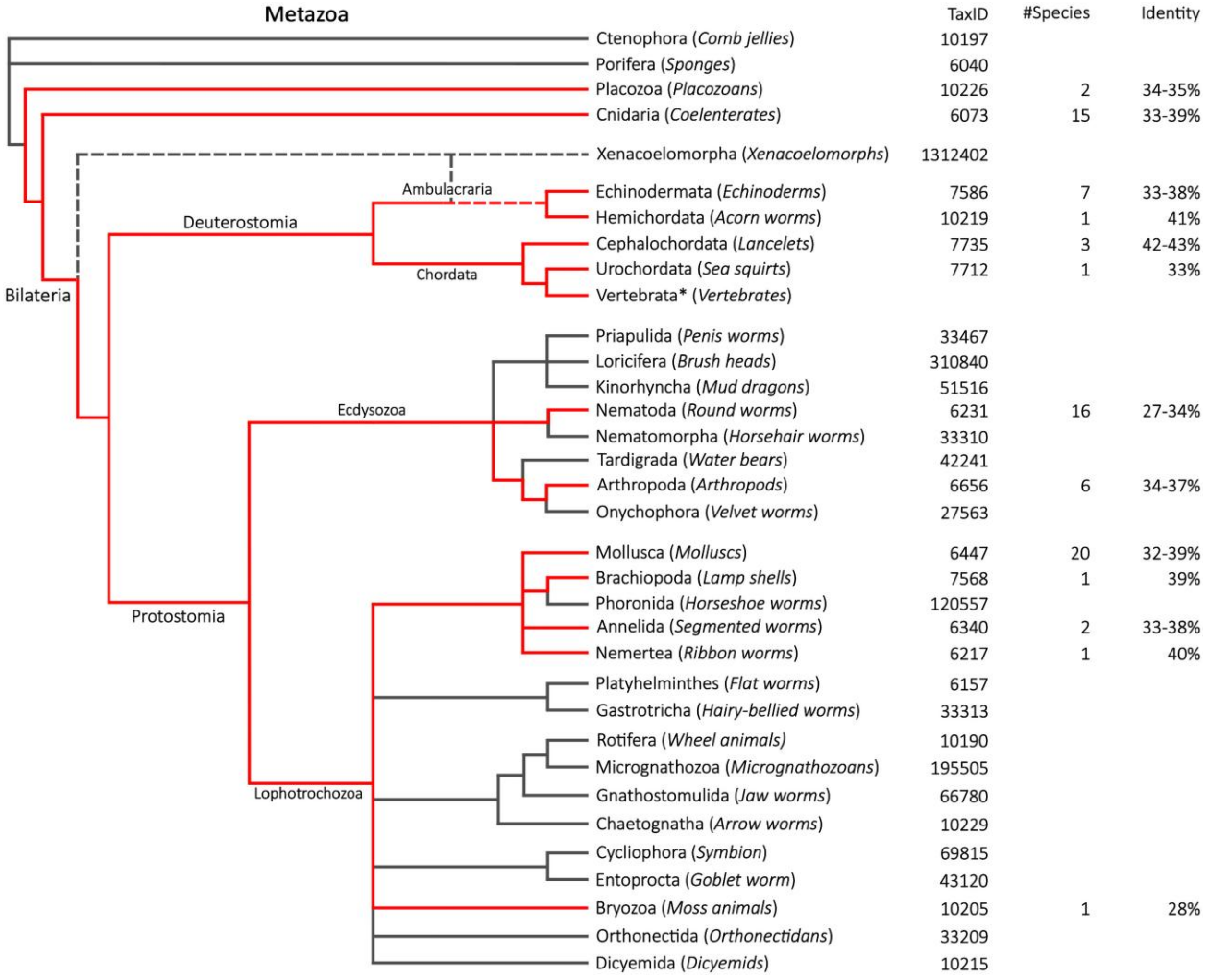


Fig. 3. Occurrence of hPAPP-A homologs in the animal (metazoan) kingdom. Branches colored red contain at least one hit from a BLASTp search with ≥ 90% sequence coverage of mature hPAPP-A. The TaxIDs used to restrict each search are listed for each phylum, subphylum, or clade together with the resulting number of species (#Species) that fulfilled the criteria of sequence coverage. Identity denotes the range of sequence identity compared to mature hPAPP-A (82-1627, NP_002572.2) measured in the BLASTp searches. Accession numbers identified as PAPP-A sequences are listed in (supplementary table S1, Supplementary Material online). Myxinidea (Hagfish) are included in Vertebrata as indicated with an asterisk (*). The vertebrate clade contained an excessive number of pappalysin sequences, including several sequences from the same species, whereby the number of hits for this clade has been omitted. The phylogenetic relationships were based on (Giribet and Edgecombe 2017; Paps et al. 2023).

homologs despite comprehensive genome coverage of these species.

No hPAPP-A homologs could be identified outside of the animal kingdom, and it was not possible to find a hypothetical intermediate of hPAPP-A, structurally between the multidomain metazoan pappalysin and the lower pappalysins ulilysin and mirolysin, composed of only the CD (Guevara et al. 2020). This observation may seem surprising but can probably be explained by experimental findings showing that several domains of hPAPP-A are known to be tightly connected to the function of the CD and substrate recognition, cf. below. Some of the identified sequences with homologous pappalysin CDs, do have sequence stretches connected to the CD, however, but these do not bear resemblance to the non-CD domains of PAPP-A/PAPP-A2 and could also be zymogen prodomains, as seen in ulilysin (Tallant et al. 2006) and mirolysin (Koneru et al. 2017).

Molecular Phylogeny of Animal Pappalysins

To investigate the molecular evolution of the animal pappalysins, a phylogenetic analysis was performed. Based on the inferred phylogenetic tree (Fig. 4), the duplication event of a single ancestral PAPP-A gene to create PAPP-A and PAPP-A2 is ancestral to both cyclostomes (hagfish and

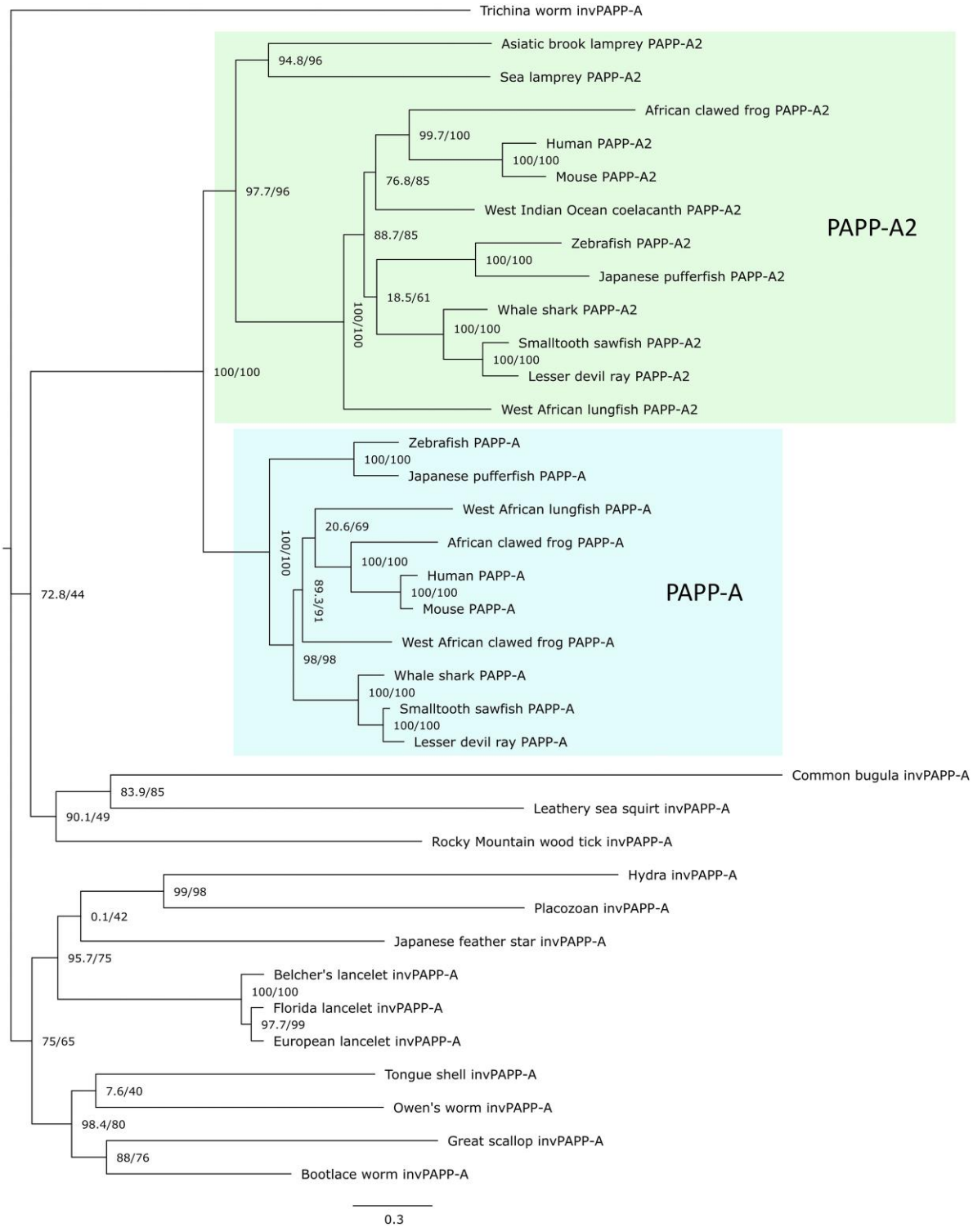


Fig. 4. Phylogenetic tree of metazoan PAPP-A, PAPP-A2, and invPAPP-A protein sequences. Selected metazoan pappalysin sequences representing all phyla, subphyla, or clades identified in the homolog search and representative sequences from vertebrates (supplementary table S2, Supplementary Material online) were subjected to MSA using Clustal Omega 1.2.4. Phylogenetic analysis was carried out using IQ-TREE 2.3.5. The scale bar represents the number of substitutions per site. The SH-like approximate likelihood ratio test (alrt) and the UltraFast Bootstrap approximation (UFBoot) support values, respectively, are indicated at each node. Compared to mature hPAPP-A (82 to 1627, NP_002572.2), the PAPP-A cluster has a global sequence identity range of 54% to 87%, the PAPP-A2 cluster a range of 38% to 45%, and the invPAPP-A sequences a range of 22% to 40%. The 2 lamprey PAPP-As have a sequence identity range of 42% to 45%. The sequence identity was measured using EMBOSS Needle pairwise sequence alignment.

