The evolutionary origin of postsynaptic signalling machineries – Insights from the single-celled sister group to the animals.

Submitted by Tarja Tamara Hoffmeyer to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Biological Sciences

in February 2020

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: 

Acknowledgement

First and foremost, I would like to thank my supervisor Dr Pawel Burkhardt, who gave me the opportunity to investigate this interesting topic. He taught me many of the techniques I used throughout my project, and always took the time for discussions and initiated some interesting collaborations. His enthusiasm about the study of synapse evolution is contagious and he inspired me a lot during my time in his lab.

I would also like to thank Prof Dr Thomas Richards, who supervised me as an external PhD student at the University of Exeter. I would like to thank him for his patience in explaining the principles of phylogeny construction and his very valuable input, specifically on the computational parts of my project.

I would also like to thank Prof Dr James Wakefield for acting as my second supervisor at Exeter.

I am very grateful for the support of Prof Dr Colin Brownlee (former director of the Marine Biological Association). I would also like to thank Prof Dr Daniel Chourrout and Carol Bruce, who together with Dr Pawel Burkhardt, found a solution for me to join the Sars Centre in Bergen, Norway, when Dr Pawel Burkhardt moved his lab there.

I am very grateful for all co-authors/ collaborators that were part of the work for individual chapters. I would like to thank Dr Fiona Savory for teaching me ancestral protein reconstruction and who provided her R scripts for this analysis. I would further like to thank Dr Mads Grønborg from the Novo Nordisk Park in Denmark for the mass spectrometry analysis of isolated protein complexes. Moreover, I would like to thank Prof Dr Radu Aricescu and Dr Thomas Walter, who offered extensive help in our X-ray crystallography trials, both at the Division for Structural Biology in Oxford and at the Diamond Light Source.

Having had the immense luck to work at three vibrant institutions during my PhD, there are many people, I would like to express my gratitude to. I would like to thank all previous and current members of the Burkhardt group, who really made this time enjoyable, who were always open for discussions and made me feel as part of a team. I would like to thank Angela Ward for the management of the molecular labs at the Marine Biological Association in Plymouth, UK, and Eva-Lena Nordmann for building up the Burkhardt lab and keeping everything running

after our move to the Sars Centre. I am very grateful for the technical and administrative staff at the MBA, the Sars Centre and the University of Exeter, who made sure that experiments could be performed smoothly and who offered a lot of support, when I moved labs in the middle of my PhD studies.

I would like to thank Dr Charlotte Walker and Dr Glen Wheeler for support with microscopes at the MBA, as well as Dr Marios Chatzigeorgiou for support with the confocal microscope at the Sars Centre.

There are many people that supported me in the last four years, personally and scientifically, and that helped me to endure even in stressful phases. I would like to thank Jenny Arns, Dr Cecilia Balestreri, Abi Bayo, Clemens Döring, Sonja Dunemann, Dr James Gahan, Ronja Göhde, Alix Harvey, Dr Katherine Helliwell, Claire Hopkins, Senna Hiensch, Ruairí Kavanagh, Friedrich Kleiner, Dr Dorothee Kottmeier, Davis Laundon, Dr Carine Le Goff, Dr Michelle Leger, Paula Miramon Puertolas, Chris and Meike Möhle, Dr Gideon Mordecai, Dr Benjamin Naumann, Eva-Lena Nordmann, Andrea Orus Alcalde, Mariano Peruzzo, Albert Pessarrodona Silvestre, Aishwarya Ravi, Amy Romer, Dr Maria Sachkova, Hania Salmonowicz, Dr Samantha Simpson, Dr Joan Soto Angel, Dr Océane Tournière, Madeleine Witt, Jezebel Yamada, Dr Birthe Zäncker, and Jordi Zwiggelaar.

I would like to especially thank Dr James Gahan and Dr Carine Le Goff for proofreading parts of my thesis, as well as Dr Océane Tournière, who was going through the writing stage at the same time and who was incredibly supportive.

I am furthermore grateful to Prof Dr Hartmut Arndt and Dr Frank Nitsche, who introduced me to choanoflagellates in the first place.

Last but not least, I would like to thank my family for their support and encouragement, in particular my father Detlef Beune, who raised me with a curious mindset that led me into research. I would also like to thank my two mothers, my mom Siggi Beune, who made me believe in myself, and my mother Sylvia Hoffmeyer, who supported me every step along the way. I would also like to thank Elzbieta Beune and my grandparents Günter and Thea Beune as well as Richard and Elsbeth Hoffmeyer. Lastly, I would like to express my thanks to my siblings, my brother Josh Hoffmeyer, my sister Tanja Anselmi and her wife Karla Anselmi, and especially Larissa Hoffmeyer, whose support and friendship mean a lot.

Abstract

Choanoflagellates are a diverse monophyletic group of aquatic heterotrophic flagellates that form the sister group to the animals (Leadbeater, 2015). Genome surveys in two closely related choanoflagellate species – *Salpingoeca rosetta* and *Monosiga brevicollis* – showed that choanoflagellates and animals share many genes that are crucial for animal biology, including genes encoding proteins with functions at animal synapses (King et al., 2008; Alié and Manuel, 2010; Fairclough et al., 2013; Burkhardt et al., 2014).

I surveyed 19 choanoflagellate transcriptomes (Richter et al., 2018) for the presence of putative homologs to a key set of synaptic proteins in order to extend our knowledge of the putative ancestral prerequisites for postsynaptic signalling machineries. Postsynaptic signalling machineries are crucial for signal reception and transduction as well as the regulation of signal transduction strength (Kennedy, 2000). Importantly, I could identify putative homologs to Shaker-like potassium channels, nitric oxide synthases and ionotropic glutamate receptors in several choanoflagellate species. The survey further showed that putative homologs of postsynaptic scaffolding proteins (Homer, Shank, and membrane associated guanylate kinases – MAGUKs including Dlg and MAGUK p55) occur in choanoflagellates that branch throughout the phylogenetic radiation of this group.

The high degree of structural conservation in *S. rosetta* Dlg, and Homer homologs suggests that these proteins are of functional importance in choanoflagellates. Furthermore, my data indicate that the scaffolding function of both of these proteins is conserved in choanoflagellates. Combining ancestral protein reconstruction with *in vitro* binding assays, allowed me to establish that the capacity of Homer to bind its synaptic binding partner Shank presumably preceded the evolution of animals and choanoflagellates. Moreover, in an experiment using co-immunoprecipitation in combination with mass spectrometry analysis, I investigated *in vivo S. rosetta* Dlg interaction partners. I found evidence that the interaction between Dlg and MAGUK p55 might be conserved in choanoflagellates. This type of interaction was observed at animal postsynapses and tight junctions (Stucke et al., 2007; Rademacher et al., 2016). My data suggest that synaptic scaffolding complexes might have preceded the evolution

of animals. Synaptic signalling machineries therefore presumably were built upon pre-existing structural scaffolds.

Table of Contents:

List of Figures:

List of tables:

1 Introduction

Postsynaptic signalling machineries are protein complexes regulating processes occurring at the site of synaptic signal transduction (Kennedy, 2000). Particularly in vertebrate animals, the composition and functionality of these complexes have been investigated in depth (for example Blackstone and Sheng, 2002; Collins et al., 2006; Voglis and Tavernarakis, 2006; Fernández et al., 2009; Klemmer et al., 2009; Kegel et al., 2013). However, little is known about the evolutionary origin of these complexes. Choanoflagellates, the single-celled sister group to the animals (Carr et al., 2008; King et al., 2008; Ruiz-Trillo et al., 2008), can be used for comparative studies to resolve which proteins and protein-protein interactions required for these signalling machineries are putatively ancestral and originated before the emergence of animal synapses and neurons.

In this introduction, I will first define and explain the functionality and significance of postsynaptic signalling machineries. I will pay particular emphasis on glutamatergic synapses, which are suggested to occur in all animal lineages with a nervous system (Kass-Simon and Pierobon, 2007; Moroz et al., 2014; Burkhardt and Sprecher, 2017). I focus on three scaffolding proteins that are known to organise signalling machineries at vertebrate glutamatergic synapses. These proteins were suggested to be conserved in animals and choanoflagellates (Alié and Manuel, 2010; Burkhardt et al., 2014) and therefore serve as good starting point to investigate ancestral signalling machineries built on this scaffold. Second, I will provide background about the current knowledge of nervous system evolution. Here, I will also discuss approaches currently used to address questions related to this topic. Third, I will introduce choanoflagellates and state how they make useful models to investigate the evolutionary origin of synaptic proteins and particular signalling machineries. Lastly, I will define the purpose of this study and will state my hypotheses that direct this thesis and the specific aims laid out to test these hypotheses.

1.1 Postsynaptic signalling machineries

1.1.1 Cellular signalling machineries

Every organism is dependent on cellular signalling machineries. Single-celled organisms need signalling processes that enable them to perceive and react to environmental changes and undergo processes such as growth, mitosis and meiosis (de Nadal et al., 2011). Multicellular organisms similarly depend on these basic machineries (de Nadal et al., 2011). On top of that, these organisms require mechanisms to coordinate cells within tissues. Mainly, they need measures for cell-cell communication and adherence (Baluška et al., 2003; Seymour et al., 2004; Richter and King, 2013). More complex organisms need systems that allow them to react effectively across the organism in a coordinated manner. These systems include hormone systems, immune systems and nervous systems. Hormone systems, enabling body wide coordination across tissues, are present in animals, plants and fungi (Huxley, 1935; Zakelj-Mavri et al., 1995; Gaspar et al., 2003). Immune systems, protecting the host from infections, are known from animals and plants, although only animals have an adaptive, body wide immune response (Ausubel, 2005). Nervous systems, allowing fast and precise perception and reaction, as well as learning and memory, were only described in animals (Brodal, 2004). All these systems require specialised cell types that fulfil specific processes, which are likewise based on signalling machineries (Achim and Arendt, 2014). A common feature of all signalling machineries is the involvement of proteins that activate, inactivate and coordinate each other (Alberts et al., 2002).

1.1.2 Synaptic signalling

In intercellular communication, signalling machineries function across cellular borders. There is local cell communication via signal molecule diffusion that work on neighbouring cells (paracrine signalling), system wide signalling via hormones (endocrine signalling) and targeted synaptic signalling through nerve cells (neurons) (Alberts et al., 2002).

Synapses are specialised structures in between two neurons (or a neuron and an effector cell such as a muscle cell), specialised for the quick transfer of

information between cells. Signal transduction can be electrical (electrical synapse via gap junctions) or chemical (chemical synapse) (Squire et al., 2008). In chemical synapses, an electrical signal in one neuron is translated into a chemical signal (neurotransmitter or neuropeptide) and then again converted into an electrical signal in the other neuron. Stimulation of a neuron through neurotransmitters leads to the opening of ion channels, inducing (excitatory synapse) or inhibiting (inhibitory synapse) membrane depolarisation. In vertebrates, glutamate, acetylcholine and aspartate are typical excitatory neurotransmitters, which bind receptors acting as cation channels; GABA and glycine are typical inhibitory neurotransmitters, which bind receptors acting as chloride channels (Squire et al., 2008). Cations (positively charged ions) can induce depolarisation of the membrane, whereas negative chloride ions lead to hypopolarisation inhibiting depolarisation through another synapse acting on the same neuron. Upon strong depolarisation, an electrical impulse (action potential) is formed and leads to the propagation of the signal to the synaptic terminal usually via the stepwise opening of voltage-gated cation channels. At the synaptic terminal the action potential induces the opening of calcium channels in the presynaptic membrane. Calcium influx then leads to the release of neurotransmitters into the synaptic cleft. Neurotransmitters bind to receptors (ligand gated ion channels) in the dendritic membrane of a second neuron (postsynaptic membrane). The ion channel opens, leading to the influx of ions in the second neuron, which may result in depolarisation (excitatory neurotransmitter) (Figure 1-1) or hypopolarisation (inhibitory neurotransmitter). These processes (electrophysiologically observed by Fatt and Katz, 1951; Fatt and Katz, 1952; Hodgkin, 1951; further described by Del Castillo and Katz, 1954; Palay, 1956; inhibitory neurotransmitter effect reviewed by Kuffler and Edwards, 1958) are relatively straightforward and found in every biology school book. However, the underlying biochemical processes are quite complex and every single step requires many proteins interacting in signalling complexes.

There are presynaptic signalling machineries in the active zone (allocated to the presynaptic membrane adjacent to the synaptic cleft) that enable the calcium-dependent fusion of synaptic vesicles with the cell membrane, in order to release the neurotransmitters inside of these vesicles (Augustine et al., 1999). Likewise, there are postsynaptic signalling machineries (allocated to the

postsynaptic membrane adjacent to the synaptic cleft) that, upon binding of neurotransmitters to the receptors, enable signal transduction and modulation of the synaptic strength by the growth of dendrites and the recruiting of more receptors (Sheng and Kim, 2011). In fact, proteins are so densely packed at the postsynaptic membrane that this region appears as very electron-dense in electron micrographs (Boeckers, 2006; Figure 1-2 A). The region is therefore called postsynaptic density (PSD).

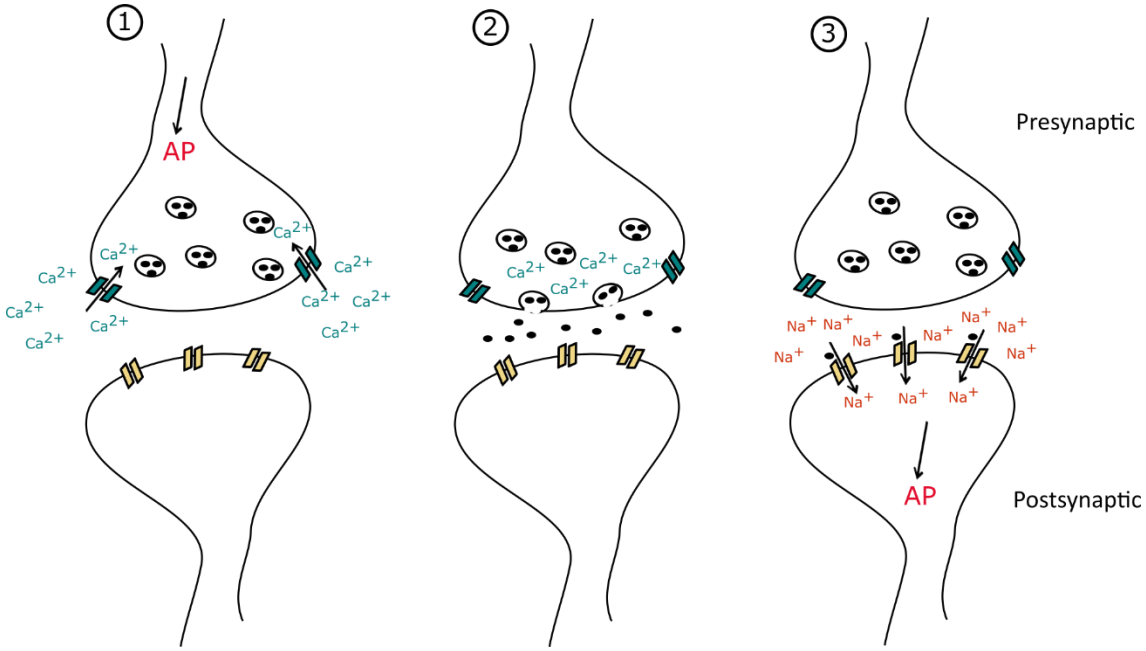


Figure 1-1: Function of chemical synapses. 1) An action potential arrives in the presynaptic terminal of the first neuron, which leads to the opening of voltage-gated calcium channels. 2) Calcium influx triggers membrane fusion between neurotransmitter-filled vesicles and the plasma membrane, resulting in the release of neurotransmitters into the synaptic cleft. 3) Neurotransmitters bind to ligand gated ion channels in the postsynaptic membrane, which opens these channels and leads to ion influx into the second neuron. Sodium ions, as shown in this example, induce depolarisation and another action potential is formed. Adapted from (Kandel et al., 2000).

1.1.3 The postsynaptic density

Complexes formed by PSD proteins are involved in many processes, such as signal transduction, calcium signalling, cytoskeleton organisation, and the regulation of synaptic strength (described in section 1.1.5) (Cho et al., 1992; Sala et al., 2001; Boeckers, 2006; Shiraishi-Yamaguchi and Furuichi, 2007; Kim et al.,

2009). *In vitro*, PSD proteins self-assemble at physiological concentrations via liquid-liquid phase separation (Zeng et al., 2018). *In vivo*, the PSD complex is attached to membrane regions with properties of lipid rafts – small, highly dynamic regions in the membrane with enrichment in cholesterol and sphingolipid content (Pike, 2006; Suzuki et al., 2011). Organisation of proteins in lipid rafts has been shown to allow signal transmission across different cellular membrane systems in calcium signalling (Weerth et al., 2007). Anchoring of the PSD complex to the membrane is carried out by scaffolding proteins (Sheng and Kim, 2011). These proteins interact with many other proteins, thereby connecting receptors in the membrane with signalling molecules, and organising the whole complex (Pawson and Scott, 1997; Sheng and Kim, 2011; Kim and Sheng, 2004; Shiraishi-Yamaguchi and Furuichi, 2007).

1.1.4 Postsynaptic scaffolding proteins

Scaffolding proteins are amongst the most abundant proteins in the PSD (Cheng et al., 2006; Hayashi et al., 2009). These proteins are the organisers of protein complexes, as they form multimeric scaffolds and also interact with many other proteins via protein-protein interaction domains (Zhu et al., 2016). The functionality of PSD scaffolding proteins has been studied in depth in vertebrate synapses (reviewed in Garner et al., 2000; Montgomery et al., 2004). Three scaffolding proteins of crucial importance for postsynaptic signalling machineries at vertebrate glutamatergic synapses are PSD-95, Shank and Homer (Figure 1-2 B). These proteins anchor receptors and adhesion molecules and interconnect receptor-mediated processes with downstream signalling via protein-protein interactions (Irie et al., 1997; Tu et al., 1998; Naisbitt et al., 1999; Sala et al., 2001; Kim and Sheng, 2004) (Figure 1-2 C).

PSD-95 is a homolog of *Drosophila* Discs large (Dlg) (te Velthuis et al., 2007). It is known as the key organiser of the PSD at glutamatergic synapses, because it anchors PSD complexes to ionotropic glutamate receptors (iGluRs) (Kim and Sheng, 2004) and seems to be one of the most abundant proteins in the PSD (Cheng et al., 2006). It belongs to the protein family of membrane associated guanylate kinases (MAGUKs). Proteins of this family do not really have guanylate kinase function, but all their domains are specialised for protein-

protein interaction (McGee et al., 2001; Verpelli et al., 2012; Won et al., 2017). Dlg proteins can recruit receptors to the membrane (a process that is regulated via phosphorylation through Ca^{2+}/calmodulin-dependent protein kinase II (CAMKII)) and furthermore bring signalling molecules such as nitric oxide synthase, kinases and phosphatases in proximity to receptors (Blackstone and Sheng, 2002; Yamada et al., 2007). Details about the Dlg/PSD-95 domain architecture and details about the binding capacities of the different domains are given in section 4.1.

The proteins Shank and Homer are crucial for the formation of a multimeric platform (Shank) (Baron et al. 2006) and the linkage of the PSD complex to signalling pathways and the cytoskeleton (Homer & Shank) (Naisbitt et al. 1999; Sala et al. 2001). Homer proteins form tetramers (Hayashi et al. 2006; 2009). They directly interact with Shank, metabotropic glutamate receptors (mGluRs) and inositole-1,4,5-triphosphate (IP3) receptors (Tu et al., 1998, 1999; Beneken et al., 2000). Homer and Shank act together as a structural framework for the binding of other PSD-proteins (Baron et al., 2006). This structural framework forms a mesh-like Shank-Homer matrix structure, which can be visualised by electron microscopy (Hayashi et al. 2009). They interconnect different receptor complexes (mainly through the linkage of PSD-95 and Shank via GKAP, CRIPT and IRSp53) and bind signalling molecules with an impact on calcium signalling and actin cytoskeleton remodelling (Naisbitt et al., 1999; Sala et al., 2001; Valtschanoff and Weinberg, 2001; Soltau et al., 2004; Baron et al., 2006; Hayashi et al., 2009). The impact on cytoskeleton remodelling and calcium signalling is described in sections 1.1.5.2 and 1.1.5.3, respectively. Details about Homer and Shank domain architectures as well as domains and motifs involved in binding are given in section 3.1.

1.1.5 Functions of PSD scaffolding proteins

The main function of PSD scaffolding proteins is the formation and organisation of protein complexes (Zhu et al., 2016). Apart from providing the structural framework for the PSD (Okabe, 2007), the many protein-protein interactions they are involved in make them crucial components for several cellular processes that enable synaptic functions.

1.1.5.1 Synapse functioning, learning and memory

At glutamatergic synapses, Dlg homologs recruit ionotropic glutamate receptors (Rumbaugh et al., 2003) and are the link between these receptors and proteins that react to cation influxes (Blackstone and Sheng, 2002) (section 0). Dlg proteins can thereby modulate synaptic plasticity, such as long term potentiation (LTP, strengthening of a synapse) and long term depression (LTD, weakening of a synapse), mechanisms crucial for learning and memory (Xu, 2011). Homer and Shank also play a role in learning and memory, as they are involved in spine head growth via cytoskeleton remodelling and intracellular calcium signalling (Sala et al., 2001) (sections 1.1.5.2 and 1.1.5.3).

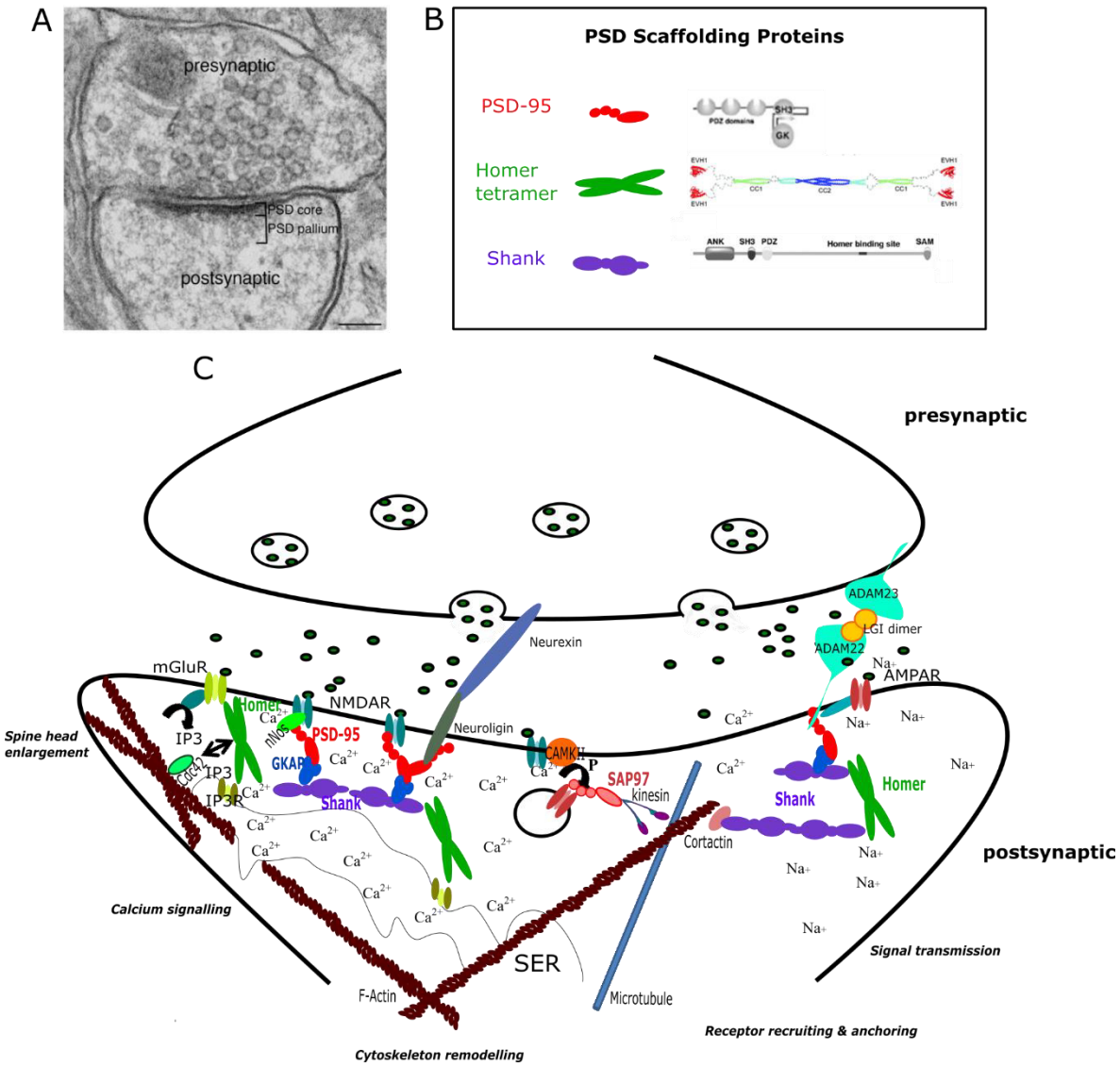


Figure 1-2: The Postsynaptic Density (PSD). A) Electron micrograph from Dosemeci et al. (2016) (Frontiers in Synaptic Neuroscience, Creative commons license) showing the presynaptic neuron with synaptic vesicles, the synaptic cleft and the electron-dense PSD adjacent to the membrane of the postsynaptic neuron. B) PSD scaffolding proteins with domain architecture (adapted from McGee et al., 2001; Hayashi et al., 2009; Burkhardt, 2015 with permission from Molecular Cell (Elsevier), Cell (Elsevier), and Journal of Experimental Biology (The Company of Biologists), respectively). C) Protein complexes of the PSD organised by PSD-95, Shank and Homer (on the basis of the following papers: Irie et al. (1997); Blackstone and Sheng (2002); Kim and Sheng (2004); Shiraishi-Yamaguchi and Furuichi (2007); Feng and Zhang (2009); Kegel et al. (2013); Zhu et al. (2016)). Shown are interactions of Homer and Shank with the actin cytoskeleton, the Homer interaction with mGluRs and IP3Rs involved in calcium signalling, and selected interactions of PSD-95 in complexes at NMDA and AMPA receptors and in interaction with proteins connecting into the presynaptic active zone. CAMKII activated by Ca^{2+} induces the exocytosis of more AMPA receptors, shown in a hypothetical process as suggested by Yamada et al. (2007) for SAP97 receptor recruitment with binding of SAP97 to a kinesin motor protein that moves along microtubules.

1.1.5.2 Cytoskeleton remodelling

Homer and Shank act together in the remodelling of the postsynaptic cytoskeleton. Shank's proline rich region can interact with the actin nucleation factor cortactin (Naisbitt et al., 1999). Homer binds F-actin in vitro and interacts with activated Cdc42 (a Rho GTPase that was shown to be involved in the restructuring of the actin cytoskeleton in spine enlargement and stability (Caroni et al., 2012)) in a heterologous expression system (Shiraishi et al., 1999). Sala et al. (2001) demonstrated that Homer and Shank are involved in spine head growth (a consequence of LTP). Two proteins interacting with PSD-95 and Shank are also connected to the cytoskeleton (Passafaro et al., 1999; Krugmann et al., 2001; Valtschanoff and Weinberg, 2001; Soltau et al., 2004). CRIPT binds microtubules (Passafaro et al., 1999) and IRSp53 interacts with Cdc42 (Krugmann et al., 2001).

1.1.5.3 Calcium signalling

Homer plays an important role in calcium signalling. With its EVH1 domain it binds to the C-terminal PPxxF motif of mGluRs (Beneken et al., 2000) that are localised to the postsynaptic membrane and are connected to G-proteins activating several cellular pathways upon glutamate binding. One of these pathways induces the production of IP3, a second messenger that can trigger the intracellular release of calcium from the endoplasmic reticulum (ER) and

subsequent activation of protein kinase C (PKC) via its binding to IP3 receptors (Szumlinski and Woodward, 2014). Interestingly, Homer also binds to IP3 receptors (same binding mechanism), coupling mGluR signalling and calcium release from the ER (Tu et al., 1998). Yuan et al. (2003) showed that Homer also binds to TRPC1 plasma membrane receptors. Activation of TRPC1 via phosphorylation through protein kinase C leads to calcium influx (Ahmmed et al., 2004). Yuan et al. (2003) suggested that Homer couples the crosstalk between the intracellular calcium machinery and signalling that induces extracellular calcium entry. Kang et al. (2016) demonstrated that Homer can also bind ryanodine (RYR) receptors *in vitro*. This interaction was already suspected by Pouliquin and Dulhunty (2009) who found PPxxF and similar motifs in RYRs. RYRs also release calcium from the ER (reviewed by Lanner et al., 2010).

1.1.5.4 Non-synaptic functions

Dlg homologs and other MAGUKs in general anchor protein complexes to receptors in the membrane at cell-cell contact sites (Ebnet, 2008). Therefore, they play important roles at adherens junctions (shown for *Caenorhabditis elegans* Dlg (Firestein and Rongo, 2001)), septate junctions and neuromuscular junctions (shown for *Drosophila melanogaster* Dlg (Woods and Bryant, 1993; Lahey et al., 1994)). Dlg is an important tumour suppressor, regulating cell polarity and proliferation in *Drosophila* and probably also in humans (Humbert et al., 2003; Bergstralh and St Johnston, 2012). In epithelia, the proteins Scribble, Lgl and Dlg form one of three modules required to achieve apico-basal cell polarity (Su et al., 2013). Cell polarity of animal epithelia is facilitated by Dlg binding of the receptor Pins, which aligns orthogonally to the orientation of other cells in the epithelium, and interacting with KHC-73, a motor protein that arranges astral microtubules, to ensure a division of the cell in the same orientation the other cells are in (reviewed by Lu and Johnston (2013). Pins and KHC-73 were both reported to bind Dlg (Johnston et al., 2012; Yu et al., 2006).

Shank3 has been shown to inhibit integrin-dependent processes by sequestering Ras and Rap G proteins in various human cell lines (Lilja et al., 2017). Selective expression of Homer homologs was suggested to promote

muscle differentiation by regulating IP3R versus RYR Ca^{2+} release, of which only the ladder activates the transcription factor NFAT (Stiber et al., 2005).

1.1.5.5 PSD scaffolding proteins in human disease

PSD scaffolding proteins have also been found to play a role in diseases of the nervous system, cancer development and T-cell activation. There are reduced levels of PSD-95 in the hippocampus due to a mutation in an RNA interacting protein in Fragile X syndrome (Bassell and Warren, 2008) or because of mutations in other involved genes leading to synapse elimination in autism (Tsai et al., 2012). A Shank3 mutation leads to neurogenetic deletion syndrome (Bonaglia et al., 2001; Boeckers et al., 2002). SAP97/Dlg-1 is an important tumour suppressor which is targeted by human papilloma viruses (Gardiol et al., 1999). Dlg-1 and Homer proteins 2 and 3 are negative regulators of T-cell activation (Xavier et al., 2004; Huang et al., 2008). Dlg-1 can be recruited to the actin cytoskeleton in T-cells, where it acts as negative regulator of their activation (Xavier et al., 2004). Homer competes with the activator calcineurin for the binding of nuclear factor of activated T-cells (NFAT), probably controlling self-reactivity, as Homer deficient mice develop autoimmune symptoms (Huang et al., 2008). Regulation of T-cell activation plays a role in allergic reactions, systemic lupus erythematosus and inflammatory bowel diseases (Neurath et al., 2002; Ling et al., 2004; Fernandez et al., 2006).

1.2 Nervous system evolution

1.2.1 Origin(s) of nervous systems

Nervous systems occur in almost all animal groups, except sponges (sessile filter feeders) and placozoans (animals with only two epithelia that crawl on surfaces and externally digest single-celled algae) (Degnan et al., 2015; Schierwater and Eitel, 2015). Sponges and placozoans belong to animal taxa that split from the lineage that evolved into the Bilateria (comprising most animal phyla including molluscs, flatworms, annelids, arthropods, echinoderms and vertebrates as well as other chordates) (Baguñà et al., 2008; Dunn et al., 2015). Two other animal taxa (cnidarians and ctenophores) are not included in the Bilateria (Dunn et al., 2015). The phylogeny between sponges, placozoans,

cnidarians and ctenophores as well as their relation to Bilateria is still debated. Historically, ctenophores were considered to have a close phylogenetic relationship with cnidarians, which was based on morphological studies (Leuckart, 1848). Dunn et al. (2008) performed a phylogenetic analysis that suggested ctenophores as the sister group to all other animals. Since then, more genomes and transcriptomes of non-bilaterian animals were sequenced (for example Srivastava et al., 2008; Ryan et al., 2013; Moroz et al., 2014; Riesgo et al., 2014). Most recent phylogenies recover Cnidaria as the sister group to the Bilateria and, depending on method, taxon sampling and genes included, suggest either sponges or ctenophores as the sister group to all other animals (Shen et al., 2017; Simion et al., 2017). Considering that animals without nervous systems might have closer relationships to animals with nervous systems has spurred the debate on the origin of nervous systems. Discussed are a single origin with potential loss in sponges and placozoans or multiple origins of the nervous system (Figure 1-3) (Jékely et al., 2015b; Ryan and Chiodin, 2015; Liebeskind et al., 2016; Moroz and Kohn, 2016; Martín-Durán and Hejnol, 2019).

Liebeskind et al. (2016) emphasised the possibility that similarities in synapses and neurons of different lineages do not necessarily imply homology of these features, but might as well be an example of homoplasy, where distinct features converge to a similar phenotype due to selective pressure. They described an evolutionary scenario where neurons and synapses evolved independently in ctenophores and the cnidarian-bilaterian lineage, for example referring to evidence for the convergent transition of voltage-gated calcium to voltage-gated sodium channels in cnidarians and bilaterians. Jékely et al. (2015b) looked at different characters such as neuropeptide signalling, ciliary photoreceptors, gap junctions and presynaptic molecules. They argue that the similarity of these characters in ctenophores and other animals with nervous systems is compatible with a single nervous system origin, irrespective of the phylogenetic position of ctenophores. Even if ctenophores are the sister group to all other animals, the animal last common ancestor could have possessed only a few protoneuron types that served as the precursors to all modern nervous systems (Jékely et al., 2015b). Nervous systems could then have undergone independent complexifications in the ctenophore and the cnidarian-bilaterian

lineages, whereas a simplification occurred in sponges and placozoans (Jékely et al., 2015b).

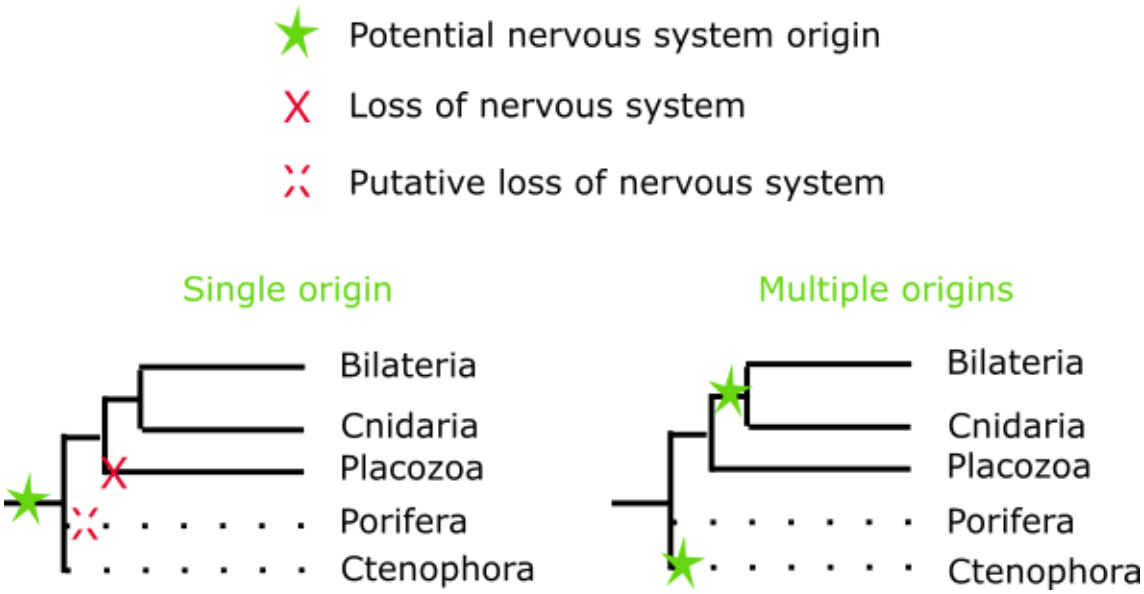


Figure 1-3: Putative nervous system origins based on latest phylogenies. Ctenophora and Porifera phyla are both possible sister groups to all other animals (Shen et al., 2017; Simion et al., 2017). Both possible scenarios are compatible with a single nervous system origin (with loss in Placozoa and putative loss in Porifera) or independent origins in the Ctenophora lineage and the bilaterian-cnidarian lineage. Figure adapted from Jékely, Paps, et al. (2015) and Martín-Durán and Hejnol (2019).

Interestingly, genome surveys showed that all animals (including sponges and placozoans) encode many proteins that are essential for the functioning of bilaterian chemical synapses (Sakarya et al., 2007; Alié and Manuel, 2010).

1.2.2 Conceptual frameworks shaping our view on the origin of nervous systems

In order to learn about the origin of neurons, we need to ask fundamental questions. Are we actually able to precisely define a neuron? In 1888, Santiago Ramón y Cajal discovered and morphologically described the cells that make up the nervous system (Cajal, 1888; López-Muñoz et al., 2006). Morphological features used to describe neurons by their polarised nature as well as the presence of projections such as an axon and dendrites. Although true for many neurons, this definition does not fit all neuron types. Neurons can be non-polarised in nerve nets (Anderson, 1985; Westfall, 1996; Satterlie, 2011) and

some interneurons do not have any obvious projections (Bucher and Anderson, 2015). Definitions that are more functional focus on abilities of neurons such as excitability and the ability to release and respond to neurotransmitters. Bucher and Anderson (2015) for instance, proposed to define a neuron functionally as a cell that electrically communicates with discrete, distant target cells via synapses with pre- and postsynaptic elements. Jékely et al. (2015a) defined neurons as electrically excitable cells with specialised projections that use electrical or sensory mechanism to influence other cells, aiming to include neuroendocrine cells that do not target specific cells via synapses, but release hormones.

In concordance with the difficulties of defining a neuron are difficulties of defining cell types in general. Sachkova and Burkhardt (2019) reviewed how the historical definition of cell types – as sharing both a specific function and a morphology distinct from other cell types – is being challenged by single cell transcriptomic data. These data show that there is a high molecular diversity within cells previously considered as cell types, which could be accredited to a hidden variety of cell types or to temporal cell states as well as cells undergoing developmental changes such as differentiation or reprogramming (Baran et al., 2019; Sachkova and Burkhardt, 2019). Gene regulatory networks control gene differentiation, which suggests that they control cell identity (Davidson, 2010; Sachkova and Burkhardt, 2019). This is demonstrated by the function of master terminal selectors (specification genes in the last step of neurogenesis for different neuron classes in *Caenorhabditis elegans*) (Hobert, 2016; Sachkova and Burkhardt, 2019). A POU homeobox gene is a neuronal terminal selector in *C. elegans* and mouse as well as in the cnidarian *Nematostella vectensis*, supporting previous studies suggesting that neurogenesis principles observed in Bilateria are also applicable to Cnidaria (Quina et al., 2009; Richards and Rentzsch, 2014; Richards and Rentzsch, 2015; Serrano-Saiz et al., 2018; Tournière et al., 2020 preprint). Malosio et al. (1999) showed that changes in posttranscriptional modifications can lead to the loss of whole functional modules as shown by the absence of all genes involved in neurosecretion in mutated cell lines. Arendt et al. (2016a) suggested an evolutionary definition of a cell type, in which the developmental cell lineage is not always equal to the evolutionary lineage. The cell's evolutionary identity is rather defined by the use of a certain core gene regulatory network that has a common evolutionary origin (Arendt et

al., 2016a). They propose a model, in which sister cell types that emerged from the same precursor cell types can be evolutionary retraced, aiding to understand the evolutionary history of neurons and other cell types (Arendt et al., 2016a).