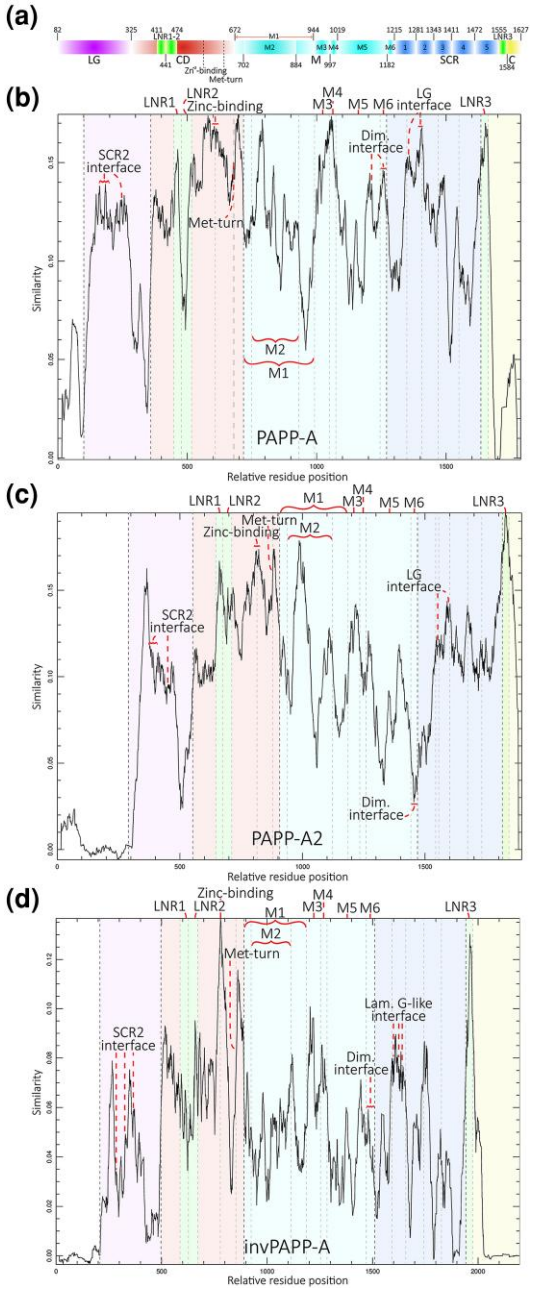


Fig. 5. Conservation plots based on MSAs for PAPP-A, PAPP-A2, and invPAPP-A with a window size of 30 residues. a) PAPP-A domain overview aligned with the plots below. b) Conservation plot of PAPP-A. Domains and selected modules and motifs are indicated with dashed lines and coloring matching the coloring scheme of the domain composition of (a). Selected features are indicated with red dashed lines or red curly brackets. Sequences used are from *R. typus, P. pectinate, M. hypostoma, D. rerio, T. rubripes, L. chalumnae, P. annectens, X. laevis, M. musculus,* and *H. sapiens.* c) Conservation plot of PAPP-A2 based on sequences from the same species as for PAPP-A (b). d) Conservation plot of invPAPP-A based on sequences from *T. adhaerens, H. symbiolongicarpus, A. japonica, B. belcheri, S. clava, T. pseudospiralis, D. andersoni, P. maximus, L. anatine, O. fusiformis, L. longissimus,* and *B. neritina.*

lampreys) and gnathostomes (jawed vertebrates) with PAPP-A subsequently being lost in cyclostomes. Of interest, this analysis revealed a broad range of hPAPP-A homologs outside of vertebrates. We name this diverse group of hPAPP-A/hPAPP-A2 homologs outside the vertebrate clade *invertebrate PAPP-A* (invPAPP-A). To account for possible multiple sequence alignment (MSA) bias, 2 additional phylogenetic inferences were performed based on MSAs using the MAFFT and T-Coffee algorithms (supplementary fig. S3 and S4, Supplementary Material online). Although the placement of the invPAPP-A sequences varies between MSA algorithms, the placement of PAPP-A, PAPP-A2, and lamprey PAPP-A2 sequences, remain largely the same in relation to each other. Due to the short branch separating lamprey PAPP-A2 from the previous node, the phylogenetic inference was also performed with a model that accounts for heterotachy to provide a more accurate phylogenetic estimation (supplementary fig. S5, Supplementary Material online). This phylogenetic estimation has slightly altered branch lengths, but the positioning of lamprey PAPP-A2 sequences remains similar to the phylogenetic estimation obtained using a homotachous model (Fig. 4). It should be noted that the phylogenetic inference was conducted using a single algorithm (IQ-TREE 2.3.5), which may introduce potential biases into the results.

Comparative Sequence Analysis of PAPP-A, PAPP-A2, and invPAPP-A

Based on the above phylogenetic analysis, a broad range of pappalysin orthologs (invPAPP-A) were found outside the vertebrate PAPP-A and PAPP-A2 clades. To identify variations or similarities in functionally or structurally important regions, conservation plots were generated, visualizing the degree of residue conservation across the aligned sequences (Fig. 5a to d). The plots were based on MSAs of sequences from species representing each of the 3 pappalysin groupings. The plots show that, as expected, the overall global sequence conservation of invPAPP-A is lower than that of PAPP-A and PAPP-A2 due to a larger evolutionary distance between the species of this grouping (Fig. 5b to d). PAPP-A and PAPP-A2 are conserved to a higher degree due to shorter evolutionary distance, but the degree of conservation of known functional regions differs markedly (Fig. 5b and c). For instance, the plots reveal that M6, containing the dimerization interface, is highly conserved in PAPP-A, but poorly conserved in PAPP-A2 and invPAPP-A (Fig. 5b to d). As expected, all 3 pappalysins exhibit a high degree of conservation of residues of, and around, the zinc-binding and Met-turn motifs, indicating that catalytic function is preserved throughout the metazoan pappalysins. Strikingly, all 3 pappalysin groupings contain extremely high conservation peaks of the LNR3 module residues that are similar to, or even extend beyond, the conservation

peaks corresponding to the zinc-binding motif of the active site (Fig. 5b to d). The functional implications of differences and similarities in these sequence elements will be further explored below.

Substrate Interface

The only known proteolytic substrates of hPAPP-A and hPAPP-A2 are IGFBP-2, -3, -4, and -5 (Fig. 1), of which IGFBP-4 cleavage by PAPP-A is IGF-dependent (Conover et al. 1993; Lawrence et al. 1999; Laursen et al. 2001). No complete PAPP-A:IGFBP structure is currently available, however there exists a partial structure of hPAPP-A:IGFBP-5 based on cryogenic electron microscopy (cryo-EM) (Judge et al. 2022). This structure shows how the CD, M2, and M5 forms a substrate binding groove that interacts with the IGFBP-5 CLD α-helix (supplementary fig. S2, Supplementary Material online), which precedes the scissile bond (Judge et al. 2022). Based on the partial hPAPP-A:IGFBP-5 structure and a structure prediction of the hPAPP-A:IGFBP-4:IGF complex (see Materials and Methods), potential key interacting hPAPP-A residues surrounding the zinc-binding motif were identified (supplementary fig. S6, Supplementary Material online). Interestingly, these residues are highly invariable between all 3 metazoan pappalysin groupings, indicating a similar substrate interface in the immediate proximity of the active site. The M2 and M5 of the M domain form β-sandwiches with the loops of one end facing toward the substrate (supplementary fig. S2, Supplementary Material online). Such a dynamic substrate interface may contribute to the difference in substrate binding and/or specificity between hPAPP-A and hPAPP-A2. The sequences of the loops of M2 and M5 facing IGFBP-4 or IGFBP-5 are shown in (supplementary figs. S7 to S9, Supplementary Material online) for all 3 pappalysin groupings. In PAPP-A, the M2 and M5 loops exhibit a high degree of conservation, especially the first loops of M2 (Fig. 5b; supplementary fig. S7, Supplementary Material online). The loops of M5 contain more insertions and differences between the PAPP-A sequences of these species than the M2 loops. This pattern is similar in PAPP-A2, albeit with even more variance in the M5 loops (supplementary fig. S8, Supplementary Material online), consistent with the finding that deletion of M5 does not impact hPAPP-A2 cleavage of IGFBP-5 (Sridar et al. 2023). The M2 loops of PAPP-A and PAPP-A2 are similar but with minor, highly conserved differences that might reflect the known substrate specificity. The second M2 loop of PAPP-A2 contains several conserved hydrophobic and/or cyclic residues that extend far into the groove, whereas these residues are much less conserved in the PAPP-A sequences (supplementary fig. S7, Supplementary Material online). In hPAPP-A2, alanine substitution of W931, S932, and P933, that are predicted to be interacting with the IGFBP-5 anchor residue, abolishes IGFBP-5 cleavage (Sridar et al. 2023). Alanine substitution of the corresponding residues in hPAPP-A (W760, S761, and P762) similarly reduces IGFBP-5 cleavage (Sridar et al. 2023). These residues are completely conserved in all investigated sequences of both PAPP-A and PAPP-A2, highlighting their importance in IGFBP-5 substrate recognition. Another interesting residue is H861 of hPAPP-A. Judge et al. (2022) notes that H861 and W738 of PAPP-A engages in hydrogen bonds and hydrophobic interactions, respectively, with a key anchoring residue of IGFBP-5. W738 is completely conserved in all PAPP-A and PAPP-A2 sequences compared across vertebrate species, while H861 is highly conserved in PAPP-A sequences. In PAPP-A2, the corresponding residue is an arginine (R1028 in hPAPP-A2) or a lysine (supplementary fig. S8, Supplementary Material online). This change from histidine to arginine/lysine potentially impacts the binding and cleavage of IGFBP-5, which may explain the reduced cleavage efficiency of IGFBP-5 by hPAPP-A2 compared to hPAPP-A (Sridar et al. 2023). The M2 and M5 loops of invPAPP-A are highly variable (supplementary fig. S9, Supplementary Material online), indicating either a lack of participation in substrate binding/specificity, or that invPAPP-A of these species have varying, yet unknown, substrates, which are not similar to the IGFBPs from the vertebrate IGF system.

Zinc-Binding and Met-turn Consensus Sequences

The CD is highly conserved in sequences of all 3 metazoan pappalysin groupings (Fig. 5b to d), and the extended zinc-binding consensus sequence (HExxHxxGxxH) is completely conserved (Fig. 6). Besides the zinc-binding consensus sequence, the function of the catalytic site is dependent on a Met-turn motif containing the canonical methionine residue. In all investigated sequences except one (*Bugula neritina*), the canonical methionine is conserved and the residues surrounding it are also highly invariable, suggesting that the active site function is conserved across all pappalysins (Fig. 6). In *B. neritina*, the single identified sequence without the canonical methionine, there is a drastic shortening of the gap between the 2 motifs, indicating that a larger deletion event may have occurred (supplementary fig. S1b, Supplementary Material online), depicting the active site of a predicted structure of *B. neritina* invPAPP-A, indicates that L483 could compensate for the loss of the canonical methionine, thus still providing the structural environment required for a functional active site.

LNR Modules

All metazoan pappalysins contains 3 LNR modules (29 to 33 residues) but with varying degrees of conservation (Fig. 5b to d; supplementary fig. S10 to S12, Supplementary Material online). LNR1 is highly conserved in PAPP-A and

Species	Accession nr.	Zinc-binding motif	Gap	Met-turn
		PAPP-A		
		H562-H572		N634-Y638
Rhincodon typus	XP_048469594.1	H E I G H S L G L Y H	-62-	N Y M S Y
Pristis pectinata	XP_051893072.1	H E I G H S L G L Y H	-62-	N Y M S Y
Mobula hypostoma	XP_062929858.1	H E I G H S L G L Y H	-62-	N Y M S Y
Danio rerio	NP_001265734.1	H E I G H S L G L Y H	-62-	N Y M S Y
Takifugu rubripes	XP_011602829.2	H E I G H S L G L Y H	-62-	N Y M S Y
Latimeria chalumnae	XP_064420709.1	H E L G H S L G L F H	-62-	N Y M S Y
Protopterus annectens	XP_043914113.1	H E I G H S L G L Y H	-62-	S Y M S Y
Xenopus laevis	XP_018085171.1	H E I G H S L G L Y H	-64-	N Y M S Y
Mus musculus	NP_067337.1	H E I G H S L G L Y H	-62-	N F M S Y
Homo sapiens	NP_002572.2	H E I G H S L G L Y H	-62-	N F M S Y
		PAPP-A2		
		H733-H743		N805-Y809
Rhincodon typus	XP_048458031.1	H E L G H I L G L Y H	-60-	N Y M S Y
Pristis pectinata	XP_051867159.1	H E L G H I L G L Y H	-60-	N Y M S Y
Mobula hypostoma	XP_062919822.1	H E L G H I L G L Y H	-60-	N Y M S Y
Danio rerio	XP_695417.4	H E M G H I F G L Y H	-60-	N Y M S Y
Takifugu rubripes	XP_029686694.1	H E M G H I L G L Y H	-60-	N Y M S Y
Latimeria chalumnae	XP_064415794.1	H E V G H V L G L Y H	-60-	N Y M S Y
Protopterus annectens	XP_043941906.1	H E L G H V L G L Y H	-60-	N Y M S Y
Xenopus laevis	XP_041446424.1	H E V G H A L G L Y H	-60-	N Y M S Y
Mus musculus	NP_001078845.1	H E V G H V L G L Y H	-60-	N Y M S Y
Homo sapiens	NP_064714.2	H E V G H V L G L Y H	-60-	N Y M S Y
		invPAPP-A		
		H554-H564		N625-Y629
Trichoplax adhaerens	XP_002109449.1	H E I G H N L G L W H	-61-	N F M S Y
Hydractinia symbiolongicarpus	XP_057305553.1	H E L G H V L G L W H	-62-	N F M G Y
Anneissia japonica	XP_033096949.1	H E I G H S L G L W H	-62-	N Y M S Y
Branchiostoma belcheri	XP_019621199.1	H E L G H N L G L W H	-60-	N F M S Y
Styela clava	XP_039270265.1	H E L G H N L G L L H	-60-	N F M S Y
Trichinella pseudospiralis	KRZ20972.1	H E F G H A L G L W H	-72-	N Y M S Y
Dermacentor andersoni	XP_050051616.1	H E L G H V L G L W H	-60-	N Y M G Y
Pecten maximus	XP_033761458.1	H E L G H V L G L W H	-60-	N Y M G Y
Lingula anatina	XP_013388811.1	H E V G H V L G L L H	-58-	N Y M S Y
Owenia fusiformis	CAH1782730.1	H E I G H N L G L L H	-60-	N Y M S Y
Lineus longissimus	XP_064650054.1	H E F G H N L G L W H	-61-	N Y M G Y
Bugula neritina	KAF6027397.1	H E L G H V L G L W H	-19-	– – – – –

Fig. 6. MSA segments of the zinc-binding motif and Met-turn sequences from selected species representing PAPP-A, PAPP-A2, and invPAPP-A are shown with majority consensus highlighted. Sequence numbering is based on hPAPP-A, hPAPP-A2, and *B. belcheri* invPAPP-A. Gap denotes the number of amino acids spanning the sequence stretch between the 2 motifs.

PAPP-A2 and they share a high degree of similarity. LNR1 of invPAPP-A is less conserved both between invPAPP-A sequences and compared to PAPP-A and PAPP-A2 LNR1. However, most residues known to be involved in calcium-binding in hPAPP-A are conserved (supplementary fig. S10, Supplementary Material online), as well as almost all cysteines, indicating a conservation of the overall structure of the module. PAPP-A LNR2 is less conserved than LNR2 of PAPP-A2, partly due to small insertions (supplementary fig. S11, Supplementary Material online). LNR2 of PAPP-A has lost the first cysteine (C618, hPAPP-A2), which seems to still be present in teleosts (*D. rerio* and *T. rubipres*) and

coelacanths (*L. chalumnae*). In all 3 pappalysin groupings, the LNR2 modules contain conserved calcium-binding residues and cysteine residue patterns, apart from the first cysteine in PAPP-A LNR2 (supplementary fig. S11, Supplementary Material online). Both PAPP-A and PAPP-A2 LNR2 contain several conserved, positively charged, and outwards-facing residues (Kobberø et al. 2022) in the first half of the module, perhaps important in substrate specificity. Strikingly, LNR3 shows an extreme degree of conservation in all pappalysins (Fig. 5b to d; supplementary fig. S12, Supplementary Material online). However, some positions differ between PAPP-A, PAPP-A2, and InvPAPP-A sequence groupings, especially in the first part of the module, but are highly conserved within the individual variants, perhaps contributing to differences in substrate specificity.

Dimerization Interface

Human PAPP-A exists as a covalent, disulfide-bonded dimer, and hPAPP-A dimerization is strictly required for cleavage of IGFBP-4, but not IGFBP-5, although disulfide bond formation is not needed for the formation of a functional dimer (Weyer et al. 2007). As mentioned, hPAPP-A2 has been reported as both a monomer and dimer. Based on the dimerization interface known from structural analysis of hPAPP-A, the corresponding amino acid stretches are present in the invPAPP-A and PAPP-A2 sequences, but the sequence is only conserved in PAPP-A (Fig. 5b to d; supplementary fig. S13, Supplementary Material online). This would support that only PAPP-A is a dimer. Intriguingly, however, in almost all of the investigated sequences, both invPAPP-A and PAPP-A2 contains a conserved cysteine residue corresponding to the dimerization cysteine of PAPP-A (C1210). Unfortunately, a recent cryo-EM structure of hPAPP-A2 (Sridar et al. 2023) does not include M6, which contains the hPAPP-A2 cysteine residue corresponding to hPAPP-A C1210. Further experimental data are therefore required to determine the function of the cysteine and the oligomerization state of PAPP-A2 and invPAPP-A.

SCR1 and SCR2 Fixation

As mentioned, SCR1 interacts in trans with M5 in a complex of hPAPP-A and human STC2 (hSTC2), and SCR2 interacts in trans with the LG domain (Fig. 2b). Both of these interactions reduce the flexibility of the SCR1 and SCR2 domains. The M5-interacting surface of SCR1 exhibits high variability between the 3 pappalysin groupings, with no clear conserved corresponding sequence in PAPP-A2 nor invPAPP-A (supplementary fig. S14, Supplementary Material online). InvPAPP-A homologs appear to lack the corresponding sequence, while it is present but significantly different in PAPP-A2, even between different PAPP-A2 sequences. A predicted structure of hPAPP-A2 does, however, seem to form similar interactions as hPAPP-A with a moderately conserved part of M5 (supplementary fig. S14, Supplementary Material online), indicating that SCR1 might be similarly fixed in PAPP-A2. In SCR2, the residues oriented toward the LG domain are fairly conserved between all pappalysins (Fig. 5b to d; supplementary fig. S15, Supplementary Material online). Likewise, the loops of the LG domain oriented toward SCR2 exhibit a similar pattern of conservation between the pappalysins (Fig. 5b to d; supplementary fig. S16, Supplementary Material online), indicating that all pappalysins likely have the same fixation of SCR2 through the LG domain, even if SCR2 of invPAPP-A and PAPP-A2, if these are monomers, interacts *in cis* with the LG domain.

SCR3 and SCR4 Cell Surface Binding

SCR3 and SCR4 of hPAPP-A mediate cell surface adhesion through GAG binding via several positively charged residues (Fig. 7a; supplementary fig. S17, Supplementary Material online). The PAPP-A2 sequences lack several of these charged key residues (Fig. 7b; supplementary fig. S17, Supplementary Material online), making surface binding unlikely, in agreement with the experimentally confirmed inability of hPAPP-A2 to bind to cells (Laursen et al. 2002; Weyer et al. 2004). This supports current models of PAPP-A2 function in the circulation (Argente et al. 2017; Oxvig and Conover 2023). InvPAPP-A from *B. belcheri* has more accessible positively charged residues on SCR3 and SCR4 compared to PAPP-A2 sequences (Fig. 7c; supplementary fig. S17, Supplementary Material online). However, the pattern of the positively charged residues varies from the PAPP-A sequences and there is high variability between invPAPP-A sequences, whereby PAPP-A most likely is the only variant that binds cell surfaces.

SCR4 and SCR5 Bend

In hPAPP-A, as present in the hPAPP-A:hSTC2 complex, there is a distinct sharp bend of ∼50° between SCR4 and SCR5 that might be important in bringing the C domain/LNR3 close to the catalytic center of hPAPP-A (Kobberø et al. 2022) (Fig. 2b). Using the structure of the hPAPP-A:hSTC2 complex, possible key residues forming the bend were identified (Fig. 7d). Compared to PAPP-A sequences in other species, there is a high degree of functional conservation (supplementary fig. S18, Supplementary Material online). Most of the key residues are hydrophobic, forming a stable hydrophobic core between the end of SCR4 and the beginning of SCR5. Of special note are hPAPP-A residues H1423 and Y1546, and V1470 and M1474 that are positioned at the outer regions of the bend (Fig. 7d). In the predicted model of hPAPP-A2, a very similar bend is observed (Fig. 7e), and potential key residues also seem to be functionally conserved between PAPP-A2 sequences (supplementary fig. S18, Supplementary Material online).