Another basic question asks whether neurons or synapses were the first instance of the nervous system. Different neural origins were suggested that would affect the answer to this question. It is assumed that the first tissues that emerged in animals were epithelia (Tyler 2003). Epithelia form cohesive sheets of cells that can line body surfaces and can form functional units for absorption or secretion (Lowe and Anderson, 2015). Epithelial cells are usually tightly connected through cellular junctions that adhere cellular membranes (adherens junction) and act as a barrier (occluding junction) (Lowe and Anderson, 2015). Harden et al. (2016) noted the similarity between synapses and other cellular junctions and suggested that all cellular junctions evolved from an ancestral septate junction (a form of occluding junction) that emerged in an epithelial context. Starting from simple animals with epithelia, two main hypotheses considering neuron and synapse evolution were suggested (Figure 1-4). The first hypothesis is based on the close functional relationship between neurons and muscle cells and states that muscles could have evolved from a contractile epithelium (myopepithelium) and could have acted as primitive receptor-effector cells, forming first synapses (Parker, 1919; Mackie, 1970; Nickel, 2010). The second hypothesis considers neurosecretion as the ancient effector activity (Grundfest, 1959; Nickel, 2010). Neurosecretion could have its origin in nutrient delivery and the same system could potentially be used for signal transmission, where different nutrients (for example glutamate or small peptides) could be used as signals as well (Varoqueaux and Fasshauer, 2017). These signals could presumably be sensed by surrounding cells, altering their behaviour in a paracrine fashion (Nickel, 2010). In the first scenario, actual neurons interconnecting sensory and muscular cells via chemical transmission could have evolved later, whereas, in the second scenario, neurosecretory cells preceded the evolution of a synapse in between separate cells.

Jékely et al. (2015a) discussed how different concepts of nervous systems influenced hypotheses about nervous system evolution. One concept is an input-output system, in which sensory information is translated to a reaction. The other concept describes nervous systems as a form of internal coordination of tissues,

such as muscles or ciliated epithelia. They conclude that both the coordination as well as the integration of and reaction to sensory information play a role in the neural control of behaviour, physiology and development. Thus, both concepts need to be considered in order to understand the evolutionary boundaries under which nervous systems evolved (Jékely et al., 2015a).

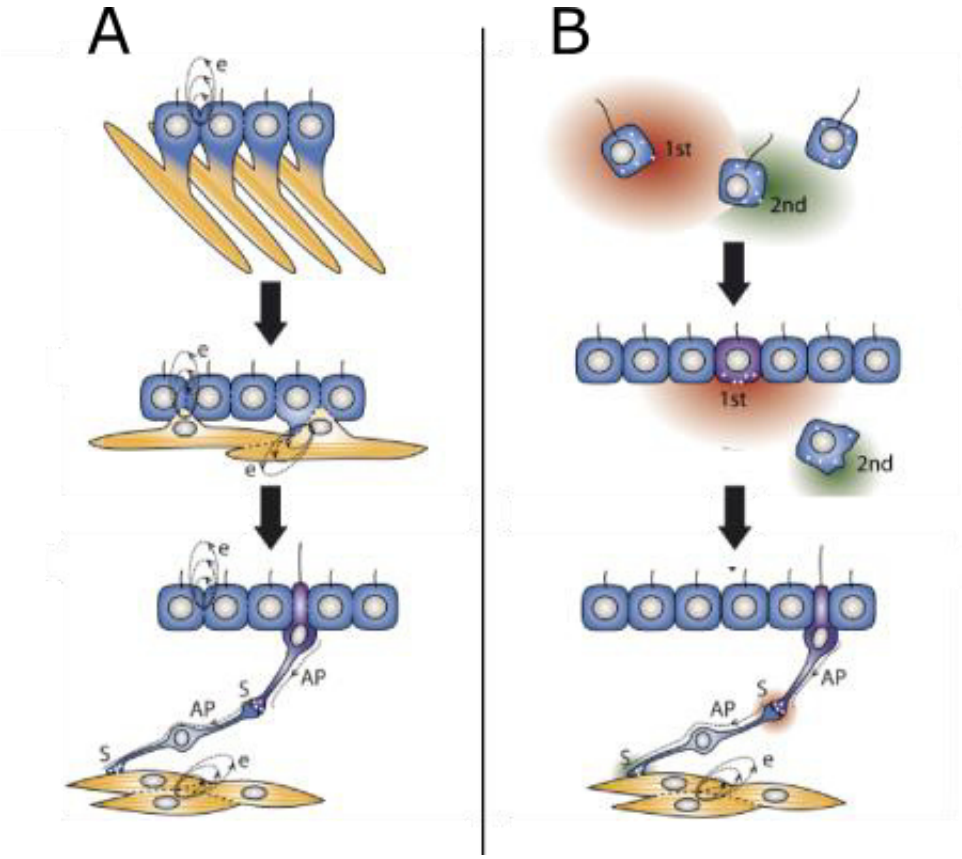


Figure 1-4: Two hypothetical scenarios for the emergence of neurons from epithelia. A) Muscle cells and neurons evolved from electrically coupled (e) myoepithelia. Muscles cells specialised first, then cells acquired means for the chemical transmission of electrical impulses (action potentials: AP) via synapses (S). B) Even before the evolution of epithelia, cells could secrete substances. First neurons differentiated from epithelia, specialising in secretion. These neurosecretory cells evolved into neurons with chemical transmission. Figure modified from Nickel (2010) with elements by Mackie (1970) and Arendt (2008) with permission from Invertebrate Biology (John Wiley & sons), The University of Chicago Press, and Nature (Macmillan Publishers Ltd, respectively

1.2.3 Comparative approaches to study the origin(s) of nervous systems

One approach to resolve nervous system origins is the comparison of nervous systems of extant animal taxa and inferring information into how an ancestral nervous system might have looked like. In the past, these comparisons

have been purely morphological, while they are getting more and more informed by sequence similarities and phylogenetic relationships (Kristan, Jr., 2016; Haen Whitmer, 2018).

Hejnol and Rentzsch (2015) collected phylogenetically informed evidence that first nervous systems were nerve nets that condensed and centralised multiple times in the evolution of animals. They argue that in nerve nets, the random outgrowth of neuronal processes might be sufficient to act on muscular tissue for coordinated movements and to connect to sensory cells allowing movements based on sensory input without target-derived neurite guidance (Hejnol and Rentzsch, 2015). Most bilaterian animals possess condensed nerve structures such as longitudinal nerves and/or centralised brains (Hejnol and Rentzsch, 2015). Many of them also have nerve nets in parts of their body (Hejnol and Rentzsch, 2015). Cnidarians and ctenophores (the two other animal groups with nervous system) also possess nerve nets (Hejnol and Rentzsch, 2015). Cnidarians have some condensed structures, such as ring nerves and larval aboral sensory organs (for settlement) in different cnidarian species (Watanabe et al., 2009). There are different hypotheses about homology of cnidarian and bilaterian nervous systems. Furness and Stebbing (2018) hypothesise that the central nervous system evolved in the bilaterian lineage. They use morphological and physiological similarities between the cnidarian nervous system and the bilaterian enteric nervous system in the gut to propose a common origin of these systems (Furness and Stebbing, 2018). Arendt et al. (2016b) suggest that the aboral sensory organ in cnidarian larvae is homologous to a similar structure in many bilaterian larvae. There is no evidence that the ctenophore aboral organ (containing mechanosensory cells, sensing gravity and coordinating beat frequency of ciliary combs) is homologous to the one found in bilaterian and cnidarian larvae (Jager et al., 2011; Arendt et al., 2016b).

Electron microscopy studies suggested the presence of synapses in different Cnidaria and Ctenophora species (Horridge and Mackay, 1962, 1964; Horridge, 1965; Hernandez-Nicaise, 1973; Westfall, 1996). Cnidarian nervous systems appear to have a wide variety of peptidergic neurons, that use neurotransmission rather than solely neurosecretion, as shown by the presence of RF/RW-amide neuropeptides in synaptic vesicles in *Anthopleura elegantissima* (Anthozoa) (Westfall and Grimmelikhuijzen, 1993; Watanabe et al.,

2009). Several neuropeptides in *Hydra magnipapillata* (e. g. LW-amides) induce muscle contractions (Yum et al., 1998; Takahashi et al., 2003; Watanabe et al., 2009). RF-amides seem to have the important role to coordinate movements, allowing phototactic behaviours and regulated gut contractions (Fujisawa, 2008; Plickert and Schneider, 2004; Watanabe et al., 2009). In addition, genome surveys in *Hydra magnipapillata* and *Nematostella vectensis* revealed the presence of genes for the synthesis and reception of acetylcholine and GABA, as well as genes with similarity to bilaterian glutamate-, epinephrine-, dopamine- and glycine-receptors (Putnam et al., 2007; Chapman et al., 2010; Watanabe et al., 2009). Apart from these receptors, the genomes also encode proteins with similarity to many synaptic scaffolding proteins and proteins involved in neurosecretion, as well as to bilaterian genes for nervous system and neuron cell type specification (Putnam et al., 2007; Chapman et al., 2010; Watanabe et al., 2009). Rentzsch et al. (2017) pointed out the similarity between generic neurogenic programmes of cnidarian and bilaterian nervous systems. This was based on functional studies, suggesting a positive regulation of neurogenesis through *SoxB* genes and a negative regulation of neurogenesis through Notch signalling in both Cnidaria and Bilateria (Richards and Rentzsch, 2014, 2015). Single cell RNA sequencing confirmed that specification genes and other markers are related to distinct neural cell populations in *Nematostella vectensis* (Sebé-Pedrós et al., 2018b).

For ctenophores, bilaterian markers for nervous system and neuron specification could not be assigned to distinct cell populations (Sebé-Pedrós et al., 2018a). Although the ctenophore *Mnemiopsis leidyi* possesses many proteins that resemble bilaterian structural components of chemical synapses, it is not certain to which degree ctenophores utilise bilaterian neurotransmitters (Ryan et al., 2013). The *M. leidyi* genome seems to lack genes encoding enzymes for the synthesis of the neurotransmitters serotonin, dopamine and norepinephrine (Ryan et al., 2013; Simmons and Martindale, 2016). The ctenophore *Pleurobrachia bachei* has a diverse set of putative ionotropic glutamate receptors (Moroz et al., 2014). Isolated, putative muscle cells of this species were shown to be reactive to glutamate, but not to GABA, histamine or acetylcholine (Moroz et al., 2014; Simmons and Martindale, 2016). Previous studies, however, showed

effects of adrenaline, acetylcholine and possibly serotonin on *M. leidyi* bioluminescence flashes (Anctil, 1985; Simmons and Martindale, 2016).

Even though sponges and placozoans lack nervous systems, both phyla have contractile cells and can sense and react to certain neurotransmitters (Nickel, 2010; Senatore et al., 2017; Armon et al., 2018). Most sponges can contract either their entire body or parts such as their water canal system (Nickel, 2010). Contractions have been shown to function in self-cleaning of the water canal system when it gets blocked (Elliott and Leys, 2007). Usually rhythmic contractions can be induced by transmitter substances such as GABA, glutamate and serotonin in the demosponge *Tethya wilhelma* or glutamate in *Oscarella lobularis* (Homoscleromorpha) and *Clathrina clathrus* (Calcispongia) (Ellwanger and Nickel, 2006; Nickel, 2010). Acetylcholine and glycine further seemed to indirectly effect the contraction rhythm in *Tethya wilhelma* (Ellwanger and Nickel, 2006; Nickel, 2010). Contractions were suggested to be mediated by actinocytes in the mesohyl and/or pinacocytes (Elliott and Leys, 2007; Nickel, 2010). A subset of pinacocytes carrying non-motile cilia and lining the inner osculum of the adult demosponge *Ephydatia muelleri* were suggested to be sensory cells (Ludeman et al., 2014). The removal of the osculum or the application of putative blocking agents for TRP channels (important for sensory responses in other organisms and also expressed in the sponge) inhibited contractions despite of glutamate stimulation (Gale et al., 2001; Praetorius and Spring, 2001; Ludeman et al., 2014).

The placozoan species *Trichoplax adhaerens* can also contract (Armon et al., 2018). Contractions occur in cells of the dorsal epithelium, whereas other cells in this epithelium are softened (Armon et al., 2018). The counteraction of these processes was suggested to maintain tissue integrity (Armon et al., 2018). *T. adhaerens* moves via the beating of cilia in its ventral epithelium, which create a gliding motion on mucus that is secreted by gland cells in this epithelium (Ueda et al., 1999; Smith et al., 2015; Mayorova et al., 2019). Other gland cells have been described that supposedly secrete neuropeptides, as they are labelled with an antibody against a putative endomorphin-like propeptide (Senatore et al., 2017). When the surface below *T. adhaerens* is covered in single celled algae, the animal pauses to externally digest them (Ueda et al., 1999; Smith et al., 2015; Senatore et al., 2017). The ciliary arrest leading to this pausing behaviour

depends on external Ca^{2+} ions and can be modulated by endomorphin-like peptide transmitters encoded in the *T. adhaerens* genome as aforementioned propeptide (Senatore et al., 2017). FMRF-amides also induce pausing in some animals at high concentrations, whereas SIFG-amides lead to a partial detachment of the animal from the substrate as well as folding and writhing at higher concentration (Senatore et al., 2017). Varoqueaux et al. (2018) elucidated that there are distinct populations of peptidergic cells in *T. adhaerens* and that the animal responds with different behaviours (crinkling, turning, flattening and churning) to different neuropeptides.

The molecular mechanisms of neurotransmission in Cnidaria and Ctenophora, as well as the function of genes with importance to bilaterian neurotransmission in all non-bilaterian animals remain elusive, but it appears that both the use of neuropeptides and traditional neurotransmitters for behavioural integration is conserved among all animal phyla. Varoqueaux and Fasshauer (2017) suggested that neurotransmission evolved from a mechanism that delivered nutrients from the lysosome to the plasma membrane. Lysosomal degradation is used in heterotrophic protists (Lancaster et al., 2019), and SNARE proteins required for secretion are conserved across eukaryotes (Kloepper et al., 2007). Calcium-dependency of secretion in animals enabled a more controlled and directed delivery mechanism that is used in neurosecretion and might be conserved in choanoflagellates (Burkhardt and Sprecher, 2017; Varoqueaux and Fasshauer, 2017). Varoqueaux and Fasshauer (2017) hypothesise that glutamate and short peptides were originally degradation products that were transported from the lysosome to the plasma membrane. The type of secreted degradation products depends on bacteria and type of nutrients taken up by the secreting cell (Varoqueaux and Fasshauer, 2017). The different molecules secreted could be sensed by surrounding cells, changing their behaviour (Varoqueaux and Fasshauer, 2017). They suggest that the same molecules and machinery subsequently were re-used in targeted neurotransmission (Varoqueaux and Fasshauer, 2017).

1.3 How single celled organisms inform our understanding of the evolutionary origin of neurons and synapses

Generally, to study the origin of synapses and neurons, it is necessary to investigate not only animals, but also lineages that branched off before the evolution of the animals (Burkhardt, 2015), mainly other Holozoa (Holozoa include all animals and their close relatives: choanoflagellates, filastereans, ichthyosporeans and corallochytreans; Figure 1-5) (Torruella et al., 2015).

Genome surveys revealed that choanoflagellates have many genes encoding proteins with high similarity to bilaterian proteins that are involved in the function of chemical synapses (Alié and Manuel, 2010; Burkhardt et al., 2014). Amongst these are proteins involved in vesicle exocytosis, adhesion and signalling molecules, receptors and transmembrane proteins as well as structural proteins important at the presynaptic active zone and the PSD (Burkhardt, 2015) (Figure 1-6).

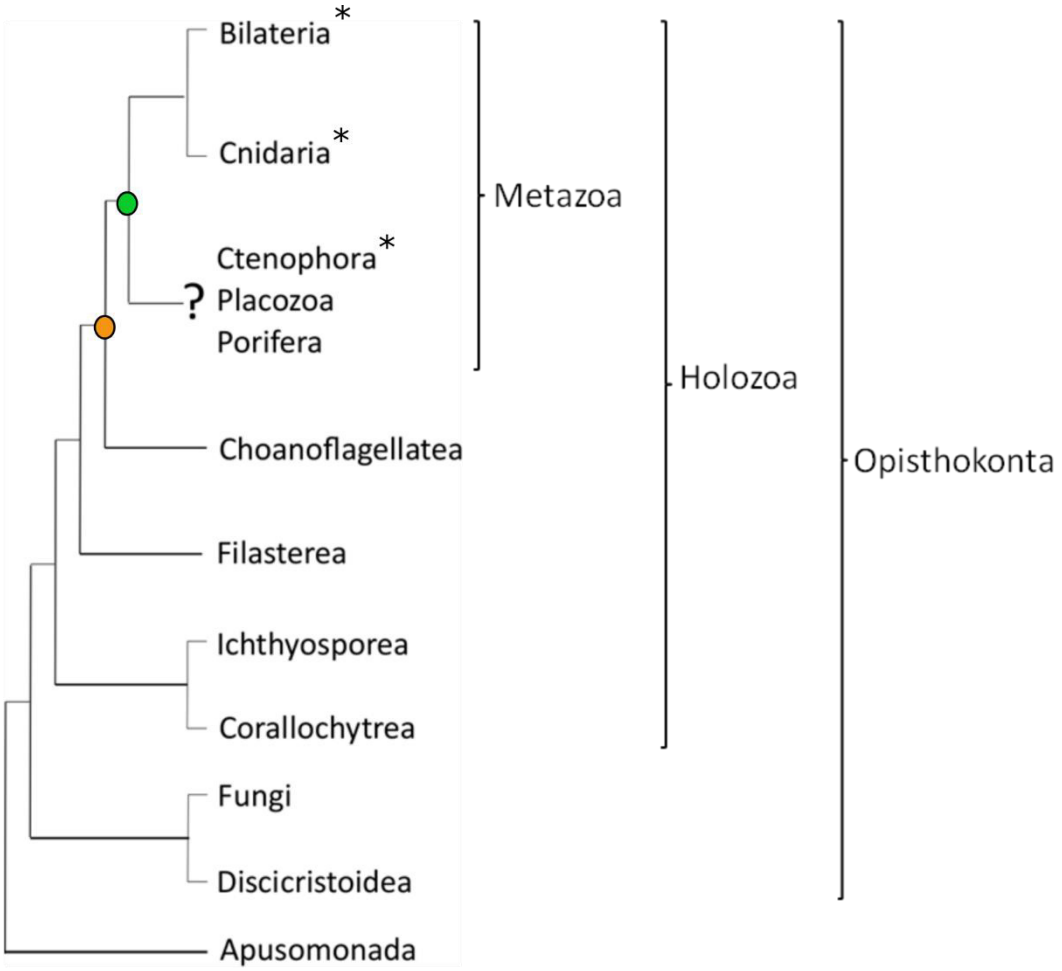


Figure 1-5: Opisthokonta phylogeny. Phylogeny adapted and simplified from Grau-Bové et al. (2017). Animal (metazoan) phylogeny shows current uncertainty about placement of Ctenophora, Porifera and Placozoa, and current consensus about the placement of Cnidaria as sister group to Bilateria. Animal groups with nervous system are labelled with an asterisk. Green circle: Urmetazoan (= last common ancestor of all animals), orange circle: Urchoanimal (= last common ancestor of animals and choanoflagellates) (definitions after Richter and King (2013).

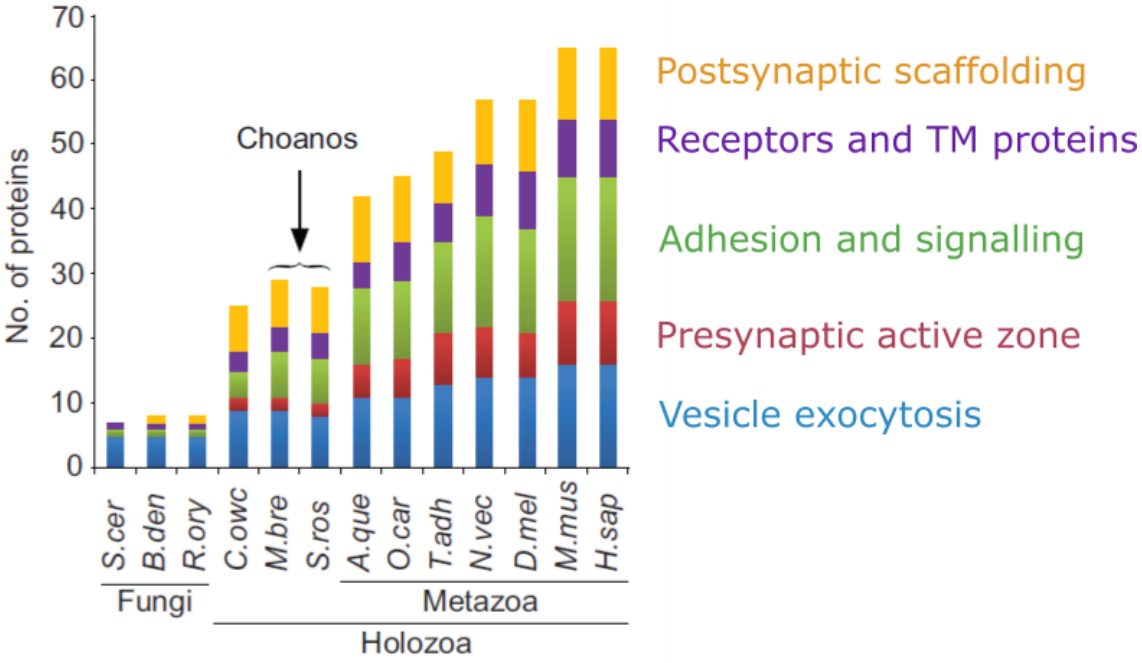


Figure 1-6: Presence of genes encoding for proteins with similarity to proteins with important functions at bilaterian chemical synapses in different species of the Opisthokonta. Presence of proteins in different categories (as indicated) are shown for the animal (metazoan) species: Bilateria: *Homo sapiens* (*H.sap*), *Mus musculus* (*M.mus*), *Drosophila melanogaster* (*D.mel*), Cnidaria: *Nematostella vectensis* (*N.vec*), Placozoa: *Trichoplax adhaerens* (*T.adh*), Porifera: *Oscarella carmela* (now: *Oscarella pearsei*; *O.car*), for the choanoflagellate (choanos) species: *Salpingoeca rosetta* (*S.ros*), *Monosiga brevicollis* (*M.bre*), the filasterean species: *Capsaspora owczarzaki* (*C.owc*), and the Fungi: *Rhizopus oryzae* (*R.ory*), *Betrachochytrium dendrobatidis* (*B.den*) and *Saccharomyces cerevisae* (*S.cer*). Figure modified from Burkhardt (2015) with permission of Journal of Experimental Biology (The Company of Biologists).

Richter and King (2013) used genome comparisons between holozoan organisms to infer the features of the Urmetazoan (the last common ancestor of all animals). They emphasised that some gene families with importance for animal innovations were already present before the evolution of animals (genes with presence in various non-animal holozoan species). On the stem lineage

leading to the Urmetazoan these gene families diversified and were likely integrated into novel contexts and functions, along with the emergence of animal specific genes and the loss of other genes (Richter and King, 2013). In terms of nervous system evolution, Richter and King (2013) hypothesised that due to the presence of synaptic genes in all animal lineages, many molecular building blocks required for the function of neurons were already present in the Urmetazoan. The evolution of these molecular building blocks has become of large interest, as the mere presence of certain proteins is not sufficient for the functioning of signalling machineries. Specific cellular functions are dependent on the interaction of many proteins and other biomolecules (Achim and Arendt, 2014). Important animal signalling pathways, as well as cell type specific functions therefore seem to have evolved in modular units, as modularity is a means of adaptive variation of a functional unit without interfering with other functional units (Achim and Arendt, 2014; Babonis and Martindale, 2017). Evolution of these modular units itself occurs stepwise, and is driven by changes in the temporal and spatial expression of existing genes and the evolution of new molecules via duplication, divergence, exon shuffling and gene fusion (Achim and Arendt, 2014; Babonis and Martindale, 2017). The stepwise evolution of complexes also allows molecular exploitation, a mechanism that recruits ancestral molecules into an interaction with newly evolved molecules (Anderson et al., 2016).

1.3.1 Choanoflagellates are at a key phylogenetic position to investigate the origin of animal traits

Choanoflagellates are the closest known relatives of animals (Carr et al., 2008; King et al., 2008; Ruiz-Trillo et al., 2008). They comprise a diverse but monophyletic group of aquatic single-celled organisms, sharing a unique cell morphology with an apical flagellum surrounded by a collar of microvilli (Leadbeater, 2015) (

Figure ***1-7***). The beat of the flagellum produces a water stream directed into the collar, which transports bacteria, the main food source of choanoflagellates, to the cell, where they can be taken up by phagocytosis (Pettitt et al., 2002). Both cell morphology and feeding mode resemble those of sponge choanocytes (Nielsen, 2008) which led to the early conception of a close

phylogenetic relationship and homology of these cells. In fact, choanoflagellate cells and sponge choanocytes have many ultrastructural similarities but also important differences (Mah et al., 2014; Laundon et al., 2019).

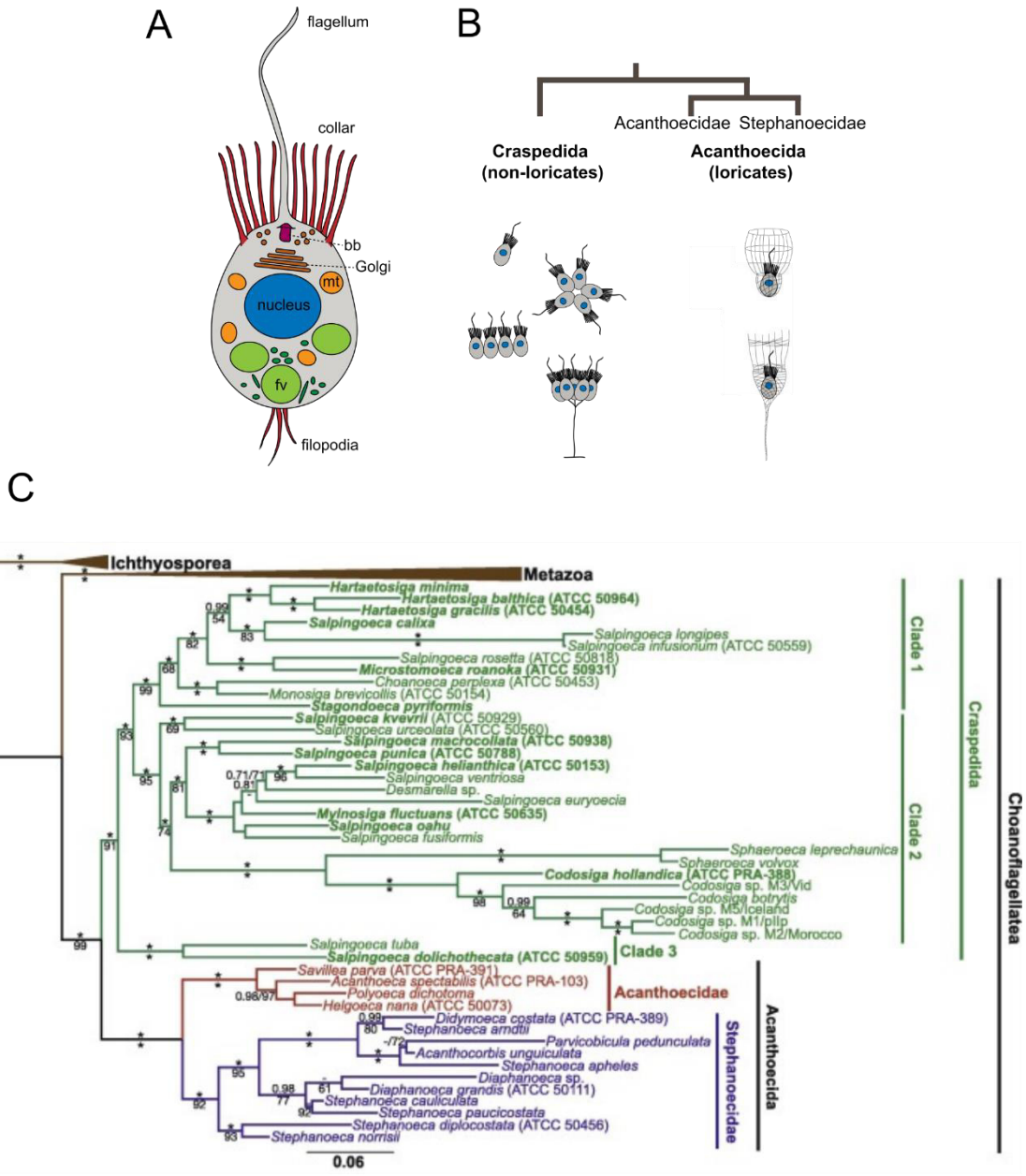


Figure 1-7: Choanoflagellate morphology and diversity. A) Choanoflagellate morphology. Picture from Hoffmeyer and Burkhardt (2016, Current Opinion in Genetics and Development, Creative commons license) with information based on (Leadbeater, 2015). bb = basal body, fv = food vacuoles. B) Phenotypical diversity of choanoflagellates of the order Craspedida (comprising species with organic covering, stalked or unstalked and many species with colonial stages) and the order Acanthoecida (comprising two families with species forming silicate-based basket-like loricas of various shapes is shown. Picture modified from Hoffmeyer and Burkhardt (2016, Current Opinion in Genetics and Development, Creative commons license) with information based on

Carr et al. (2008); Nitsche et al. (2011). C) Most recent six-gene choanoflagellate phylogeny (from Carr et al. (2017, Molecular Phylogenetics and Evolution, Creative commons license).

Choanoflagellates are emerging as important model organisms to understand the evolution of animal multicellularity (Fairclough, 2015; Hoffmeyer and Burkhardt, 2016). Many choanoflagellate species have the ability to form multicellular colonies by incomplete cytokinesis (Leadbeater and Morton, 1974; Carr et al., 2008; Dayel et al., 2011; Stoupin et al., 2012; Leadbeater, 2015). Genome surveys also showed the presence of many choanoflagellate genes with high similarity to genes that were previously considered animal-specific and are involved in multicellularity-related processes (Abedin and King, 2008; King et al., 2008; Manning et al., 2008). Genes with occurrence in choanoflagellates are for example related to cell signalling (e. g. tyrosine kinases) and cell-cell adhesion (e. g. cadherins) in animals (Abedin and King, 2008; Manning et al., 2008).

1.3.2 The choanoflagellate model *Salpingoeca rosetta*

The choanoflagellate species *Salpingoeca rosetta* has a well characterised life cycle (Dayel et al., 2011, illustrated in Figure 1-8) with at least three single celled life cycle stages (attached cells, fast swimmers and slow swimmers) and two colonial life cycle stages (chain colonies and rosette colonies), which can only be formed by slow swimmers via incomplete cytokinesis. The whole genome and transcriptomes assemblies for the different life cycle stages are available (Fairclough et al., 2013). It is further known that *S. rosetta* single cells can be induced to form rosette colonies with a sulfonolipid of the prey bacterium *Algoriphagus machipongonensis*, called rosette inducing factor 1 (RIF-1) (Alegado et al., 2012). Additional to the asexual life cycle (described above) it has been shown that the usually haploid *S. rosetta* cells can form distinct gametes and then mate by merging into one diploid cell, meiosis then leads to a return to the haploid state (Levin and King, 2013). Gamete formation was shown to be triggered by EroS, a chondroitin lyase of the bacterium *Vibrio fischeri* (Woznica et al., 2017). Triggering the sexual life cycle is useful to perform forward genetic screens, where random mutations are introduced into the cells' genome (via X-ray or EMS exposure), mutants are screened for a

specific phenotype and then outcrossed to the wildtype and sequenced in order to find the responsible gene (Levin et al., 2014). This led to the discovery, that a C-type lectin localised to the *S. rosetta* extracellular matrix, is required for rosette colony formation (Levin et al., 2014). The approach has been improved since, making use of bulk segregation methods, allowing the easier detection of genes responsible for the mutation phenotype, identifying two more proteins (two glycosyltransferases) with importance for the inhibition of cell clumping and the proper formation of rosette colonies (Wetzel et al., 2018). Recently, Booth et al. (2018) established a method to transfect *S. rosetta* with plasmids, allowing the recombinant expression of fluorescently tagged proteins in order to determine their subcellular localisation (Booth et al., 2018; Wetzel et al., 2018). This method allows the transient transfection of *S. rosetta* without integration of the gene into the genome (Booth et al., 2018) and was recently adapted for CRISPR/Cas9 mediated genome editing (Booth and King, 2020).

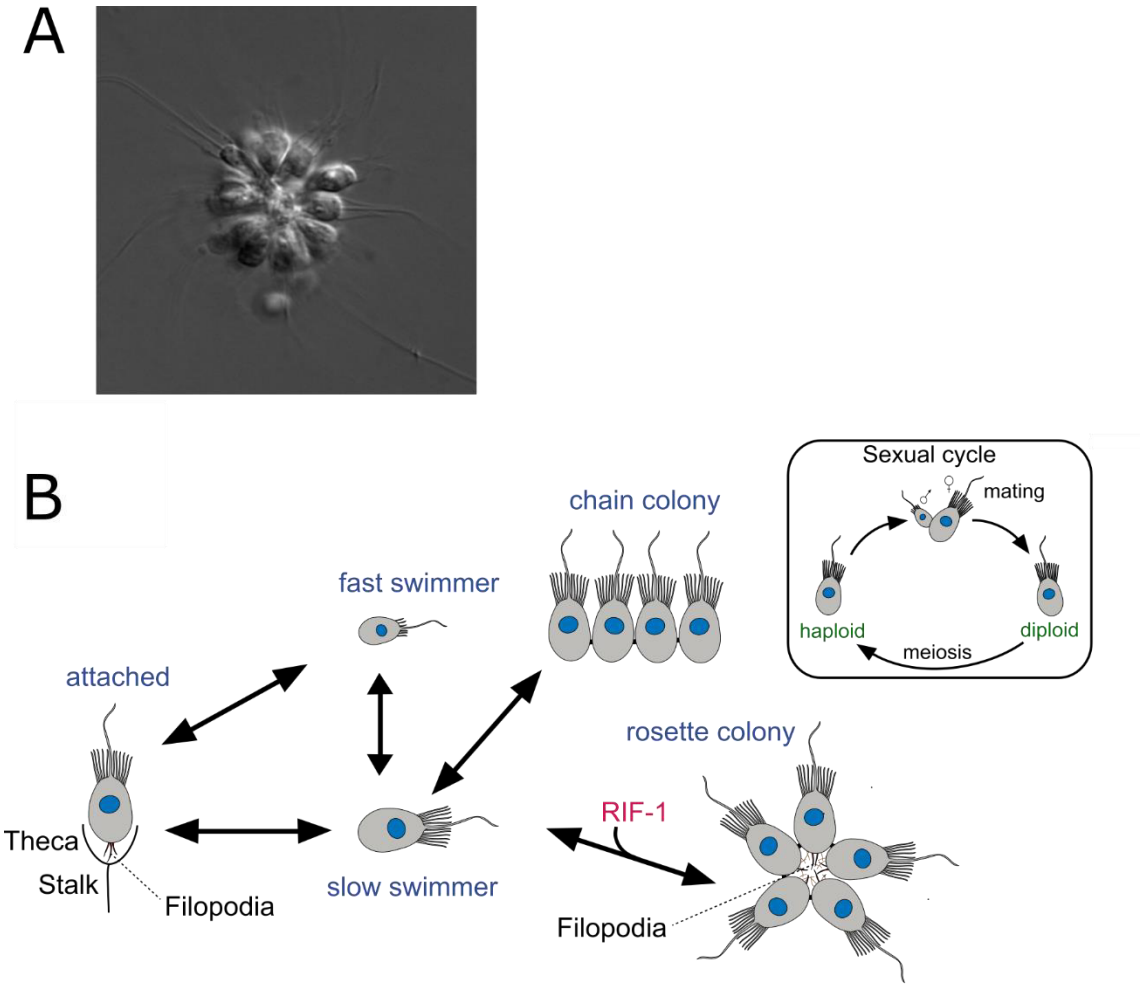


Figure 1-8: The choanoflagellate species *Salpingoeca rosetta*. A) DIC photography of a rosetta colony. B) *S. rosetta* life cycle. Picture modified from Hoffmeyer and Burkhardt (2016, Current opinion in Genetics and Development, Creative commons license) based on descriptions by Dayel et al. (2011) and including rosette colony formation induction by the rosette inducing factor (RIF-1) (Alegado et al., 2012) and the sexual cycle described by Levin and King (2013).

1.3.3 Origin of synaptic proteins

Proteins with domain architectures attributed to postsynaptic scaffolding proteins are abundant in animals without nervous system and in choanoflagellates (Sakarya et al., 2007; Alié and Manuel, 2010; Burkhardt et al., 2014). MAGUKs, for instance, are encoded in all animal groups as well as choanoflagellates and filastereans, but not in other protists, plants and fungi (de Mendoza et al., 2010; te Velthuis et al., 2007). Dlg-like proteins are only found in animals and choanoflagellates (de Mendoza et al., 2010). Homer and Shank proteins are found in ichthyosporeans, filastereans, choanoflagellates and all animals, the motifs presumably required for binding evolved in the lineage leading to choanoflagellates and animals (Burkhardt and Sprecher, 2017).

All studied sponges have proteins with homology to metabotropic glutamate receptors, whereas there are only some species encoding ionotropic glutamate receptors (Riesgo et al., 2014). The genomes of the two choanoflagellate model organisms *Monosiga brevicollis* and *Salpingoeca rosetta* on the other hand do not seem to encode glutamate receptors at all (King et al., 2008; Fairclough et al., 2013; Burkhardt et al., 2014), although ionotropic glutamate receptors are found in plants (Lam et al., 1998). Chiu et al. (1999) created phylogenetic evidence that ionotropic glutamate receptors in animals and plants have a common origin. If this is true, the receptors were lost in at least some choanoflagellates and many sponges.

On the other hand, choanoflagellates do express voltage-gated ion channels (Fairclough et al., 2013), such as voltage-gated calcium channels (Ca_v1 and Ca_v3) and even voltage-gated channels probably reacting to both calcium and sodium (with similarity to the invertebrate channel Na_v2) (Zhou et al., 2004; Liebeskind et al., 2011; Zakon, 2012; Moran et al., 2015). The function and localisation of these ion channels in choanoflagellates is not clear, but it was shown in the marine diatom *Odontella sinensis* (Stramenopiles) (Taylor, 2009) that some protist voltage-gated ion channels can generate fast Na^+/Ca^{2+} based action potentials. Brunet and Arendt (2016) hypothesised that a depolarisation of the membrane evolved in the last eukaryotic common ancestor (LECA) to trigger an emergency response to the influx of calcium occurring due to membrane damage. Membrane rupture has been shown to lead to a fast reaction in algae and animals involving a contraction of an actomyosin ring around the damaged part of the membrane (Goddard and La Claire II, 1991; Bement et al., 1999; Brunet and Arendt, 2016). In animals this then leads to the fast exocytosis of vesicles that fuse with the membrane to seal the rupture (in a molecular mechanism that is also used for the exocytosis of neurotransmitters) (Steinhardt et al., 1994). Brunet and Arendt (2016) suggest that the coupling of depolarisation to contraction and secretion occurred at the evolutionary radiation of the animals and allowed alternative cellular responses, potentially first for amoeboid movement, then for ciliary movement. The coupling of depolarisation to cellular responses along with the evolution of voltage-gated calcium (and later sodium) channels could then ultimately be used for muscle contraction and long range propagation in neurons (Brunet and Arendt, 2016).

1.3.4 Current knowledge about the origin of synaptic complexes

Most studies on the origin of synaptic proteins are based on genome or transcriptome surveys. Some case studies give phylogenetic evidence for the homology of individual proteins, such as for animal and other holozoan MAGUKs (te Velthuis et al., 2007; de Mendoza et al., 2010) and animal glutamate receptors (Riesgo et al., 2014; Ramos-Vicente et al., 2018). Richter et al. (2018) addressed the fact that the only two sequenced choanoflagellate genomes belong to two closely related choanoflagellate lineages in the same family (Carr et al., 2017) (

Figure ***1-7*** C), by sequencing the transcriptomes of 19 choanoflagellate species. They subsequently used OrthoMCL clustering of the assembled gene sequences from these species and the 2 sequenced choanoflagellate genomes and gene sequences of 21 animal species (Richter et al., 2018) to group the genes into cluster groups approximating ortholog groups (Li et al., 2003). Although the study by Richter et al. (2018) certainly underestimates the occurrence of some choanoflagellate genes, because transcriptomes lack genes that were not expressed at the conditions chosen for sequencing, they included many species and were therefore able to give an overall prediction of which choanoflagellate proteins might be homologous to which animal proteins. More phylogenetic studies are required to confirm true homology and the functions of most choanoflagellate proteins are unknown.

The first study, showing conservation of a whole signalling machinery between animals and choanoflagellates, involved co-crystallisation and other binding assays. It showed that the vesicle exocytosis machinery used for synaptic neurotransmitter release at the presynapse is conserved in the choanoflagellate species *Monosiga brevicollis* (Burkhardt et al., 2011).

Less is known about the origin of postsynaptic signalling machineries. In situ hybridisation studies in sponges showed co-expression of five postsynaptic scaffolding proteins (Dlg, GKAP, GRIP, Homer and CRIPT) in the same cell type in larvae of the demosponge *Amphimedon queenslandica* (Sakarya et al., 2007). Recently, a combination of single-cell transcriptomics and electron microscopy of the adult demosponge *Spongilla lacustris* showed the co-expression of

postsynaptic genes (Dlg, Homer and Shank) in certain choanocytes, whereas presynaptic genes were co-expressed in an amoeboid cell type (Musser et al., 2019 preprint). Electron microscopy analysis revealed that an amoeboid cell contacts microvilli of choanocytes (Musser et al., 2019 preprint). These studies suggest that postsynaptic scaffolding proteins might act together in demosponge signalling machineries and might potentially be involved in cell-cell communication in *S. lacustris* (Sakarya et al., 2007; Musser et al., 2019 preprint). There is however a lack of studies on protein level in non-bilaterian animals.