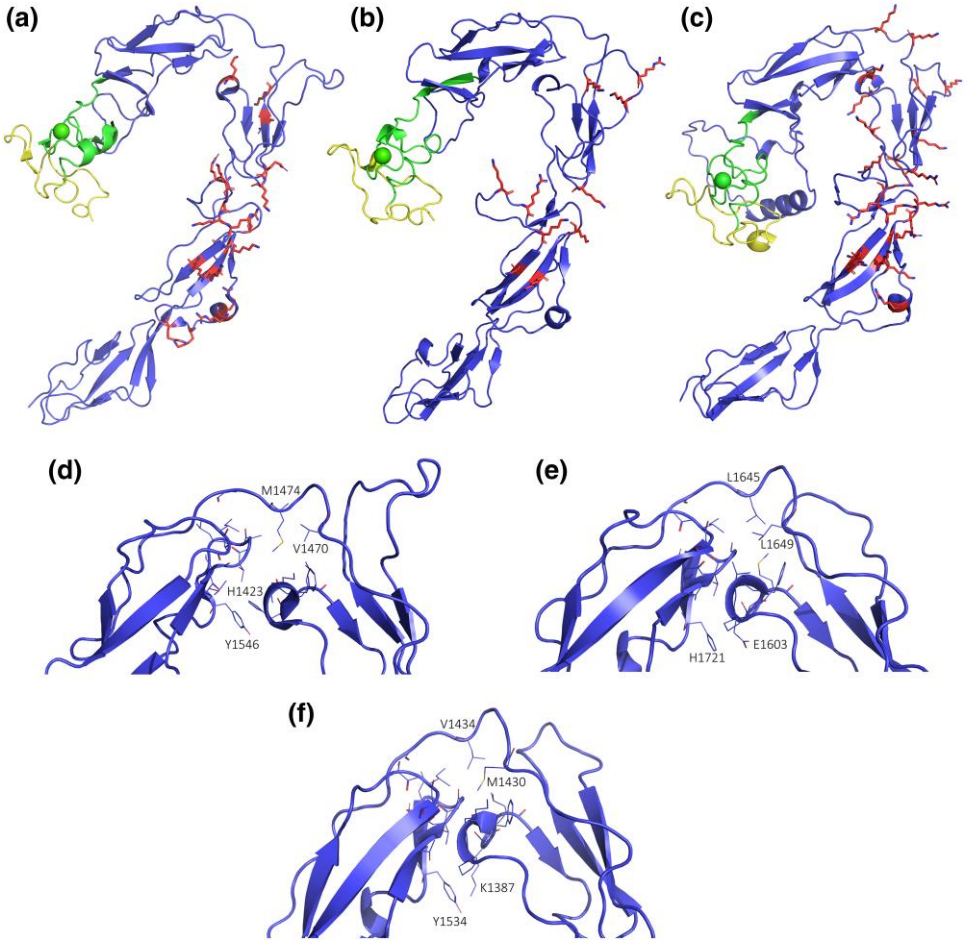


Fig. 7. Partial structures corresponding to SCR2-5 and the C domain (SCR2-5-C) and SCR4 to SCR5 (SCR4-5) with residues potentially involved in formation of the bend. a) The structure of SCR2-5-C of hPAPP-A based on a the experimental structure of the hPAPP-A:hSTC2 complex (PDB 8A7E). b) The structure of SCR2-5-C of the hPAPP-A2 (NP_064714.2) structure predicted using AlphaFold 3. c) The structure of SCR2-5-C of the *B. belcheri* (XP_019621199.1) invPAPP-A predicted using AlphaFold 3. All positively charged residues in a-c are shown as red sticks. d) The structure of SCR4-5 of hPAPP-A based on the structure of the hPAPP-A:hSTC2 complex (PDB 8A7E). e) The structure of SCR4-5 of the hPAPP-A2 (NP_064714.2) structure predicted using AlphaFold 3. f) The structure of SCR4-5 of the *B. belcheri* (XP_019621199.1) invPAPP-A predicted using AlphaFold 3. In (d-f), residues potentially involved in formation of the bend between SCR4 and SCR5 are visualized as sticks, and selected residues are numbered.

The position and identity of the involved residues are closely related to those of PAPP-A, indicating the likely presence of a similar bend in PAPP-A2. The predicted *B. belcheri* invPAPP-A model also contains a similar bend (Fig. 7f). The *B. belcheri* invPAPP-A key residues are comparable to both hPAPP-A and hPAPP-A2, but there is a higher degree of variance between the invPAPP-A sequences compared to the variance seen between PAPP-A sequences, as expected due to the diversity of the invPAPP-A sequences. However, the predicted *B. neritina* invPAPP-A model, which only shares 28% sequence identity with hPAPP-A, also exhibits a similar bend (not shown). This indicates that the bend is likely conserved in all pappalysins, but the key residues involved in the formation of the bend can be variable. Thus, the bend between SCR4 and SCR5 is probably conserved in all 3 pappalysin variants, possibly permitting the positioning of their C domain/LNR3 in close proximity to the active site.

Occurrence of the STC-PAPP-A-IGFBP-IGF Axis in Chordates

A tight functional relationship between PAPP-A/PAPP-A2 and the components of the STC-PAPP-A-IGFBP-IGF axis has been demonstrated genetically and functionally in mice and humans (Oxvig and Conover 2023). In mice, PAPP-A or PAPP-A2 knockout, as well as STC1 and STC2 overexpression, results in dwarf or growth restricted phenotypes (Varghese et al. 2002; Conover et al. 2004, 2011; Gagliardi et al. 2005), while STC2 knockout, but not STC1 knockout, results in increased size (Varghese et al. 2002; Chang et al. 2008), indicating that a regulated

balance between active and inhibited PAPP-A and PAPP-A2 exists in vivo. Furthermore, PAPP-A2 knockout in mice results in increased total IGFI (Christians et al. 2015), consistent with findings in humans deficient in PAPP-A2 (Dauber et al. 2016; Argente et al. 2017; Babiker et al. 2021). Additionally, an STC2 R44L mutation in humans, that shows compromised ability to bind hPAPP-A, increases height by ~2.1 cm per allele (Marouli et al. 2017). Thus, functional cooperation of the STC-PAPP-A-IGFBP-IGF axis proteins is strongly supported in these species.

As metazoan pappalysins are ancient in the animal kingdom, we sought to pinpoint the emergence of the STC-PAPP-A-IGFBP-IGF axis to address whether the animal pappalysins exists separate from the context of this axis. To do this, the phylogenetics of each of the relevant components was explored in the existing literature combined with molecular phylogenetic analysis based on information of the current sequence databases, thus, capturing a broader range of chordate species than previously investigated.

IGFI, IGFII, and INS

IGFI and IGFII belong to the insulin superfamily that includes INS and insulin-like peptides (ILPs) (Rao et al. 2024). Cephalochordates (lancelets) contain a single gene encoding an ILP that both resembles IGF and INS (Chan et al. 1990), whereby ILP likely is the ancestor of both IGFI, IGFII, and INS through tandem duplication and subsequent loss of 1 INS copy (Patton et al. 1998; Li 2022). This is in accordance with the phylogenetic analysis of the relationship of ILP, INS, IGFI, and IGFII based on the currently available sequences (supplementary fig. S19, Supplementary Material online). It is evident that ancestral chordate ILP, as represented in modern lancelets, duplicated early during vertebrate evolution based on the presence of INS and proto-IGF ancestor sequences in Myxinidae (hagfish) and Hyperoartia (lampreys). The presence of 4 distinct IGF sequences in chondrichthyans (cartilaginous fishes) (supplementary fig. S19, Supplementary Material online) suggests that duplication events of the proto-IGF occurred after the divergence of gnathostomes (supplementary fig. S19, Supplementary Material online). In contrast, most osteichthyans only possess 2 distinct IGF sequences, suggesting a loss of 2 IGF genes following the divergence of chondrichthyans. Additionally, a duplication of IGFI has occurred in teleost fishes, resulting in IGFIII, which has been shown to play an important role in fish reproduction (Wang et al. 2008; Li et al. 2021). Nevertheless, caution should be exercised when interpreting the phylogenetic relationships of ILP, INS, and IGFs, as the short length of these sequences limits the robustness of the analysis.

IGFIR, IGFIIR, and IR

In accordance with the presence of ILP in lancelets, they also have a single ILP receptor (ILPR). ILPR shows resemblance to both human IR and IGFIR and can be autophosphorylated by incubation with either IGFI or INS (Pashmforoush et al. 1996). It has previously been found that a single ILPR gene, represented by orthologs in modern lancelets and sea squirts, duplicated during early vertebrate evolution to create IGF1R, IR and the IR-related receptor (IRR) (Hernández-Sánchez et al. 2008). Phylogenetic analysis of the relationship between ILPR, IGFIR, IR, and IRR, based on the currently available sequences, reveals that the ancestral ILPR, as represented in modern lancelets, duplicated during early vertebrate evolution as IR and an ancestral IGFIR/IRR appear in lampreys. After the divergence with cyclostomes, an ortholog of the IGFIR/IRR gene underwent duplication in a common ancestor to gnathostomes forming the IRR and IGFIR sister group (Fig. 8; supplementary fig. S20, Supplementary Material online). Additionally, as mentioned in Hernández-Sánchez et al. 2008, it was not possible to identify IRR in any teleost fish (Hernández-Sánchez et al. 2008), although in this analysis, IRR could be identified in Chondrostei (Sturgeons and paddlefish), indicating a loss after the divergence of chondrostean fishes from other Actinopterygii lineages.

IGFBPs

In lancelets, a single ancestral IGFBP exists that is 1 amino acid change (F70L) from being able to bind either IGFI or IGFII (Zhou et al. 2013). After the divergence with cephalochordates, an ortholog of this gene underwent duplication in a common ancestor to vertebrates resulting in 2 adjacent IGFBP genes, which then underwent 2 successive rounds of tetraploidization resulting in 8 vertebrate IGFBP genes, of which 2 have been lost (Daza et al. 2011; Allard and Duan 2018). From phylogenetic analysis based on the currently available sequences, 3 paralog IGFBP sequences could be identified in lampreys (Fig. 8; supplementary fig. S21, Supplementary Material online), one of which has been experimentally shown to have both IGF-independent and IGF-dependent functions (Zhong and Duan 2017). In the analysis, one of the lamprey IGFBP sequences was placed as an ortholog of the vertebrate IGFBP-2 with strong support (supplementary fig. S21, Supplementary Material online), whereby the duplication event leading to IGFBP-2 and the common ancestor of IGFBP-1/4 must have predated the divergence of cyclostomes and gnathostomes. The IGFBP-1/4 gene may have subsequently been lost in cyclostomes before a gnathostome duplication event leading to separate IGFBP-1 and IGFBP-4 genes (Fig. 8; supplementary fig. S21, Supplementary Material online). The 2 other lamprey IGFBP paralogs are placed as sister clades to the vertebrate IGFBP-3 clade, but the branching support is only moderate for both lamprey paralogs (Fig. 8; supplementary fig. S21, Supplementary Material online). Furthermore, the

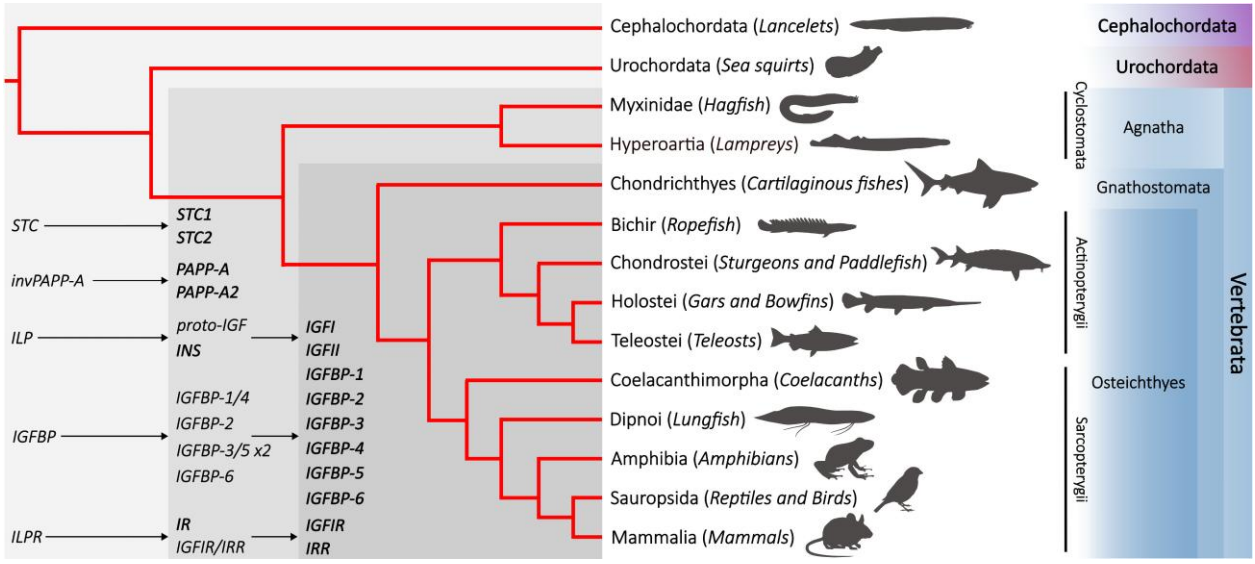


Fig. 8. Emergence of the STC-PAPP-A-IGFBP-IGF axis components in the Chordata phylum. The phylogenetic relationships were based on (Pascual-Anaya et al. 2022; Marlétaz et al. 2024).

branching of cyclostome IGFBP sequences requires that the origin of vertebrate IGFBP-6, as a sister clade to IGFBP-3 and IGFBP-5, predates the divergence of cyclostomes and gnathostomes, indicating a loss of IGFBP-6 in cyclostome evolution.