Burkhardt et al. (2014) identified the localisation and the binding partners to the *Salpingoeca rosetta* homolog of the protein Homer. This was the first study looking at a postsynaptic scaffolding protein homolog in choanoflagellates on protein level. The authors showed that *S. rosetta* Homer is localised to the nucleus, interacting with IP3 receptors and flotillins (Burkhardt et al., 2014). The interaction with IP3 receptors is conserved at vertebrate glutamatergic synapses (Tu et al., 1998). Nuclear localisation and flotillin binding was subsequently found to be conserved in rat astrocytes (Burkhardt et al., 2014). This study suggests that the ancient function of Homer might have been in a nuclear complex that still exists in choanoflagellates and certain animal cell types. No evidence could be found however, showing that choanoflagellate homologs of postsynaptic scaffolding proteins are part of a scaffold as it is found at vertebrate postsynapses (Burkhardt et al., 2014).

1.4 Purpose of this thesis

The main goal of this thesis is to improve the understanding for the origin of postsynaptic signalling machineries. It is known that neurotransmitter release machineries at the presynapse were built upon existing mechanisms for vesicle exocytosis (Burkhardt et al., 2011). Signalling machineries on the postsynaptic site are indispensable to support cell adhesion and synapse formation, as well as enable signal transduction and activity dependent regulation of the synapse (Irie et al., 1997; Naisbitt et al., 1999; Sala et al., 2001; Xu, 2011). Although genome surveys suggested that many postsynaptic proteins preceded animal origins (Alié and Manuel, 2010; Burkhardt et al., 2014), little is known about the evolution of protein-protein interactions required for the formation of postsynaptic complexes.

Only recently, Richter et al. (2018) sequenced a variety of choanoflagellate transcriptomes and showed that even more previously labelled animal-unique genes are actually older than the animal lineage, because they are found in some choanoflagellate species. Therefore, one aim of this thesis was to survey the new choanoflagellate transcriptomes for postsynaptic genes to provide a better picture of which postsynaptic proteins were present in the choanoflagellate-animal ancestor.

Scaffolding proteins are key to the formation of postsynaptic signalling machineries, as they form platforms for other proteins to bind and thereby form a link between receptors and downstream signalling cascades (Naisbitt et al., 1999; Sala et al., 2001; Baron et al., 2006; Hayashi et al., 2009). One crucial interaction at vertebrate glutamatergic synapses is the interaction between the scaffolding proteins Homer and Shank (Tu et al., 1999; Naisbitt et al., 1999; Sala et al., 2001). These proteins interconnect complexes anchored to different receptors and ion channels and link them to the calcium signalling machinery and modulators of the actin cytoskeleton (Naisbitt et al., 1999; Sala et al., 2001). Even though Shank was not co-precipitated with Homer in an immunoprecipitation experiment from *Salpingoeca rosetta* cell lysate (Burkhardt et al., 2014), it remained uncertain, whether the two proteins actually have the capacity to bind in choanoflagellates. Based on the similarity of the Homer EVH1 domain in choanoflagellates, we hypothesised that the Homer and Shank interaction is ancestral and was present in the last common ancestor of animals and choanoflagellates.

The protein known as the key regulator of the postsynaptic scaffold at vertebrate glutamatergic synapses is PSD-95 (Chen et al., 2011). This protein anchors ionotropic glutamate receptors in the membrane and integrates them into various signalling machineries (Kim and Sheng, 2004). Identifying interaction partners of the *S. rosetta* homolog to PSD-95 might give insights into which components of postsynaptic signalling machineries are conserved in choanoflagellates.

I hypothesise that:

1) **Surveying a broader variety of choanoflagellate species for postsynaptic protein homologs will reveal new insights into the choanoflagellate and holozoan ancestral gene content.**

2) **Homer and Shank were binding partners in the last common ancestor of choanoflagellates and animals.**
3) **The *S. rosetta* Dlg homolog has scaffolding function and forms a scaffold similar to the one found in the postsynapse. This scaffold is ancestral and was present in the last common ancestor of choanoflagellates and animals. The scaffold was expanded in the animal lineage with the evolution of crucial interactions for individual signalling machineries that led to the emergence of neuronal cell types and synapses.**

Testing the hypotheses will reveal, if in addition to the conservation of many structural postsynaptic density proteins, protein-protein interactions are conserved between animals and choanoflagellates. Finding evidence supporting these hypotheses would suggest that some of the structural framework required for postsynaptic signalling machineries in animal neurons might have preceded the evolutionary origin of animals. If, on the contrary, these protein-protein interactions cannot be identified, this would suggest that the evolution of these complexes might have been a defining moment in the evolution of synapses and neurons.

The following specific aims were designed in order to test the hypotheses:

1) Develop a better picture of the presence of proteins with statistically significant sequence similarity to postsynaptic proteins in the diverse group of choanoflagellates by surveying 19 choanoflagellate transcriptomes for the presence of corresponding sequences. This will be a first step to predict which postsynaptic proteins might have been present in the last common ancestor of animals and choanoflagellates. Identified proteins are candidates that might have been prerequisites for the emergence of neurons and synapses (Chapter 2).
2) Reconstruct in which lineage the binding capacity of Homer and Shank proteins originated by combining ancestral protein reconstruction with binding assays based on isothermal titration calorimetry measurements (Chapter 3).
3) Investigate complexes formed by the *Salpingoeca rosetta* Dlg/PSD-95 homolog making use of custom-made antibodies against the protein, to identify its subcellular localisation and native interaction partners. Further,

structural and biochemical properties of the *S. rosetta* Dlg/PSD-95 homolog with focus on features enabling its function as scaffolding protein will be investigated to complement these insights and help to understand how conserved important functional features of this protein are (Chapter 4).

The results of our studies will lay a foundation for the understanding of the putative origin of postsynaptic protein complexes. This will allow hypothesising which components of postsynaptic signalling machineries are of ancestral nature and which important interactions evolved in the animal lineage. Very recently, Booth and King (2020 preprint) developed a protocol to use the CRISPR/Cas9 system for genome editing in *S. rosetta*, which offers new opportunities for testing the functions of genes in this choanoflagellate. The fundamental understanding of protein complexes will help to predict and understand consequences of manipulating postsynaptic protein homologs in *Salpingoeca rosetta*.

2 Data chapter 1: Survey of diverse choanoflagellate transcriptomes for the presence of protein sequences with statistically significant similarity to postsynaptic proteins

Tarja T. Hoffmeyer[1,2,3] and Pawel Burkhardt[2,3].

[1]University of Exeter; [2]Marine Biological Association of the UK (Plymouth); [3]Sars International Centre of Marine Molecular Biology, University of Bergen (Norway).

Text: Tarja Hoffmeyer

Survey and data analysis: Tarja Hoffmeyer

Project idea: Pawel Burkhardt

2.1 Introduction

Only animals have a canonical nervous system (Brodal, 2004), but it remains unclear in which lineage the first neurons and synapses evolved. Bilaterian animals, cnidarians and ctenophores have nervous systems, whereas sponges (Porifera) and placozoans do not. The debated phylogenetic position of the latter three phyla evoked different hypotheses on the origin of nervous systems (Shen et al., 2017; Simion et al., 2017). There might have been a single evolutionary origin of synapses and neurons or two independent origins in ctenophores and in the lineages comprising bilaterians and cnidarians (Ryan and Chiodin, 2015; Jékely et al., 2015b; Liebeskind et al., 2016; Moroz and Kohn, 2016). It is possible that poriferan and placozoan lineages either never had synapses and neurons or lost them (Ryan and Chiodin, 2015; Liebeskind et al., 2016; Moroz and Kohn, 2016). Surveys in non-bilaterian animals and choanoflagellates showed that many proteins with important synaptic functions are conserved in all animal lineages independent of the presence of a nervous system (Sakarya et al., 2007; Alié and Manuel, 2010). Many of these proteins even occur in the single-celled sister group to the animals, the choanoflagellates (Alié and Manuel, 2010; Burkhardt et al., 2014). These studies, which were restricted to the two choanoflagellate species with sequenced genomes *Monosiga brevicollis* and *Salpingoeca rosetta* (King et al., 2008; Fairclough et al., 2013), showed that many proteins involved in presynaptic vesicle exocytosis and postsynaptic scaffolding, as well as some receptors and synaptic signalling

components are conserved in choanoflagellates and therefore evolved before the emergence of animal neurons and synapses (Burkhardt et al., 2014).

Choanoflagellates are a diverse monophyletic group with about 250 named species (Leadbeater, 2015). 47 choanoflagellate species included in the most recent multigene phylogeny have been divided into three clades of craspedid species (family Craspedida; species with an organic cell covering) and two clades of acanthoecid species (family Acanthoecida, species with an inorganic cell covering) (Carr et al., 2008; Nitsche et al., 2011; Carr et al., 2017). Recently, Richter et al. (2018) sequenced and assembled the transcriptomes of 19 choanoflagellate species, with representatives in each of these clades. The authors made large scale comparisons between the genes of the in total 21 sequenced choanoflagellate species and the genes of 21 animal species. They grouped the genes of all species into orthologous groups, giving an indication of their origin. Furthermore, they annotated the similar domains of every protein in the dataset. We made use of these resources to extend the survey of postsynaptic protein homologs to a greater variety of choanoflagellate species.

This survey aims to unravel the origin of specific synaptic proteins. We focussed on proteins that are involved in postsynaptic signalling at glutamatergic synapses. This type of synapse is the main excitatory synapse in vertebrate brains (Meldrum, 2000). Glutamatergic synapses probably occur not only in bilaterians but also in cnidarians and ctenophores (Burkhardt and Sprecher, 2017; Kass-Simon and Pierobon, 2007; Moroz et al., 2014). Glutamate is released from vesicles in the presynapse and binds to different kinds of receptors in the postsynaptic membrane – ionotropic glutamate receptors (iGluRs) that act as sodium and potassium or calcium ion channels upon activation by glutamate binding, and metabotropic glutamate receptors (mGluRs) that are G protein coupled (Sugiyama et al., 1989; Gasic and Hollmann, 1992). Sodium influx can lead to a depolarisation which results in signal transduction, whereas both calcium influx and G protein activation through glutamate binding of mGluRs influence synaptic plasticity (regulating synaptic strength) (Bortolotto et al., 1999; Zucker, 1999; Voglis and Tavernarakis, 2006).

Scaffolding proteins organise signalling complexes in the postsynaptic density (PSD) – the region adjacent to the postsynaptic membrane (Boeckers,

2006; Zhu et al., 2016). Members of the protein family of membrane associated guanylate kinases (MAGUKs) – including the Discs large (Dlg) and MAGUK p55 subfamily proteins – anchor receptors (such as ionotropic glutamate receptors – iGluRs) in vertebrate postsynaptic membranes and bind proteins involved in the adhesion of pre- and postsynaptic membrane (Kim et al., 1995; Irie et al., 1997; Kim and Sheng, 2004; Kegel et al., 2013; Chen et al., 2015; Rademacher et al., 2016). Lin-7 is a small protein known to mediate the interaction between different MAGUKs in *Drosophila melanogaster* and humans (Bohl et al., 2007; Bachmann et al., 2010). Another known binding partner of vertebrate PSD-95 is the voltage gated potassium channel Shaker (Kim et al., 1995). This channel regulates excitability of the neuron through its action in repolarisation (Xing and Wu, 2018). Homer and Shank proteins form a structured network, connecting the different signalling complexes at receptors and interconnecting them to the calcium signalling machinery and proteins that can remodel the cytoskeleton (Naisbitt et al., 1999; Sala et al., 2001; Baron et al., 2006; Hayashi et al., 2009). Network formation is mediated by the Homer EVH1 domain that binds PPxxF motifs of Shank proteins, but also of metabotropic glutamate receptors (mGluRs) and IP3 receptors (IP3Rs) (Tu et al., 1998, 1999; Beneken et al., 2000; Barzik et al., 2001). Upon glutamate binding, mGluRs in the postsynaptic membrane initiate the synthesis of IP3 (Masu et al., 1991; Abe et al., 1992; Aramori and Nakanishi, 1992; Knöpfel et al., 1995). This molecule, as well as cytosolic calcium can activate IP3Rs in the endoplasmic reticulum membrane, leading to calcium release from ER stores into the cytoplasm (reviewed by Taylor and Tovey, 2010). The Homer and Shank network also interconnects PSD-95 anchored receptor-complexes, which is mediated through a GKAP protein that binds both Shank and PSD-95 (Kim et al., 1997; Naisbitt et al., 1999). Ca^{2+}/calmodulin-dependent protein kinase II (CAMKII) phosphorylates MAGUKs and can regulate synaptic targeting (Mauceri et al., 2004). Nitric oxide synthase catalyses the synthesis of nitric oxide from L-arginin (Boucher et al., 1999). Nitric oxide can modulate neurotransmitter release of neighbouring synaptic terminals (Prast and Philippu, 2001).

In this study, we surveyed a diverse collection of choanoflagellate transcriptomes for the presence of proteins with statistically significant sequence

similarity (candidate homologs) to proteins of these described signalling machineries. We hypothesised that:

Surveying a broader variety of choanoflagellate species for protein homologs will reveal new insights into the choanoflagellate and holozoan ancestral gene content.

We surveyed the transcriptomes of 19 choanoflagellate species both for proteins that are known to be present in *S. rosetta* and *M. brevicollis*, and for proteins that were not detected in these species. Using this approach, we aimed to identify and test for increased distribution of synaptic proteins across choanoflagellates. Because we are surveying transcriptome data, we will likely not detect all proteins, as only proteins that were expressed in the species prior to sequencing were detected. However, the breadth of the study will likely generate data that can inform our understanding of the utilisation of synaptic like gene networks in these protists and enable us to explore the possibility that more postsynaptic proteins predate the emergence of animal neurons and synapses.

2.2 Materials and methods

2.2.1 General survey procedure to identify sequences with statistically significant similarity to postsynaptic proteins

Foundation for these studies were supplementary data published by Richter et al. (2018). We used associated transcriptome datasets made available at https://dx.doi.org/10.6084/m9.figshare.5686984.v2: dataset 2,4,5 and 7.

We used two strategies for the initial identification of candidate homologs in the amino acid translated choanoflagellate transcriptomes (summarised in Figure 2-1). Strategy A was based on sequence similarity. We performed a BLASTp search (Altschul et al., 1990; Camacho et al., 2009) (with a relaxed expectation value (e-value) cut-off of 0.1 as a gathering threshold) querying the predicted amino acid sequences of a known homolog of this protein from different animal species (and if available choanoflagellate species) to protein databases of each surveyed choanoflagellate transcriptome. All query sequences used for each protein are listed in supplementary Table 6-1. This strategy was used for Shank proteins, as these proteins have divergent domain architectures in

choanoflagellates and animals (Figure 2-3). The strategy was also used in addition to strategy B for several proteins (listed in the blue box in Figure 2-1).

Strategy B was based on domain architecture conservation. We defined diagnostic domains for proteins of interest by comparing known animal domain architectures corresponding to these proteins (diagnostic domain architectures given in Table 2-1). Dataset 7 (associated with Richter et al. (2018); https://dx.doi.org/10.6084/m9.figshare.5686984.v2) – containing the names of domains associated with each protein identified through PANTHER HMM Scoring (Thomas et al., 2006) (v. 1.03, e-value cut-off: 10^{-23}) against PANTHER Hmm library (v. 7.2) (Mi et al., 2010) – was searched for the presence of proteins with diagnostic domains. This strategy was used for all proteins with the exception of Shank proteins.

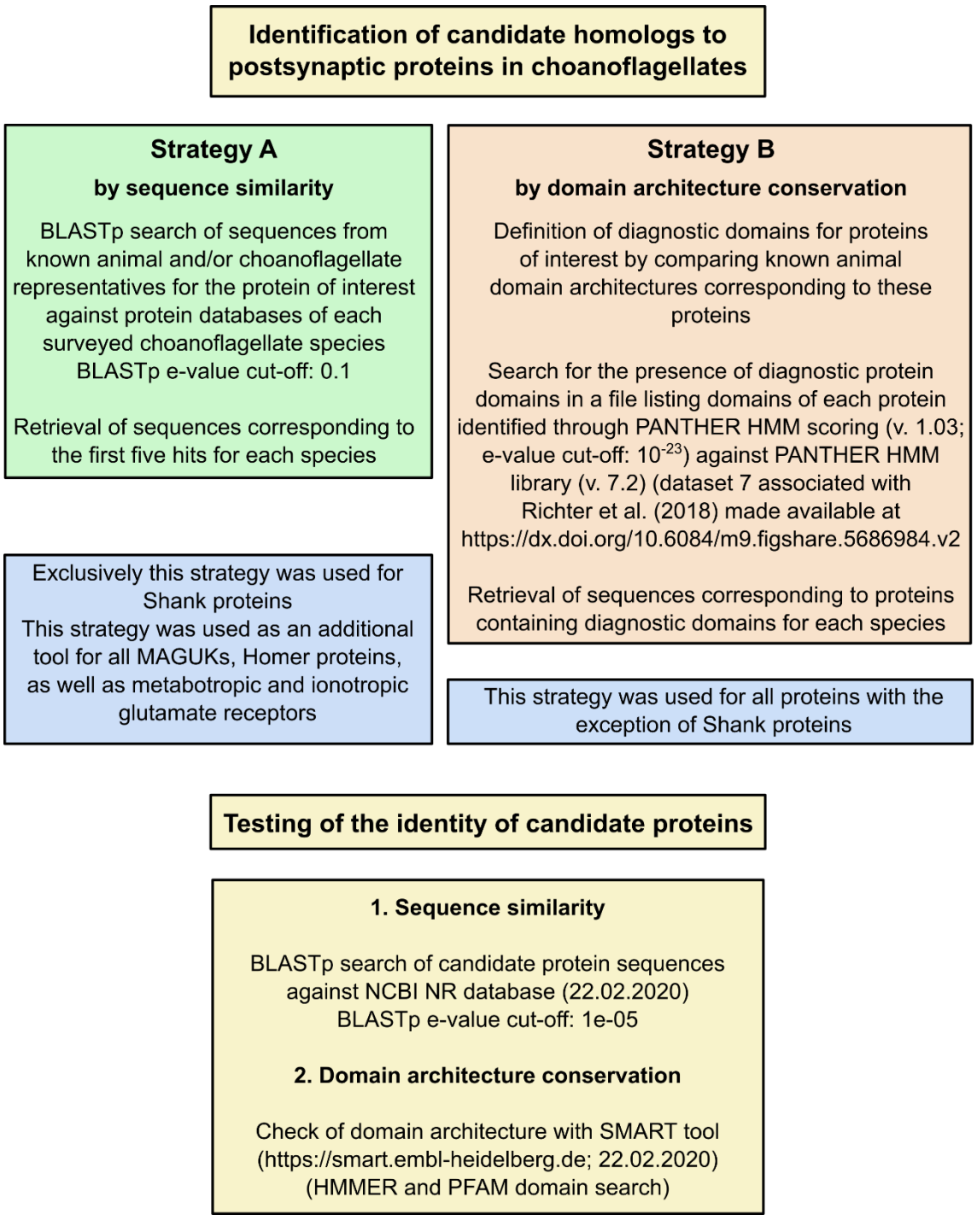


Figure 2-1: Summary of procedures used for the identification of sequences with statistically significant similarity to postsynaptic proteins in choanoflagellate transcriptomes.

Two approaches were used to test the identity of each candidate protein sequence identified with strategies A and B (summarised in Figure 2-1). First, each sequence was used as query in a BLASTp search of the non-redundant (NR) protein database of NCBI (ncbi.nlm.nih.gov; last accessed 22.02.2020) with an e-value cut-off of 10^{-5}. Second, we checked for domain architecture

conservation (using specific criteria for each protein listed in Table 2-1) with the SMART tool (http://smart.embl-heidelberg.de; last accessed 22.02.2020) (Schultz et al., 1998; Letunic and Bork, 2018) using HMMER and PFAM domain search. Sequences that had statistically significant similarity with another sequence that was classified as a protein of interest (highest-ranking BLASTp hit with an e-value of 10^{-5} or lower) and had a conserved or partially conserved domain architecture (based on our criteria listed in Table 2-1) were inferred to be homologous to the protein of interest in at least parts of their sequence (i. e. the simplest explanation for statistically significant similarity after Pearson, 2013). These sequences were used to confirm presence or partial presence of a protein of common ancestry with a particular postsynaptic protein. The accession numbers of all these sequences are listed in Supplementary Table 6-2. In cases, where the highest-ranking BLASTp hit was not annotated, we used the same approaches (BLASTp search and SMART domain architecture check) to confirm the identity of this protein. It was noticed that some domains were not listed in the dataset of Richter et al. (2018) even though those domains were found with the SMART tool, probably due to differences in the threshold used.

2.2.2 Modifications in the approaches to test the identity of candidate MAGUK and GKAP proteins.

For the particular case of MAGUKs, we tested the identity through a highest-ranking BLASTp hit against another MAGUK and the presence of the MAGUK module (PDZ domain, SH3 domain, GuK domain – in this order). MAGUK proteins were sorted into the gene families Dlg, MAGUK p55, or choanoflagellate-unique MAGUK according to ortholog clusters defined by Richter et al. (2018) through OrthoMCL (Li et al., 2003) analysis (Figure 2-5). This was done, because BLASTp searches gave some ambiguous results concerning the placement of proteins into these families. Candidate GKAP homologs were BLASTp searched against only the bilaterian fraction of NCBI NR protein database, as many of the non-bilaterian hits were not annotated. They were confirmed, if a GKAP domain was present and the highest-ranking BLASTp search hit was a DLGAP protein. In the same way as with MAGUK proteins sorting into the DLGAP1-4 group (synaptic GKAP) or the DLGAP5 group (epithelial protein) was done based on ortholog clusters defined by Richter et al.

(2018). Gene families according to Richter et al. (2018) for the identified proteins are listed in Table 6-2. The sorting is preliminary and only future phylogenetic analysis can reveal evolutionary relationships, clarifying if homology (common ancestry) can be confirmed and whether it can be explained through gene duplication (paralog) or through speciation (ortholog) (Koonin, 2005).

2.2.3 Additional sequences included in our survey

For comparison, animal representatives of Porifera (*Oscarella pearsei*), Placozoa (*Trichoplax adhaerens*), Ctenophora (*Mnemiopsis leidyi*), Cnidaria (*Nematostella vectensis*), invertebrate Bilateria (*Drosophila melanogaster*) and vertebrate Bilateria (*Mus musculus*) were included in this survey. Additionally, *Capsaspora owczarzaki* was included. *C. owczarzaki* belongs to the Filasterea, which form the sister group to choanoflagellates and animals (Torruella et al., 2015; Grau-Bové et al., 2017). Sequences of all these species were identified via strategy B described above. Sequence absences were additionally confirmed via strategy A (with the exception of Shank proteins). The identity of all but the annotated bilaterian sequences were confirmed via BLASTp searches and SMART domain predictions (as described in section 2.2.1). Predicted and annotated protein sequences of the two bilaterian species were retrieved from NCBI. Peptide translated transcriptome sequences of *Oscarella pearsei* were retrieved from compagen.org (OCAR_T-PEP_130911.fa). Peptide translated genome sequences of *Nematostella vectensis* and *Trichoplax adhaerens* were retrieved from the joint genome institute server (genome.jgi.doe.gov). Peptide translated genome sequences of *Mnemiopsis leidyi* were retrieved from the *M. leidyi* genome browser (research.nhgri.nih.gov/mnemiopsis/). *C. owczarzaki* genome translated protein sequences were retrieved from NCBI. The accession numbers of all identified sequences are listed in Table 6-2.

2.2.4 Specific criteria implemented for every protein of interest in order to identify candidate homologs of postsynaptic proteins with both statistically significant sequence similarity and comparable domain architecture

Different proteins show a different level of conservation. Shank proteins for example, have a variety of domain architectures even among different animal species (described in more detail in section 3). Therefore, specific criteria had to be implemented for every single protein to identify candidate homologs with statistically significant sequence similarity and comparable domain architecture (Table 2-1).

Table 2-1: Criteria assigned to evaluate presence or partial presence of protein sequences with statistically significant sequence similarity to postsynaptic proteins.

Protein	Assigned criteria for each protein to define presence and partial presence
Dlg	Present if domain architecture comprises at least PDZ-PDZ-PDZ-SH3-GuK, with or without N-terminal L27 domain, and in OrthoMCL ortholog cluster with animal Dlg proteins. Partially present if at least MAGUK module PDZ-SH3-GuK is present, and in OrthoMCL ortholog cluster with animal Dlg proteins.
MAGUK p55	Present if domain architecture comprises at least L27-PDZ-SH3-GuK, with or without a second L27 domain and in one of the OrthoMCL ortholog clusters with animal MAGUK p55 proteins. Partially present if at least MAGUK module PDZ-SH3-GuK is present and in one of the OrthoMCL ortholog clusters with animal MAGUK p55 proteins.
Choano-flagellate specific MAGUK	Present if MAGUK module PDZ-SH3-GuK is present and in OrthoMCL ortholog clusters distinct from the animal MAGUK OrthoMCL ortholog clusters.
Homer	Present if domain architecture comprises at least EVH1/WH1 and one or two coiled coil domains and highest-ranking BLASTp search hit against the NCBI non-redundant protein (NR) database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as a Homer homolog. Partially present if domain architecture comprises at least EVH1 domain and the highest-ranking BLASTp search hit against the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as a Homer homolog*.
Shank	Present if domain architecture comprises at least ankyrin repeats and either a PDZ or a SH3 domain and highest-ranking BLASTp search hit against the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as a Shank homolog. Partially present if domain architecture comprises at least ankyrin repeats and highest-ranking BLASTp search hit against the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as a Shank homolog.
GKAP	Present if GKAP domain present and highest-ranking BLASTp search hit against only the fraction of bilaterian proteins in the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as a DLGAP1-4 homolog. Partially present if GKAP domain present and highest-ranking BLASTp search hit against only the fraction of bilaterian proteins in the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as a DLGAP5 homolog. DLGAP5 is an epithelial protein.
Lin-7	Present if domain architecture L27-PDZ and highest-ranking BLASTp search hit against the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as Lin-7/ Mals/ Veli homolog.

	Partially present if at least one of the two domains is present and highest ranking BLASTp search hit against the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as Lin-7/ Mals/ Veli homolog.
iGluR	Present if ANF_receptor or Peripla_6-bp domain is in combination with Lig_chan domain and highest-ranking BLASTp search hit against the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as an ionotropic glutamate receptor/ receptor subunit. Partially present if only Lig_chan domain present and highest-ranking BLASTp search hit against the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as an ionotropic glutamate receptor/ receptor subunit.
mGluR	Present if ANF_receptor, NC3DG and 7-transmembrane domain present and highest-ranking BLASTp search hit against the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as an mGluR homolog.
IP3R	Present if at least the domains: MIR, RYDR_ITPR, RIH_assoc, and Ion_trans are present and highest-ranking BLASTp search hit against the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a progein classified as an IP3 receptor. Partially present, if the majority of the named domains are present and highest-ranking BLASTp search hit against the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as an IP3 receptor.
Shaker/ Shal potassium channel	Present if the domains BTB and Ion_trans are in the protein and highest-ranking BLASTp search hit against the NCBI NR database (last accessed 24.02.2020) with cut-off $1e^{-5}$ is a protein classified as a voltage gated potassium channel and a protein classified as Shaker/Shal/Shaw homolog is among the five highest ranking BLASTp search hits.
CAMKII	Present if serine-threonine protein kinase domain (S_TK_c) and CAMKII_AD domain occur in the protein and highest-ranking BLASTp search hit against the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as a CAMKII. Partially present if only CAMKII_AD domain occurs in the protein and highest-ranking BLASTp search hit against the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as a CAMKII.
NOS	Present if domains NO_synthase, Flavodoxin-1, FAD-binding-1 and NAD-binding-1 are in the protein and highest-ranking BLASTp search hit against the NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as a nitric oxide synthase (even if the hit corresponds to a cyanobacterial sequence). Partially present if domains NO-synthase and Flavodoxin-1 are in the protein and highest-ranking BLASTp search hit against the NCBI NR database (last accessed 22.02.2020) with cut-off $1e^{-5}$ is a protein classified as a nitric oxide synthase (even if the hit corresponds to a cyanobacterial sequence).

*If in a single species two different proteins both had a highest-ranking BLASTp search hit to a protein classified as Homer homolog (of which one comprised the EVH1 and the other comprised a coiled coil domain) were found, these proteins were also placed in the category “partially present”, even though they might be fully present and are just in different contigs due to the transcriptome assembly. The same principle was used for the presence check of Shank homologs.

2.3 Results

2.3.1 Surveying a broader range of choanoflagellate transcriptomes gives increased insight about the evolutionary ancestry of postsynaptic proteins

Prior studies surveying for homologs of synaptic proteins included the closely related choanoflagellate species *Salpingoeca rosetta* and *Monosiga brevicollis*. New transcriptomic data (Richter et al. 2018) enabled us to expand the survey to more choanoflagellate species (Figure 2-2). Previous evidence suggested that the majority of postsynaptic scaffolding proteins are conserved in choanoflagellates (Burkhardt et al., 2014). Our survey supports this, as we could detect expression of these proteins in most choanoflagellate species. Choanoflagellates in general possess few postsynaptic receptors and signalling proteins, such as suggested by *M. brevicollis* and *S. rosetta* genome surveys (Alié and Manuel, 2010; Burkhardt et al., 2014). Sequences with statistically significant similarity to some of these proteins could now however be identified in some choanoflagellate species (described in section 2.3.4).

2.3.2 All assessed choanoflagellates express proteins with statistically significant sequence similarity to postsynaptic scaffolding proteins

We conducted sequence similarity survey searches to identify sequences with statistically significant similarity to six scaffolding proteins with known function in postsynaptic complexes. Proteins with sequence similarity to the membrane associated guanylate kinases (MAGUKs), Homer, and Shank, have previously been identified in *Capsaspora owczarzaki*, *Salpingoeca rosetta*, and *Monosiga brevicollis* (Burkhardt et al., 2014). Our survey shows that these proteins are expressed in choanoflagellates that branch throughout the phylogenetic radiation of this group. We observed a similarly broad distribution of proteins with GKAP domain in choanoflagellates, although a majority of these proteins have higher similarity to epithelial proteins containing this domain.

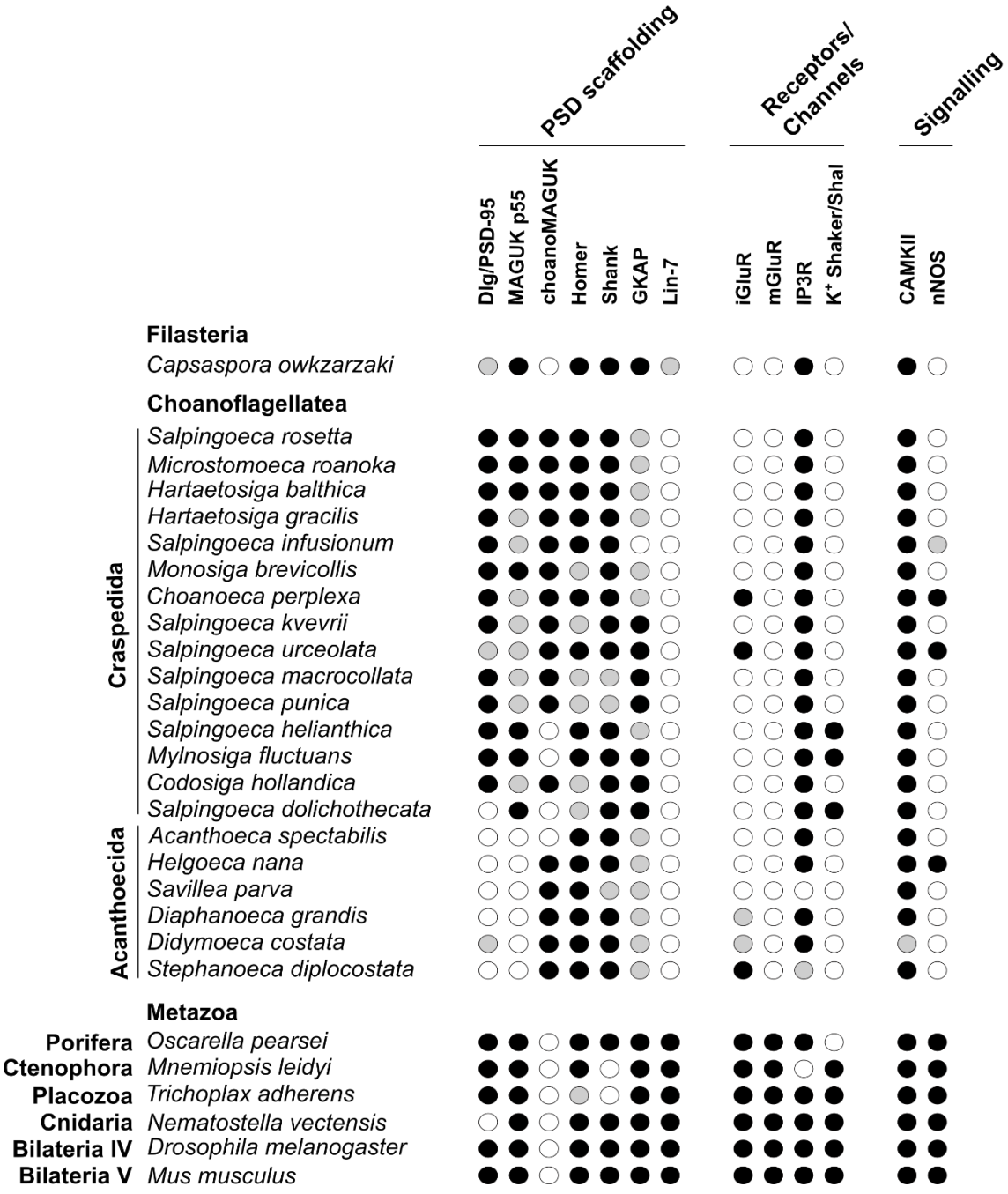


Figure 2-2: Survey of a subset of postsynaptic proteins in a broader range of choanoflagellate transcriptomes. Shown are presence (black filled circle) with canonical conservation regarding the criteria determined for each postsynaptic protein of interest, partial presence (grey filled circle) with highly similar sequences missing some canonical domains, and absence (white filled circle) lacking any evidence for the presence of a homologous sequence to these proteins in the sampled transcriptome of these species. Presence, partial presence and absence for the surveyed choanoflagellate species (sorted into families) as well as for other included choanoflagellates (*M. brevicollis* and *S. rosetta*) with sequenced genome, the filasterean *Capsaspora owczarzaki* and representatives of different animal phyla are shown.

As seen also in animals, domain architectures of Shank homologs are very variable, but many have PPxxF motifs that were described to be bound by the vertebrate Homer EVH1 domain (Tu et al., 1999; Beneken et al., 2000; Barzik et al., 2001) (Figure 2-3). Homer EVH1 domains are retained in all choanoflagellate species, but in some of these species they did not co-occur with coiled-coil domains that are known to mediate Homer tetramerisation in rat and *S. rosetta* (Hayashi et al., 2006, 2009; Burkhardt et al., 2014) (Figure 2-4). However, in many cases there is a second transcript encoding this domain. It is possible that the two transcripts are actually connected, which cannot be detected due to potential sequencing artefacts. Alternatively, this gene could have been subjected to a gene fission resulting in two independent open reading frames. Lin-7 proteins, comprising an L27 and a PDZ domain, occur in all animal phyla (Figure 2-2), and a gene encoding a highly similar protein, which only comprises the PDZ domain, was detected in *C. owczarzaki* (Burkhardt et al., 2014). Lin-7 is absent in the genomes of *S. rosetta* and *M. brevicollis* (Burkhardt et al., 2014), and could also not be detected in any of our surveyed choanoflagellate transcriptomes (Figure 2-2).

2.3.3 Transcriptome data suggest that species of the family Craspedida retained animal-like membrane associated guanylate kinases

Membrane associated guanylate kinases (MAGUKs) were identified in all choanoflagellate transcriptomes but the one of *Acanthoeca spectabilis* (Figure 2-5). Most species of the family Craspedida were shown to express animal-like MAGUKs of the Dlg family and of the p55 family. For each detected MAGUK, we checked for their placement according to OrthoMCL predictions calculated by Richter et al. (2018), which provides a first approximation of the evolutionary relationship of these genes. According to this approximation and in agreement with the predicted domain architectures, it was observed that – respecting the limitations of our study arising by the survey of transcriptomes that might lack genes that are present in the genome but were not expressed at the timepoint of sequencing – we did not detect homologous MAGUK sequences with canonical Dlg or MAGUK p55 domain architecture in any surveyed species of the family Acanthoecida.

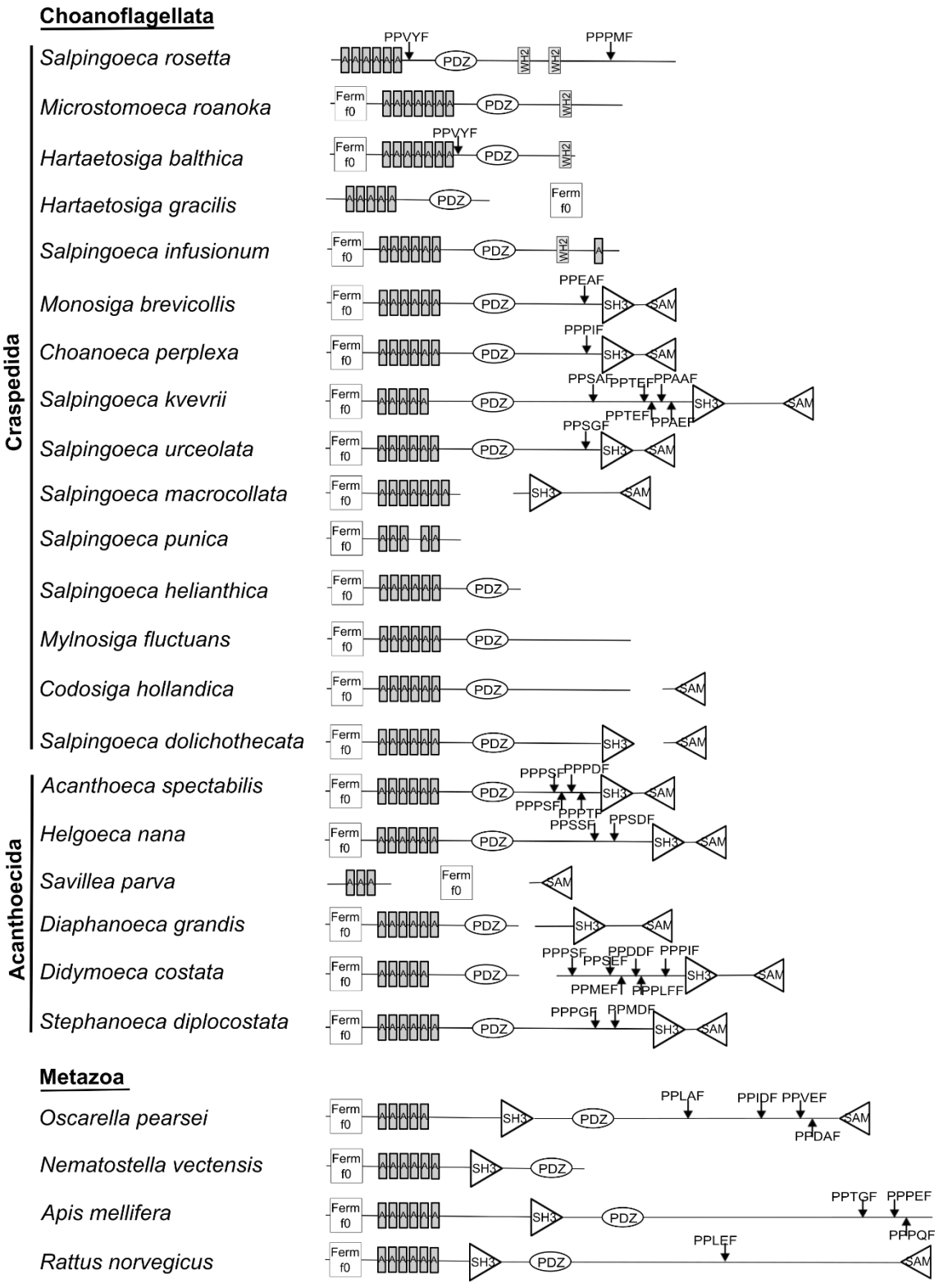


Figure 2-3: Shank domain architectures in a diversity of choanoflagellate and animal species. Shown are the Pfam Ferm_f0 domain, Ankyrin repeats (A), SH3, PDZ, SAM and WH2 domains. Shown are also all detected PPxxF motifs. In some cases, several transcripts with high similarity to Shank proteins were found. They are all shown in the figure, clearly indicated as separate proteins.