During evolution of the jawed vertebrates, 6 paralog IGFBPs sequences are present before the divergence of cartilaginous fishes. Five of the 6 cartilaginous fish IGFBPs group with IGFBP-1 to IGFBP-5, but the 6th IGFBP sequences is placed as the sister group to the IGFBP-3/5 clade of vertebrates (supplementary fig. S21, Supplementary Material online). However, this is likely due to phylogenetic uncertainty as the branch support values are weak (Fig. 8; supplementary fig. S21, Supplementary Material online).

Outside of Chordata, IGFBP homologs have been identified in crustaceans (Rosen et al. 2013), insects (Roed et al. 2018), and oysters (Zhang et al. 2017). However, it was not possible to identify invPAPP-A in crustaceans, insects, or oysters, suggesting that regulation of ILPs in invertebrates function differently compared to the pappalysin-based proteolytic regulation of vertebrates. In *Drosophila,* insulin-like polypeptide binding proteins (IBPs) were previously considered as homologs to IGFBPs, but structures of free and ligand-bound IBP Imp-12 have revealed that its structure and ligand-binding mode are unrelated to the vertebrate IGFBPs, and therefore likely represents an alternative ILP-bioavailability regulating system (Roed et al. 2018). Some of the described invertebrate IGFBP homologs contain an insulin-like binding domain like that of the vertebrate IGFBPs suggesting a common ancestor. However, through specialization, the ILP system in invertebrates evolved to serve pleiotropic roles with different ways of regulation, while a distinct vertebrate IGF system evolved from the invertebrate ILP-system with its own specialized regulation of bioavailability (Huang et al. 2015).

STC1 and STC2

STC has been found to be ancient in the animal kingdom (Roch and Sherwood 2011). Three STC paralogs have been described in lancelets (STCa, STCb, and STCc), where STCc is similar to the STC sequences of other invertebrate species, while STCa and STCb more resemble human STC1 (hSTC1) and hSTC2 (Roch and Sherwood 2011). Phylogenetic analysis based on the currently available STC sequences suggests that the monophyletic STCa and STCb in lancelets occurred due to a duplication event after the divergence of cephalochordates from the other chordate lineages. In cyclostomes, the single identified STC sequence is monophyletic to STC2 of jawed vertebrates. This indicates that STC duplicated in the common ancestor after the divergence of urochordates, with the subsequent loss of STC1 in the cyclostome ancestor (Fig. 8; supplementary fig. S22, Supplementary Material online). During the evolution of jawed vertebrates, more STC paralogs appeared. In cartilaginous fishes, 3 STC paralogs (STC1, STC2, and STC-like) could be identified, while 2 paralogs (STC1 and STC2) are present in osteichthyans (bony fishes) (Fig. 8; supplementary fig. S22, Supplementary Material online), suggesting that the STC-like sequence was lost in the common ancestor to bony fishes.

As mentioned, inhibition of hPAPP-A by hSTC2 depends on the formation of a disulfide bond between the

proteinase and inhibitor, mediated by C120 of hSTC2 (Jepsen et al. 2015) and C732 of hPAPP-A (Kobberø et al. 2022). In the absence of this covalent linkage, hSTC2 cannot inhibit hPAPP-A (Jepsen et al. 2015; Kobberø et al. 2022). However, the evolutionary emergence of the involved cysteine residues remains unexplored. C732 (C903 in hPAPP-A2) is present in all the investigated PAPP-A and PAPP-A2 sequences, including lamprey PAPP-A2 (Fig. 8; supplementary fig. S23, Supplementary Material online). Thus, C732 emerged before the divergence of lampreys from the other vertebrate lineages. For STC, C120 is conserved in all STC2 sequences, together with the STC variant found in cyclostomes (hagfish and lampreys) and some STC-like variants found in cartilaginous fishes, whereby C120 emerged early in vertebrate evolution before the divergence of agnathans and gnathostomes lineages (Fig. 8; supplementary fig. S23, Supplementary Material online). Interestingly, this suggests that the covalent inhibition of PAPP-A by STC2 co-evolved in the early vertebrates. Furthermore, hSTC2 also interacts with LNR3, as evident from the known structure of the hPAPP-A:hSTC2 complex (Kobberø et al. 2022). Since LNR3 is highly conserved across pappalysin homologs, and STC homologs are found throughout Metazoa (Roch and Sherwood 2011), there is a possibility that invPAPP-A and invertebrate STC homolog proteins could interact. However, the invertebrate STC homologs do not contain C120 and exhibit low sequence conservation besides their overall cysteine pattern (Roch and Sherwood 2011), thus making their interaction with invPAPP-A unlikely.

To summarize the emergence of the individual components of the STC-PAPP-A-IGFBP-IGF axis, the duplication event of invPAPP-A resulting in PAPP-A and PAPP-A2, both containing C732, is ancestral to both cyclostomes (hagfish and lampreys) and gnathostomes (jawed vertebrates) with PAPP-A subsequently being lost in the cyclostomes (Figs. 4 and 8). As for IGF, ancestral chordate ILP duplicated in a common vertebrate ancestor, producing proto-IGF and INS. The subsequent duplication of proto-IGF to IGFI and IGFII occurred prior to the divergence of the chondrichthyans from the other gnathostome lineages (Fig. 8; supplementary fig. S19, Supplementary Material online). For IGFIR, the ancestral ILPR duplicated during early vertebrate evolution to IR and an ancestral IGFIR/IRR, which, after the divergence with cyclostomes, underwent duplication in a common ancestor to gnathostomes forming the IRR and IGFIR sister group (Fig. 8; supplementary fig. S20, Supplementary Material online). For IGFBPs, a single IGFBP sequence is present in cephalochordates and urochordates, while 5 ancestral IGFBPs could be identified in lampreys, and 6 paralog IGFBP sequences exist before the divergence of cartilaginous fishes from the other gnathostome lineages (Fig. 8; supplementary fig. S21, Supplementary Material online). Lastly, the STC sequence identified in cyclostomes containing C120, is monophyletic to STC2 of jawed vertebrates, indicating that STC duplicated in the common ancestor after the divergence of urochordates, with the subsequent loss of STC1 in the cyclostome ancestor. Furthermore, in cartilaginous fishes, 3 STC paralogs (STC1, STC2, and STC-like) could be identified, while 2 paralogs (STC1 and STC2) are present in osteichthyans (bony fishes) (Fig. 8; supplementary fig. S22, Supplementary Material online), suggesting that the STC-like sequence was lost in the common ancestor to bony fishes. This demonstrates that before the division of chondrichthyans from the other gnathostome lineages, all components of the STC-PAPP-A-IGFBP-IGF axis are present (Fig. 8).

Conclusions and Open Questions

Phylogenetic analysis of metazoan pappalysins revealed the evolutionary history of pappalysins. While invPAPP-A is present in diverse invertebrate phyla, PAPP-A and PAPP-A2 are specific to vertebrates. Despite the evolutionary distance between PAPP-A, PAPP-A2 and their single invertebrate pappalysin orthologs, they exhibit a remarkable degree of sequence identity, particularly in their active site environments and LNR3. Additionally, structural motifs such as the LNR modules, along with residues involved in SCR2 fixation and in the bend between SCR4 and SCR5, are conserved. However, certain features, such as dimerization and cell surface binding, appear to be unique to PAPP-A. The substrate interface shows variability, not only between invPAPP-A and the PAPP-A/PAPP-A2 clusters, but also among the invPAPP-A sequences.

The high degree of conservation between the pappalysins underscores the conservation of a common, complex domain architecture. Metazoan pappalysins, even those evolutionarily very distant from vertebrate pappalysins, retain this domain composition in which all components act interdependently and cooperatively to facilitate substrate recognition and maintain core structure integrity, rather than functioning as independent, sequential domains. For example, the SCR modules exists as multiple tandem repeats of variable length in other proteins; L-selectin has 2 SCR domains, and complement receptor 1 has 30 SCR domains (or splice variants with 23, 37, or 44 repeats) and -2 has 15 SCR domains (Santos-López et al. 2023), thus allowing a wide range of potential interdomain orientations (Perkins et al. 2002). In the metazoan pappalysins, the domain architecture is restricted to exactly 5 repeats to necessitate the required interdomain orientation between the active site and the C domain/LNR3. This conservation points to a strong co-evolutionary relationship between the metazoan pappalysin domains, enabling the evolution of highly specific substrate recognition. It is remarkable that so many individual domains are necessary for catalytic function and

substrate recognition, when it is taken into account that no transitional multidomain pappalysin could be identified to bridge the monodomain lower pappalysins ulilysin or mirolysin and the multidomain metazoan pappalysins.

In mammals, both PAPP-A and PAPP-A2 have a tight functional relationship with components of the IGF system. From the phylogenetic analyses, it was evident that the complete STC-PAPP-A-IGFBP-IGF axis first appears in gnathostomes. Besides IGFBPs, PAPP-A and PAPP-A2 have no known substrates. Interestingly, this analysis shows that invPAPP-A in all likelihood is an active proteolytic enzyme and is found in invertebrate species that have no characterized IGFBP homologs, thus raising the question of possible substrates beyond the IGFBPs. It is an intriguing question whether such hypothetical substrates are still present in, e.g. vertebrates, and, in particular, if PAPP-A and/or PAPP-A2 exhibit proteolytic activity against these. Of particular interest is the conservation of the LNR modules. In hPAPP-A, all 3 LNR modules are strictly necessary for IGFBP-4 cleavage, while IGFBP-5 cleavage by hPAPP-A is independent of any of these modules (Boldt et al. 2004). Thus, the LNR modules are not necessary for proteolytic activity, but rather impact substrate specificity in a selective manner. However, since invPAPP-A is present in invertebrate species lacking vertebrate IGFBP homologs, it appears not to be the IGFBP-LNR interactions that drive the evolutionary conservation of the LNR modules. In PAPP-A, the LNR modules likely serve both as an exosite for IGFBP-4 recognition (Weyer et al. 2007), and as an interaction site for hSTC2 (Kobberø et al. 2022). The conservation of the LNR modules in invPAPP-A implies that these modules, prior to the mechanisms regulating IGF-bioavailability in vertebrates, still may facilitate exosite-mediated substrate or inhibitor recognition mechanisms. However, there might also exist unknown substrates that are independent of the exosite, as seen with IGFBP-5 cleavage by hPAPP-A (Boldt et al. 2004). Furthermore, even when the exosite of hPAPP-A is occupied by hSTC2, in the inhibited PAPP-A:STC2 complex, that cannot cleave intact IGFBPs, the active site is still active and has been shown to be able to cleave a 26-residue peptide derived from IGFBP-4 (Kobberø et al. 2022).