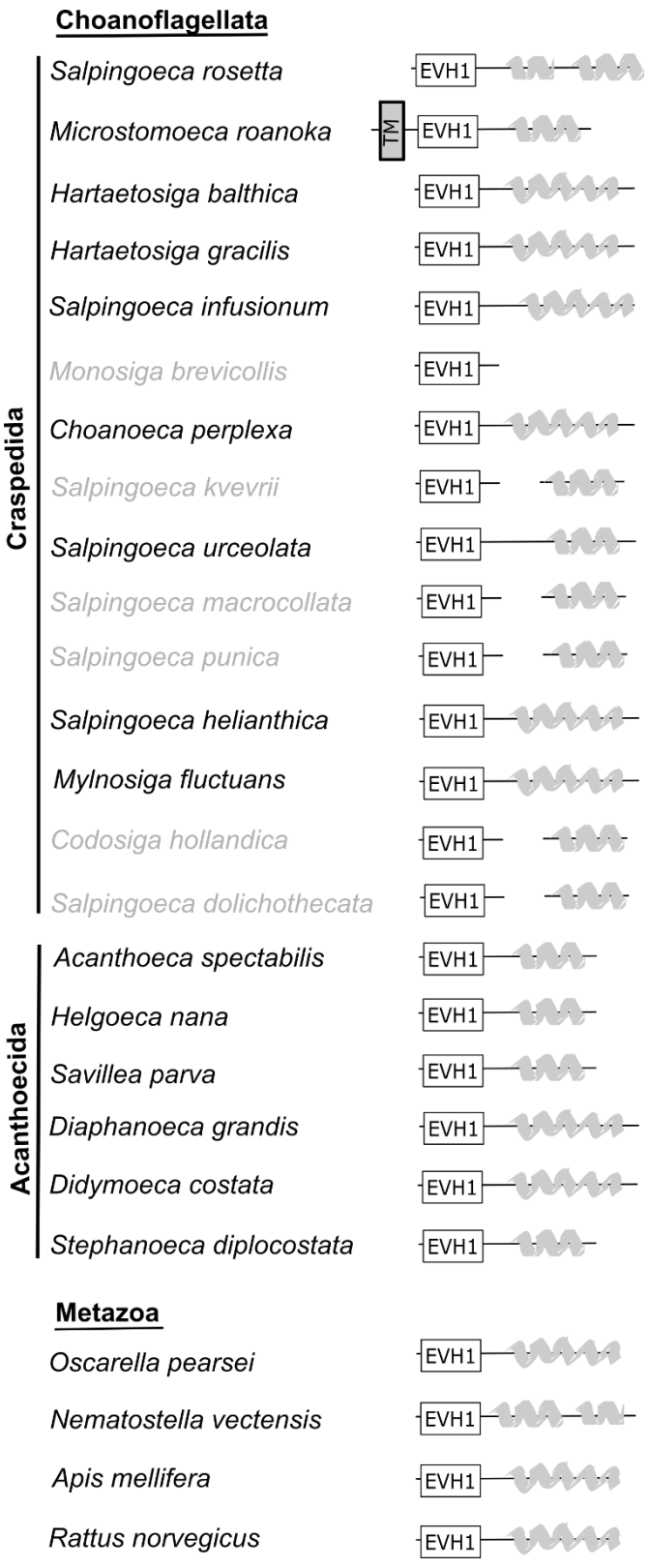


Figure 2-4: Homer domain architectures in different choanoflagellate and animal species. The EVH1 domains and coiled coil domains (spiral) (not to scale) are shown, as well as a transmembrane domain TM in one case. Regions of low complexity are not shown. The second coiled coil domain of *N. vectensis* was labelled as DKPG coiled coil in SMART search. Some species showed only the EVH1 domain, or two transcripts (one containing encoding the EVH1 domain and one encoding a coiled coil domain) (all transcripts shown had the highest-ranking BLASTp search hit to a protein classified as Homer). Species with these non-canonical domain architectures are written in grey.

As described by de Mendoza et al. (2010), choanoflagellates also possess choanoflagellate-unique MAGUKs. Most of these MAGUKs are extended, containing many PDZ domains. Some of them also contain a combination of a BAND-41 (B41) and a c-terminal FERM (FERM_C) domain. These domains are found in animal proteins that are localised to membranes and associate with the cytoskeleton (Chishti et al., 1998). A SMART search of animal proteins with a combination of these domains and the MAGUK module could only detect such a protein in the nematode *Pristionchus pacificus* (frm-2 protein; NCBI: PDM65888.1) (Figure 2-5). Accordingly, it is possible that the ancestor of animals and choanoflagellates had such a protein, which was then lost in almost all animals, but it is also possible that proteins with this domain architecture emerged independently in choanoflagellates and in *P. pacificus*.

2.3.4 Proteins with statistically significant sequence similarity to ionotropic glutamate receptors, animal Shaker/Shal-like voltage gated potassium channels and nitric oxide synthase are expressed in some choanoflagellate species

We also conducted sequence similarity searches to identify postsynaptic receptors and ion channels as well as proteins involved in postsynaptic signalling machineries. Proteins with statistically significant sequence similarity to IP3 receptors were detected in the majority of choanoflagellate transcriptomes. In one species, *Savillea parva*, no IP3 receptor could be detected, but it is possible that it is present nevertheless and is simply missing from the transcriptome assembly or could not be detected with the sequence search strategies applied. Another species, *Stephanoeca diplocostata*, was missing MIR repeats and therefore did not have the conserved domain architecture as expected for this protein family. This could be an artefact in the transcriptome assembly, or it could mean that although conserved in most species, MIR repeats are not essential for some aspects of the functionality of this protein. Proteins with statistically significant sequence similarity to CAMKII were detected in all choanoflagellate transcriptomes. The detected CAMKII sequence of *Didymoeca costata* is lacking the protein kinase domain.


Choanoflagellatea
Dlg gene family
MAGUK p55 gene family
Choanoflagellate unique MAGUKs
Craspedida
Salpingoeca rosetta
Microstomoeca roanoka
Hartaetosiga balthica
Hartaetosiga gracilis
Salpingoeca infusionum
Monosiga brevicollis
Choanoeca perplexa
Salpingoeca kvevrii
Salpingoeca urceolata
Salpingoeca macrocollata
Salpingoeca punica
Salpingoeca helianthica
Mylnosiga fluctuans
Codosiga hollandica
Salpingoeca dolichothecata
Acanthoecida
Acanthoeca spectabilis
Helgoeca nana
Savillea parva
Diaphanoeca grandis
Didymoeca costata
Stephanoeca diplocostata
Pristionchus pacificus protein frm-2
Legend domains
L27
PDZ
SH3
GuK
Ring finger
Band 4.1
FERM_C
FERM_adjacent

Figure 2-5: MAGUK domain architectures in choanoflagellates. Shown are domain architectures for MAGUKs of the Dlg family (all in the same gene family as animal Dlg 1-4 proteins), of the MAGUK p55 family (all in either one or another animal gene family containing animal MAGUK p55 proteins) and choanoflagellate unique MAGUKs (in gene families distinct from animal MAGUK families). Gene families correspond to OrthoMCL approximations by Richter et al. (2018). Species names are listed in the left. Domains are defined in the bottom right corner. The box at the very bottom shows the domain architecture of the only found animal protein with a combination of Band 4.1 and MAGUK module in *Pristionchus pacificus*.

Most choanoflagellates did not show any evidence for expression of ionotropic glutamate receptors, metazoan Shaker/Shal-like voltage-gated potassium channels and nitric oxide synthase. However, we found sequences with statistically significant similarity to these proteins for the first time in some choanoflagellate species. Occurrences of these genes are distributed over both choanoflagellate families, suggesting that the last common ancestor of animals and choanoflagellates had these genes. From transcriptome data, it is not possible to know if these genes are present in the genomes of these species without being expressed, however, it is known that they are missing in the two choanoflagellate species with sequenced genomes, *M. brevicollis* and *S. rosetta*. Metabotropic glutamate receptors are detected in all animal phyla but not in any available choanoflagellate genomes and transcriptomes, and neither in the *C. owczarzaki* genome.

We detected the expression of putative nitric oxide synthases (NOS) in some choanoflagellate species, comprising the canonical domain architecture of this enzyme in cyanobacteria and animals (NO_synthase, Flavodoxin-1, FAD-binding-1, NAD-binding-1). Vertebrate neuronal NOS (nNOS) has distinctive features that make the enzyme calcium-dependent and enable its binding to PSD-95 (nNOS has an N-terminal PDZ domain enabling binding to PSD-95). Out of all choanoflagellate and animal candidate NOS sequences studied here, the PDZ domain was only detected in *Mus musculus* and *Oscarella pearsei* NOS proteins.

2.4 Discussion

2.4.1 Extended transcriptome sequence survey refines the model of ancestral components of the animal postsynapse proteomes

Richter et al. (2018) revealed animal-like genes in an extended survey of choanoflagellate transcriptomes that were not detected in the genomes of the closely related choanoflagellate species *Monosiga brevicollis* and *Salpingoeca rosetta*. We conducted a sequence similarity survey of postsynaptic proteins in these choanoflagellate transcriptome datasets to test the hypothesis that the ancestral proteome of the animals and the choanoflagellates contained an extended range of these gene families.

Our survey for the first time detected proteins with statistically significant sequence similarity to animal Shaker/Shal-like voltage-gated potassium channels, ionotropic glutamate receptors, and nitric oxide synthase in choanoflagellate transcriptomes. Shaker/Shal-like voltage-gated potassium channels were identified in three craspedid choanoflagellate species (*Salpingoeca helianthica*, *Mylnosiga fluctuans*, and *Salpingoeca dolichothecata*). Other choanoflagellates have bacterial-like potassium channels without the voltage-gating BTB2 domain (Craspedida: *Salpingoeca rosetta*, *Salpingoeca infusionum*; Acanthoecida: *Stephanoeco diplocostata*) and might therefore not have a need for animal-like channels. Proteins with statistically significant sequence similarity to ionotropic glutamate receptors (iGluRs) were detected in *Choanoeca perplexa*, *Salpingoeca urceolata* (both Craspedida) and *Stephanoeca diplocostata* (Acanthoecida). Additionally, one diagnostic domain was detected in two other acanthoecidan species (*Diaphanoeca grandis* and *Didymoeca costata*). Ancestry of ionotropic glutamate receptors was suggested before, as iGluRs occur in plants (Lam et al., 1998). Interestingly, plant iGluRs react to a broader range of amino acids (Forde and Roberts, 2014). Putative nitric oxide synthases were detected in *Choanoeca perplexa*, *Salpingoeca urceolata* (both Craspedida) and *Helgoeca nana* (Acanthoecida). A protein with nitric oxide synthase domain, but missing some other diagnostic domains was identified in the craspedidan *Salpingoeca infusionum*. Nitric oxide synthases (NOS) occur in eukaryotes, prokaryotes and archaea (Santana et al., 2017). Nitric oxide signalling initiates protective cellular responses in cyanobacteria and plants in response to stressors (Beligni and Lamattina, 1999; Chen et al., 2003). Mice have

three NOS proteins, neuronal NOS (nNOS), endothelial NOS (eNOS) and inducible NOS (iNOS) (Moncada, 1999; Villanueva and Giulivi, 2010). iNOS is induced during immune response, whereas nNOS and eNOS are constitutively expressed, but calcium-dependent (Moncada, 1999; Villanueva and Giulivi, 2010). The enzymes produce nitric oxide (NO) in a variety of tissues (Villanueva and Giulivi, 2010). eNOS, for example, induces vascular relaxation through activation of guanylate cyclase via NO in smooth muscle cells (Moncada, 1999; Villanueva and Giulivi, 2010). nNOS activity is calcium dependent and is therefore activated through the NMDA receptor (an iGluR that acts as ligand gated calcium channel) activation (Brenman et al., 1996; Blackstone and Sheng, 2002). The enzyme binds PSD-95 via its PDZ domain (in this study a PDZ domain was only found in *Mus musculus* nNOS and the putative NOS of the sponge *Oscarella pearsei*), which brings it into proximity of NMDA receptor calcium influx. Although no metabotropic glutamate receptor was detected in all available choanoflagellate genomes and transcriptomes, a receptor that was phylogenetically placed in the animal mGluR family was detected in *Dictyostelium discoideum*, (Taniura et al., 2006), arguing that mGluR receptors could be ancestral and might have been lost in *Salpingoeca rosetta* and *Monosiga brevicollis* and potentially other choanoflagellate species. It is however difficult to draw conclusions for the absence of proteins in some or all choanoflagellate species, because we here only surveyed transcriptomes (with the exception of *S. rosetta* and *M. brevicollis*).

The putative presence of Shaker/Shal-like voltage-gated potassium channels, iGluRs and nitric oxide synthase in the animal ancestor has some implications for the origin of synapses, because it further supports that all three proteins did not evolve independently in the animal lineage but were one prerequisite for the evolution of synaptic function. Future work should conduct phylogenetic analysis in order to test if these proteins have a shared ancestry with the animal proteins. Within the frame of this analysis, it will be important to consider all protein domains separately (given each amino acid domain sequence is large enough and contains enough phylogenetic signal to allow for useful phylogenetic analysis) in order to test for the possibility that only single domains share common ancestry, whereas the full-length protein evolved independently in the different lineages by the shuffling of pre-existing protein domains. Sequence similarity

searches will produce data suggesting that these sequences are homologous, because the BLASTp algorithm creates local sequence alignments, identifying the most similar region between two sequences (Pearson, 2013). Creating separate phylogenies for every domain can help to unravel the evolutionary history for every domain by reconciling the domain trees with gene trees that were previously reconciled with species trees (Stolzer et al., 2015).

2.4.2 Components of the structural framework for postsynaptic signalling machineries originated before the emergence of animals

Even though some elements important for postsynaptic reception and signalling were present before animal origins, there are still many missing components and there is no evidence suggesting that postsynaptic-like signalling machineries existed in the ancestor of animals and choanoflagellates. Interestingly, however, most scaffolding proteins necessary to organise postsynaptic signalling machineries seem to be ancestral and are conserved in choanoflagellates and animals (Alié and Manuel, 2010; Burkhardt et al., 2014). This suggests that the framework required to build signalling machineries already existed in their last common ancestor. The high level of conservation proposes that scaffolding proteins are of high importance in both lineages and are probably implicated in important signalling complexes. Scaffolding proteins are pleiotropic even in animals (Woods et al., 1996; Worley et al., 2007; Burkhardt and Sprecher, 2017). Their function varies in different tissues depending on the receptors and signalling proteins they are associated with (Montgomery et al., 2004). Therefore, it makes sense that a protein scaffold could be re-used and elaborated for different functions in different species as well, enabling the evolution of postsynaptic and other animal signalling machineries.

Obviously, scaffolding proteins evolved over time in both choanoflagellates and animals. This can be observed for membrane associated guanylate kinases. Choanoflagellates seem to have evolved own unique MAGUK families – given current genome/transcriptome sampling (de Mendoza et al., 2010 and this study). Canonical animal-like MAGUKs (Dlg and MAGUK p55-like) were detected only in transcriptomes of craspedidan species and not in transcriptomes of acanthoecidan species. Because we here only surveyed transcriptomes, it is

possible that animal-like MAGUKs do occur in Acanthoecida, but were simply not expressed at the time point of sequencing. Nevertheless, differences in the detection of expression of Dlg and MAGUK p55-like MAGUKs between the two families would also be interesting. Craspedida species produce organic cell coverings and stalks and many species of this family are capable of forming colonies with cell-cell contact via incomplete cell division (Carr et al., 2008; Nitsche et al., 2011; Leadbeater, 2015). Acanthoecida on the other hand form inorganic silicate based loricas and were never described in colonies with cell-cell contact (Carr et al., 2008; Nitsche et al., 2011; Leadbeater, 2015). Therefore, the need for Dlg and p55 MAGUK proteins exclusively in Craspedida species suggests that these proteins could be involved in Craspedida unique processes. Alternatively, it is possible that choanoflagellate-unique MAGUKs can take over similar functions to animal-like MAGUKs.

2.5 Outlook

By extending the amount of choanoflagellate species used for sequence similarity-based surveys of postsynaptic protein homologs, we could reveal new insights about the ancestral gene content. We further observed a high degree of sequence conservation of candidate postsynaptic scaffolding proteins, which suggests that they are functionally important in choanoflagellates. For future analysis, we aim to include more closely related species to choanoflagellates and animals than only *C. owkzarzaki*, as recently many genomes and transcriptomes of more holozoans were sequenced (de Mendoza et al., 2015; Torruella et al., 2015; Grau-Bové et al., 2017). One important future step for our analysis will be the phylogenetic analysis of the identified proteins with statistically significant sequence similarity to synaptic proteins in order to elucidate their evolutionary history. Furthermore, here we looked at a key subset of postsynaptic protein homologs. Including more proteins might reveal additional insights.

3 Second data chapter: Homer and Shank, two proteins organising signalling machineries in the postsynapse were putatively ancestral binding partners prior to the evolution of animals

Tarja T. Hoffmeyer[1,2,3], Fiona R. Savory[1], Thomas S. Walter[4], A. Radu Aricescu[4,5], Thomas A. Richards[1] and Pawel Burkhardt[2,3]

[1]University of Exeter; [2]Marine Biological Association of the UK (Plymouth);[3]Sars International Centre of Marine Molecular Biology, University of Bergen (Norway); [4]Division of Structural Biology, University of Oxford (UK); [5]University of Cambridge (UK).

Text: Tarja Hoffmeyer

Experimental work (lab and PC): Tarja Hoffmeyer

Antibody generation and immunostaining: Dr Pawel Burkhardt

Project idea: Tarja Hoffmeyer and Dr Pawel Burkhardt

X-ray Crystallography: Dr Thomas Walter, Tarja Hoffmeyer, and Prof Dr Radu Aricescu

Ancenstral protein reconstruction: Dr Fiona Savory, Tarja Hoffmeyer, and Prof Dr Thomas Richards

3.1 Introduction

In chemical synapses signals are transmitted from one neuron to another neuron or effector cell via neurotransmitters. These transmitters are released from the presynaptic membrane of an axon terminal and bind to receptors in the postsynaptic membrane (Squire et al., 2008). Some of these receptors act as sodium/potassium ion channels and can induce a depolarisation of the postsynaptic membrane, which leads to a signal transduction (Kennedy, 2000). Other receptors act as calcium channels or influence the cellular calcium signalling, which is involved in the regulation of synaptic strength (synaptic plasticity) (Kennedy, 2000). The strength of a synapse can be increased via recruiting more receptors, dendritic spine growth through a remodelling of the actin cytoskeleton, or presynaptic modulations (Prast and Philippu, 2001; Sala et al., 2001; Kruijssen and Wierenga, 2019).

Postsynaptic scaffolding proteins are of major importance to couple receptor-mediated processes with downstream signalling machineries, by providing binding platforms via many protein-protein interaction sites (Kennedy, 2000). The scaffolding proteins Homer and Shank are particularly important at

vertebrate glutamatergic synapses. Together, they form a multimeric platform (Baron et al., 2006; Hayashi et al., 2009), linking complexes coupled to different receptors to components of the intracellular calcium machinery and regulators of the actin cytoskeleton (Naisbitt et al., 1999; Sala et al., 2001).

Vertebrate Shank proteins are comprised of ankyrin repeats, an SH3 domain, a PDZ domain, a proline rich region and a SAM domain (Naisbitt et al., 1999) (Figure 3-1). Multimerisation of Shank proteins is facilitated via its SAM domain; and the PDZ domain of Shank binds the C-terminus of GKAP (Naisbitt et al., 1999) (links the Shank platform to PSD-95). In its proline rich region Shank carries a PPxxF motif, which is bound by the EVH1 domain of Homer (Barzik et al., 2001; Beneken et al., 2000; Tu et al., 1999). In addition to this EVH1 domain Homer proteins carry a coiled coil domain (Ponting and Phillips, 1997; Xiao et al., 1998). This domain is required for the tetramerisation of Homer proteins (Hayashi et al., 2006, 2009) (Figure 3-1).

Apart from Shank, EVH1 domains also bind receptors, such as group I metabotropic glutamate receptors (mGluR1 and mGluR5) (Beneken et al., 2000). These receptors are localised to the postsynaptic membrane and are connected to G-proteins, which induce several cellular pathways upon glutamate binding, such as the synthesis of inositol triphosphate (IP3) (Abe et al., 1992; Aramori and Nakanishi, 1992; Masu et al., 1991; Knöpfel et al., 1995). Interestingly, Homer also binds to IP3 receptors, coupling mGluR signalling and calcium release from the ER (Tu et al., 1998). Homer and Shank both play a role in cytoskeleton remodelling and spine head growth, facilitated by Homer interaction with Rho GTPase Cdc42 and Shank interaction with actin nucleation factor cortactin (Caroni et al., 2012; Naisbitt et al., 1999; Sala et al., 2001; Shiraishi et al., 1999).

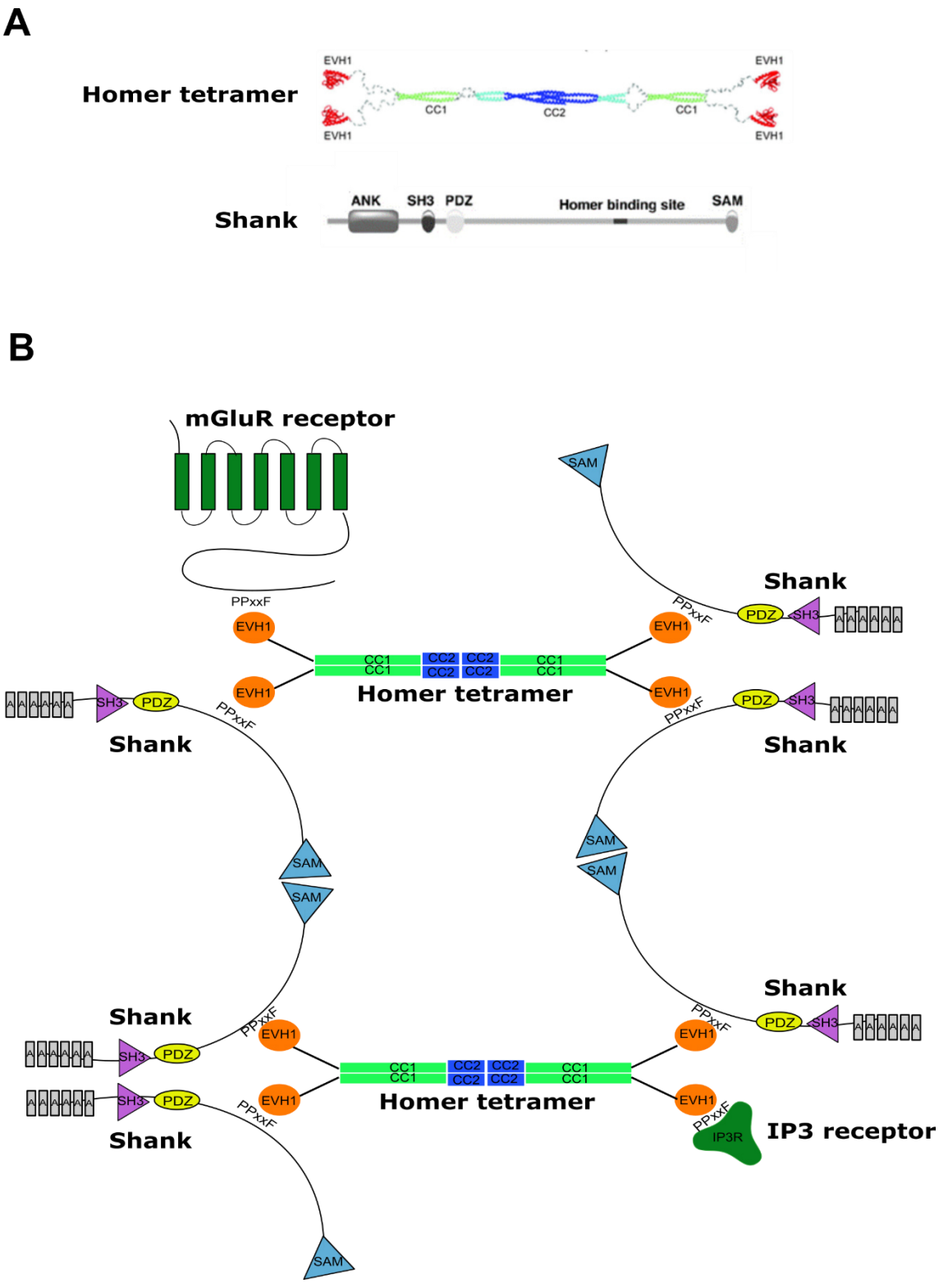


Figure 3-1: Domain architectures and complex of rat Homer and Shank proteins. A) Homer occurs as tetramer (one monomer consists of an EVH1 domain and two coiled coil (CC1 and CC2) domains). The coiled coil domains interact, forming the tetramer. Shank is composed of ankyrin repeats (ANK/A), an SH3 domain, a PDZ domain, a proline-rich Homer binding site with PPxxF motif, and a SAM domain. Pictures adapted from Hayashi et al. (2009) and Burkhardt (2015), with permission from Cell (Elsevier) and Journal of Experimental Biology (The Company of Biologists), respectively. B) Homer and Shank proteins interact and form a polymeric network structure (after Hayashi et al. 2009). Shown are Homer EVH1 interactions with PPxxF motifs of Shank proteins, metabotropic glutamate receptors (mGluR receptor) and IP3 receptors, as well as Shank multimerisation via SAM domains.

Homer and Shank protein homologs are found in all animals, as well as in their close unicellular relatives, the choanoflagellates, filastereans and ichthyosporeans (holozoan phyla) (Burkhardt and Sprecher, 2017). The actin nucleation factor cortactin occurs in bilaterians, cnidarians, a variety of sponges, as well as choanoflagellates and filastereans, but was not detected in non-holozoan eukaryotes (Sebé-Pedrós et al., 2013a; Riesgo et al., 2014). The choanoflagellate species *Salpingoeca rosetta* encodes a protein annotated as cortactin (NCBI accession: PTSG_10805), which has the diagnostic domain architecture (HS1 repeat, coiled-coil, SH3) (Fairclough et al., 2013). Burkhardt et al. (2014) showed that tetramer formation of Homer is conserved in the choanoflagellate species *Salpingoeca rosetta*, but that Homer probably plays another role in choanoflagellates, as it is found localised to the nucleus with interaction partners mainly distinct from the ones known from vertebrate neurons. In contrast to the Shank homolog found in the closely related choanoflagellate species *Monosiga brevicollis*, *S. rosetta* Shank does not possess a SAM domain and can therefore not multimerise via this domain (Fairclough et al., 2013; King et al., 2008). Although *S. rosetta* Shank has a PPxxF binding motif, it was not co-precipitated with Homer from *S. rosetta* colony lysate (Burkhardt et al., 2014). This might have been due to experimental conditions, yielding the possibility that Homer and Shank can bind in *S. rosetta* under specific circumstances (Burkhardt et al., 2014). Alternatively, it is also possible that Homer and Shank do not bind in choanoflagellates. Nevertheless, the presence of binding motifs in both proteins suggests that at least the binding capacity is conserved among choanoflagellates and animals. We therefore hypothesise:

Homer and Shank were binding partners in the last common ancestor of choanoflagellates and animals.

The following specific aims were set in order to test the hypothesis:

1) Computationally reconstruct putative Homer EVH1 sequences from the animal and the choanoflagellate ancestors.
2) Use isothermal titration calorimetry to test if Homer-Shank binding is conserved within animals, within choanoflagellates and among their putative ancestors.

3.2 Materials and methods

3.2.1 Alignment and phylogeny construction

For alignments and phylogenies new choanoflagellate transcriptome sequences were included (Richter et al., 2018). Proteins were identified via BLASTp search (Altschul et al., 1990) of the *Salpingoeca rosetta* Homer or Shank protein sequence against a peptide database of each of the choanoflagellate species. Candidate sequences were then subjected to a reciprocal BLASTp search against the NCBI database (https://www.ncbi.nlm.nih.gov/) and domain organisation was evaluated with the SMART tool (Schultz et al., 1998) (http://smart.embl-heidelberg.de/) (detailed procedures, cut-offs and date of data base access described in section 2.2). For Homer, only proteins with a predicted domain arrangement of EVH1 domain and 1-2 coiled coil domains with a highest-ranking reciprocal BLASTp search hit to another protein classified as Homer were included. For Shank, the domain arrangement even in animals is more variable. Two different domain arrangements for choanoflagellate proteins were accepted, when the reciprocal BLASTp search highest-ranking hit was a protein classified as Shank and when the protein contained at least ankyrin repeats and a PDZ domain and either an SH3 and a SAM domain or 1-2 WH2 domains (according to described domain architectures of *Monosiga brevicollis* and *Salpingoeca rosetta* Shank, respectively). Domain arrangements of identified putative Shank and Homer proteins are provided in section 2 (Figure 2-3 and Figure 2-4, respectively). References for choanoflagellate sequences are provided in supplementary Table 6-2. References for other sequences used for alignments and phylogenies are provided in supplementary Table 6-3. For both alignments sequences were chosen to cover a range of animal and choanoflagellate lineages. Homer sequences used for the alignment are slightly more comprehensive, as they were used for the phylogeny. Initially, they covered even more species, however, poorly aligned sequences, as well as sequences likely to cause long-branch attraction artefacts were excluded (resulting in, for example, the exclusion of ctenophore and placozoan sequences for this sub-analysis).

Alignments were made with Seaview v. 4.7 (Gouy et al., 2010) using the MUSCLE alignment tool (Edgar, 2004). Shank sequences were then further surveyed for PPxxF motifs (Figure 2-3). We generated several alternative alignments that differed in the length of the sequences and the species included.

For instance, we generated alignments that contained only the EVH1 domain region of sampled sequences (labelled: short alignment). Additionally, we generated extended alignments (labelled: long alignment) that contained additional conserved sequence regions from Homer coiled coil domain regions. These extended alignments helped separate choanoflagellate from animal clades in subsequent phylogeny constructions and were therefore chosen for further analysis. Alternative phylogenies based on different alignments are provided and discussed in section 6.2.4. Importantly, in the extended alignment, we excluded non-conserved regions in between conserved Homer EVH-1 and coiled-coil regions. The full-length alignment of Homer including all species is provided in the supplement (Figure 6-2). Regions that were used for the analysis are in grey boxes. The final alignment used for analysis is shown in Figure 3-2. This alignment contains variable sites that were masked for phylogeny construction (grey marked areas in the alignment of Figure 3-2).

Phylogenies were calculated with a bootstrap analysis using IQ tree (Nguyen et al., 2014) (maximum likelihood analysis, 1000 bootstraps, model LG+G4 according to BIC, AIC and AICc, as selected by IQ tree analysis). Tree topology of the created phylogeny and the resulting ancestral sequences (section 3.2.3) were dependent on varying sequence length sampled, the inclusion and exclusion of certain taxa and the inclusion of an outgroup (alternative phylogenies and ancestral sequences are presented in supplementary section 6.2.4).

3.2.2 X-ray crystallography and structure prediction

The *S. rosetta* Homer EVH1 domain (aa 1-110 of Homer3, NCBI: XP_004998981.1) was expressed in *E. coli* and purified via poly-histidine (His-) tag affinity purification and ion exchange chromatography (as described in section 3.2.4). Subsequent size exclusion chromatography in 150 mM NaCl and 10 mM Tris pH 7.4 on a HighPrep™ 16/60 Sephacryl™ S-100 HR column on an Äkta prime plus system (GE Healthcare Biosciences AB, Uppsala, Sweden) was performed to select for only one mono- or oligomeric state of the protein before crystallisation. The protein was concentrated to 43.59 mg/mL in Ultra-4 centrifugal devices for proteins larger than 3 kDa (Amicon, Kent, UK). Most conditions in the pre-crystallisation test (PCT test Kit (Hampton Research, Aliso

Viejo, CA, USA) stayed clear even using these high concentrations. Crystallisation trials were set up (Block 1, Block 2, Block 3, PACT premier, Morpheus (Molecular dimensions, Sheffield, UK), Index, PEGRx, SaltRx (Hampton Research, Aliso Viejo, CA, USA)) in MRC plates with the sitting drop method, mixing 0.75 µl of the protein with 0.75 µl of the reservoir solution for the sitting drop. Drops were prepared with a Cartesian #2 sqII2673 robot (Digilab, Marlborough, MA, USA) at the division for Structural Biology (Oxford, UK). Plates were incubated at room temperature and at 4 °C (trials at 4 °C were set up with fresh protein at a concentration of 43.20 mg/mL). Because crystallisation was unsuccessful under these conditions, we prepared fresh protein and subjected it to lysine methylation (Walter et al., 2006). The protein was dialysed into 50 mM HEPES (pH 7.5), 250 mM NaCl at 4 °C. At protein concentrations of 1 mg/mL or less, 20 µl of freshly prepared 1 M dimethylamine-borane complex (ABC; Fluka product 15584) and 40 µl of 1 M formaldehyde (made form 37 % stock; Fluka product 33220) were added per mL protein solution. The reactions were gently mixed and incubated at 4 °C for 2 hours, before adding an additional 20 µl 1M ABC and 40 µl 1 M formaldehyde per mL protein solution. After another 2 hours incubation at 4 °C, a final 10 µl of 1M ABC per mL protein solution were added and the reaction was incubated over night at 4 °C. Precipitate was removed by centrifugation, following size exclusion chromatography as described above for the non-methylated protein. Due to high loss of protein, we were only able to concentrate the protein to 20.44 mg/mL and trials were set up (Block 1, Morpheus and SaltRx). One condition (1.6 M tri sodium citrate at a pH of 6.5) produced crystals. Optimisation screens on this condition (3-row optimisation with decreasing concentration of the reagent and protein:reservoir ratios of 1:1 vs 2:1 vs 3:1; pH – tri sodium citrate concentration grid) as well as additive screens on this condition and addition of sodium citrate to Block 1 vs SaltRx were tried. Most crystals appeared in the pH grid screen. These crystals were subjected to X-ray diffraction (Diamond Light Source, Didcot, UK).

S. rosetta Homer EVH1 structures in this thesis were modelled computationally (https://swissmodel.expasy.org) via alignment to the rat Homer EVH1 structure (Beneken et al., 2000; PDB ID: 1DDV). Structures were visualised and aligned with Pymol 2.3 (Schrödinger).

3.2.3 Ancestral protein reconstruction

Ancestral sequences were calculated via PAML4 (Yang, 1997, 2007). The input tree was the phylogeny rooted on the Homer EVH1 sequence of the filasterean *Capsaspora owczarzaki* as the outgroup (when included) or on the animal/choanoflagellate division (when no outgroup was included). Input alignment was the Homer alignment including indels (insertion/deletion sites). These indels, present only in a subset of the species sampled, were judged unlikely to be ancestral characters and were removed from the ancestral sequence following ancestral sequence reconstruction. The PAML output contains information about the ancestral sequence (amino acids and probabilities for these amino acids at every site). R code (courtesy of Dr Fiona Savory) was used to recover this information, providing ancestral sequences in .txt format and simplifying the identification of variant amino acid positions and plausible alternative residue characters in these sequences. This code (provided in the supplement with permission of Dr Savory (Section 6.2.3)) made use of the R language and working environment R 3.4.0 (R Core Team, 2017) and the Stringr package 1.2.0 (Wickham, 2017) and was executed in Rstudio 1.0.44 (Rstudio Team, 2015). We used these data to identify one candidate ancestral sequence equivalent to the most likely ancestral sequence (given current taxon sampling) and then an alternative amino acid sequence representing an amalgamation of the variant amino acid positions with the second highest probability that is larger than 20 %.

3.2.4 Proteins and peptides

Expression constructs for Homer EVH1 and Shank constructs were codon optimised, synthesised and subcloned into pET28a plasmid vectors with NdeI and XhoI restriction enzymes (as outlined in Table 3-1). The domain boundaries for the Homer EVH1 domain were determined with the SMART tool (Schultz et al., 1998) (http://smart.embl-heidelberg.de/) for the rat and the *Salpingoeca rosetta* protein, and adapted according to the published rat structure of Homer EVH1 (Beneken et al., 2000; PDB ID: 1DDV). Domain boundaries for the other used EVH1 domains were determined according to assessment of conservation patterns across the multiple sequence alignment (Figure 3-2). Rat Shank used

for our experiments is a construct by Hayashi et al. (2009) (Shank1CΔPEST from the *Rattus norvegicus* Shank1A sequence (NCBI ref AAD29417.1 without amino acids 1-576 upstream of the PDZ domain and without PEST sequences from amino acids 700-1463 and 1522-2010). A similar construct was cloned for the Shank protein identified in the choanoflagellate species *Stephanoeca diplocostata*, using the Shank alignment (Figure 6-1) as reference. However, this protein expressed from this construct was insoluble and was not used for subsequent experiments.

Table 3-1: Information about protein constructs produced.

Construct name	Construct description	Cloning
Homer EVH1 *Rattus norvegicus*	Homer 1 NCBI ref. NP_113895.1 aa 1-111	pET28+ vector (Synbio Tech, NJ, USA) Cloning with NdeI and XhoI restriction enzymes through the company.
Homer EVH1 *Oscarella pearsei*	Compagen.org: OCAR_TPEP_130911.fa (m.8451) aa 3-111	
Homer EVH1 *Stephanoeca diplocostata*	Homer EVH1 aa 3-111 Dataset 2 (Richter et al. 2018 paper) FR isolate reference: m.51812 https://dx.doi.org/10.6084/m9.figshare.5686984.v2	
choanc Homer EVH1	Reconstructed ancestral choanoflagellate Homer EVH1	
Anianc Homer EVH1	Reconstructed ancestral animal Homer EVH1	
Shank *Rattus norvegicus*	Shank1CΔPEST (Hayashi et al. 2009) from Shank1A sequence NCBI ref. AAD29417.1 (excluding aa 1-576, 700-1463 and 1522-2010)	
Shank *Stephanoeca diplocostata*	similar construct as the rat construct according to alignment (turned out to be insoluble)	
Homer EVH1	NCBI XP_004998981.1 aa 7-116	pET28+ vector (Novagen, Merck, Darmstadt, Germany)

Salpingoeca rosetta	Cloning with NdeI and XhoI restriction enzymes through Dr Pawel Burkhardt.

Expression vectors were transformed into *E. coli* BL21 (DE3) (pLysS for all but *S. rosetta* Homer EVH1) (Novagen, Merck, Darmstadt, Germany). Bacteria were grown in 3x 1 L autoclaved Terrific Broth (TB) medium (48.2 g Terrific Broth Ezmix powder microbial growth medium (MERCK, NJ, USA), and 4 mL glycerol per L MilliQ H_2O) with 30 µg per mL kanamycin sulphate at 37 °C and 200 rpm to an optical density of 1.2-1.5. Expression was induced with 0.5 mM IPTG for 3 hours at 37 °C and 200 rpm. Bacteria were pelleted and redissolved in 50 mL wash buffer (500 mM NaCl, 20 mM Tris pH7.4).

Proteins were extracted from *E. coli* via 20 min incubation with lysozyme and 0.4 mM PMSF at room temperature, 3x 30 seconds sonication on ice with 30 % amplitude and 2-1 pulse, and 10 min incubation with DNase1, 2 % Triton X-100, 0.4 mM PMSF and 2 mM MgCl2. Bacterial debris was pelleted at 25,000 g and 4 ˚C for 40 minutes, leaving the protein of interest in the supernatant.

Affinity chromatography was performed with HisPurTM Cobalt beads (Thermo Scientific, Rockford, IL, USA), rotating at 4 ˚C for 2 hours. Beads were washed with wash buffer and the protein was eluted from cobalt beads with 400 mM imidazole in wash buffer. Proteins were dialysed against 100 mM NaCl (50 mM for *S. rosetta* Homer EVH1), 20 mM Tris pH 7.4 and the polyhistidine tag was cleaved off with thrombin with a specific activity of 1500 U/mg protein at a concentration of 10 U/mL (MP Biomedicals, Santa Ana, CA, USA). Ion exchange chromatography was performed on an Äkta prime plus (GE Healthcare Biosciences AB, Uppsala, Sweden) with HiTrapTM Q HP column (for the rat Shank construct (isoelectrical point (pI) 5.04) and *S. rosetta* Homer EVH1 (pI 7.03)) or HiTrapTM SP HP column (for rat Homer EVH1 (pI 9.65); *O. pearsei* Homer EVH1 (pI 9.26); *S. diplocostata* Homer EVH1 (pI 9.63); choanc Homer EVH1 (pI 9.35) and anianc Homer EVH1 (pI 8.98)). Proteins were concentrated in Ultra-4 Centrifugal Devices for proteins larger than 3 kDa (Ami-con, Kent, UK).

Protein concentration was determined with a Nanodrop One spectrophotometer (Thermo Fisher Scientific, Rockford, IL, USA) to an accuracy of ± 0.1 mg/mL.