Finally, PAPP-A is known to be involved in multiple pathologies, including cancer, cardiovascular disease, kidney disease, and visceral obesity (Conover and Oxvig 2017, 2018, 2023, 2024; Steffensen et al. 2019). Because of its suspected role in the development of these conditions, PAPP-A is an emerging therapeutic target. In mouse models, for example, experimental treatment with a monoclonal antibody targeting the exosite of PAPP-A, and thus selectively blocking the cleavage of IGFBP4, has been shown to inhibit ovarian cancer growth (Becker et al. 2015), to reduce atherosclerotic plaque progression (Conover et al. 2016), and to ameliorate autosomal dominant polycystic kidney disease (Kashyap et al. 2020). Since the active site of PAPP-A remains active while the exosite is occupied, the cleavage of potential unknown, exosite-independent substrates is most likely not affected by exosite targeting. It is therefore important, depending on PAPP-A targeting strategies, to understand any potential effects on broader physiological and regulatory systems that are potentially affected upon targeting of PAPP-A. Such systems may potentially be identified by experimental studies of invPAPP-A.

Materials and Methods

Homolog search

We assessed the presence of all components of the STC-PAPP-A-IGFBP-IGF axis currently contained in the NCBI BLAST non-redundant protein sequence database (nr, searched between July 1 and July 18, 2024). All homolog proteins were identified using BLASTp (Altschul et al. 1990; Camacho et al. 2009) with standard parameters (BLOSUM62, word size 5, E-value threshold 0.05), except otherwise stated. The homolog search for each protein was performed against each phylum or class of the animal kingdom (TaxIDs used are listed in Fig. 3). Pappalysin homologs were identified using mature hPAPP-A (82-1627, NP_002572.2) as query sequence. Hits with ≥90% sequence coverage rounded to the nearest integer, of mature hPAPP-A were accepted as true hits. In addition, similar homolog search was performed using the pappalysin sequence of a more distant species (*Hydractinia symbiolongicarpus,* XP_057305553.1), and we also carried out a tBLASTn search, but no sequences from additional species were identified that fulfilled the specified criterion. IGFI, IGFII, and INS homologs were identified using human IGFI (NP_001104753.1) as query. The ancestral *Branchiostoma belcheri* ILP (XP_019634785.1) was also investigated as a query sequence, but this did not result in the identification of additional homolog sequences. For IGFIR and IR homologs, the ancestral *Branchiostoma floridae* ILPR (XP_035668283.1) was used. IGFBP homologs were identified using the ancestral *B. floridae* IGFBP (XP_035666669.1) as query, supplemented with human IGFBP-4 (NP_001543.2) query hits. STC homologs were identified using human STC1 (NP_003146.1) and a BLOSUM45 scoring matrix, supplemented with hits from *Lingula anatina* STC (XP_013419942).

MSA and conservation plot

MSA of pappalysin sequences (accession numbers are listed in supplementary table S2, Supplementary Material online) and selected pappalysin variant sequences was performed using COBALT (Papadopoulos and Agarwala 2007) and visualized in AliView (Larsson 2014) with majority rule consensus characters highlighted. For the selected pappalysin variant, MSAs the following sequences and accession

numbers were used: PAPP-A: *Rhincodon typus* (XP_048469594.1), *Pristis pectinata* (XP_051893072.1), *Mobula hypostoma* (XP_062929858.1), *Danio rerio* (NP_001265734.1), *Takifugu rubripes* (XP_011602829.2), *Latimeria chalumnae* (XP_064420709.1), *Protopterus annectens* (XP_043914113.1), *Xenopus laevis* (XP_018085171.1), *Mus musculus* (NP_067337.1) and *Homo sapiens* (NP_002572.2); PAPP-A2: *Rhincodon typus* (XP_048458031.1), *Pristis pectinata* (XP_051867159.1), *Mobula hypostoma* (XP_062919822.1), *Danio rerio* (XP_695417.4), *Takifugu rubripes* (XP_029686694.1), *Latimeria chalumnae* (XP_064415794.1), *Protopterus annectens* (XP_043941906.1), *Xenopus laevis* (XP_041446424.1), *Mus musculus* (NP_001078845.1) and *Homo sapiens* (NP_064714.2); invPAPP-A: *Trichoplax adhaerens* (XP_002109449.1), *Hydractinia symbiolongicarpus* (XP_057305553.1), *Anneissia japonica* (XP_033096949.1), *Branchiostoma belcheri* (XP_019621199.1), *Styela clava* (XP_039270265.1), *Trichinella pseudospiralis* (KRZ20972.1), *Dermacentor andersoni* (XP_050051616.1), *Pecten maximus* (XP_033761458.1), *Lingula anatina* (XP_013388811.1), *Owenia fusiformis* (CAH1782730.1), *Lineus longissimus* (XP_064650054.1), and *B. neritina* (KAF6027397.1). Conservation plots of each pappalysin variant MSA were made using EMBOSS Plotcon (Rice et al. 2000), with a window size of 30 residues, and visualized using Inkscape v1.3.2 based on the domains of hPAPP-A.

Phylogenetic analysis

A phylogenetic analysis was performed on the homologs of PAPP-A and PAPP-A2, ILP, IGF, and INS, IGFIR, IR, and ILPR, IGFBP, and STC, which were identified in the homolog searches, with a focus on sequences from a diverse array of chordates and selected invertebrate relatives. Amino acid sequences were aligned with Clustal Omega 1.2.4 (Sievers et al. 2011) using standard parameters, and for PAPP-A sequences MAFFT FFT-NS-i v7.525 (Katoh and Standley 2013) and T-Coffee v11.00 (Di Tommaso et al. 2011) was additionally performed using standard parameters. Sequences with poorly aligned regions or missing sequence stretches were removed from the dataset and the final alignments were repeated without these sequences. Phylogenetic analysis was performed by the maximum likelihood method of IQ-TREE 2.3.5 (Minh et al. 2020) using 1000 ultrafast bootstrap approximations (UFBoot) (Hoang et al. 2018), 1000 SH-like approximate likelihood ratio tests (alrt) (Guindon et al. 2010), and 2 threads (on an AMD Ryzen 7 4700 U 2.00 GHz CPU). To find an appropriate substitution model, ModelFinder (Kalyaanamoorthy et al. 2017) was employed. The following substitution models were used in each phylogenetic analysis: PAPP-A and PAPP-A2: Clustal Omega: JTT + I + R5 (seed 69871, general heterogeneous evolution on a single topology (GHOST): seed 874688), MAFFT: JTT + R5 (seed 55610), T-Coffee: JTT + I + R5 (seed 336910); ILP, IGF, and INS: Q.mammal + I + R3 (seed 411737); IGFIR, IR, and ILPR: JTTDCMut + F + R6 (seed 521133); IGFBP: JTT + I + R4 (seed 90740); and STC: JTT + I + R4 (seed 802453). Species and sequences accession number used are listed in (supplementary tables S2 to S6, Supplementary Material online). Due to the short branch separating lamprey PAPP-A from the previous node, an additional phylogenetic inference was performed using the GHOST model (Crotty et al. 2020) that accounts for heterotachy. To find the appropriate number of classes, the analysis was run with unlinked parameters, with an increasing number of classes (2 to 6 classes). Since the Akaike Information Criterion score kept decreasing with increasing classes, the best fitting number of classes was determined based on the lowest class above (or equal to) the lowest Bayesian Information Criterion score that only contained non-negligible class weights (the lowest weight being 0.1677), resulting in 4 classes being the best fit. Using 4 classes with linked parameters, with either empirical or inferred base frequencies, did not provide a better fit. The resulting trees were visualized in FigTree v1.4.4 (Rambaut 2018) and Inkscape v1.3.2.

Structure prediction

Protein structure predictions were performed using AlphaFold 3 accessed via the AlphaFold Server (Abramson et al. 2024). Structure prediction of *B. neritina* invPAPP-A was performed using a single pappalysin sequence (KAF6027397.1), 6 Ca^{2+} ions, and 1 Zn^{2+} ion (seed 28612464). Prediction of the hPAPP-A2 structure was based on a single, mature hPAPP-A2 sequence (235-1791, NP_002572.2), 8 Ca^{2+} ions, and 1 Zn^{2+} ion (seed 1399085819). Structure prediction of *B. belcheri* invPAPP-A was based on a single pappalysin sequence (81-1651, XP_019621199.1), 9 Ca^{2+} ions, and 1 Zn^{2+} ion (seed 1782406552). Structure prediction of the hPAPP-A:IGFBP-4:IGF complex was performed using 2 mature hPAPP-A sequences (82-1627, NP_002572.2), 2 mature human IGFBP-4 sequences (22-258, NP_001543.2), 2 mature human IGFI sequences (49-118, P05019), 16 Ca^{2+} ions, and 2 Zn^{2} ions (seed 2126702040). Structural figures were prepared by using PyMOL Molecular Graphics System v3.0.3 (Schrödinger 2024).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Funding

This work was supported by grants to CO from Independent Research Fund Denmark and the Novo Nordisk Foundation.

Data Availability

All accession numbers used in the analyses are listed in the main text or in tables of the supplementary information. The MSA generated from these are uploaded as supplementary material.

Literature Cited

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024:630(8016):493–500. https://doi.org/10.1038/s41586-024-07487-w.

Allard JB, Duan C. IGF-Binding Proteins: why do they exist and why are there so many? Front Endocrinol (Lausanne). 2018:9:117. https://doi.org/10.3389/fendo.2018.00117.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990:215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Argente J, Chowen JA, Pérez-Jurado LA, Frystyk J, Oxvig C. One level up: abnormal proteolytic regulation of IGF activity plays a role in human pathophysiology. EMBO Mol Med. 2017:9(10):1338–1345. https://doi.org/10.15252/emmm.201707950.

Babiker A, Al Noaim K, Al Swaid A, Alfadhel M, Deeb A, Martín-Rivada Á, Barrios V, Pérez-Jurado LA, Alfares A, Al Alwan I, et al. Short stature with low insulin-like growth factor 1 availability due to pregnancy-associated plasma protein A2 deficiency in a Saudi family. Clin Genet. 2021:100(5):601–606. https://doi.org/10.1111/cge.14030.

Becker MA, Haluska P Jr, Bale LK, Oxvig C, Conover CA. A novel neutralizing antibody targeting pregnancy-associated plasma protein-a inhibits ovarian cancer growth and ascites accumulation in patient mouse tumorgrafts. Mol Cancer Ther. 2015:14(4):973–981. https://doi.org/10.1158/1535-7163.MCT-14-0880.

Belfiore A, Frasca F, Pandini G, Sciacca L, Vigneri R. Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease. Endocr Rev. 2009:30(6):586–623. https://doi.org/10.1210/er.2008-0047.

Bode W, Gomis-Rüth FX, Stöckler W. Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and met-turn) and topologies and should be grouped into a common family, the 'metzincins'. FEBS Lett. 1993:331(1-2):134–140. https://doi.org/10.1016/0014-5793(93)80312-I.

Boldt HB, Kjaer-Sorensen K, Overgaard MT, Weyer K, Poulsen CB, Sottrup-Jensen L, Conover CA, Giudice LC, Oxvig C. The Lin12-notch repeats of pregnancy-associated plasma protein-A bind calcium and determine its proteolytic specificity. J Biol Chem. 2004:279(37):38525–38531. https://doi.org/10.1074/jbc.M405222200.

Boldt HB, Overgaard MT, Laursen LS, Weyer K, Sottrup-Jensen L, Oxvig C. Mutational analysis of the proteolytic domain of pregnancy-associated plasma protein-A (PAPP-A): classification as a metzincin. Biochem J. 2001:358(Pt 2):359–367. https://doi.org/10.1042/bj3580359.