Peptides were ordered (GeneCust, Boynes, France) for the known rat Homer binding site and putative *S. rosetta* Homer binding sites from Shank homologs. Peptides included PPxxF motifs as well as 8 amino acids before and after the motif (Table 3-2). Peptides were dissolved in ITC buffer to a concentration of 25 µM assuming all peptide is dissolved.

Table 3-2: Peptides used for PPxxF binding experiments.

Protein of origin	Peptide sequence	Amino acid positions
Rat Shank1a	FLFAEPLPPPLEFSNSFEKPE	1487-1507
S. rosetta Shank	ADAAPVKAPPVYFARTRTSSV	259-279
S. rosetta Shank	EASTSPFIPPPMFLADVQMTT	980-1000

3.2.5 Choice of binding assay

First binding experiments were GST pull-downs, performed with *S. rosetta* Shank peptide (amino acids 980-1000) (Table 3-2) and *S. rosetta* Homer EVH1. The peptide was recombinantly expressed with GST-tag and affinity purified. *Escherichia coli* strain BL21 (DE3) competent cells (Novagen, Madison, WI, USA) were transformed with a pGEX-6P-1 plasmid vector (Synbio Technologies, Monmouth Junction, NJ, USA), containing a DNA sequence encoding purely GST or GST with the *S. rosetta* Shank peptide EASTSPFIPPPMFLADVQMTT. The linker sequence between GST and the peptide is SDLEVLFQGPLGSPEF. After the peptide sequence four amino acids (AAAS) follow before the stop codon. These sequences arise from the vector, using EcoRI and Not1 restriction enzymes for cloning (cloning through Synbio Technologies, Monmouth Junction, NJ, USA). The GST protein without peptide ended with the following sequence using the same restriction enzymes: SDLEVLFQGPLGSPEFPGRLERPHRD. Bacteria were grown in 2.5 L Lysogeny Broth (LB) medium (10 g tryptone, 5 g yeast extract, 10 g NaCl and 1 mL 1M NaOH per litre MilliQ water (autoclaved) with 100 µg/mL ampicillin sodium salt to an optical density (OD) of 0.8. Expression was induced with 0.5 mM IPTG overnight at 16 ˚C and 200 rpm.

Bacteria were pelleted and dissolved in 50 mL glutathione beads wash buffer I (500 mM NaCl, 50 mM Tris pH 7.4). Proteins were extracted from *E. coli* with 20 min incubation with lysozyme and 0.4 mM PMSF (for protein stabilisation) at room temperature, 3x 30 seconds sonication on ice with 30 % amplitude and 2-1 pulse, and 10 min incubation with DNase1, 2 % Triton X-100, 0.4 mM PMSF and 2 mM $MgCl_2$ (to disrupt lipid membranes and degrade DNA). Bacterial debris was pelleted at 15,000 g and 4 ˚C for 40 minutes, leaving the protein of interest in the supernatant. Affinity chromatography was performed with glutathione-agarose beads (Thermo Fisher Scientific, Rockford, IL, USA), rotating at 4 ˚C for 2 hours. Beads were washed with wash buffer II (10 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 2 mM $MgCl_2$, 1 mM DTT) and the protein was eluted from glutathione beads with 50 mM glutathione in wash buffer II. Proteins were dialysed against 10 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 2 mM $MgCl_2$, 1 mM DTT. 100-200 µg peptide were bound to 10 µL glutathione agarose beads and then incubated with different concentrations of the EVH1 region of *S. rosetta* Homer for 1 hour at room temperature. Each sample was washed 3x with binding buffer (10 mM Tris pH 7.4, 150 mM NaCl, 1 mM EDTA, 2 mM MgCl2, 1 mM DTT) and then eluted with SDS sample buffer (containing β-Mercaptoethanol) at 95 ˚C for 5 minutes. Binding was confirmed via SDS-PAGE. With this assay, no reliable difference between the treatment and the negative control could be shown (Figure 6-4).

Next, we recombinantly expressed a rat Shank construct and rat Shank Homer EVH1 (as described in section 3.2.4) and tried other binding assays with these proteins as positive control to evaluate the binding assay before testing choanoflagellate and ancestral sequences. We performed gel filtration (size exclusion chromatography on a SuperdexTM 75 Increase 10/30 GL column on AEKTA in 20 mM sodium phosphate buffer pH 7.4; 150 mM NaCl; results in Figure 6-3) and isothermal titration calorimetry (ITC; further explained in section 3.2.6). ITC showed clear binding of the rat proteins. Thus, it was used for all further experiments.

3.2.6 Isothermal titration calorimetry

Proteins were dialysed twice into Isothermal titration calorimetry (ITC) buffer (20 mM sodium phosphate buffer pH 7.4; 150 mM NaCl) at 4 ˚C. Lyophilised peptides were dissolved in 1-2 mL of the same buffer (filtered before through 0.2 µm pore filter). Dilutions were prepared with this buffer and proteins were concentrated (as described in 3.2.4) as appropriate to reach experimental concentrations. Experimental ITC buffer was filtered through 0.2 µm pore filter and proteins were centrifuged for 10 min at 16,000 g, before degassing all samples at 400 mm Hg vacuum at 25 ˚C for 15 minutes. A Nano ITC (TA Instruments, Newcastle, DE, USA) was used for measurements. The sample cell was filled with approximately 300 µL (results in 170 µL in the actual sample cell) and the injection syringe was filled with approximately 55 µL. For all experiments, Homer (in syringe) was titrated into Shank (in sample cell). Tests with different protein concentrations showed that 25 µM Shank and 200 µM Homer EVH1 were appropriate concentrations and were therefore applied to all treatments. Homer EVH1 was titrated into ITC buffer as a negative control. The stirring rate during ITC was set to 250. A first injection of 0.5 µL (real: 0.47 µL) was made after auto equilibration, to be excluded from the dataset later; followed by 21 injections of either 2.2 µL (real: 2.19 µL) (used for first experiments with positive controls and the first *S. rosetta* test) or 1.9 µL (real: 1.89 µL) every 300 seconds. Incremental titration was chosen, and expected heats was set to medium. The resulting raw data graph was adjusted in TA instruments software, limiting integration regions to the main heat peak and manually adjusting the baseline. Data fitting was performed with the TA instruments software with the Independent model. A blank constant was subtracted, so that the curve approaches zero.

3.3 Results

3.3.1 Choanoflagellates possess the protein regions required for Homer and Shank binding

Homologs to the protein Homer are present across the animal and the choanoflagellate lineage with very similar domain architectures (1 EVH1 domain and 1-2 coiled coil domains) (Figure 2-4). Beneken et al. (2000) described four amino acid residues in the Homer EVH1 binding region that are involved in the

binding of the PPxxF motif. An alignment of animal and choanoflagellate Homer EVH1 sequences (Figure 3-2) showed that these four residues are conserved in all sampled animal and choanoflagellate species. An additional two amino acids that are generally conserved in EVH1 domains are also present. An exception is the valine residue in the 6th marked position, which in the choanoflagellates *Salpingoeca rosetta* and *Microstomoeca roanoka* is replaced by leucine, and in the sponges *Amphimedon queenslandica* and *Oscarella pearsei*, as well as in the honey bee *Apis mellifera*, is replaced by isoleucine. Valine, leucine and isoleucine are three amino acids, which share similar properties as they all have a hydrophobic, aliphatic, more or less branched side chain. The substitution might therefore still allow for binding. Furthermore, it was shown that a substitution of valine with alanine in rat Homer EVH1 still allows binding to Shank3 PPxxF and only diminished binding to mGluR PPxxF (Beneken et al., 2000).

The sequence and domain architecture of Shank proteins is more variable even among closely related animal species. The same is true for Shank homologs found in choanoflagellates (Figure 6-1 showing the Shank alignment, and Figure 2-3 showing Shank domain architectures of animals and choanoflagellates). Not all choanoflagellate species possess clearly attributable Shank homologs (Figure 2-3), and among those that possess putative Shank proteins, domain architectures and sequences are very variable. Most, but not all animal and choanoflagellates species have PPxxF motifs, which are required for Homer EVH1 domain binding (indicated in Figure 2-3).

3.3.2 The *S. rosetta* Homer EVH1 domain is structurally similar to the rat Homer EVH1 domain

We modelled the *S. rosetta* Homer EVH1 structure against the rat Homer EVH1 structure (Beneken et al., 2000) and can show that the two domains are structurally highly similar (Figure 3-3). The structures slightly diverge in some areas, but – at least in this reconstruction – the PPxxF ligand binding region appears identical (Figure 3-3B). This is coherent with the fact that important amino acids at the ligand binding site are conserved between animals and choanoflagellates (Figure 3-2). Computational models can be useful to get an idea about structure similarity of protein homologs, but they always represent

approximations. In order to resolve the true *S. rosetta* Homer EVH1 structure, we set up trials for X-ray crystallography. The choanoflagellate domain did not seem to crystallise in any of the conditions we tried (Material and methods section 3.2.2). Using lysine methylation, crystals appeared in some conditions. Most of these crystals turned out to be salt crystals, but one condition with 1.6 M tri sodium citrate at a pH of 6.5 gave protein crystals. The resolution of 3.5-4 Å X-ray diffraction is not sufficient to solve the structure of *S. rosetta* Homer EVH1. Furthermore, crystals in this condition appeared very late (after 6 months), indicating protein degradation, upon which crystals can form from partially degraded protein. We did not manage to solve these problems yet, but nevertheless present here the results of our optimisation, which could be used for future crystallisations.

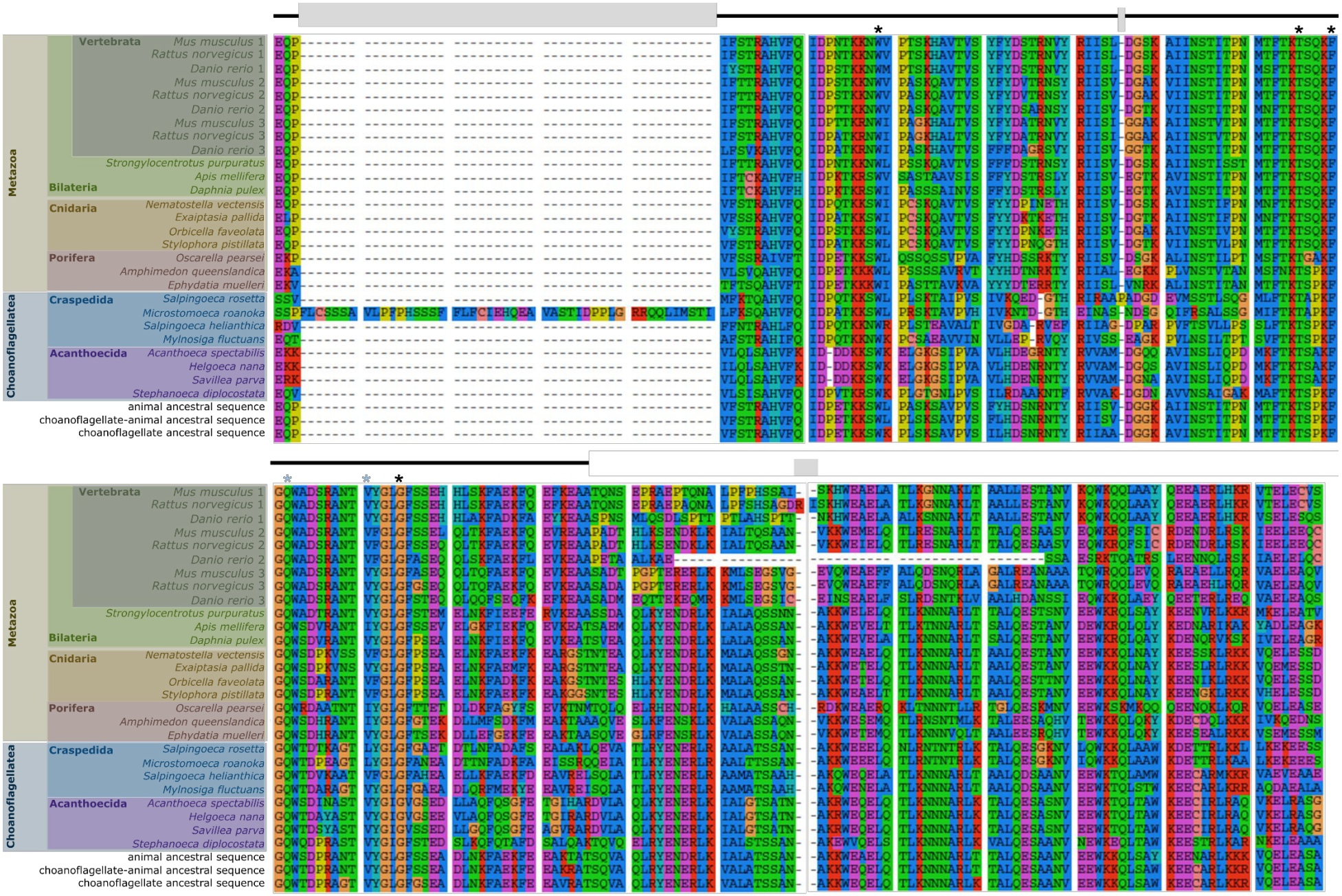

Metazoa
Vertebrata
Mus musculus 1
Rattus norvegicus 1
Danio rerio 1
Mus musculus 2
Rattus norvegicus 2
Danio rerio 2
Mus musculus 3
Rattus norvegicus 3
Danio rerio 3
Strongylocentrotus purpuratus
Bilateria
Apis mellifera
Daphnia pulex
Cnidaria
Nematostella vectensis
Exaiptasia pallida
Orbicella faveolata
Stylophora pistillata
Porifera
Oscarella pearsei
Amphimedon queenslandica
Ephydatia muelleri
Choanoflagellatea
Craspedida
Salpingoeca rosetta
Microstomoeca roanoka
Salpingoeca helianthica
Mylnosiga fluctuans
Acanthoecida
Acanthoeca spectabilis
Helgoeca nana
Savillea parva
Stephanoeca diplocostata
animal ancestral sequence
choanoflagellate-animal ancestral sequence
choanoflagellate ancestral sequence

Figure 3-2 Alignment of Homer domains used for phylogeny construction. Included in the alignment were the EVH1 domain (indicated by black line), 5 amino acids following the EVH1 domain with particular importance for the separation of animals and choanoflagellates in the phylogeny, and conserved regions in the C-terminal stretch containing coiled coil domains. Non-conserved regions in between these conserved regions were excluded. Indicated on the left are the species encoding the sequences and their taxonomic affiliation. The grey bars indicate regions that were excluded for phylogeny reconstruction but included in ancestral protein reconstruction analysis. Black asterisks indicate regions in the Homer EVH1 domain that were found to be important for the binding of Shank PPXXF motif; blue asterisks indicate amino acids important for FPPPP binding by Mena/EVL EVH1 domains; that are also conserved in Homer EVH1 domains (Beneken et al. 2000). The three sequences at the bottom are the calculated ancestral sequences that were subsequently added to the alignment.

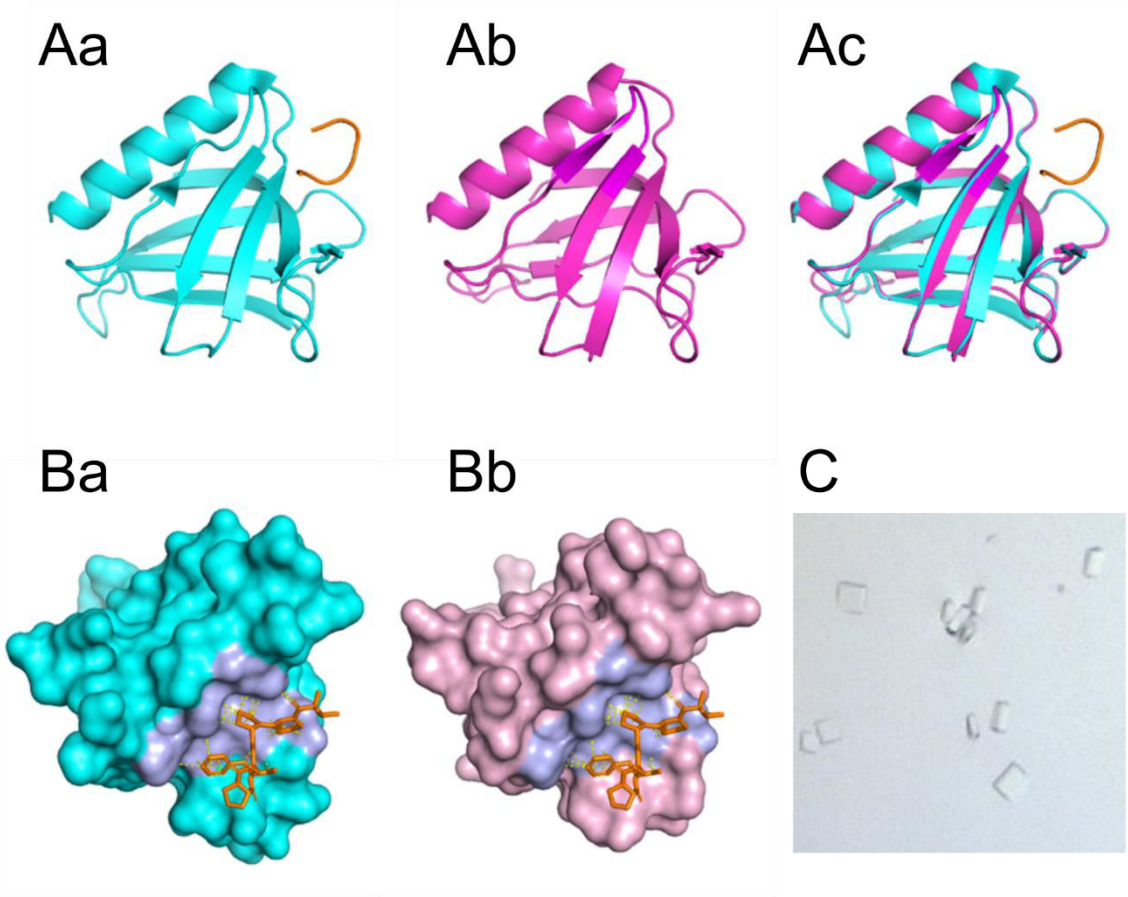


Figure 3-3 Structural similarity of rat and *Salpingoeca rosetta* Homer EVH1. A) Cartoon Homer EVH1 structures showing α-helices (shown as spirals) and β-sheets (shown as arrows). Aa) Rat Homer EVH1 structure (cyan) co-crystallised with TPPSPF ligand (orange) (Beneken et al. 2000; PDB ID: 1DDV; modified in pymol 2.3). Ab) Computational model of *S. rosetta* Homer EVH1 (Swiss Model alignment against rat crystal structure). Ac) Pymol alignment of the two structures in Aa and Ab. B. Surface Homer EVH1 structures showing the ligand amino acids as sticks (main and side chains, orange), conserved Homer EVH1 amino acids at binding site in purple, and sites of interaction in yellow. Ba) Rat Homer EVH1 structure of un-ligated protein (PDB: 1DDW) vs ligand (PDB: 1DDV) (modified in pymol 2.3 from Beneken et al. 2000). Bb) S. *rosetta* Homer EVH1 structure (same model as in Ab) vs ligand (PDB: 1DDV; modified from Beneken et al. 2000). C. Lysine methylated *S. rosetta* Homer EVH1 crystals that diffracted at 3.5-4 Å. Images of structures re-used (from PDB) with permission of Neuron (Elsevier)).

3.3.3 Homer and Shank binding seems to be conserved across choanoflagellates, animals and their common ancestor

In order to solve the question in which lineage the Homer-Shank binding evolved, we performed ancestral protein reconstruction for the Homer EVH1 domain. We constructed a phylogeny (Figure 3-4) based on the alignment of choanoflagellate and animal Homer sequences (Figure 3-2). Sequences calculated by ancestral reconstruction were added to the sequence alignment for comparison (Figure 3-2).

Binding of Homer EVH1 domains from different organisms and putative ancestors to Shank was tested via isothermal titration calorimetry. We tested the putative choanoflagellate ancestral Homer EVH1, the putative animal ancestral Homer EVH1, as well as Homer EVH1 from extant animal and choanoflagellate species. Chosen were Homer EVH1 domains from one bilaterian animal (*Rattus norvegicus*), one non-bilaterian animal (the sponge *Oscarella pearsei*), and one choanoflagellate species of each choanoflagellate family (*Salpingoeca rosetta* and *Stephanoeca diplocostata*). All Homer EVH1 domains were expressed in *E. coli*, purified and used for isothermal titration calorimetry measurements. All Homer EVH1 proteins were tested for binding to rat Shank1CΔPEST. Rat Homer EVH1 was used as a positive control. The technique chosen here could show that rat Homer and Shank bind with an affinity at a similar range to previous tests (here: Kd (dissociation constant) = 1.1 ± 0.2 µM (same value in two separate experiments; 95 % confidence interval) (Fig. 3.5A) vs Zeng et al., 2018: Kd = 1.35 µM). *Oscarella pearsei* Homer EVH1 binds to rat Shank with slightly lower affinity (Kd = 2.8 ± 1.0 µM / 5.0 ± 2.1 µM in two separate experiments) (Figure 3-5). Likewise, choanoflagellate Homer EVH1 domains bind to rat Shank (Figure 3-6) (*S. rosetta* Kd = 4.2 ± 1.7 µM / 4.1 ± 1.7 µM in two separate experiments; *S. diplocostata* Kd = 6.6 ± 3.0 µM / 3.9 ± 2.0 µM in two separate experiments). The prospective Homer EVH1 of the ancestor of animals as well as the ancestor of choanoflagellates could also bind rat Shank (Figure 3-7) (animal ancestor Kd = 2.2 ± 0.3 µM / 7.0 ± 1.9 µM in two separate experiments; choanoflagellate ancestor Kd = 3.7 ± 1.7 µM / 3.5 ± 1.0 µM in two separate experiments).

It is noticeable that all affinities lie approximately in the same range (with only the affinity of rat Homer to rat Shank being slightly higher), whereas there seem to be differences in the free enthalpies, which occur to be slightly higher in

animals and animal ancestor. This might be due to the fact that an animal Shank construct was used. Possibly, regions in the Shank protein apart from the PPxxF motif and surrounding amino acids are involved in Homer EVH1 binding. These regions could have evolved differently in the choanoflagellate lineage. However,affinities and free enthalpies have to be considered with caution, as the two replicates show some variation in the experiments and the stoichiometry, which was assumed to be n = 1 (for a 1:1 interaction), in many cases calculates to below 0.8. This could be due to inaccuracies in protein concentration measurements that would also impact Kd and ΔH values (Nanodrop measurements have an accuracy of ± 0.1 mg/mL). It is also possible that concentrations are correct, but that n values reflect the amount of active protein with competence for binding (which can be incongruent with the total amount of protein used to determine the concentration (Paketurytė et al., 2019). We furthermore regularly observed protein precipitation when Homer EVH1 domains were titrated to the rat Shank protein (but not when they were titrated to buffer), which likely caused further inaccuracies in Kd and ΔH values, as the precipitation itself can release energy (Grossoehme et al., 2010). Nevertheless, the data strongly suggest that the Shank binding capacity of Homer EVH1 is conserved among choanoflagellates and animals and therefore likely predated this division and the evolution of animals.

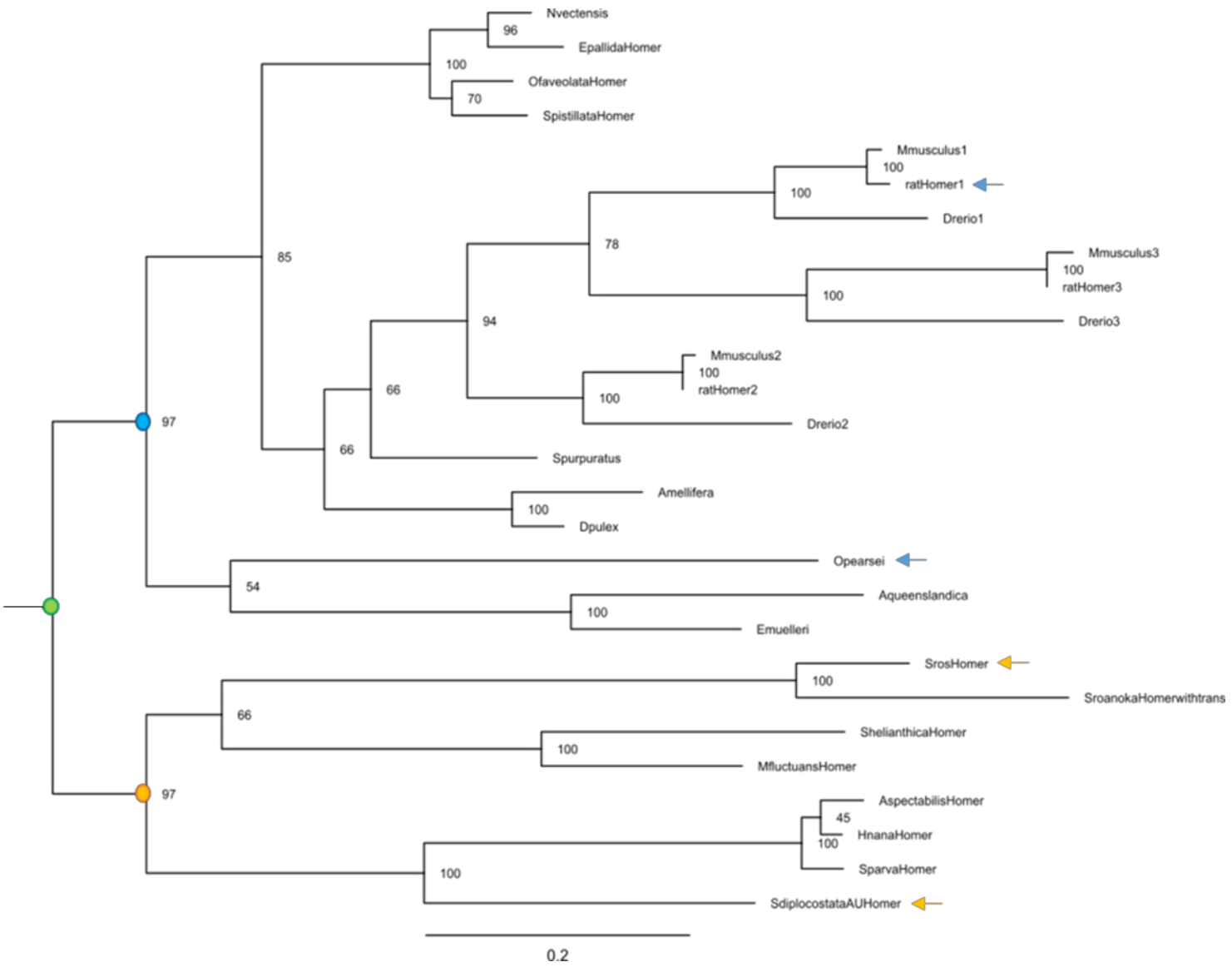

Nvectensis
96
EpallidaHomer
100
OfaveolataHomer
70
SpistillataHomer
Mmusculus1
100
ratHomer1
100
Drerio1
78
Mmusculus3
100
ratHomer3
100
Drerio3
85
94
Mmusculus2
100
ratHomer2
66
100
Drerio2
97
66
Spurpuratus
Amellifera
100
Dpulex
Opearsei
54
Aqueenslandica
100
Emuelleri
SrosHomer
100
SroanokaHomerwithtrans
66
ShelianthicaHomer
100
MfluctuansHomer
97
AspectabilisHomer
45
HnanaHomer
100
SparvaHomer
100
SdiplocostataAUHomer
0.2

Figure 3-4 Homer phylogeny. The phylogeny is based on the alignment in Figure 3-2, using the masked sequence alignment of the conserved Homer sequences of 27 species with exclusion of the regions marked with a grey bar (resulting in 188 amino acid characters used for phylogeny construction). Numbers on nodes give the bootstrap support (using 1000 bootstraps). Branch lengths are proportionate to the number of nucleotide substitutions as indicated by the scale bar. Blue circle: animal ancestor; yellow circle: choanoflagellates ancestor; green circle: choanoflagellate-animal ancestor; blue arrows: animal sequences used for ITC analysis; yellow arrows: choanoflagellates sequences used for ITC analysis.

In order to address the question if choanoflagellate Homer EVH1 can also bind choanoflagellate Shank, a Shank construct for the species *Stephanoeca diplocostata* was constructed. Because this construct was insoluble, we chose to tackle the question using peptides including the PPxxF motif of rat Shank, as well as peptides including either one or the other PPxxF motif of *S. rosetta* Shank Table 3-2). For each peptide 8 amino acids before and 8 amino acids after the PPxxF motif were included, as for the rat peptide some residues with importance for binding are in this region (Barzik et al., 2001). Rat Homer EVH1 can bind the rat Shank peptide FLFAEPLPPPLEFSNSFEKPE with an affinity comparable to that of its binding to the longer rat Shank construct (Figure 3-8A) (Kd = 4.0 ± 1.6 µM/ 2.9 ± 2.0 µM in two separate experiments, errors describe the 95 % confidence interval). Likewise, *S. rosetta* Homer EVH1 can bind the rat Shank peptide (Fig. 3.8B) (Kd = 1.9 ± 1.3 µM / 3.7 ± 2.7 µM in two separate experiments). *S. rosetta* Homer EVH1 can also bind to one of its own Shank peptides (peptide II: EASTSPFIPPPMFLADVQMTT) (Figure 3-9A) (Kd = 10.9 ± 16.8 µM / 7.6 ±14.8 µM in two separate experiments). Interestingly, the affinity of *S. rosetta* Homer EVH1 towards the PPxxF motif peptide of its own Shank is lower than that of rat Shank. In order to get a clear result, concentrations for both Homer EVH1 and the Shank peptide had to be strongly increased (600 µM Homer EVH1 vs 75 µM Shank peptide), which might be due to the low affinity and small peaks resulting from a weaker interaction. Even at these concentrations, the errors in Kd and ΔH values corresponding to the 95 % confidence interval are high. As in the measurements with the recombinant Shank protein, concentrations might have been slightly different, according to the Nanodrop precision error for the EVH1 domains and potential incongruity in the amount of lyophilised peptide used and the amount that actually dissolved in the buffer. *S. rosetta* Homer EVH1 does not seem to bind *S. rosetta* Shank peptide I (ADAAPVKAPPVYFARTRTSSV) under

the same conditions (Figure 3-9B; heat rate as well as enthalpy are comparable to that of the negative control (Figure 3-9C)). Moreover, rat Homer EVH1 presumably does not bind either of the *S. rosetta* Shank peptides (Figure 6-5).

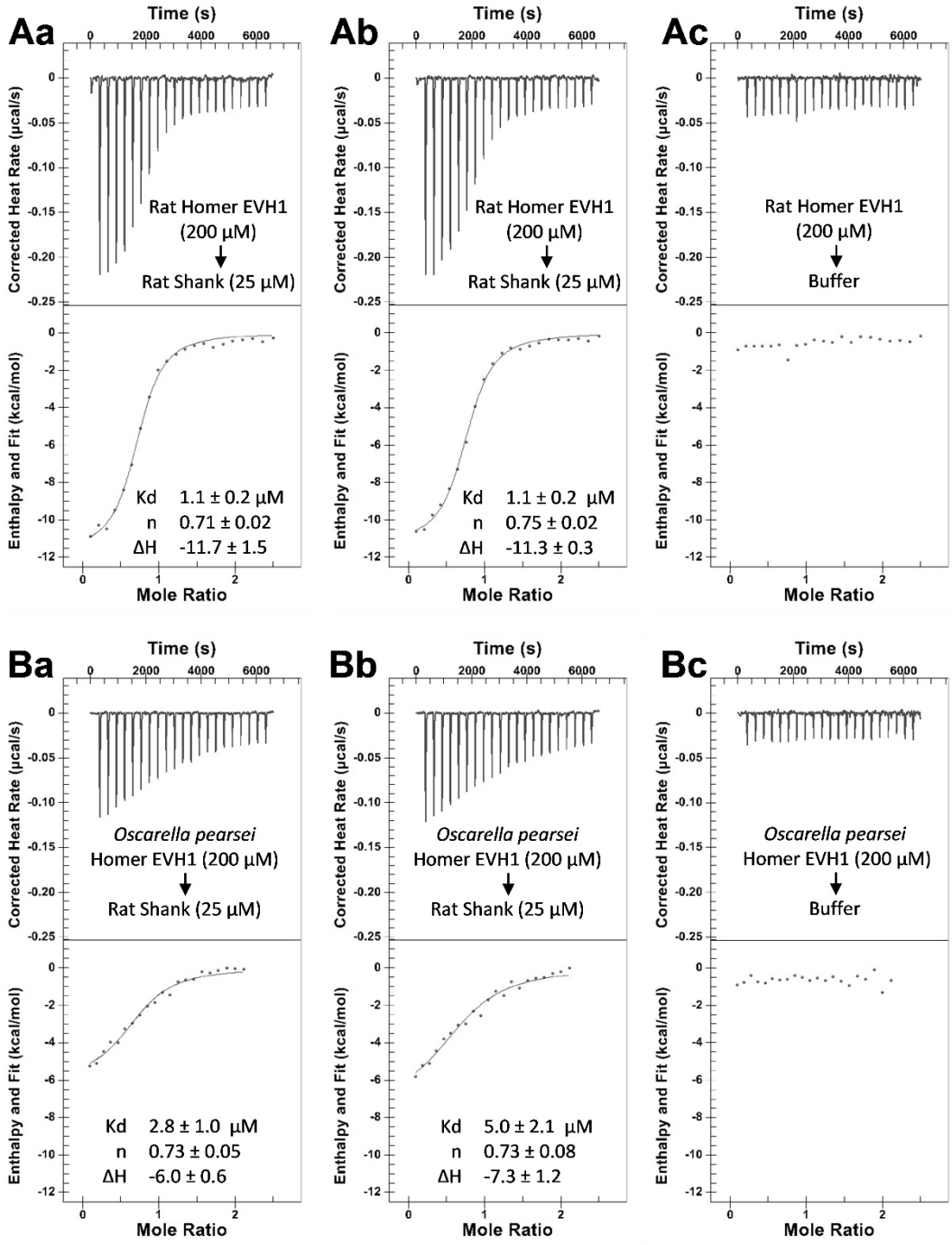


Figure 3-5: Binding kinetics of animal Homer EVH1 to Shank. All diagrams show calorimetric titrations; the upper panel of the diagram describes the corrected heat rate in µcal/s over the time course of the incremental titration. The lower panel describes the integrated areas normalised to the amount of Shank vs the molar ratio of Homer to Shank. The solid line represents the best fit to the data using a single site model. A) Binding of rat Homer EVH1 to rat Shank1ΔCPEST. Aa & Ab) 2 replicates injecting rat Homer EVH1 at a concentration of 200 µM into rat Shank1ΔCPEST at a concentration of 25 µM (22 injections, first injection 0.5 µl (not included in fit), 21 injections of 2.2 µl). Ac) Control injection of rat Homer EVH1 at concentration of 200 µM to experimental buffer (same injections as in Aa and Ab). B) Binding of *Oscarella pearsei* Homer EVH1 to rat Shank1ΔCPEST. Ba & Bb) 2 replicates injecting *O. pearsei* Homer EVH1 at a concentration of 200 µM into rat Shank1ΔCPEST at a concentration of 25 µM (22 injections, first injection 0.5 µl (not included in fit), 21 injections of 1.9 µl). Bc) Control injection of *O. pearsei* Homer EVH1 at concentration of 200 µM to experimental buffer (same injections as in Ba and Bb). Kd = equilibrium dissociation constant describing binding affinity; n = stoichiometry of binding; ΔH = free enthalpy difference in kcal/mol.

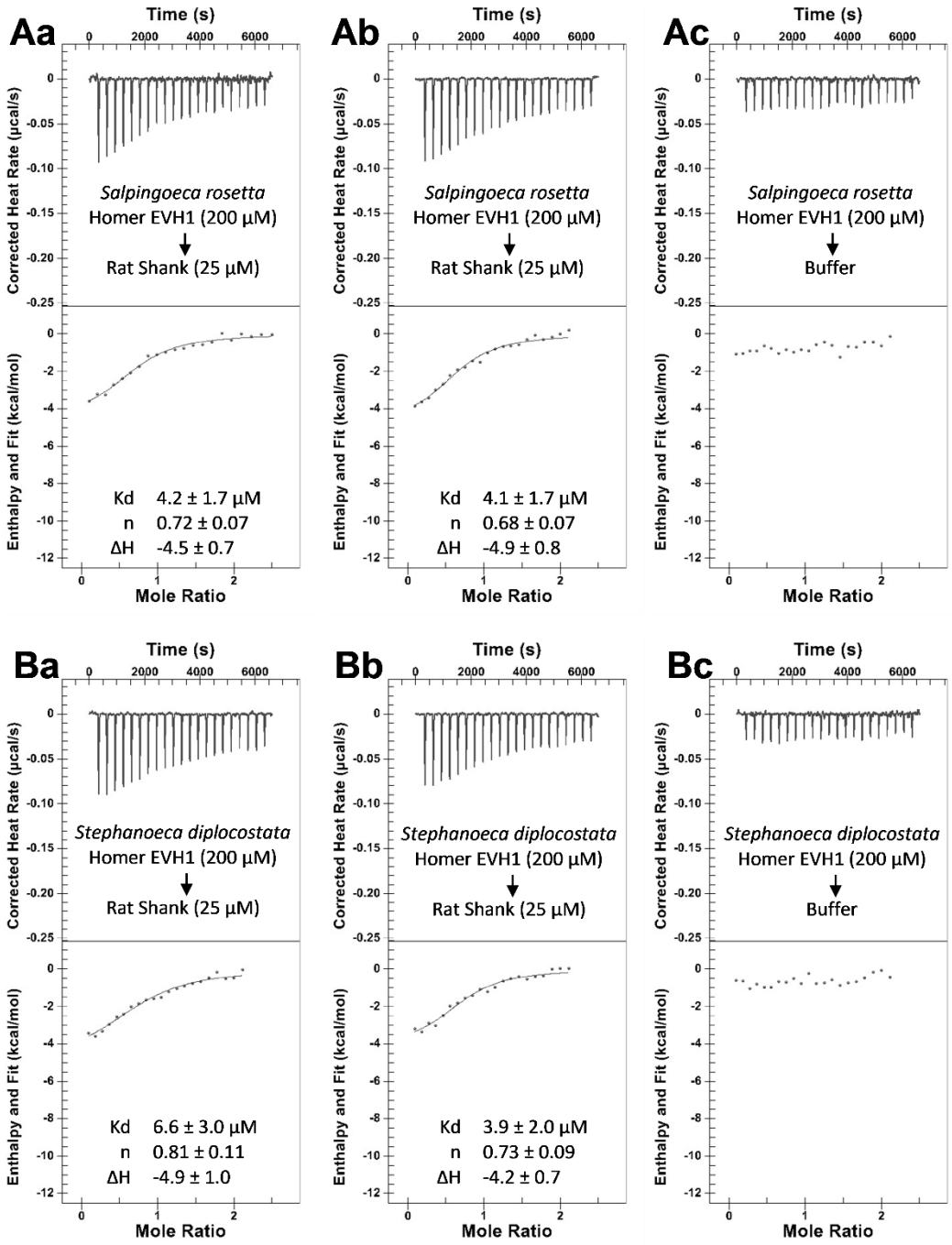


Figure 3-6: Binding kinetics of choanoflagellate Homer EVH1 to Shank. All diagrams show calorimetric titrations; the upper panel of the diagram describes the corrected heat rate in µcal/s over the time course of the incremental titration. The lower panel describes the integrated areas normalised to the amount of Shank vs the molar ratio of Homer to Shank. The solid line represents the best fit to the data using a single site model. A) Binding of *Salpingoeca rosetta* Homer EVH1 to rat Shank1ΔCPEST. Aa & Ab) 2 replicates injecting *S. rosetta* Homer EVH1 at a concentration of 200 µM into rat Shank1ΔCPEST at a concentration of 25 µM (22 injections, first injection 0.5 µl (not included in fit), 21 injections of 2.2 µl (Aa)/1.9 µl (Ab)). Ac) Control injection of *S. rosetta* Homer EVH1 at concentration of 200 µM to experimental buffer (same injections as in Ab). B) Binding of *Stephanoeca diplocostata* Homer EVH1 to rat Shank1ΔCPEST. Ba & Bb) 2 replicates injecting *S. diplocostata* Homer EVH1 at a concentration of 200 µM into rat Shank1ΔCPEST at a concentration of 25 µM (22 injections, first injection 0.5 µl (not included in fit), 21 injections of 1.9 µl). Bc) Control injection of *S. diplocostata* Homer EVH1 at concentration of 200 µM to experimental buffer (same injections as in Ba and Bb). Kd = equilibrium dissociation constant describing binding affinity; n = stoichiometry of binding; ΔH = free enthalpy difference in kcal/mol.

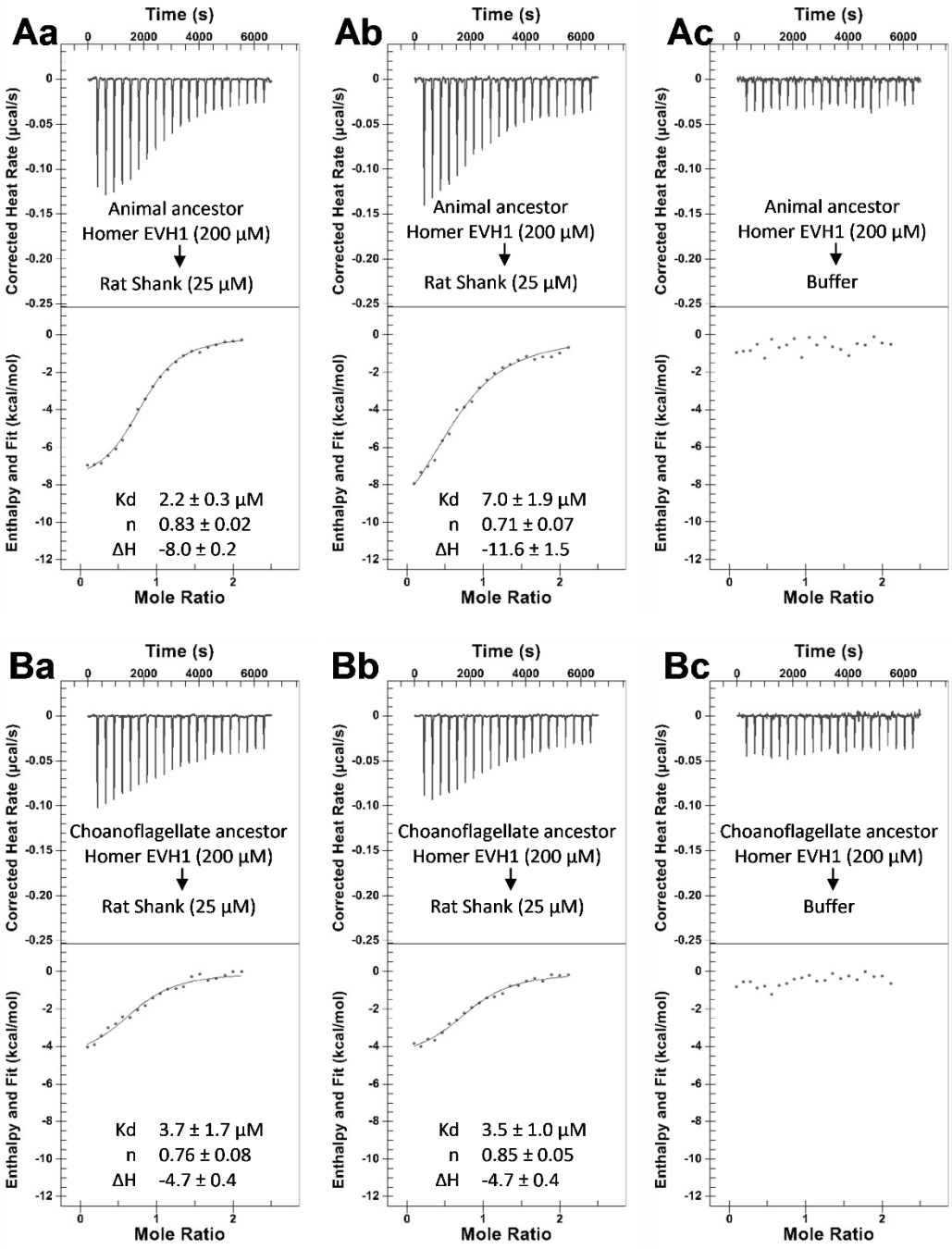


Figure 3-7: Binding kinetics of ancestrally reconstructed Homer EVH1 to Shank. All diagrams show calorimetric titrations; the upper panel of the diagram describes the corrected heat rate in µcal/s over the time course of the incremental titration. The lower panel describes the integrated areas normalised to the amount of Shank vs the molar ratio of Homer to Shank. The solid line represents the best fit to the data using a single site model. A) Binding of animal ancestral (anianc) Homer EVH1 to rat Shank1ΔCPEST. Aa & Ab) 2 replicates injecting anianc Homer EVH1 at a concentration of 200 µM into rat Shank1ΔCPEST at a concentration of 25 µM (22 injections, first injection 0.5 µl (not included in fit), 21 injections of 1.9 µl). Ac) Control injection of anianc Homer EVH1 at concentration of 200 µM to experimental buffer (same injections as in Aa and Ab). B) Binding of choanoflagellate ancestral (choanc) Homer EVH1 to rat Shank1ΔCPEST. Ba & Bb) 2 replicates injecting choanc Homer EVH1 at a concentration of 200 µM into rat Shank1ΔCPEST at a concentration of 25 µM (22 injections, first injection 0.5 µl (not included in fit), 21 injections of 1.9 µl). Bc) Control injection of choanc Homer EVH1 at concentration of 200 µM to experimental buffer (same injections as in Ba and Bb). Kd = equilibrium dissociation constant describing binding affinity; n = stoichiometry of binding; ΔH = free enthalpy difference in kcal/mol.

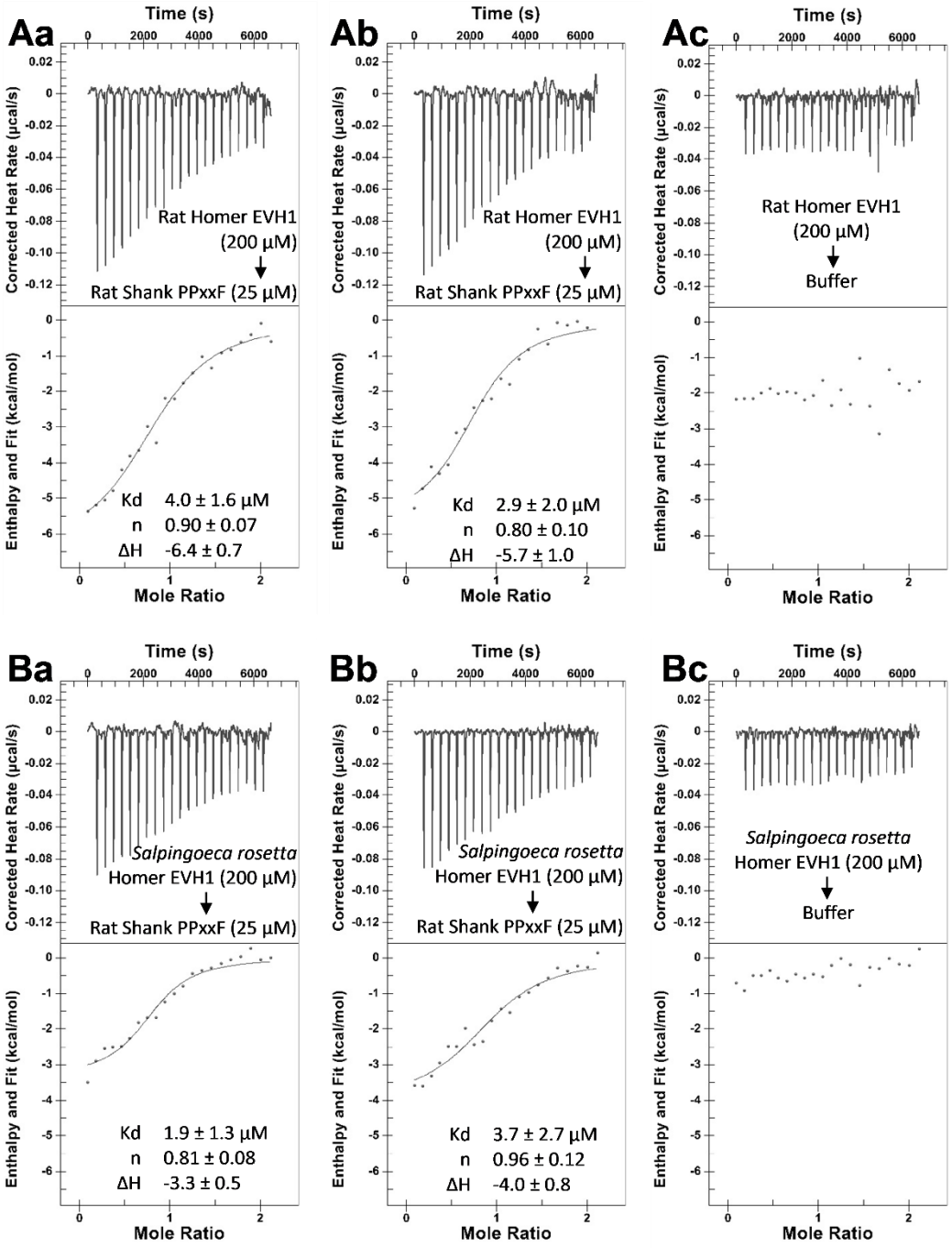


Figure 3-8: Both rat Homer EVH1 and *Salpingoeca rosetta* Homer EVH1 bind the rat Shank PPXXF motif. All diagrams show calorimetric titrations; the upper panel of the diagram describes the corrected heat rate in µcal/s over the time course of the incremental titration. The lower panel describes the integrated areas normalised to the amount of Shank vs the molar ratio of Homer to Shank. The solid line represents the best fit to the data using a single site model. A) Binding of rat Homer EVH1 to rat Shank PPXXF motif. Aa & Ab) 2 replicates injecting rat Homer EVH1 into a PPXXF motif containing rat Shank peptide. Ac) Control injection of rat Homer EVH1 into experimental buffer (same injections as in Aa and Ab). B) Binding of *S. rosetta* Homer EVH1 to rat Shank PPXXF motif. Ba & Bb) 2 replicates injecting *S. rosetta* Homer EVH1 into a PPXXF motif containing rat Shank peptide. Bc) Control injection of *S. rosetta* Homer EVH1 into experimental buffer (same injections as in Ba and Bb). All titrations: Homer EVH1 concentration 200 µM; peptide concentration 25 µM; 22 injections, first injection 0.5 µl (not included in fit), 21 injections of 1.9 µl. Kd = equilibrium dissociation constant describing binding affinity; n = stoichiometry of binding; ΔH = free enthalpy difference in kcal/mol.

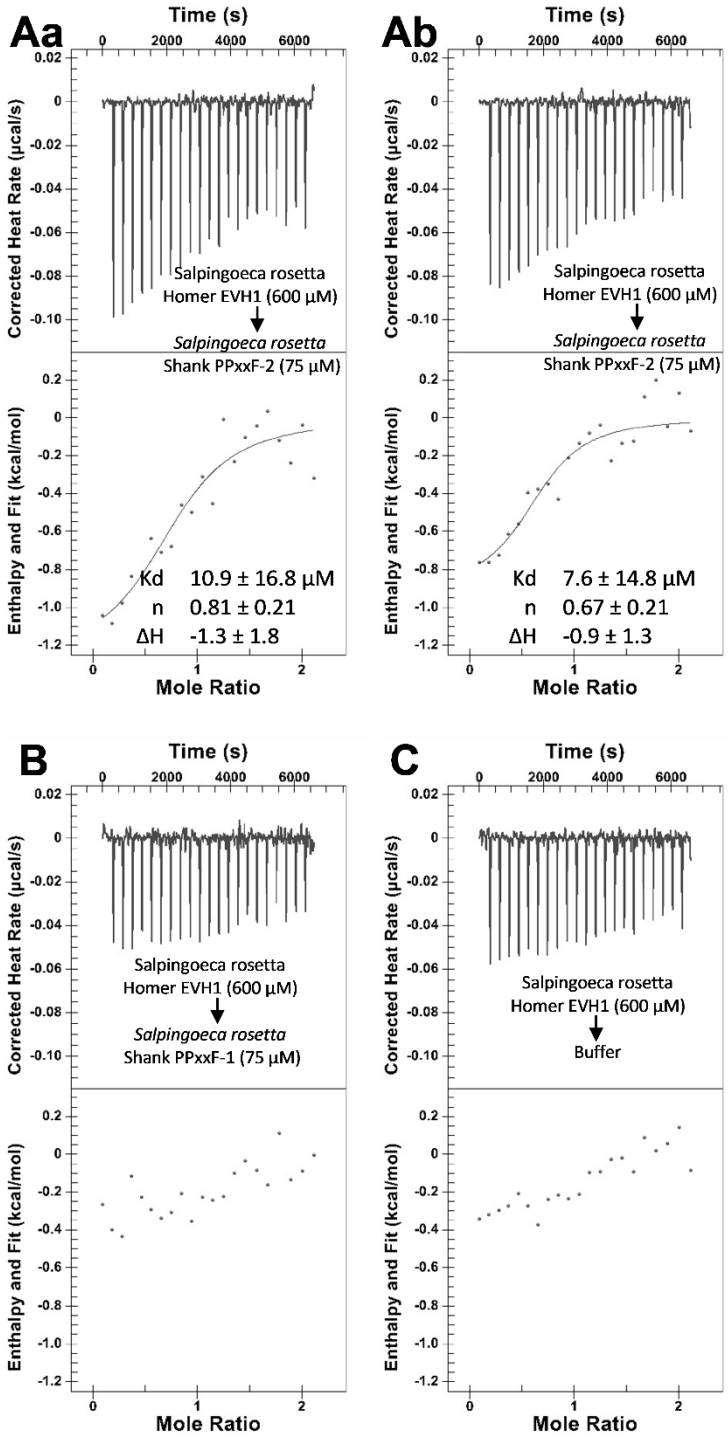


Figure 3-9: *Salpingoeca rosetta* Homer EVH1 binds to one of two PPXXF motifs of *S. rosetta* Shank. All diagrams show calorimetric titrations; the upper panel of the diagram describes the corrected heat rate in µcal/s over the time course of the incremental titration. The lower panel describes the integrated areas normalised to the amount of Shank vs the molar ratio of Homer to Shank. The solid line represents the best fit to the data using a single site model. A) *S. rosetta* Homer EVH1 binds to peptide II with PPXXF motif. Aa & Ab) 2 replicates injecting S. *rosetta* Homer EVH1 at a concentration of 600 µM into the PPXXF motif containing S. rosetta Shank peptide II at a concentration of 75 µM (22 injections, first injection 0.5 µl (not included in fit), 21 injections of 1.9 µl). B) *S. rosetta* Homer EVH1 does not bind to peptide I with PPXXF motif under the same conditions. C) Control injection of *S. rosetta* Homer EVH1 at a concentration of 600 µM into experimental buffer. The control and the injection into peptide I give similar peaks, therefore under current conditions *S. rosetta* Homer EVH1 binding to S. rosetta Shank peptide I cannot be shown. Kd = equilibrium dissociation constant; n = stoichiometry of; ΔH = free enthalpy difference in kcal/mol.

3.4 Discussion

Homer and Shank binding is crucial in vertebrate glutamatergic synapses for synaptic plasticity and the interconnection of glutamate receptor signalling and other signalling such as cytoskeleton remodelling and calcium signalling (Sala et al., 2001; Tu et al., 1998). It was previously shown that non-bilaterian animals as well as choanoflagellates have Homer and Shank homologs (Burkhardt et al., 2014; Sakarya et al., 2007), but it was never tested if these proteins can actually bind. This study reveals that Homer EVH1 domains from different animal species (*Rattus norvegicus* and *Oscarella pearsei*), and different choanoflagellates species (*Salpingoeca rosetta* and *Stephanoeca diplocostata*) can bind to a *Rattus norvegicus* Shank construct including a known Homer binding site with PPxxF motif. Additionally, we used ancestral protein reconstruction to investigate if this binding capacity is plausibly ancestral. Anderson et al. (2016) demonstrated how the combination of ancestral protein reconstruction and functional assays was sufficient to predict the evolutionary assembly of a protein complex through molecular exploitation, by recruiting proteins with an old function into a functional interaction with an evolutionarily younger molecule. Here, we used this method to show that the prospective animal and choanoflagellate ancestral Homer EVH1 domains could bind to rat Shank PPxxF. Furthermore, we could show that *S. rosetta* Homer EVH1 has the capacity to bind the rat Shank PPxxF motif as well as one of two PPxxF motifs in the *S. rosetta* Shank homolog. Rat Homer EVH1 on the other hand can bind to rat Shank PPxxF but does not seem to be capable of binding either of the *S. rosetta* Shank PPxxF motifs.

PPxxF motifs are very short motifs that only support EVH1 binding if the amino acids around the motif have certain properties. It was suggested that charged amino acids to both sides of the PPxxF motif could engage in further interactions with the EVH1 domain (Barzik et al., 2001). Even considering this suggestion, it is hard to predict which PPxxF motifs have binding potency and which will not bind EVH1 domains. Furthermore, different EVH1 domains of different species might support binding to different motifs, as there could be a co-evolution effect of the two proteins (de Juan et al., 2013). It can also not be excluded that other motifs could additionally allow binding. For the first time it could be shown that *S. rosetta* Homer and Shank have the capacity to bind *in vitro*. Moreover, this study provides new insights about the binding mode of *S.*

rosetta Homer and Shank and it extends the knowledge about the properties of rat PPxxF motifs that are required for Homer EVH1 binding. To date it is not known, if Homer-Shank binding is physiologically relevant in *S. rosetta*. Shank was not co-immunoprecipitated with Homer from *S. rosetta* colony lysate (Burkhardt et al., 2014). In this study we show that binding of the two proteins is possible but of low affinity. This could mean that binding is lost during cell lysis. It is also possible that the two proteins only bind under certain circumstances (or life history stages). One animal Homer isoform Homer1a (an isoform without coiled coil domain) is known to regulate the binding of other Homer isoforms to metabotropic glutamate receptors by competing for the binding site (Tu et al., 1998, 1999; Xiao et al., 1998). It is possible that something similar could be happening in choanoflagellates, potentially avoiding a Homer-Shank complex to be formed under given circumstances. Lastly, it is possible that although the two proteins can bind, this binding might be a remnant from a machinery used in the ancestor of choanoflagellates and animals that was subsequently lost from choanoflagellates.

The data presented in this chapter support the hypothesis that Homer and Shank were binding partners in the last common ancestor of choanoflagellates and animals. Future studies about the physiological relevance of Homer and Shank binding in choanoflagellates and non-bilaterian animals could be beneficial in order to learn more about the functional origin of this binding. This study is clearly showing that homologs to postsynaptic scaffolding proteins were not only present in the last common ancestor, but also were capable of binding in a manner that is relevant in synapse function. Moreover, the ancestral protein reconstruction demonstrates that the study of choanoflagellate proteins can reveal insights about the origin of synaptic protein complexes.

3.5 Outlook

We conducted this study in order to test the hypothesis that Homer and Shank binding preceded the animal-choanoflagellate split in evolution. The fact that recombinant Homer EVH1 domains from both selected animals and choanoflagellates have the capacity to bind Shank supports this hypothesis. Furthermore, Homer EVH1 domains of reconstructed ancestral sequences for the

animal and the choanoflagellate ancestor also bound rat Shank. For further confidence, we will conduct a control experiment. For this experiment, the robustness of the binding capacity of Homer EVH1 domains based on ancestral sequences will be tested. This will be done by exchanging all amino acids of the putative ancestral Homer EVH1 domain at sites with another plausible variant at a probability of more than 20 % with their variant and testing this Homer EVH1 domain for its binding to Shank. Eick et al. (2017) demonstrated that this increases the robustness of the results.

The reconstructed choanoflagellate-animal ancestral sequence was not tested, because the phylogeny used here was rooted on the animal/choanoflagellate division and did not include an outgroup. This phylogeny was chosen, because the branch of the outgroup was very long and changed the topology of the tree. However, without an outgroup reference the probability of every amino acid to be either as in the choanoflagellate ancestor or as in the animal ancestor is 50 %, resulting in slightly different choanoflagellate-animal ancestral sequences upon every new calculation. Nevertheless, the capacity of both choanoflagellate and animal ancestral Homer EVH1 gives a good indication that binding was possible in their common ancestor as well, especially, if plausible alternative ancestral Homer EVH1 domains as described above, retain their binding capacities.

As mentioned in section 3.3.3, variations in between replicates, n values below the expected stoichiometries and the observation of protein precipitation during isothermal titration calorimetry experiments lower the confidence in the calculated Kd and ΔH values. This could be addressed by complementing the results with another method that allows the determination of these values, such as fluorescent anisotropy. For this approach a peptide is labelled with a fluorescent probe (Jameson and Mocz, 2005; Lea and Simeonov, 2011). Fluorescent probes polarise light and this polarisation increases with increasing molecular size and therefore increases upon binding to a larger protein (Weigert, 1920; Jameson and Mocz, 2005; Lea and Simeonov, 2011). The method is very sensitive and allows the quantification of the complex in solution even in the presence of free ligand (Jameson and Mocz, 2005). Furthermore, future experiments should include at least three replicates in order to allow statistical evaluation of the variation between experiments.

Within the time frame of this thesis, it was not possible to acquire high quality *S. rosetta* Homer EVH1 crystals for structural reconstruction from X-ray diffraction data. However, we were able to produce crystals using lysine methylation. Further optimisation might lead to high quality crystals in the future, which would allow to make more accurate assumptions about the structure than with the computational model. Furthermore, it would enable us to aim at co-crystallisation with PPxxF ligand as performed for the rat Homer EVH1 domain (Beneken et al., 2000). One approach to improve crystallisation could include a more thorough determination of domain boundaries. For this study domain boundaries for the Homer EVH1 domain were determined using the SMART tool and information from a published crystal structure of the rat Homer EVH1 domain as well as from a multiple sequence alignment. In order to improve crystallisation, it can be helpful to reduce the domain to be crystallised to its core, removing all flexible linkers (Cohen et al., 1995). This can be achieved by limited proteolysis through proteases such as chymotrypsin or trypsin (Cohen et al., 1995; Dong et al., 2007; Wernimont and Edwards, 2009). The sequence of the core domain can subsequently be determined by mass spectrometry analysis (Cohen et al., 1995). Crystals can be produced by recombinantly expressing this core domain or by applying trace amounts of protease to the crystallisation (in situ proteolysis) (Dong et al., 2007; Wernimont and Edwards, 2009). Studies showed that this can rescue crystallisation in many proteins that did not crystallise or only produced low quality crystals without this limited proteolysis step. Considering computational methods in order to determine domain boundaries, Kirillova et al. (2009) suggested to use different methods in order to make useful predictions, as they found no single method to be reliable on its own. Computational methods to complement the SMART predictions include tools for secondary structure prediction (Busetta and Barrans, 1984) as well as tools, such as Domaination that are based on iterative PSI-BLAST searches that use multiple sequence alignments to predict domain boundaries (George and Heringa, 2002).

We further plan to transform *S. rosetta* with plasmids encoding fluorescently tagged Homer and Shank, respectively. This will be done after a protocol established by Booth et al. (2018), which allows to bring plasmids into choanoflagellates, where they are expressed without being inserted into the genome. Prior immunostainings of *S. rosetta* with antibodies against Homer and

Shank (polyclonal rabbit antibody against *S. rosetta* Homer coiled coil region (Burkhardt et al., 2014) with suitability for co-IP, Western blot and immunostaining and a peptide antibody against *S. rosetta* Shank with suitability for Western Blot and immunostaining but not for co-IP (antibodies produced by Pawel Burkhardt and Nicole King through Covance, Princeton, NJ, USA)) have shown that both proteins localise to the nucleus (Figure 3-10; Homer: Burkhardt et al., 2014 and Shank (Pawel Burkhardt, unpublished). Transformations will allow to confirm this localisation, and potentially identify further cellular localisations and investigate if Homer and Shank co-localise in living choanoflagellates.

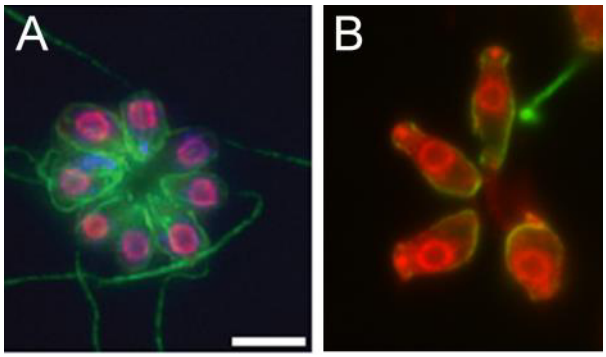


Figure 3-10: Homer and Shank immunostainings. A. Homer localisation to the nucleus of *S. rosetta* colonies (from Burkhardt et al. 2014, Molecular Biology and Evolution, Oxford University Press, Creative Commons license): Tubulin stain (green), DAPI stain (blue) overlapping with far red secondary antibody bound to Homer in the nucleus. B. Shank localisation to the nucleus of *S. rosetta*. Tubulin stain (green) and Shank stain (far red) (by Pawel Burkhardt, unpublished).

4 Third data chapter: A protein scaffold mediated by a PSD-95 homolog in choanoflagellates reveals insights into the origin of postsynaptic signalling machineries

Tarja T. Hoffmeyer[1,2,3], Mads Grønborg[4], Thomas S. Walter[5], A. Radu Aricescu[5,6], and Pawel Burkhardt[2,3]

[1]University of Exeter; [2]Marine Biological Association of the UK (Plymouth); [3]Sars International Centre of Marine Molecular Biology, University of Bergen (Norway); [4]Novo Nordisk Park, Måløv (Denmark) [5]Division of Structural Biology, University of Oxford (UK); [6]MRC Laboratory of Molecular Biology, Cambridge (UK).

Text: Tarja Hoffmeyer

Experimental work (lab and PC): Tarja Hoffmeyer

Antibody generation: Tarja Hoffmeyer and Pawel Burkhardt

Project idea: Pawel Burkhardt

X-ray Crystallography: Thomas Walter, Tarja Hoffmeyer, and Radu Aricescu

Mass spectrometry analysis: Mads Grønborg

4.1 Introduction

In chemical synapses, an electrical signal in a neuron is converted into a chemical signal via controlled neurotransmitter release from the presynapse. Binding of these neurotransmitters to receptors in the membrane of another neuron or effector cell (postsynapse) initiates signalling cascades that enable synaptic transmission as well as the modulation of synaptic strength (synaptic plasticity) (reviewed by Kennedy, 2016). Scaffolding proteins in postsynapses organise protein complexes involved in signalling machineries that enable these processes (Funke et al., 2005). PSD-95 is a scaffolding protein with a key regulatory function at glutamatergic synapses (Chen et al., 2011). This type of synapse is the main excitatory synapse in the vertebrate brain (Meldrum, 2000). Moreover, glutamatergic synapses seem to be more broadly distributed across the animal lineage, as there is evidence suggesting the presence of glutamatergic synapses in cnidarians and ctenophores (Burkhardt and Sprecher, 2017; Kass-Simon and Pierobon, 2007; Moroz et al., 2014). Understanding more about the origin of PSD-95 signalling complexes is therefore key to unravel the evolutionary history of synapses and neurons.

PSD-95 is a Dlg (*Drosophila* discs large) homolog (Cho et al., 1992). Invertebrate animals have one Dlg homolog, whereas vertebrates contain four Dlg duplicates (Garner and Kindler, 1996). These vertebrate paralogs are referred to as Dlg-1 (SAP97), Dlg-2 (Chapsyn-110/ PSD-93), Dlg-3 (SAP102) and Dlg-4 (SAP90/ PSD-95). Dlg homologs show conserved domain architecture with two PDZ domains and a C-terminal region made up of another PDZ domain and a super-module formed by an SH3 domain and a domain homologous to yeast guanylate kinase (GuK) without enzymatic activity (Figure 4-1) (Kistner et al., 1995; Garner and Kindler, 1996; McGee et al., 2001). The N-terminal region of Dlg proteins is most variable, and can contain a double cysteine palmitoylation site and/ or an L27 domain (Won et al., 2017; Schlüter et al., 2006). All Dlg domains are known as protein-protein interaction domains (McGee et al., 2001; Verpelli et al., 2012; Won et al., 2017). The C-terminal region is not restricted to Dlg proteins but is found in a whole family of proteins called membrane associated guanylate kinases (MAGUKs), important scaffolding proteins at the postsynapse and other cell-cell junctions (Anderson, 1996). Dlg proteins anchor complexes to receptors and other proteins in the membrane. The formation of these complexes is regulated bi-directionally. Dlg homologs recruit proteins to the membrane (e. g. Yamada et al., 2007), whereas alternative splicing and posttranslational modifications of Dlg homologs determine their recruiting into specific complexes (e. g. Topinka and Bredt, 1998).

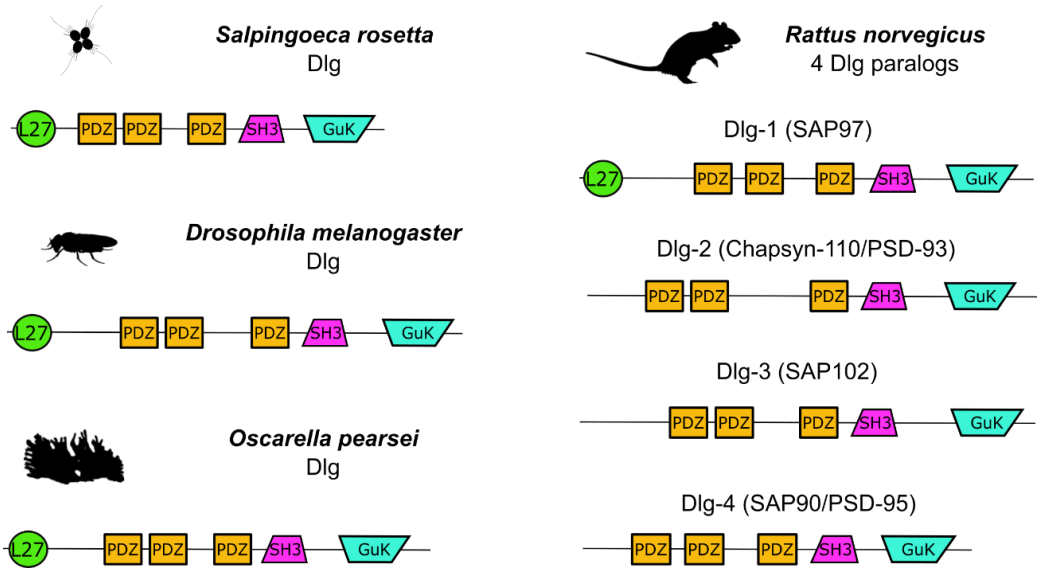


Figure 4-1: Domain architectures of Dlg homologs from different species. Reference sequences: *S. rosetta* Dlg: XP_004997111.1; *D. melanogaster* Dlg: NP_996406.1; *Oscarella pearsei* Dlg: m.306928 from OCAR_T-PEP_130911 (compagen.org); *Rattus norvegicus* Dlg-1: Q62696.1; Dlg-2: Q63622.1; Dlg-3: Q62936.1; Dlg-4: P31016.1. Animal phylum pictures from phylopic.org.

At vertebrate glutamatergic synapses, PSD-95 is crucial for the anchoring of AMPA and NMDA complexes (two ionotropic glutamate receptors (iGluRs)) (Chen et al., 2015). AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid) receptors mainly act as sodium and potassium channels, causing membrane depolarisation and signal transduction (Voglis and Tavernarakis, 2006). NMDA (N-methyl-D-aspartate) receptors (activated through membrane depolarisation via Mg2+ release (Mayer et al., 1984)) mainly act as calcium ion channels, thereby activating Ca2+/calmodulin-dependent protein kinase II (CAMKII). This enzyme increases the sensitivity of AMPA receptors and supports the exocytosis of more AMPA receptors from the endosome (increase of synaptic strength, a process important for learning and memory) (Voglis and Tavernarakis, 2006). PSD-95 binds NMDA receptors directly with its second PDZ domain, which interacts with the NR2 subunits' region with the conserved motif (E-S/T-D/E-V) (Niethammer et al., 1996). Bringing proteins in proximity to NMDARs is one important role of PSD-95 (Blackstone and Sheng, 2002). PSD-95 also clusters potassium channels and in addition, postsynaptic adhesion proteins, thus linking the postsynaptic density (PSD) and the presynaptic active zone (Kim et al., 1995; Irie et al., 1997; Kegel et al., 2013).

Animal Dlg homologs are involved in different complexes fulfilling diverse functions not only in neurons but in various animal cell types (Montgomery et al., 2004). Functional diversity is achieved through posttranslational modifications, alternative splicing, as well as the co-occurrence of different Dlg paralogs in vertebrates (Topinka and Bredt, 1998; Montgomery et al., 2004). At vertebrate glutamatergic synapses, these paralogs can compensate many of PSD-95 functions (Xu, 2011). Apart from its synaptic functions, Dlg-1/SAP97 has reported functions in epithelia for cell polarity establishment in mammals and birds, where it acts in protein complexes that are conserved in *Drosophila melanogaster* (KHC-73-Dlg-Pins/LGN and LGL-Scribble-Dlg complex) (Bilder, 2001; Siegrist and Doe, 2005; Bergstralh and St Johnston, 2012; Saadaoui et al., 2014). At *D. melanogaster* cellular junctions, Dlg recruits and anchors Fasciclin (an adhesion

protein) (Thomas et al., 1997; Szafranski and Goode, 2007; Harden et al., 2016). Loss of Dlg at *D. melanogaster* epithelia disrupts cellular polarity and correct distribution of septate junction proteins (Woods and Bryant, 1991; Franz and Riechmann, 2010). Dlg also plays a role at the *D. melanogaster* neuromuscular junction, where its loss leads to structural changes at synaptic type I boutons that entail the expansion of the presynaptic active zone and the reduction of postsynaptic complexity (Lahey et al., 1994).

The MAGUK family of scaffolding proteins evolved before the first animals. MAGUKs are encoded in choanoflagellate and filasterean genomes (de Mendoza et al., 2010). The genomes of other sampled protists, plants and fungi do not encode MAGUKs (te Velthuis et al., 2007). Dlg-like domain architecture is conserved in choanoflagellates and animals only (de Mendoza et al., 2010). Many homologs to postsynaptic proteins that were identified as interaction partners of PSD-95, are also encoded in choanoflagellate genomes of the species *Monosiga brevicollis* and *Salpingoeca rosetta* (Burkhardt et al., 2014), but it was never investigated if these proteins bind to the *S. rosetta* Dlg/PSD-95 homolog. The *S. rosetta* Dlg homolog contains an L27 domain and shares most similarity with the vertebrate Dlg-4 (PSD-95) paralog (35.23 % identity with human PSD-95 (Accession number: AAH40533.1) over 82 % of the sequence; BLASTp search against *Homo sapiens* sequences of the NCBI non-reduntant protein database, accessed: 19.02.2020). Investigating *S. rosetta* Dlg interaction partners could elucidate the origin of epithelial and synaptic protein complexes. *Salpingoeca rosetta* has neither epithelia nor synapses, however, the single-celled choanoflagellate can form colonies with cell-cell contact via incomplete cytokinesis (Dayel et al., 2011). Many receptors with important synaptic signalling functions (such as ionotropic and metabotropic glutamate receptors) are not encoded in the *S. rosetta* genome (Fairclough et al., 2013; Burkhardt et al., 2014). Synaptic scaffolding proteins on the other hand are conserved among animals and choanoflagellates (Burkhardt et al. 2014). Based on these prerequisites, we hypothesise:

The *S. rosetta* Dlg homolog has scaffolding function and forms a scaffold similar to the one found in the postsynapse. This scaffold is ancestral and was present in the last common ancestor of choanoflagellates and animals. The scaffold was expanded in the animal

lineage with the evolution of crucial interactions for individual signalling machineries that led to the emergence of neuronal cell types and synapses.

Finding an interaction of *S. rosetta* Dlg with other scaffolding proteins that are known PSD-95 interaction partners at postsynapses, would support our hypothesis. If, on the contrary, we do not find such interaction partners, this would argue for the emergence of the entire postsynaptic scaffold in the animal lineage or its loss in some or all choanoflagellates.

The following specific aims were designed in order to test the hypothesis:

1) Identify the interaction partners of *S. rosetta* Dlg.
2) Investigate structural and biochemical properties of *S. rosetta* Dlg with focus on multimerisation modules.
3) Elucidate the intracellular localisation of Dlg in *S. rosetta*.

4.2 Materials and methods

4.2.1 Choanoflagellate cultures

Salpingoeca rosetta was cultured at 18 °C in 20 % cereal grass medium in sterile artificial seawater (with a weekly 1:10 inoculation). PxI cultures (ATCC PRA-366) are rosette colonies enriched and solely cultured with the bacterium *Algoriphagus machipongonensis*. Col$^-$ cultures (ATCC 50818) are enriched for single-celled *S. rosetta* containing mainly attached cells and some fast swimmers. Col- contains a mixture of environmental bacteria. Cell lysate was produced by pelleting 3x 10 mL of *S. rosetta* culture, lysing the pellet with 600 µL lysis buffer (20 mM potassium phosphate buffer pH 7.4, 150 mM EDTA, 1 mM EGTA, 1 % Triton X-100) with 15 µL protease inhibitor cocktail P8465 (dissolved in 20 % DMSO per manufacturer instruction, Sigma Aldrich, Saint Louis, MO, USA) and immediate centrifugation for 10 min at 4 ˚C and 13,000 g. The lysate (supernatant) was kept at 4 ˚C and utilised immediately.

4.2.2 Proteins and antibodies

Expression constructs were cloned into pET28a plasmid vectors for poly-histidine (His)-tagged proteins and into pMAL-2c plasmid vectors for proteins bound to maltose binding protein (MBP). Sequences encoding *S. rosetta* Dlg full-length, Dlg SH3-HOOK-GuK, Dlg PDZ1-2, and Dlg L27 were codon optimised for *E. coli*, synthesised and cloned into pET28a vectors with NdeI and XhoI restriction enzymes (Novagen, Madison, WI, USA; cloning through Dr Pawel Burkhardt). Sequences encoding *S. rosetta* Dlg PDZ 1-3, Dlg peptides and MPP7 full-length were codon optimised for *E. coli*, synthesised and cloned into pET28a vectors with restriction enzymes NdeI and XhoI for Dlg PDZ1-3 and MPP7, and NdeI and SacI for Dlg peptides (Synbio Technologies, Monmouth Junction, NJ, USA, cloning through company). Sequences encoding Dlg peptides were additionally cloned into pMAL-2c vectors with restriction enzymes EcoRI and SalI for the Dlg peptides (Synbio Technologies, Monmouth Junction, NJ, USA, cloning through company). All plasmids were transformed into *E. coli* strain BL21(DE3) or BL21(DE3) pLysS (Novagen, Madison, WI, USA). Bacteria were grown in Terrific Broth (TB) medium (12 g tryptone, 24 g yeast extract, 4 mL glycerol per 900 mL MilliQ water (autoclaved), mixed with 100 mL of a sterile solution of 0.17 M KH2PO4 and 0.72 M K2HPO4) or Lysogeny Broth (LB) medium (10 g tryptone, 5 g yeast extract, 10 g NaCl and 1 mL 1M NaOH per litre MilliQ water (autoclaved)) with 30 µg/mL kanamycin for bacteria with pET28a plasmid and 100 µg/mL ampicillin plus 2g/L glucose for bacteria with pMAL plasmid. Bacteria in TB were grown to an optical density (OD) of 1.2-1.5; bacteria in LB were grown to an OD of 0.6-0.8 at 37 °C at 200 rpm. Expression was induced with 0.5 mM IPTG and was performed for 3 hours at 200 rpm at 25 or 37 °C (as indicated in Table 4-1). Bacteria were pelleted and dissolved in 50 mL Cobalt beads wash buffer (500 mM NaCl, 20 mM Tris pH 7.4) or in 50mL MBP column buffer (20 mM Tris, pH7.4; 200 mM NaCl, 1 mM EDTA).

Proteins were extracted from *E. coli* via 20 min incubation with lysozyme and 0.4 mM PMSF at room temperature, 3x 30 seconds sonication on ice with 30 % amplitude and 2-1 pulse, and 10 min incubation with DNase1, 2 % Triton X-100, 0.4 mM PMSF and 2 mM MgCl2. Bacterial debris was pelleted at 15,000 g at 4 ˚C for 40 minutes, leaving the protein of interest in the supernatant.

Table 4-1: Proteins that were expressed and purified.