Brown J, Delaine C, Zaccheo OJ, Siebold C, Gilbert RJ, van Boxel G, Denley A, Wallace JC, Hassan AB, Forbes BE, et al. Structure and functional analysis of the IGF-II/IGF2R interaction. EMBO J. 2008:27(1):265–276. https://doi.org/10.1038/sj.emboj.7601938.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009:10(1):421. https://doi.org/10.1186/1471-2105-10-421.

Chan SJ, Cao QP, Steiner DF. Evolution of the insulin superfamily: cloning of a hybrid insulin/insulin-like growth factor cDNA from amphioxus. Proc Natl Acad Sci U S A. 1990:87(23):9319–9323. https://doi.org/10.1073/pnas.87.23.9319.

Chang AC-M, Hook J, Lemckert FA, McDonald MM, Nguyen M-AT, Hardeman EC, Little DG, Gunning PW, Reddel RR. The murine stanniocalcin 2 gene is a negative regulator of postnatal growth. Endocrinology. 2008:149(5):2403–2410. https://doi.org/10.1210/en.2007-1219.

Christians JK, Bath AK, Amiri N. Pappa2 deletion alters IGFBPs but has little effect on glucose disposal or adiposity. Growth Horm IGF Res. 2015:25(5):232–239. https://doi.org/10.1016/j.ghir.2015.07.001.

Conover CA, Bale LK, Overgaard MT, Johnstone EW, Laursen UH, Füchtbauer E-M, Oxvig C, van Deursen J. Metalloproteinase pregnancy-associated plasma protein A is a critical growth regulatory factor during fetal development. Development. 2004:131(5):1187–1194. https://doi.org/10.1242/dev.00997.

Conover CA, Bale LK, Oxvig C. Targeted inhibition of pregnancy-associated plasma protein-A activity reduces atherosclerotic plaque burden in mice. J Cardiovasc Transl Res. 2016:9(1):77–79. https://doi.org/10.1007/s12265-015-9666-9.

Conover CA, Boldt HB, Bale LK, Clifton KB, Grell JA, Mader JR, Mason EJ, Powell DR. Pregnancy-associated plasma protein-A2 (PAPP-A2): tissue expression and biological consequences of gene knockout in mice. Endocrinology. 2011:152(7):2837–2844. https://doi.org/10.1210/en.2011-0036.

Conover CA, Kiefer MC, Zapf J. Posttranslational regulation of insulin-like growth factor binding protein-4 in normal and transformed human fibroblasts. Insulin-like growth factor dependence and biological studies. J Clin Invest. 1993:91(3):1129–1137. https://doi.org/10.1172/JCI116272.

Conover CA, Oxvig C. PAPP-A: a promising therapeutic target for healthy longevity. Aging Cell. 2017:16(2):205–209. https://doi.org/10.1111/acel.12564.

Conover CA, Oxvig C. PAPP-A and cancer. J Mol Endocrinol. 2018:61(1):T1–T10. https://doi.org/10.1530/JME-17-0236.

Conover CA, Oxvig C. The pregnancy-associated plasma protein-A (PAPP-A) story. Endocr Rev. 2023:44(6):1012–1028. https://doi.org/10.1210/endrev/bnad017.

Conover CA, Oxvig C. The insulin-like growth factor system and aging. Endocr Rev. 2024:46(2):214–223. https://doi.org/10.1210/endrev/bnae029.

Crotty SM, Minh BQ, Bean NG, Holland BR, Tuke J, Jermiin LS, Haeseler AV. GHOST: recovering historical signal from heterotachously evolved sequence alignments. Syst Biol. 2020:69(2):249–264. https://doi.org/10.1093/sysbio/syz051.

Dauber A, Muñoz-Calvo MT, Barrios V, Domené HM, Kloverpris S, Serra-Juhé C, Desikan V, Pozo J, Muzumdar R, Martos-Moreno GÁ, et al. Mutations in pregnancy-associated plasma protein A2 cause short stature due to low IGF-I availability. EMBO Mol Med. 2016:8(4):363–374. https://doi.org/10.15252/emmm.201506106.

Daza DO, Sundström G, Bergqvist CA, Duan C, Larhammar D. Evolution of the insulin-like growth factor binding protein (IGFBP) family. Endocrinology. 2011:152(6):2278–2289. https://doi.org/10.1210/en.2011-0047.

Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang J-M, Taly J-F, Notredame C. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. 2011:39(Web Server issue):W13–W17. https://doi.org/10.1093/nar/gkr245.

Forbes BE, McCarthy P, Norton RS. Insulin-like growth factor binding proteins: a structural perspective. Front Endocrinol. 2012:3:38. https://doi.org/10.3389/fendo.2012.00038.

Frasca F, Pandini G, Scalia P, Sciacca L, Mineo R, Costantino A, Goldfine ID, Belfiore A, Vigneri R. Insulin receptor isoform a, a newly recognized, high-affinity insulin-like growth factor ii receptor in fetal and cancer cells. Mol Cell Biol. 1999:19(5):3278–3288. https://doi.org/10.1128/MCB.19.5.3278.

Gagliardi AD, Kuo EYW, Raulic S, Wagner GF, DiMattia GE. Human stanniocalcin-2 exhibits potent growth-suppressive properties in transgenic mice independently of growth hormone and IGFs. Am J Physiol Endocrinol Metab. 2005:288(1):E92–105. https://doi.org/10.1152/ajpendo.00268.2004.

Giribet G, Edgecombe GD. Current understanding of ecdysozoa and its internal phylogenetic relationships. Integr Comp Biol. 2017:57(3):455–466. https://doi.org/10.1093/icb/icx072.

Gomis-Rüth FX. Catalytic domain architecture of metzincin metalloproteases. J Biol Chem. 2009:284(23):15353–15357. https://doi.org/10.1074/jbc.R800069200.

Guevara T, Rodriguez-Banqueri A, Ksiazek M, Potempa J, Gomis-Rüth FX. Structure-based mechanism of cysteine-switch latency and of catalysis by pappalysin-family metallopeptidases. IUCrJ. 2020:7(1):18–29. https://doi.org/10.1107/S2052252519013848.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate Maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010:59(3):307–321. https://doi.org/10.1093/sysbio/syq010.

Guler H-P, Zapf J, Schmid C, Froesch ER. Insulin-like growth factors I and II in healthy man. Estimations of half-lives and production rates. Acta Endocrinol (Copenh). 1989:121(6):753–758. https://doi.org/10.1530/acta.0.1210753.

Hernández-Sánchez C, Mansilla A, de Pablo F, Zardoya R. Evolution of the insulin receptor family and receptor isoform expression in vertebrates. Mol Biol Evol. 2008:25(6):1043–1053. https://doi.org/10.1093/molbev/msn036.

Hoang DT, Chernomor O, Von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018:35(2):518–522. https://doi.org/10.1093/molbev/msx281.

Huang X, Ye H, Feng B, Huang H. Insights into insulin-like peptide system in invertebrates from studies on IGF binding domain-containing proteins in the female mud crab, *Scylla paramamosain*. Mol Cell Endocrinol. 2015:416:36–45. https://doi.org/10.1016/j.mce.2015.08.019.

Ipsa E, Cruzat VF, Kagize JN, Yovich JL, Keane KN. Growth hormone and insulin-like growth factor action in reproductive tissues. Front Endocrinol (Lausanne). 2019:10:777. https://doi.org/10.3389/fendo.2019.00777.

Jepsen MR, Kløverpris S, Mikkelsen JH, Pedersen JH, Füchtbauer E-M, Laursen LS, Oxvig C. Stanniocalcin-2 inhibits mammalian growth by proteolytic inhibition of the insulin-like growth factor axis. J Biol Chem. 2015:290(6):3430–3439. https://doi.org/10.1074/jbc.M114.611665.

Judge RA, Sridar J, Tunyasuvunakool K, Jain R, Wang JCK, Ouch C, Xu J, Mafi A, Nile AH, Remarcik C, et al. Structure of the PAPP-ABP5 complex reveals mechanism of substrate recognition. Nat Commun. 2022:13(1):5500. https://doi.org/10.1038/s41467-022-33175-2.

Junnila RK, List EO, Berryman DE, Murrey JW, Kopchick JJ. The GH/IGF-1 axis in ageing and longevity. Nat Rev Endocrinol. 2013:9(6):366–376. https://doi.org/10.1038/nrendo.2013.67.

Juul A. Serum levels of insulin-like growth factor I and its binding proteins in health and disease. Growth Horm IGF Res. 2003:13(4):113–170. https://doi.org/10.1016/S1096-6374(03)00038-8.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017:14(6):587–589. https://doi.org/10.1038/nmeth.4285.

Kashyap S, Hein KZ, Chini CC, Lika J, Warner GM, Bale LK, Torres VE, Harris PC, Oxvig C, Conover CA, et al. Metalloproteinase PAPP-A regulation of IGF-1 contributes to polycystic kidney disease pathogenesis. JCI Insight. 2020:5(4):e135700. https://doi.org/10.1172/jci.insight.135700.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013:30(4):772–780. https://doi.org/10.1093/molbev/mst010.

Kjaer-Sorensen K, Engholm DH, Jepsen MR, Morch MG, Weyer K, Hefting LL, Skov LL, Laursen LS, Oxvig C. Papp-a2 modulates development of cranial cartilage and angiogenesis in zebrafish embryos. J Cell Sci. 2014:127(Pt 23):5027–5037. https://doi.org/10.1242/jcs.152587.

Kjaer-Sorensen K, Engholm DH, Kamei H, Morch MG, Kristensen AO, Zhou J, Conover CA, Duan C, Oxvig C. Pregnancy-associated plasma protein A (PAPP-A) modulates the early developmental rate in zebrafish independently of its proteolytic activity. J Biol Chem. 2013:288(14):9982–9992. https://doi.org/10.1074/jbc.M112.426304.

Kløverpris S, Mikkelsen JH, Pedersen JH, Jepsen MR, Laursen LS, Petersen SV, Oxvig C. Stanniocalcin-1 potently inhibits the proteolytic activity of the metalloproteinase pregnancy-associated plasma protein-A. J Biol Chem. 2015:290(36):21915–21924. https://doi.org/10.1074/jbc.M115.650143.

Kobberø SD, Gajhede M, Mirza OA, Kløverpris S, Kjær TR, Mikkelsen JH, Boesen T, Oxvig C. Structure of the proteolytic enzyme PAPP-A with the endogenous inhibitor stanniocalcin-2 reveals its inhibitory mechanism. Nat Commun. 2022:13(1):6084. https://doi.org/10.1038/s41467-022-33698-8.

Koneru L, Ksiazek M, Waligorska I, Straczek A, Lukasik M, Madej M, Thøgersen IB, Enghild JJ, Potempa J. Mirolysin, a LysargiNase from Tannerella forsythia, proteolytically inactivates the human cathelicidin, LL-37. Biol Chem. 2017:398(3):395–409. https://doi.org/10.1515/hsz-2016-0267.

Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014:30(22):3276–3278. https://doi.org/10.1093/bioinformatics/btu531.

Laursen LS, Overgaard MT, Søe R, Boldt HB, Sottrup-Jensen L, Giudice LC, Conover CA, Oxvig C. Pregnancy-associated plasma protein-A (PAPP-A) cleaves insulin-like growth factor binding protein (IGFBP)-5 independent of IGF: implications for the mechanism of IGFBP-4 proteolysis by PAPP-A. FEBS Lett. 2001:504(1-2):36–40. https://doi.org/10.1016/S0014-5793(01)02760-0.

Laursen LS, Overgaard MT, Weyer K, Boldt HB, Ebbesen P, Christiansen M, Sottrup-Jensen L, Giudice LC, Oxvig C. Cell surface targeting of pregnancy-associated plasma protein A proteolytic activity. Reversible adhesion is mediated by two neighboring short consensus repeats. J Biol Chem. 2002:277(49):47225–47234. https://doi.org/10.1074/jbc.M209155200.

Lawrence JB, Oxvig C, Overgaard MT, Sottrup-Jensen L, Gleich GJ, Hays LG, Yates JR III, Conover CA. The insulin-like growth factor (IGF)-dependent IGF binding protein-4 protease secreted by human fibroblasts is pregnancy-associated plasma protein-A. Proc Natl Acad Sci U S A. 1999:96(6):3149–3153. https://doi.org/10.1073/pnas.96.6.3149.

Li J, Liu Z, Kang T, Li M, Wang D, Cheng CHK. Igf3: a novel player in fish reproduction†. Biol Reprod. 2021:104(6):1194–1204. https://doi.org/10.1093/biolre/ioab042.

Li M. The origination of growth hormone/insulin-like growth factor system: a story from ancient basal chordate Amphioxus. Front Endocrinol (Lausanne). 2022:13:825722. https://doi.org/10.3389/fendo.2022.825722.

Marlétaz F, Timoshevskaya N, Timoshevskiy VA, Parey E, Simakov O, Gavriouchkina D, Suzuki M, Kubokawa K, Brenner S, Smith JJ, et al. The hagfish genome and the evolution of vertebrates. Nature. 2024:627(8005):811–820. https://doi.org/10.1038/s41586-024-07070-3.

Marouli E, Graff M, Medina-Gomez C, Lo KS, Wood AR, Kjaer TR, Fine RS, Lu Y, Schurmann C, Highland HM, et al. Rare and low-frequency coding variants alter human adult height. Nature. 2017:542(7640):186–190. https://doi.org/10.1038/nature21039.

Marshall CR. Explaining the Cambrian "explosion" of animals. Annu Rev Earth Planet Sci. 2006:34(1):355–384. https://doi.org/10.1146/annurev.earth.33.031504.103001.

Miller BS, Rogol AD, Rosenfeld RG. The history of the insulin-like growth factor system. Horm Res Paediatr. 2022:95(6):619–630. https://doi.org/10.1159/000527123.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020:37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015.

Monget P, Mazerbourg S, Delpuech T, Maurel M-C, Manière S, Zapf J, Lalmanach G, Oxvig C, Overgaard MT. Pregnancy-associated plasma protein-A is involved in insulin-like growth factor binding protein-2 (IGFBP-2) proteolytic degradation in bovine and porcine preovulatory follicles: identification of cleavage site and characterization of IGFBP-2 degradation. Biol Reprod. 2003:68(1):77–86. https://doi.org/10.1095/biolreprod.102.007609.

Nagao H, Cai W, Wewer Albrechtsen NJ, Steger M, Batista TM, Pan H, Dreyfuss JM, Mann M, Kahn CR. Distinct signaling by insulin and IGF-1 receptors and their extra- and intracellular domains. Proc Natl Acad Sci U S A. 2021:118(17):e2019474118. https://doi.org/10.1073/pnas.2019474118.

Oberholzer AE, Bumann M, Hege T, Russo S, Baumann U. Metzincin's canonical methionine is responsible for the structural integrity of the zinc-binding site. Biol Chem. 2009:390(9):875–881. https://doi.org/10.1515/BC.2009.100.

Overgaard MT, Boldt HB, Laursen LS, Sottrup-Jensen L, Conover CA, Oxvig C. Pregnancy-associated plasma protein-A2 (PAPP-A2), a novel insulin-like growth factor-binding protein-5 proteinase. J Biol Chem. 2001:276(24):21849–21853. https://doi.org/10.1074/jbc.M102191200.

Oxvig C. The role of PAPP-A in the IGF system: location, location, location. J Cell Commun Signal. 2015:9(2):177–187. https://doi.org/10.1007/s12079-015-0259-9.

Oxvig C, Conover CA. The stanniocalcin-PAPP-A-IGFBP-IGF axis. J Clin Endocrinol Metab. 2023:108(7):1624–1633. https://doi.org/10.1210/clinem/dgad053.

Papadopoulos JS, Agarwala R. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics. 2007:23(9):1073–1079. https://doi.org/10.1093/bioinformatics/btm076.

Paps J, Rossi ME, Bowles AMC, Álvarez-Presas M. Assembling animals: trees, genomes, cells, and contrast to plants. Front Ecol Evol. 2023:11:1185566. https://doi.org/10.3389/fevo.2023.1185566.

Pascual-Anaya J, D'Aniello S, Bertrand S. Editorial: new approaches in chordate and vertebrate evolution and development. Front Cell Dev Biol. 2022:10:917101. https://doi.org/10.3389/fcell.2022.917101.

Pashmforoush M, Chan SJ, Steiner DF. Structure and expression of the insulin-like peptide receptor from amphioxus. Mol Endocrinol. 1996:10(7):857–866. https://doi.org/10.1210/mend.10.7.8813726.

Patton SJ, Luke GN, Holland PW. Complex history of a chromosomal paralogy region: insights from amphioxus aromatic amino acid hydroxylase genes and insulin-related genes. Mol Biol Evol. 1998:15(11):1373–1380. https://doi.org/10.1093/oxfordjournals.molbev.a025865.

Perkins SJ, Gilbert HE, Aslam M, Hannan J, Holers VM, Goodship THJ. Solution structures of complement components by X-ray and neutron scattering and analytical ultracentrifugation. Biochem Soc Trans. 2002:30(Pt 6):996–1001. https://doi.org/10.1042/bst0300996.

Rambaut A. Figtree v1.4.4. Edinburgh: Institute of Evolutionary Biology, University of Edinburgh; 2018.

Rao SS, Kundapura SV, Dey D, Palaniappan C, Sekar K, Kulal A, Ramagopal UA. Cumulative phylogenetic, sequence and structural analysis of Insulin superfamily proteins provide unique structure-function insights. Mol Inform. 2024:43(9):e202300160. https://doi.org/10.1002/minf.202300160.

Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. Trends Genet. 2000:16(6):276–277. https://doi.org/10.1016/S0168-9525(00)02024-2.

Roch GJ, Sherwood NM. Stanniocalcin has deep evolutionary roots in eukaryotes. Genome Biol Evol. 2011:3:284–294. https://doi.org/10.1093/gbe/evr020.

Roed NK, Viola CM, Kristensen O, Schluckebier G, Norrman M, Sajid W, Wade JD, Andersen AS, Kristensen C, Ganderton TR, et al. Structures of insect Imp-L2 suggest an alternative strategy for regulating the bioavailability of insulin-like hormones. Nat Commun. 2018:9(1):3860. https://doi.org/10.1038/s41467-018-06192-3.

Rosen O, Weil S, Manor R, Roth Z, Khalaila I, Sagi A. A crayfish insulin-like-binding protein: another piece in the androgenic gland insulin-like hormone puzzle is revealed. J Biol Chem. 2013:288(31):22289–22298. https://doi.org/10.1074/jbc.M113.484279.

Santos-López J, de la Paz K, Fernández FJ, Vega MC. Structural biology of complement receptors. Front Immunol. 2023:14:1239146. https://doi.org/10.3389/fimmu.2023.1239146.

Schrödinger L. The PyMOL Molecular Graphics System, Version 3.0.3. 2024. Available from: https://www.schrodinger.com/platform/products/pymol/ (downloaded July 10, 2024).

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011:7(1):539. https://doi.org/10.1038/msb.2011.75.

Søe R, Overgaard MT, Thomsen AR, Laursen LS, Olsen IM, Sottrup-Jensen L, Haaning J, Giudice LC, Conover CA, Oxvig C. Expression of recombinant murine pregnancy-associated plasma protein-A (PAPP-A) and a novel variant (PAPP-Ai) with differential proteolytic activity. Eur J Biochem. 2002:269(8):2247–2256. https://doi.org/10.1046/j.1432-1033.2002.02883.x.

Sridar J, Mafi A, Judge RA, Xu J, Kong KA, Wang JCK, Stoll VS, Koukos G, Simon RJ, Eaton D, et al. Cryo-EM structure of human PAPP-A2 and mechanism of substrate recognition. Commun Chem. 2023:6(1):1–14. https://doi.org/10.1038/s42004-023-01032-y.

Steffensen LB, Conover CA, Oxvig C. PAPP-A and the IGF system in atherosclerosis: what's up, what's down? Am J Physiol Heart Circ Physiol. 2019:317(5):H1039–H1049. https://doi.org/10.1152/ajpheart.00395.2019.

Tallant C, García-Castellanos R, Seco J, Baumann U, Gomis-Rüth FX. Molecular analysis of ulilysin, the structural prototype of a new family of metzincin metalloproteases. J Biol Chem. 2006:281(26):17920–17928. https://doi.org/10.1074/jbc.M600907200.

Twigg SM, Baxter RC. Insulin-like growth factor (IGF)-binding protein 5 forms an alternative ternary complex with igfs and the acid-labile subunit. J Biol Chem. 1998:273(11):6074–6079. https://doi.org/10.1074/jbc.273.11.6074.

Varghese R, Gagliardi AD, Bialek PE, Yee S-P, Wagner GF, Dimattia GE. Overexpression of human stanniocalcin affects growth and

reproduction in transgenic mice. Endocrinology. 2002:143(3):868–876. https://doi.org/10.1210/endo.143.3.8671.

Wang D-S, Jiao B, Hu C, Huang X, Liu Z, Cheng CHK. Discovery of a gonad-specific IGF subtype in teleost. Biochem Biophys Res Commun. 2008:367(2):336–341. https://doi.org/10.1016/j.bbrc.2007.12.136.

Weyer K, Boldt HB, Poulsen CB, Kjaer-Sorensen K, Gyrup C, Oxvig C. A substrate specificity-determining unit of three Lin12-notch repeat modules is formed in trans within the pappalysin-1 dimer and requires a sequence stretch C-terminal to the third module. J Biol Chem. 2007:282(15):10988–10999. https://doi.org/10.1074/jbc.M607903200.

Weyer K, Overgaard MT, Laursen LS, Nielsen CG, Schmitz A, Christiansen M, Sottrup-Jensen L, Giudice LC, Oxvig C. Cell surface adhesion of pregnancy-associated plasma protein-A is mediated by four clusters of basic residues located in its third and fourth CCP module. Eur J Biochem. 2004:271(8):1525–1535. https://doi.org/10.1111/j.1432-1033.2004.04061.x.

Zhang H, Shi Y, He M. Molecular identification of an insulin growth factor binding protein (IGFBP) and its potential role in an insulin-like peptide system of the pearl oyster, *Pinctada fucata*. Comp Biochem Physiol B Biochem Mol Biol. 2017:214:27–35. https://doi.org/10.1016/j.cbpb.2017.09.003.

Zhong Y, Duan C. Lamprey IGF-binding protein-3 has IGF-dependent and -independent actions. Front Endocrinol (Lausanne). 2017:7:174. https://doi.org/10.3389/fendo.2016.00174.

Zhou J, Xiang J, Zhang S, Duan C. Structural and functional analysis of the amphioxus IGFBP gene uncovers ancient origin of IGF-independent functions. Endocrinology. 2013:154(10):3753. https://doi.org/10.1210/en.2013-1201.

Associate editor: Beatriz Mello