S. rosetta proteins (NCBI No)	Expressed region	Expression conditions
Dlg/PSD-95 (PTSG_01141)	full-length[1] aa 1-839	3h in TB at 25 ˚C, 200 rpm
Dlg/PSD-95 (PTSG_01141)	SH3-HOOK-GuK[1] aa 509-839	3h in TB at 37 ˚C, 200 rpm
Dlg/PSD-95 (PTSG_01141)	PDZ1-2[1] aa 111-341	3h in TB at 37 ˚C, 200 rpm
Dlg/PSD-95 (PTSG_01141)	PDZ1-3[1] aa 111-508	3h in TB at 37 ˚C, 200 rpm
Dlg/PSD-95 (PTSG_01141)	L27[1] aa 1-116[3]	3h in LB at 25 ˚C, 200 rpm
Dlg/PSD-95 (PTSG_01141)	Peptide1[1] aa 578-628	3h in TB at 37 ˚C, 200rpm
Dlg/PSD-95 (PTSG_01141)	Peptide2[1] aa 329-377	3h in TB at 37 ˚C, 200 rpm
Dlg/PSD-95 (PTSG_01141)	Peptide1[2] aa 578-628	3h in LB +2g glucose/L at 37 ˚C, 200rpm
Dlg/PSD-95 (PTSG_01141)	Peptide2[2] aa 329-377	
MPP7 (PTSG_09863)	full-length[1] aa 1-547	3h in TB at 37 ˚C, 200 rpm

[1] with poly-histidine tag; [2] with maltose binding protein; [3] 1-79 after thrombin cleavage; internal cleavage site, but L27 domain itself is in the first 79 aa

Affinity chromatography for His-tagged proteins was performed with 0.5 mL HisPur™ Cobalt beads per 3L (Thermo Scientific, Rockford, IL, USA), rotating at 4 ˚C for 2 hours. Beads were washed with Cobalt beads wash buffer and the protein was eluted from cobalt beads with 400 mM imidazole in wash buffer. Affinity chromatography for MBP bound proteins was performed with amylose resin (BioNordika, Oslo, Norway). Per lysate of 6 L *E. coli* 2.5 mL of resin in 20 % ethanol were washed twice via 5 min centrifugation at 2000 g and replacement with 10 mL MBP column buffer. The lysate was incubated with the beads for 1 hour, rotating at 4 ˚C. Beads were washed with MBP column buffer and the protein was eluted with 10 mM maltose in MBP column buffer. MBP-bound Dlg peptide 2 was not further purified. All other proteins were prepared for ion exchange chromatography. MBP-bound Dlg peptide 1 in MBP column buffer was

diluted with an equal volume of 20 mM Tris pH 7.4, 1 mM EDTA buffer to bring it to a concentration of 100 mM NaCl. Full-length MPP7 was dialysed into 500 mM NaCl, 20 mM Tris pH 7.4 and 2 mM DTT and diluted with 20 mM Tris pH 7.4 and 2 mM DTT buffer to a salinity of 250 mM just before ion exchange. The other proteins were dialysed (buffers for each protein in Table 6-4) over-night at 4 ˚C while 50-100 units of thrombin (MP biomedicals, LLC, Irvine, CA, USA) were added to cleave off the His-tag. Ion exchange chromatography was performed on an Äkta prime plus (GE Healthcare Biosciences AB, Uppsala, Sweden) with HiTrap™ Q HP column or HiTrap™ SP HP column (as depicted in Table 6-4). To select only one mono- or oligomeric state of a protein before crystallisation, gel filtration was performed on a HighPrep™ 16/60 SephacrylTM S-100 HR column (GE Healthcare Biosciences AB, Uppsala, Sweden). Proteins were concentrated in Ultra-4 Centrifugal Devices (Ami-con, Kent, UK). Protein concentration was determined with a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Rockford, IL, USA).

12 antibodies against recombinant proteins/peptides of different *S. rosetta* Dlg regions were generated (Table 4-2). PDZ1-2 and L27 domain antibodies were generated by Pawel Burkhardt and Nicole King in 2014 (through Covance, Princeton, NJ, USA). Antibodies against regions outside of conserved domains were generated as part of this thesis (through Covalab, Villeurbanne, France). All antibodies were produced in duplicates in two rabbits, two guinea pigs and two rats against the PDZ1-2 region of *S. rosetta* Dlg and two rabbits each against the L27 domain and the two peptides. Blood serum containing polyclonal antibodies was frozen at -80 ˚C. Antibodies were tested on recombinant proteins as well as on *S. rosetta* cell lysate.

Table 4-2: Origin of the different *S. rosetta* Dlg/PSD-95 antibodies. Where stated the His-tag remained in the protein for injection.

S. rosetta Dlg/PSD-95 AB	Dlg/PSD-95 region the AB was raised against	Generated by
Rabbit 8857 L27 AB	L27 aa 1-116 (plus His-tag)	Pawel Burkhardt & Nicole King through Covance[1]
Rabbit 8858 L27 AB	L27 aa 1-116 (plus His-tag)	
Rabbit CA5681 PDZ1-2 AB	PDZ1-2 aa 111-341	
Rabbit CA5682 PDZ1-2 AB	PDZ1-2 aa 111-341	
Guinea Pig CA5683 PDZ1-2 AB	PDZ1-2 aa 111-341	
Guinea Pig CA5684 PDZ1-2 AB	PDZ1-2 aa 111-341	
Rat CA5685 PDZ1-2 AB	PDZ1-2 aa 111-341	
Rat CA5686 PDZ1-2 AB	PDZ1-2 aa 111-341	
Rabbit 1742004 peptide 1 AB	Pep 1 aa 578-628 (plus His-tag)	Tarja Hoffmeyer & Pawel Burkhardt through Covalab[2]
Rabbit 1742008 peptide 1 AB	Pep 1 aa 578-628 (plus His-tag)	
Rabbit 1742005 peptide 2 AB	Pep 2 aa 329-377 (plus His-tag)	
Rabbit 1742007 peptide 2 AB	Pep 2 aa 329-377 (plus His-tag)	

[1] Covance, Denver, PA, USA. [2] Covalab, Villeurbanne, France.

Antibody specificity was tested via Western blot analysis. Recombinant proteins and cell lysates were separated by sodium dodecyl sulphate-polyacrylamid gel electrophoresis (SDS-PAGE) and blotted onto a nitrocellulose membrane (Novex by life technologies, Carlsbad, CA, USA). The membrane was blocked in blocking buffer (1% (w/v) skim milk powder in 1x PBS (8g NaCl, 0.2g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4 in 1 L MilliQ water, pH 7.4)) for 2 hours and incubated with the respective antibody for 1 hour (1:1000 diluted in blocking buffer), washed 3x with 1x PBS, incubated with a secondary antibody coupled to horse radish peroxidase (Abcam, Cambridge, UK) for 1 hour (1:10,000 diluted in blocking buffer), washed 3x with 1x PBS and developed with Western Sure ECL Substrate (Premium for higher sensitivity when used on cell lysate or dilutions of recombinant protein) stable peroxide and luminol enhancer solution (LI-COR, Lincoln, NE, USA) for 5 minutes and then imaged on a LI-COR C-Digit Scanner. Antibodies with specificity for the protein of interest were further affinity purified on AminoLink Plus Immobilisation columns (Thermo Fisher Scientific, Rockford, IL, USA) coupled with the antigen at pH 10, following manufacturer instructions.

4.2.3 Identification of protein interaction partners

To identify *in vivo* protein interaction partners of *S. rosetta* Dlg, 50 µg of affinity purified rabbit antibodies from serum CA5681 were bound to 5 mg M-270 Epoxy resin magnetic dynabeads (Invitrogen, Thermo Fisher Scientific, Carlsbad, CA, USA) and incubated with 200 mL of Col- cell lysate (1 hour incubation of 1 mg antibody coated beads with 500 µL lysate under rotation at 4 ˚C). A negative control was treated in the same way, using a purified rabbit polyclonal control antibody (Genscript, Piscataway, NJ, USA). After 3 washes with lysis buffer, proteins were eluted with either 1x Laemmli buffer (Bio-rad, Hercules, CA, USA containing SDS and added β-Mercaptoethanol) for Western blot analysis or 8M guanidine hydrochloride (dissolved in molecular grade water) for mass spectrometry analysis. Mass spectrometry (LC-MS/MS) analysis performed by Mads Grønborg (Novo Nordisk Park, Måløv, Denmark). The raw data was searched against the NCBI *S. rosetta* FASTA-protein database (accessed 4th April 2017). Proteins unique to the Dlg antibody treatment were searched against the NCBI protein database and domain structure was analysed with SMART (smart.embl-heidelberg.de).

4.2.4 Protein binding assays

Two different methods to evaluate the binding capacity between *S. rosetta* Dlg and *S. rosetta* MPP7 were tested. For native polyacrylamide gel electrophoresis (Native PAGE), samples were prepared the night before. All samples (MPP7 in buffer, Dlg in buffer and different mixtures of MPP7 and Dlg) were incubated on ice over-night and for another 30 minutes at room temperature before addition of an equal volume of Native sample buffer (Bio-rad, Hercules, CA, USA). Native PAGE was run at 4 °C at 100 V in 4-20% Polyacrylamide gels in running buffer without SDS (3.4 g Tris base, 15.9 g glycine dissolved in 1 L MilliQ H_2O).

For isothermal titration calorimetry (ITC) proteins were dialysed twice in ITC buffer (20 mM sodium phosphate buffer pH 7.4; 250 mM NaCl, 1mM TCEP) at 4 ˚C. Proteins were concentrated to reach experimental concentrations in Ultra-4 Centrifugal devices (Ami-con, Kent, UK). Experimental ITC buffer was filtered through 0.2 µm pore filter and proteins were centrifuged for 10 min at 16,000 g,

before degassing all samples at 400 mm Hg vacuum at 25 ˚C for 15 minutes. A Nano ITC (TA Instruments, Newcastle, DE, USA) was used for measurements. The sample cell was filled with approximately 300 µL (results in 170 µL in the actual sample cell) and the injection syringe was filled with approximately 55 µL. Experiments were started with a buffer vs buffer run. Three different *S. rosetta* Dlg ligands were tested for binding with full-length *S. rosetta* MPP7. As a negative control the ligands were titrated into ITC buffer. For binding tests the ligands were titrated into the sample cell filled with *S. rosetta* MPP7 in ITC buffer. The stirring rate during ITC was set to 250. A first injection of 0.5 µL (real: 0.47 µL) was made after auto equilibration, to be excluded from the dataset later; followed by either 21 injections of 1.9 µL (real: 1.89 µL) or 12 injections of 3.5 µL (real: 3.49 µL) every 300 seconds. Incremental titration was chosen, and expected heats were set to medium. The resulting raw data graph was adjusted in TA instruments software, limiting integration regions to the main heat peak and manually adjusting the baseline. Data fitting was performed with the TA instruments software with the Independent model. A blank constant was subtracted, so that the curve approaches zero.

4.2.5 Structural characterisation of *S. rosetta* Dlg

S. rosetta Dlg SH3-HOOK-GuK region (aa 509-839) was recombinantly expressed in *E. coli*, purified and concentrated (as described above). In collaboration with Radu Aricescu and Tom Walter, proteins were crystallised at the Division of Structural Biology (University of Oxford, UK) and X-ray diffracted (Diamond Light Source, Didcot, UK). Favourable conditions were tested with the PCT Pre-Crystallisation Test kit (Hampton Research, Aliso Viejo, CA, USA). A protein concentration of 6.23 mg/mL was used. With the sitting drop method 960 different reservoir solutions were tested (the protein drop consisted of 50 % protein solution and 50 % reservoir solution). A Cartesian #2 sqII2673 robot (Digilab, Marlborough, MA, USA) was used to prepare the sitting drops. Crystals were grown at 20 ˚C in two conditions (20 % (w/v) PEG 3350 & 0.2 M KSCN; and 1.6 M (NH4)2SO4 & 0.1 M MES pH 6.0). These conditions were optimised to collect enough crystals for X-ray diffraction. Crystals were measured at the Diamond Light Source (Didcot, UK) and diffracted to 3.5 Ångstrom.

S. rosetta Dlg and MPP7 structures shown in this thesis were modelled computationally (https://swissmodel.expasy.org) via alignment to the rat Dlg SH3-HOOK-GuK structure (McGee et al., 2001; PDB ID: 1KJW), the structure of the human tripartite complex between Dlg L27, MPP7 L27 domain tandem and Mals3 L27 domains (Yang et al., 2010; PDB ID: 3LRA) and the structure of the *D. melanogaster* Dlg L27 homodimer structure (Ghosh et al., 2018, PDB ID: 4RP5). Structures were visualised and aligned with Pymol 2.3 (Schrödinger).

4.2.6 Determination of protein localisation

Subcellular localisation of *S. rosetta* Dlg was determined via immunostaining. Choanoflagellate cultures were concentrated via centrifugation at 500 g and transferred onto poly-L lysine coated glass bottom dishes, 15 minutes prior fixation. For fixation cells were incubated for 5 min with 6 % ice-cold acetone in seawater for microtubule stabilisation and 15 min incubation with 4 % paraformaldehyde in seawater. After 3 washes with 1x PBS, cells were blocked and permeabilised with 5% normal goat serum and 0.3 % Triton X-100 in sterile filtered PEM buffer (0.1 M PIPES pH 6.9, 1 mM MgCl2, 1 mM EGTA) for 30 min at room temperature. The antibody was applied 1:100 diluted in blocking buffer, together with mouse monoclonal antibody against β-tubulin (E7, 1:250; Developmental studies Hybridoma Bank, Iowa City, IO, USA) for 1 hour. After three washes with PEM, the secondary antibodies (Alexa fluor 647 goat anti rabbit IgG (H+L) and Alexa fluor 488 goat anti mouse IgG (H+L), both 1:200, Life technologies, Eugene, OR, USA) were incubated for 1 hour in the dark. After two washes with PEM and two washes with 1xPBS, the buffer was removed and cells were mounted with ProLong Gold Antifade Mountant with DAPI (Thermo Fisher Scientific, Rockford, IL, USA). A Leica DMi8 fluorescence microscope and a Leica SP5 confocal microscope (Leica, Wetzlar, Germany), respectively, were used to visualise stained choanoflagellates.

4.3 Results

4.3.1 Generation and validation of custom-made antibodies against *Salpingoeca rosetta* Dlg/PSD-95

In order to identify *Salpingoeca rosetta* Dlg interaction partners and reveal its subcellular localisation, we attempted finding a specific antibody against this protein suitable for the use in Western blot analysis, immunostaining and co-immunoprecipitation. Twelve custom-made antibodies were tested (Figure 4-2), four of which were generated within the scope of this study. Upon initial testing (Figure 4-2 and supplementary section 6.3.1), we chose a polyclonal rabbit antibody (CA5681) that bound *S. rosetta* Dlg in Western blots on *S. rosetta* cell lysate produced from cultures enriched in single attached cells. We chose these cultures according to differential expression data that reveal Dlg upregulation in single cells as opposed to colonies (Fairclough et al., 2013) (Figure 6-8). Antibody CA5681 specifically binds to a protein on Western blots, which according to its approximate size could correspond to *S. rosetta* Dlg (Figure 4-2 B). In a subsequent experiment (described in section 4.3.2), we could confirm that *S. rosetta* Dlg is bound by this antibody. Binding of the antibody to *S. rosetta* Dlg can be blocked by prior incubation with the antigen (Figure 4-2 D).

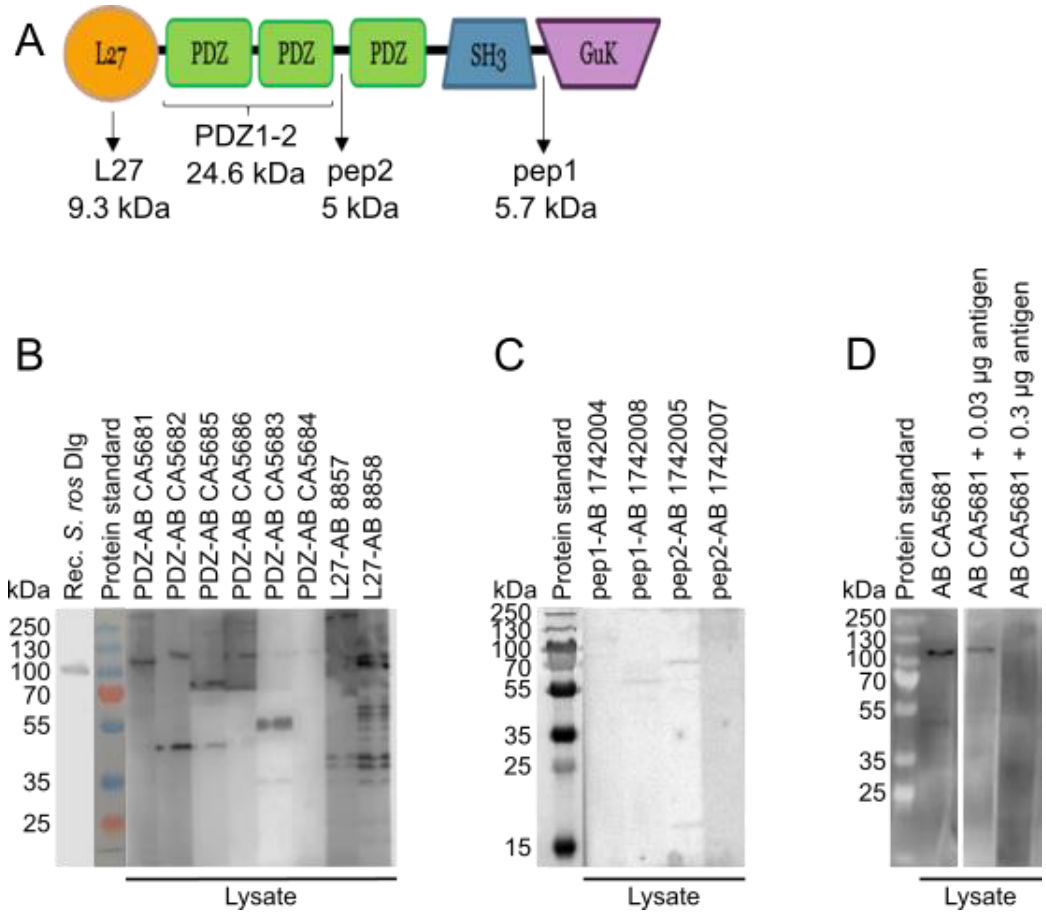


Figure 4-2: Validation of 12 custom made *Salpingoeca rosetta* Dlg/PSD-95 antibodies. A) *S. rosetta* Dlg with indication of locations and sizes of the different antigens. B) Six PDZ1-2 antibodies (PDZ-AB) and two L27 antibodies (L27-AB) were incubated on blot of separated *S. rosetta* single cell lysate proteins. First lane: recombinant full-length (FL) *S. rosetta* Dlg/PSD-95 incubated with guinea pig (Gpig) PDZ1-2 antibody, showing in which height recombinant full-length *S. rosetta* Dlg/PSD-95 runs on SDS PAGE. The native protein is expected to run slightly higher (between 100 and 130 kDa, as the strongest band given by Rabbit1 PDZ antibody). C) Four peptide antibodies (pep1 from two different rabbits and pep2 from two different rabbits) were incubated on blot of separated *S. rosetta* single cell lysate proteins. D) PDZ1-2 antibody from rabbit 1 (CA5681) can be blocked by prior incubation with the antigen (*S. rosetta* PDZ1-2). 0.3 µg PDZ1-2 are sufficient to block antibody binding to *S. rosetta* Dlg/PSD-95. Lysate was won from single *S. rosetta* cells in culture (col- culture). Western blots were cropped, and the order of lanes were changed in blot D. Original Western blots are shown in Figure 6-12.

4.3.2 Identification of *S. rosetta* Dlg interaction partners

Co-immunoprecipitation (co-IP) was used to identify *in vivo* binding partners of *S. rosetta* Dlg. Lysate was prepared from a culture enriched for *S. rosetta* single cells, because transcriptome data showed an upregulation of Dlg in single cells versus colonies (Figure 6-8 based on data from Fairclough et al., 2013). Western Blot analysis of co-IP trials revealed, that an elution with typical elution buffer (0.2 M glycine pH 2.8) or with 1x SDS sample buffer, incubated for

5 min at 50 ˚C was not sufficient to elute Dlg from the beads, but instead eluted one of the smaller proteins detected by the antibody (as shown in Figure 4-3 A for 1xSDS elution at 50 ˚C). Only strong denaturing conditions that removed the antibody from the beads eluted Dlg together with this antibody (Figure 4-3 B), suggesting a strong binding of antibody and Dlg. This worked for 1xSDS elution for 5 min at 100 ˚C and for 5 min 8 M guanidine hydrochloride at 95 ˚C. Differences in elution with 1xSDS at 50 versus 100 ˚C were captured by silver staining of a gel (Figure 4-3 C).

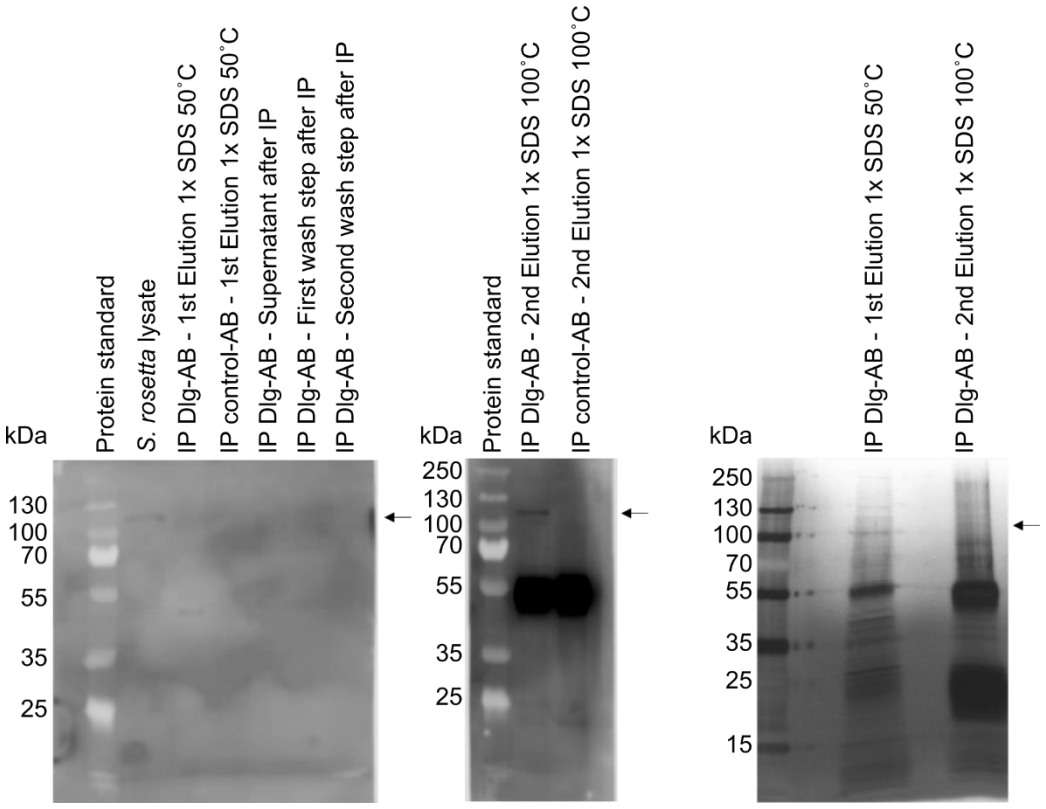


Figure 4-3: Co-immunoprecipitation conditions for Rabbit CA5681 antibody against *S. rosetta* PDZ1-2 domain of Dlg. A) Western blot showing that *S. rosetta* Dlg antibody is bound by the antibody (no presence in supernatant or washes), but is not eluted from the beads by elution with 1x SDS sample buffer with 5 minute incubation at 50 ˚C. B) Western blot showing that the antibody pulls down *S. rosetta* Dlg using an elution of 1x SDS sample buffer with an incubation of 5 minutes at 100 ˚C. The control antibody does not pull down Dlg under the same conditions. C) Silver stained SDS PAGE gel comparing the proteins released from the antibody under the tested elution conditions. The arrows show bands corresponding to Dlg. Additional bands could correspond to proteins non-specifically bound by the antibody or to degradation products of Dlg. Bands at approximately 25 and 55 kDa correspond to light and heavy chains of the antibody (respectively), eluted from the beads under the strong denaturing conditions required to elute Dlg.

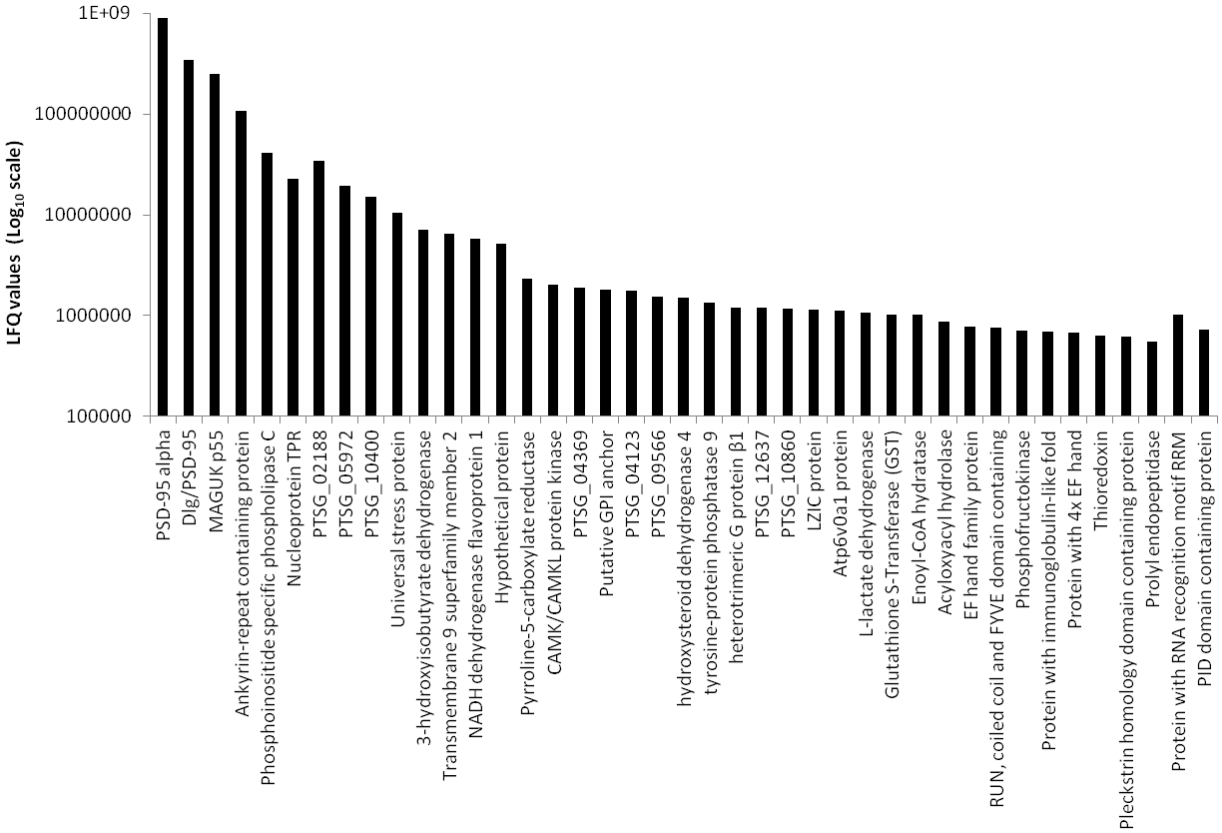


Figure 4-4: The *S. rosetta* Dlg antibody pulls down Dlg and potential interaction partners from *S. rosetta* single cell lysate. Proteins found in the Dlg co-IP that were unique to this co-IP and did not appear in the control co-IP and their label free quantification values visualised in $\log_{10}$ scale in a bar chart. Proteins with annotation were labelled accordingly; proteins without annotation were labelled with their NCBI number.

LC-MS/MS mass spectrometry data revealed that the antibody did detect *S. rosetta* Dlg at a high quantity (Figure 4-4). This protein was not found in the control, showing that the antibody targeted the correct protein. Another protein, annotated as *S. rosetta* PSD-95 alpha (a protein with two coiled coil domains and 5 PDZ domains), was detected in higher quantities than Dlg. An alignment of *S. rosetta* Dlg and PSD-95 alpha showed that the two PDZ domains of Dlg (that the antibody was raised against) had high similarity with PSD-95 alpha PDZ domains (Figure 4-5). Accordingly, there is a possibility that the antibody targeted both proteins and that all found interaction partners might correspond to either of these proteins. Western blot analysis does not show a band for the theoretical size of PSD-95 alpha according to genomic data.

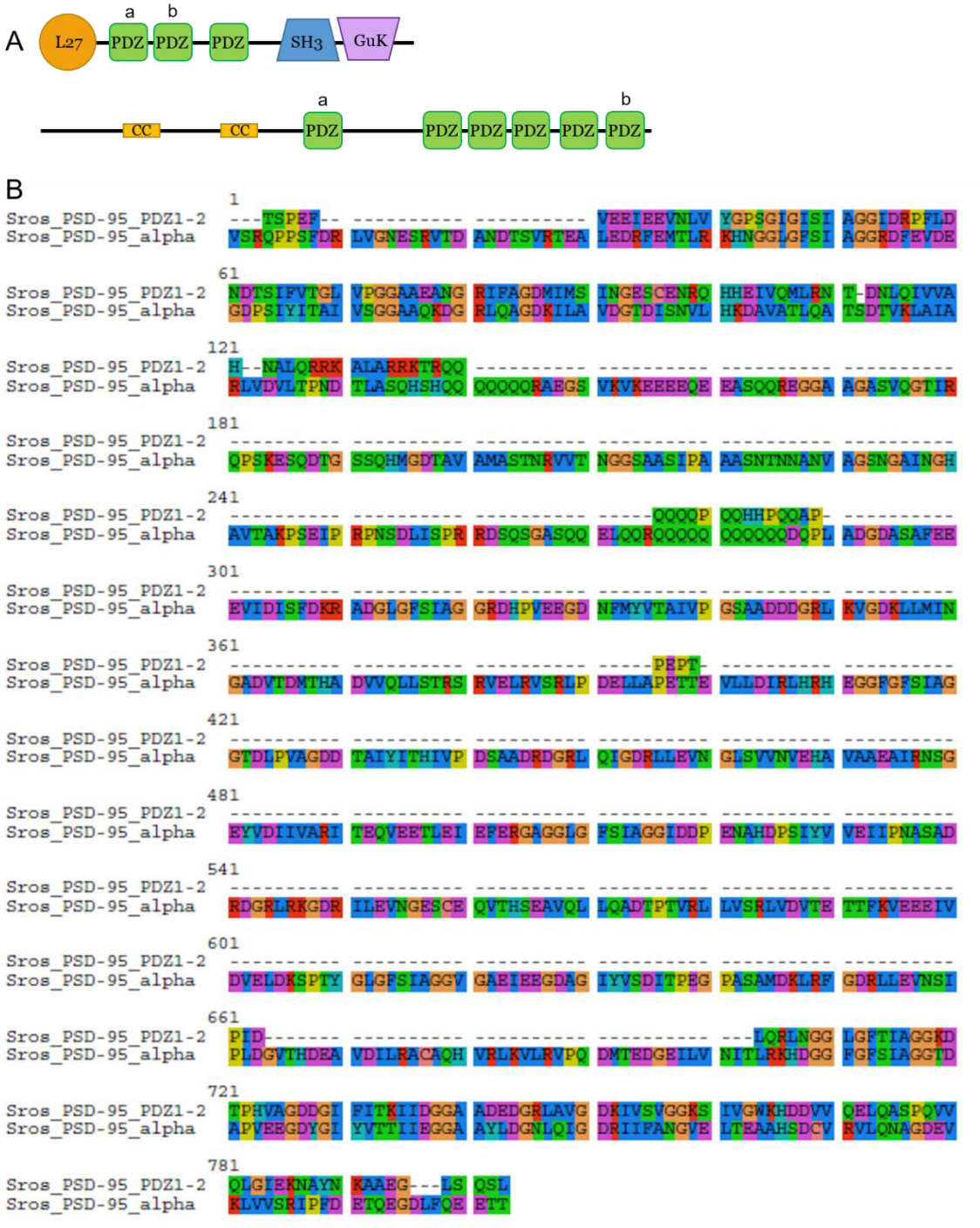


Figure 4-5: The multiple PDZ domain containing protein PSD-95 alpha – potential binding partner of *S. rosetta* Dlg or another antibody target. A) Domain architectures of *S. rosetta* Dlg (top) and *S. rosetta* PSD-95 alpha (below). PDZ domains that share the highest percent identities between the two proteins are labelled with a and b, respectively. B) Alignment of the PDZ1-2 region of *S. rosetta* Dlg that the antibody was raised against and the region of *S. rosetta* PSD-95 alpha including all six PDZ domains of the protein.

Proteins unique to the Dlg antibody co-IP (versus the control co-IP) were annotated and their quantities were compared (Figure 4-4). Few other proteins had some peptide hits in the control but did not show intensities after normalisation (label free quantification values) (a full list of these proteins is given in supplement Table 6-5). An interesting candidate in this group was a subunit of a voltage gated potassium channel.

4.3.3 *S. rosetta* Dlg might form a scaffold through homo-multimerisation and interaction with a membrane associated guanylate kinase of the p55 family

The *S. rosetta* protein that was pulled down with the *S. rosetta* Dlg antibody showed a conserved domain architecture and sequence similarity with MPP7, a membrane associated guanylate kinase of the family p55. Proteins of this family interact with Dlg family proteins in the postsynapse and at epithelial cell junctions in human cell lines (Stucke et al., 2007; Rademacher et al., 2016). This rendered MPP7 an interesting candidate for an *S. rosetta* Dlg interaction partner. We expressed and purified *S. rosetta* full-length MPP7 and Dlg and additionally three regions of *S. rosetta* Dlg (the L27 domain, aa 1-79; PDZ1-3, aa 111-508; SH3-HOOK-GuK, aa 509-839) to test for interaction.

First, we used native polyacrylamide gel electrophoresis (native PAGE) to test for aggregation between the two proteins (Figure 4-6). The principle of native PAGE is that protein complexes usually migrate differently than individual proteins, caused by a change of the overall charge of the proteins upon interaction. Adding full-length MPP7 does not result in a single band. It appears rather spread out within the lane (which is visible only at sufficiently high concentration) (Figure 4-6 A, lane 2; Figure 4-6 B, lane 1). Application of Dlg (in full length and the SH3-HOOK-GuK module) results in a single band. In lanes where full-length Dlg and MPP7 were added in combination (following incubation of the two proteins) bands appear at different positions in the gel, depending on the amount of full-length Dlg that was added to MPP7 (Figure 4-6 A, lanes 3-5). Adding varying amounts of full-length MPP7 to Dlg SH3-HOOK-GuK however, did not lead to changes in the position of the band (Figure 4-6 A, lanes 8-10).

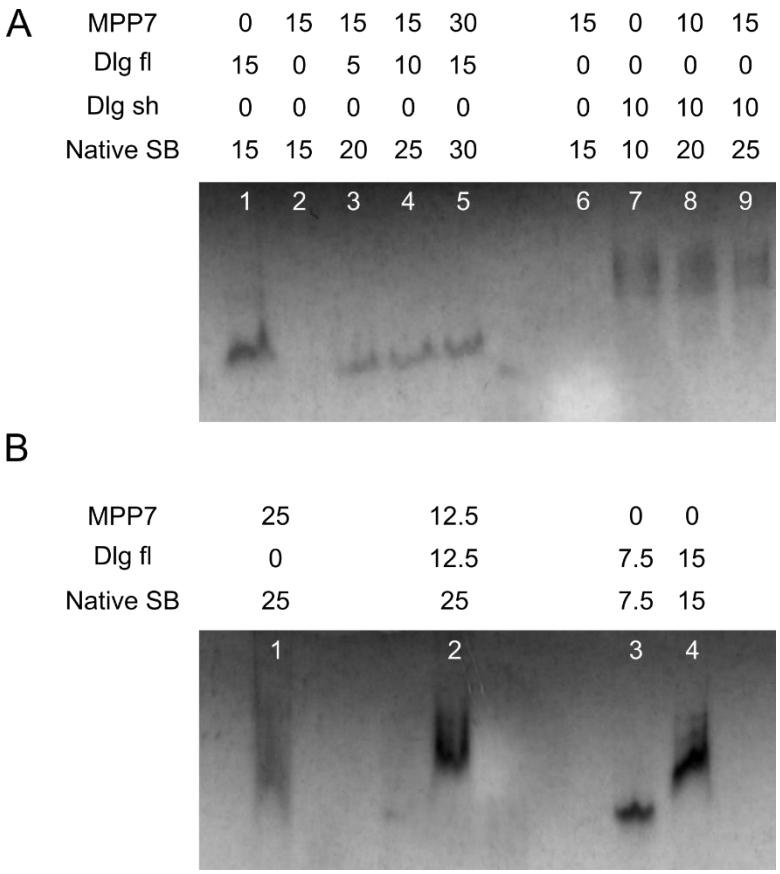


Figure 4-6: Native PAGE showing running behaviour of S. rosetta Dlg, MPP7 and the mixture of the two. Volumes in µL are indicated above the respective gel picture. Volumes in one experiment (A or B) correspond to equal concentrations, concentrations between A and B were different. Fl = full-length; Sh = short (SH3-HOOK-GuK), SB = sample buffer. After mixing the proteins and before adding the sample buffer the proteins were incubated together at room temperature for 30 minutes. Samples were loaded on 4-20 % polyacrylamide native gels and were conducted to gel electrophoresis at 4 °C.

Further native PAGE experiments were conducted in order to test if the running behaviour of Dlg changes due to concentration changes independent of MPP7 (Figure 4-6 B). In fact, we observed that the position of the band in lanes, where we added Dlg but not MPP7, changed depending on the amount of Dlg added (Figure 4-6 B, lanes 3-4). Therefore, our native PAGE experiments could not confirm if Dlg and MPP7 are indeed binding partners. However, the experiment suggested a concentration dependent oligomerisation of Dlg alone. Notably, as mentioned before, there is a chance that thrombin cleavage resulted in the removal of the L27 domain from the full-length Dlg protein. Our isothermal titration calorimetry data described below indicate that binding of L27 domains of Dlg and MPP7 might be responsible for binding of the two proteins. This interaction might have been excluded in our native PAGE experiments.

Next, we chose isothermal titration calorimetry (ITC) to investigate the Dlg-MPP7 interaction, as ITC can detect also low affinity protein-protein interactions. This method accurately measures thermodynamics and kinetics of protein-protein interactions. Titration of the Dlg L27 domain to full-length MPP7 results in slightly higher heat rates in comparison to control treatments, as well as negative enthalpies, indicating that the Dlg L27 domain potentially binds to full-length MPP7 in an exothermic reaction. Due to constraints in the expression of full-length MPP7, we worked with rather low protein concentrations (55 µM L27 versus 10-12 µM MPP7). This rendered correct fitting of the data difficult, leading to the calculation of different stoichiometries in different experiments, affecting the values for binding affinity and enthalpy (Figure 4-7). At a stoichiometry close to 1 (n = 0.84 ± 0.37 / 0.83 ± 0.08), indicating a 1:1 interaction, the K_d (describing the binding affinity) was 1.0 ± 12.8 and 0.4 ± 0.4 µM, with binding enthalpies of ΔH = -5.8 ± 11.8 and -4.7 ± 0.6, respectively (Fig. 2.7 C and D). At a stoichiometry close to 0.5 (n = 0.43 ± 0.20), indicating the interaction of the Dlg L27 domain with two regions of the full-length MPP7 molecule, the binding affinity was lower (higher K_d = 3.7 ± 2.9 µM), whereas the binding enthalpy was higher (ΔH = -25.0 ± 8.6).

The PDZ domain tandem (PDZ1-3) and the SH3-HOOK-GuK region of Dlg were tested for binding to MPP7 as well (Figure 4-8). At the concentrations used, these Dlg constructs did not show binding to MPP7, as apparent by the similarity between the heat rates and enthalpies between treatments and controls. However, at this stage it cannot be excluded that higher concentrations might show binding of MPP7 with these Dlg constructs.

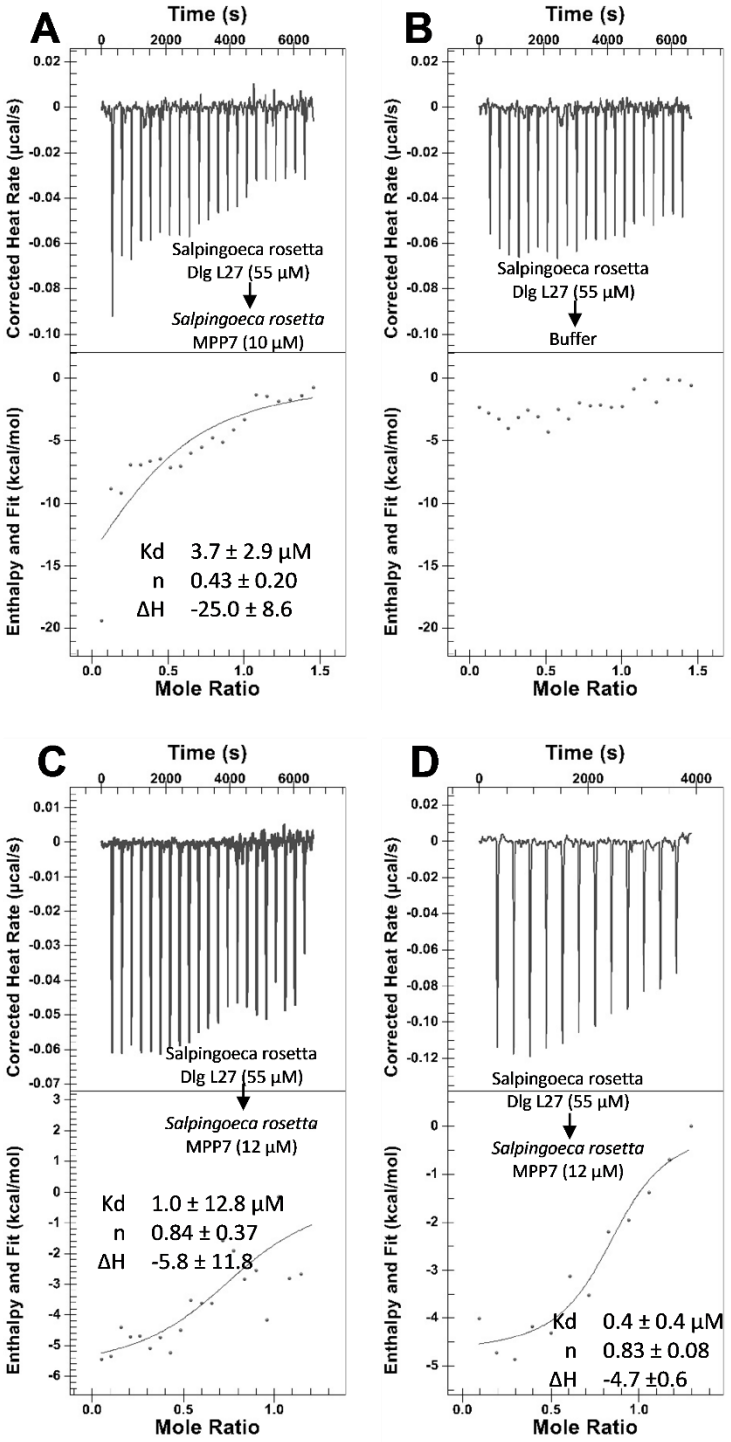


Figure 4-7: Potential binding of *Salpingoeca rosetta* MPP7 to Dlg via their L27 domains. Calorimetric titrations of Dlg L27 into full-length MPP7 (A, C and D) or in buffer (B). The upper panel of the diagram describes the corrected heat rate in µcal/s over the time course of the incremental titration. The lower panel of the diagram describes the integrated areas normalised to the amount of L27 domain (kcal/mol) versus the molar ratio of Dlg L27 domain to MPP7. The solid line represents the best fit to the data using a single binding site model. All isothermal calorimetric experiments were performed at 25 °C in 20 mM sodium phosphate buffer pH 7.4; 250 mM NaCl, 1 mM TCEP. A) Dlg L27 (conc. ~55 µM) titration into MPP7 (conc. 10 µM). B) Dlg L27 (conc. ~55 µM) titration into buffer. C) Dlg L27 (conc. ~55 µM) titration into MPP7 (conc. 12 µM). D) Dlg L27 (conc. ~55 µM) titration into MPP7 (conc. 12 µM). A-C titrations: 22 injections, first injection 0.5 µl (not included in fit), 21 injections of 1.9 µl; D titrations: 13 titrations: first injection 0.5 µl (not included in fit), 12 injections of 3.5 µl. Kd = equilibrium dissociation constant describing binding affinity (the smaller the value, the higher the affinity); n = stoichiometry of binding with n = 1 describing a 1:1 interaction; ΔH = free enthalpy difference in kcal/mol.

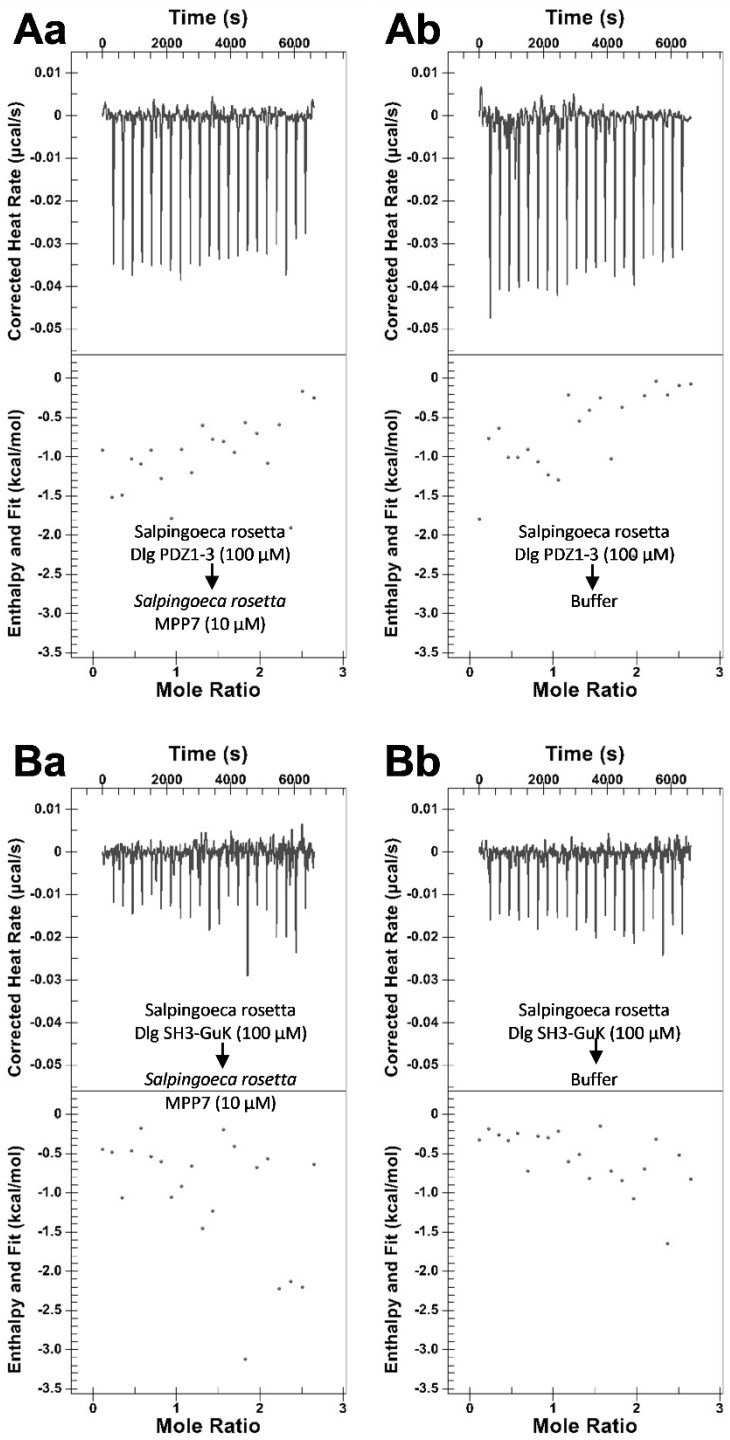


Figure 4-8: Salpingoeca rosetta MPP7 and Dlg do not seem to bind via PDZ domains or the SH3-GuK module. Calorimetric titrations of Dlg fractions into full-length MPP7 (Aa, Ba) or in buffer (Ab, Bb). The upper panel of the diagram describes the corrected heat rate in µcal/s over the time course of the incremental titration. The lower panel of the diagram describes the integrated areas normalised to the amount of Dlg fraction (kcal/mol) versus the molar ratio of Dlg fraction to MPP7. All isothermal calorimetric experiments were performed at 25 °C in 20 mM sodium phosphate buffer pH 7.4; 250 mM NaCl, 1 mM TCEP. A) Binding of Dlg PDZ1-3 to MPP7 cannot be demonstrated under these conditions. Aa) Titration of Dlg PDZ1-3 at a concentration of 100 µM into MPP7 at a concentration of 10 µM. Ab) Control titration of Dlg PDZ1-3 at a concentration of 100 µM into experimental buffer. B) Binding of Dlg SH3-GuK to MPP7 cannot be demonstrated under the used conditions. Ba) Titration of Dlg SH3-GuK at a concentration of 100 µM into MPP7 at a concentration of 10 µM. Bb) Control titration of Dlg SH3-GuK at a concentration of 100 µM into experimental buffer. Titrations: 22 injections, first injection 0.5 µl (not included in fit), 21 injections of 1.9 µl.

4.3.4 *S. rosetta* Dlg possesses multimerisation domains that could drive scaffold formation

The *S. rosetta* Dlg homolog shares a conserved domain architecture with bilaterian Dlg molecules (Figure 4-1). In order to learn, if some of these domains enable the protein to homo- or heteromultimerise with other membrane associated guanylate kinases, we performed homology reconstruction of the structures of S*. rosetta* Dlg L27 and SH3-HOOK-GuK (Figure 4-9 and Figure 4-10). SH3 and GuK domains of membrane associated guanylate kinases can interact intramolecularly (closed confirmation) or intermolecularly (open confirmation) (McGee et al., 2001). The switch between the different confirmations was suggested to be modulated by the binding of proteins to the flexible HOOK region (intramolecular binding of the N-terminal region in SAP97, binding of calmodulin) (Wu et al., 2000; Paarmann et al., 2002). The model of the *S. rosetta* SH3-HOOK-GuK structure, shows high similarity with the crystal structure of the same module from *Rattus norvegicus* PSD-95 (Figure 4-9 A-C). The backbone structure of the GuK domain appears to be nearly identical. There are some differences in flexible regions of the SH3 domain, but the overall structure appears highly similar. Major differences are visible in the HOOK region. The model suggests that this region is longer in *S. rosetta* Dlg than in rat PSD-95. Additionally, we used X-ray crystallography with the aim to solve the structure of *S. rosetta* SH3-HOOK-GuK. We purified the SH3-HOOK-GuK region of *S. rosetta* Dlg and obtained crystals in 1.6 M $(NH_4)_2SO_4$ and 0.1 M MES buffer at pH 6 (Figure 4-9 D). Crystals diffracted X-rays with a resolution of 3.5 Å only, and thus did not allow for a detailed investigation of the amino acid side chains. This resolution did not allow for detailed investigations of the side chain structures, but enabled to solve the backbone structure with molecular replacement (data not shown). Interestingly, we were not able to solve the SH3 domain and HOOK region of S. rosetta GuK, hinting towards possible degradation of the protein. Indeed, when crystals were collected and dissolved in Urea, we could observe degradation via SDS PAGE (Figure 4-9 E).

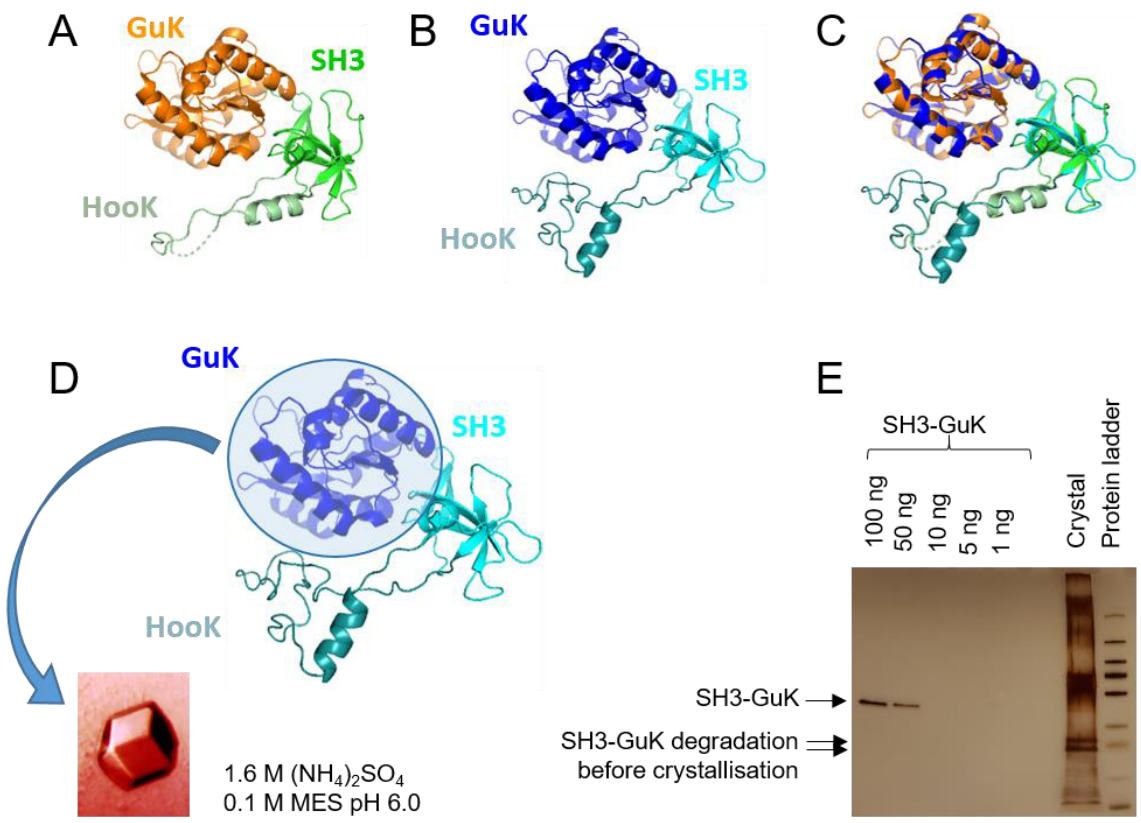


Figure 4-9: Exploring structural similarities and differences between *Salpingoeca rosetta* Dlg/PSD-95 and *Rattus norvegicus* PSD-95 SH3-HOOK-GuK region. A. Rat structure of SH3-HOOK-GuK module from McGee et al. 2001 (PDB ID: 1KJW, image of structure (from PDB) reused with permission of Molecular Cell (Elsevier)). B. Computational model of *S. rosetta* SH3-HOOK-GuK module. C. Alignment of rat vs *S. rosetta* SH3-HOOK-GuK structures. The corresponding sequence alignment is shown in Fig S2.5). D. Crystals of *S. rosetta* GuK domain. E) Silver staining gel showing degradation of the protein in the crystal.

Due to our results suggesting that in *S. rosetta* Dlg and MPP7 interact via their L27 domains, we also became interested in the structures of these domains. We computationally aligned the *S. rosetta* MPP7 L27 tandem domain and the *S. rosetta* Dlg L27 domain to the crystal structure of the human epithelial tripartite complex between hDlg (Dlg1 / SAP97), MPP7 and Mals3 (Lin7C) (Figure 4-10). In *S. rosetta* no Lin7 homolog (a small protein with one L27 domain and one PDZ domain) was identified and our mass spectrometry data set did not include another L27 domain containing protein that co-immunoprecipitated with the *S. rosetta* Dlg antibody. Interestingly, the structure alignment shows that the Dlg and MPP7 N-terminal L27 domain align well with the human structures of the heterodimer, whereas the C-terminal MPP7 L27 domain, which in human

interacts with the Mals3 L27 domain (Yang et al., 2010), has more structural differences in the *S. rosetta* protein.

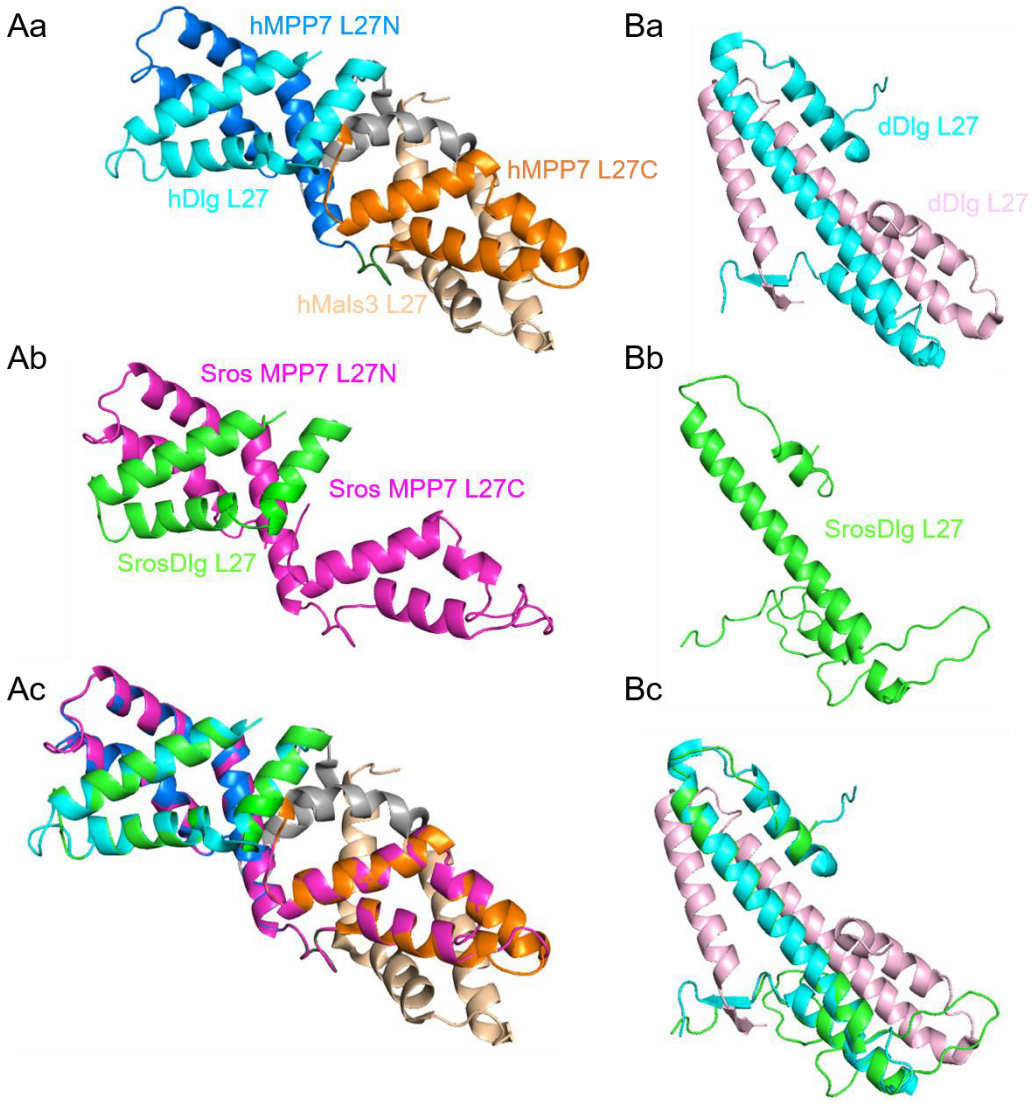


Figure 4-10: *Salpingoeca rosetta* structural requirements for L27 domain hetero- and homo-oligomerisation. A) Computational alignment of *S. rosetta* Dlg L27 and MPP7 L27-tandem structures to crystal structures of the human hDlg-hMPP7-hMals3 complex. Aa) Crystal structure of the human hDlg-hMPP7-hMals3 L27 tetramer complex by Yang et al. 2010 (PDB ID: 3LRA, image of structure from PDB re-used with permission from Federation of American Societies for Experimental Biology (John Wiley & sons)). Artificial linkers used in the study to connect all three polypeptides in a single protein are visualised in grey. The endogenous linker of MPP7 L27 domains is visualised in dark green. Ab) Theoretical structure of *S. rosetta* Dlg L27 and MPP7 L27 tandem resulting from the alignment to the human structures of hDlg L27 and hMPP7 L27-tandem from Aa. Ac) Alignment of the theoretical *S. rosetta* structure from Ab. to the human structure of the hDlg-hMPP7-hMals3 L27 tetramer from Aa. B) Computational alignment of *S. rosetta* Dlg L27 to *D. melanogaster* Dlg L27 structure in homo-dimer. Ba) *D. melanogaster* Dlg L27 crystal structure in homo-dimer by Ghosh et al. 2018 (PDB ID: 4rp5, structure illustration from PDB re-used with permission from American Chemical Society). Bb. Theoretical structure of *S. rosetta* Dlg L27 resulting from the alignment to the *D. melanogaster* structure of the Dlg L27 homodimer from Ba. Bc) Alignment of the theoretical *S. rosetta* Dlg L27 structure from Bb and one *D. melanogaster* Dlg L27 domain of the Dlg L27 homodimer from Ba. All structure cartoons were visualised with Pymol 2.3. Abbreviations: h = human; d = *D. melanogaster*; Sros = *S. rosetta*; L27N = L27 domain N-terminal in MPP7 L27 domain tandem; L27C = L27 domain C-terminal in MPP7 L27 domain tandem.

The *Drosophila melanogaster* Dlg protein has structural differences in the L27 domain. The domain includes more amino acids, and the second alpha helix is larger and covers the region of helix 2 and 3 of vertebrate L27 domains. There is a third alpha helix which extends into a region that in vertebrates does not belong to the L27 domain. When aligned to *D. melanogaster* Dlg L27, the second and third alpha helix of *S. rosetta* appear to be one long alpha helix, such as seen in *D. melanogaster*. This shows the ambiguity of computational structure alignments. A closer look into the alignment to the human protein reveals that *S. rosetta* does not have a kink between alpha helices 2 and 3 there, thus it is likely, that in *S. rosetta* both helices are fused. The first and third alpha helix of *S. rosetta* when aligned to *D. melanogaster* Dlg L27, seem to be distorted, suggesting that in *S. rosetta* the Dlg L27 domain is not extended. The Dlg L27 domain of *S. rosetta* seems to include features of both deuterostome and protostome domains.

4.3.5 *S. rosetta* Dlg seems to localise to the nucleus

Immunostaining with the affinity purified Rabbit antibody CA5681 against *S. rosetta* Dlg PDZ1-2 revealed staining of the nucleus, which was consistent in replicated experiments (Figure 4-11). The staining pattern was observed in single cells as well as in colonial cells (not shown). The staining seems to be concentrated along the nuclear membrane. It remains to be investigated whether the staining is on the inner side of the nuclear membrane, as shown for the protein Homer (Burkhardt et al., 2014), or on the outer side. In several experiments, we

also observed few cells with staining of the cell membrane at the basal pole of the cell. Some other staining patterns (that were observed in single experiments but were not observed in replicated experimens) as well as additional immunnostainings made with the other *S. rosetta* Dlg antibodies are shown in the appendix (supplementary Figure 6-9, Figure 6-10, and Figure 6-11).

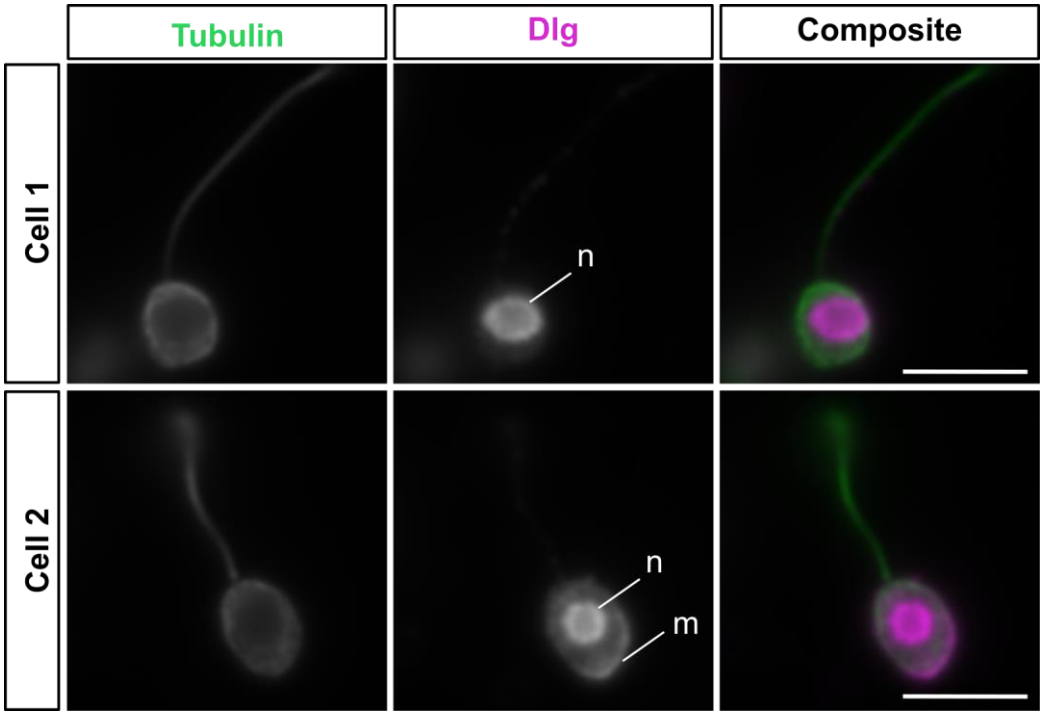


Figure 4-11: *Salpingoeca rosetta* Dlg/PSD-95 immunostaining shows staining of the nucleus (and additional membrane staining in some but not all cells). Green: Tubulin-staining highlightes microtubules in the flagellum and the cell body. Magenta: Dlg/PSD-95 staining. n = nucleus; m = membrane. Scalebar: 5 µm.

4.4 Discussion

Postsynaptic scaffolding proteins provide a platform to spatially organise key proteins of the postsynaptic density (PSD) (protein complexes adjacent to the postsynaptic membrane). These proteins originated before the evolution of the first synapse and some even before the origin of the first animals (Alié and Manuel, 2010). The key organiser at the vertebrate glutamatergic synapse is the scaffolding protein PSD-95. PSD-95 is one of four vertebrate Dlg paralogs that are homologous to *Drosophila melanogaster* Dlg. The genome of the choanoflagellate species *Salpingoeca rosetta* also encodes a Dlg homolog. We aimed to identify *S. rosetta* Dlg interaction partners, binding capacities, and

intracellular localisation in order to compare the characteristics of the *S. rosetta* Dlg complex to the synaptic PSD-95 complex at vertebrate glutamatergic synapses.

4.4.1 Indications for the presence of a Dlg protein scaffold in *S. rosetta*

We generated twelve antibodies against the *Salpingoeca rosetta* Dlg/PSD-95 homolog. Most of these antibodies bound to additional proteins in choanoflagellates and / or bacteria, but we were able to purify one antibody with high specificity for its antigen (the first two PDZ domains of *S. rosetta* Dlg). Mass spectrometry analysis of proteins co-immunoprecipitated with this affinity purified antibody revealed the presence of Dlg, but additionally also PSD-95 alpha, a protein, containing a sequence with high similarity to the targeted sequence. It is possible that the antibody targeted both proteins (implying that the identified interaction partners might bind to either of both proteins). Nevertheless, PSD-95 alpha could still occur in a complex with Dlg in *S. rosetta*. In the PSD of vertebrate glutamatergic synapses, PSD-95 is bound to Shaker type potassium channels, which are also bound by multiple PDZ domain containing proteins (Kim et al., 1995; Cohen and Brenman, 1996; Poliak et al., 2003; Leonoudakis et al., 2004). A direct interaction between Dlg and PSD-95 alpha in *S. rosetta* via PDZ domains is possible, for example through a PDZ β-finger interaction as described for the nNOS-PSD-95 interaction (Hillier et al., 1999; Tochio et al., 2000), and this will be tested in the future. However, they could also indirectly interact via another protein, which makes it more challenging to test if Dlg and PSD-95 alpha occur in the same complex.

The protein with the third highest quantities in the data set is a scaffolding protein of the membrane associated guanylate kinase (MAGUK) p55 family, most similar to the homolog MPP7. MAGUKs of this family have been shown to interact with Dlg/PSD-95 homologs in animal tissues (Figure 4-12) (Stucke et al., 2007; Rademacher et al., 2016). Rademacher et al. (2016) performed co-IP experiments from rat brain lysate and established that PSD-95 and MAGUK p55 homolog MPP2 both occur in complexes at AMPA receptors and interact via the same module used for PSD-95 multimerisation, the SH3-HOOK-GuK module (McGee et al., 2001; Rademacher et al., 2016). MPP2 also associates with the

PSD-95 binding partner GKAP and with SynCAM, a synaptic adhesion molecule interconnecting into the presynaptic active zone (Rademacher et al., 2016). Stucke et al. (2007) performed experiments on the human Caco2 cell line (human epithelial colorectal adenocarcinoma cells), finding Dlg-1/SAP97 to interact with the MAGUK p55 homolog MPP7. The two proteins localised to the lateral surface of the cells and knock down of either of the proteins slowed down tight junction formation (Stucke et al., 2007). In contrast to the interaction of PSD-95 and MPP2 via the SH3-GuK module, Dlg-1 and MPP7 were shown to interact via N-terminal L27 domains (Stucke et al., 2007). MPP7 was also shown to be recruited to the membrane by Crumbs, potentially via its interaction with PALS1 (MPP5) (Stucke et al., 2007).

Figure 4-12: MAGUK p55 interaction with Dlg found in *Salpingoeca rosetta* is conserved in epithelia and glutamatergic synapses. A) MPP7 (MAGUK p55 homolog) and SAP97 (Dlg homolog) interact in animal epithelia as shown by Stucke et al. (2007) in the human Caco2 cell line and support the function of tight junctions. B) MPP2 (MAGUK p55 homolog) and PSD-95 (Dlg homolog) interact with each other in the glutamatergic postsynapse and are important for adherens in this junction as well (MPP2 interaction with SynCAM that interconnects into the presynaptic active zone) (Image recreated from Rademacher et al. (2016, Scientific Reports, Nature, Creative commons license).

In order to confirm binding of *S. rosetta* Dlg and MPP7, we used isothermal titration calorimetry to test the binding capacity of three heterologous Dlg expression constructs (L27 domain; PDZ1-3 tandem; SH3-HOOK-GuK module) against full-length MPP7. Whereas, we did not find evidence for PDZ1-3 or SH3-

HOOK-GuK to be involved in binding MPP7, we detected heat release upon titrating Dlg L27 to MPP7, which suggests binding of the two proteins via L27 domains. The experiment requires further refinement to produce reliable proof for the interaction. Accordingly, there are four possible scenarios to explain these results. Scenario 1, Dlg and MPP7 could be direct binding partners in choanoflagellates, interacting in a similar fashion to their counterparts in animal epithelia. Scenario 2, Dlg and MPP7 could also be part of the same complex without direct interaction, or require another protein to enable their interaction, such as in the epithelial tripartite complex Dlg-MPP7-Mals3 (Bohl et al., 2007; Yang et al., 2010). Scenario 3, the possibility that MPP7 is a direct binding partner of PSD-95 alpha in choanoflagellates. This scenario requires further investigation. Lastly, scenario 4, MPP7 could be a contaminant in the co-IP and not be part of the complex. Even though the isothermal titration data does not provide satisfying evidence for the binding of *S. rosetta* Dlg and MPP7, the co-precipitation of the two proteins in combination with the data, hinting towards a possible binding between the two proteins, support the first scenario, although scenarios 2 and 3 remain a possibility. The fourth scenario is unlikely, but for stronger confidence, replicates produced for the co-IP are required.

4.4.2 Additional *S. rosetta* Dlg interactions are mainly distinct from synaptic PSD-95 interactions

The three scaffolding proteins Dlg/PSD-95, PSD-95 alpha and a MAGUK p55 protein have the highest quantities in the mass spectrometry dataset. Other proteins with still relatively high label free quantification values include an otherwise unspecified ankyrin repeat containing protein, which could be part of a signalling complex as ankyrin repeats are known to mediate protein-protein interaction (Li et al., 2006), a phosphoinositide specific phospholipase C (PLC) and nucleoprotein TPR. PLC is an enzyme that catalyses the reaction of phosphatidylinositol-4,5-biphosphate (PIP2) to inositol-1,4,5-triphosphate (IP3) and diacylglycerol, two second messengers (Clapham, 2007). Nucleoprotein TPR binds to components of the nuclear pore complex and is involved in the retention of unspliced RNA in the nucleus (Cordes et al., 1997; Rajanala et al., 2012). These proteins could be of interest as they are involved in signalling or are found in the nucleus, the *S. rosetta* Dlg localisation suggested by our

immunostaining experiments. Further proteins were detected with unique peptide counts in the Dlg co-IP dataset, some of which were not annotated. Other proteins were annotated as metabolic enzymes, which are presumably contaminants (for example enzymes of different metabolic pathways). Considering the identification of putative contaminating proteins, suggests that all the other proteins identified at these lower relative quantities have to be considered carefully and therefore more replicates are required. Furthermore, we found endosomal proteins (such as a protein with a RUN and a FYVE domain, and a transmembrane 9 superfamily member with endosomal integral membrane domain), suggesting that a subset of *S. rosetta* Dlg localises to endosomes (potentially with other proteins that are not usually in the Dlg complex) and might be in the process of being transported to other cellular compartments or the plasma membrane.

With these doubts in mind, we still looked for other interesting potential *S. rosetta* Dlg interactors. We detected a protein with an RNA recognition motif found in splicing factors and ribonucleoproteins. Moreover, we detected proteins with general importance in signalling cascades, amongst which were a heterotrimeric G protein and a protein with a Pleckstrin homology domain that is targeted by G proteins, as well as phosphatases and kinases. We identified a cAMP dependent serine/threonine kinase like protein (protein kinase A – PKA). This protein is important for example for the phosphorylation of receptors and is found in complexes with SAP97 via AKAP proteins (Gardner et al., 2007). No AKAP protein was detected in *S. rosetta*, which suggests that PKA could bind via another linker. Another interesting candidate is an LZIC like protein with an ICAT domain, but missing a leucine zipper domain (in human the LZIC ICAT domain inhibits beta-catenin binding with Tcf transcription factor without disrupting catenin/cadherin interactions; the leucine zipper domain is found in transcription factors and is important for DNA binding). LZIC proteins are involved in processes of neurogenesis, cell adhesion and transcription regulation (Tago et al., 2000; Graham et al., 2002; Shin et al., 2014). We further identified membrane anchored proteins and a protein with an immunoglobulin-like fold, found in plexins and some receptor types (Bork et al., 1999), and tumour necrosis factor repeats, also found in receptors (Banner et al., 1993).

Notably, we also found some other Dlg bound transmembrane protein candidates in an extended dataset, including proteins unique to the Dlg IP

according to their LFQ values, but with some peptide counts in the control IP. Among these candidates was a protein annotated as a subunit of a voltage gated potassium channel. PSD-95 has been shown to bind and cluster various Shaker-type and inward rectifying K+ channels (Kim et al., 1995; Cohen and Brenman, 1996; Leonoudakis et al., 2004). Although the identity and function of this protein in *Salpingoeca rosetta* cannot be known for certain as it has higher similarity to other choanoflagellate specific proteins, it represents a candidate for potential *S. rosetta* transmembrane proteins that could bind to Dlg. In synapses, PSD-95 also binds to ionotropic glutamate receptors (Niethammer et al., 1996), which were not identified in *S. rosetta*. In the extended dataset, we also identified actin and the motor protein dynein. If these proteins are indeed part of the *S. rosetta* Dlg complex, this might allow Dlg to move along microtubules in a similar way suggested for human SAP97 (Yamada et al., 2007). All these findings require future testing.

Our mass spectrometry dataset allowed us to conclude that many known interaction partners with homologs encoded in the *S. rosetta* genome (such as GKAP, CRIPT, Citron, Spar, and SynGAP) (Burkhardt et al., 2014) do not interact with Dlg in the choanoflagellate under the experimental conditions used.

4.4.3 Multimerisation via the SH3-HOOK-GuK module might not be conserved in *S. rosetta*

In order to learn more about the ability of *S. rosetta* Dlg to homo- and heteromultimerise, we investigated the structure of *S. rosetta* Dlg and MPP7 by computational homology reconstruction, aligning to solved animal crystal structures of the same proteins. Two regions of membrane associated guanylate kinases known to be involved in multimerisation are the SH3-HOOK-GuK module and the L27 domain. Rat PSD-95 has been shown to multimerise via the SH3-HOOK-GuK module by swapping between intra- and intermolecular interaction of SH3 and GuK domains (McGee et al., 2001). Computational modelling of the structure in alignment to rat PSD-95 SH3-HOOK-GuK module (McGee et al., 2001; PDB ID: 1KJW) showed that the backbone of the *S. rosetta* Dlg GuK domain is nearly identical to the backbone of the rat PSD-95 GuK domain. The *S. rosetta* SH3 domain, although similar in structure, seems to be slightly more derived. The largest difference between the rat and the *S. rosetta* molecules

seems to be the flexible HOOK region. Although both *S. rosetta* and rat HOOK contain a coiled coil region, these regions do not seem to align in the computational model.

For hDlg it was shown that phosphorylation at the N-terminus regulates the confirmation switch (Wu et al., 2000), furthermore it has been suggested that a ligand binding the HOOK region might regulate SH3-GuK interaction (McGee et al., 2001). Therefore, the presence of the N-terminus in the full-length molecule or the addition of a HOOK ligand might as well alter the conformation in choanoflagellates. Differences in the HOOK region between SAP97 and PSD-95 have been suggested to be responsible for different conformations of SH3-HOOK-GuK in the two proteins, leading to PSD-95 occurring more in the open conformation and SAP97 occurring more in the closed conformation (Vandanapu et al., 2009). The authors hypothesised that the longer HOOK region in SAP97 sterically brings SH3 and GuK domain closer together (Vandanapu et al., 2009). Potentially, the choanoflagellate HOOK region sterically hinders an interaction between the two domains intramolecularly and maybe also intermolecularly.

4.4.4 The L27 domain of *S. rosetta* Dlg is potentially responsible for homo- and heteromultimerisation of the protein

L27 domains are known to heterodimerise with L27 domains from other proteins (Feng et al., 2004; Li et al., 2004; Yang et al., 2010) and in the special case of Dlg-1/*Drosophila* Dlg to homodimerise with the L27 domain of another Dlg copy (Nakagawa et al., 2004; Ghosh et al., 2018). Dlg/SAP97 interacts with MPP7 in epithelia by forming a tripartite complex with Mals3 (LIN7C) (Bohl et al., 2007). The structure of the complex was solved for the human proteins (Yang et al., 2010). The crystal structures showed that the four L27 domains of the three proteins are responsible for the interaction with Dlg interacting with the N-terminal MPP7 L27 domain and Mals3 interacting with the more C-terminal MPP7 L27 domain (Yang et al., 2010). Bohl et al. (2007) found that Dlg and MPP7 require Mals3 to bind to each other. Yang et al. (2010) expanded Dlg-MPP7 binding assays and found in *in vitro* experiments that Dlg and MPP7 interact weakly even without Mals3 presence, but the interaction was more than 10-fold increased by Mals3. *In vivo* they could not detect an interaction between Dlg and MPP7 without

the presence of Mals3. Yang et al. (2010) suggested that the tripartite complex forms synergistically, with a simultaneous formation of the two L27 heterodimers.

S. rosetta does not encode a Mals3 (Lin7) protein in its genome and the structure of the more C-terminal L27 domain of *S. rosetta* MPP7 seems to be less conserved. Nevertheless, MPP7 was co-immunoprecipitated with *S. rosetta* Dlg. The different conclusions this finding allows were discussed above. Considering that Dlg and MAGUK p55 homologs interact in different animal cell types, and that the choanoflagellate species *S. rosetta* possesses orthologous proteins, it stands to reason that the two proteins might in fact interact. Our isothermal titration data suggest that they can interact individually without a third protein, however these data were not entirely unambiguous and require further investigation. It is possible that no other protein is needed to support the interaction in *S. rosetta*. Structure homology modelling showed that the *S. rosetta* Dlg L27 domain shows differences to both the human and *D. melanogaster* Dlg L27 domain (Figure 4-10). As discussed above, in mammals Mals3/Lin7C binding to MPP7 promotes the binding between MPP7 and Dlg (Bohl et al., 2007; Yang et al., 2010). Similarly, Lin-7 promotes binding of Metro (a MAGUK p55 protein) and Dlg at *Drosophila melanogaster* neuromuscular junctions (Bachmann et al., 2010). Potentially, the differences in the MPP7 and Dlg L27 domains in *S. rosetta* allow the two proteins to bind without this controlled mechanism. It is also possible that *S. rosetta* requires another protein to support the interaction.

4.4.5 The potential subcellular localisation of *S. rosetta* Dlg is the nuclear membrane and/or the plasma membrane at the basal side of the cell

We further performed immunostaining experiments with the same antibody used for co-IP. We observed consistent nuclear staining in replicated experiments. However, some cells also showed staining of the plasma membrane at the basal side of the cell (side opposed to the flagellum) (Figure 4-11). Considering, that the antibody might bind both *S. rosetta* Dlg and PSD-95 alpha, it is possible that at least one of the localisations actually corresponds to PSD-95 alpha. These data suggest, that the *S. rosetta* Dlg complex localises to the nucleus, or more particularly to the nuclear membrane (most probable considering replicability and strength of the signal) or to the plasma membrane at the basal side of the cell. It is also possible that *S. rosetta* Dlg is found at

different intracellular localisations including both the nucleus and the membrane and potentially at other localisations as suggested by other stainings with this antibody that could only be observed in a single experiment, as well as staining with other *S. rosetta* Dlg antibodies (Figure 6-9, Figure 6-10, and Figure 6-11). A change of localisation is typical for Dlg proteins, which are targeted to the plasma membrane via palmitoylation and into different complexes via phosphorylation (Craven et al., 1999; Xu, 2011). Yamada et al.(2007) showed that hDlg (human SAP97) binds to KIF1α, a kinesin motor protein, which led them to the suggestion that hDlg is transported to the membrane by KIF1α along microtubules, potentially already bound to receptors in vesicles that are transported to the membrane via this mechanism.

A nuclear localisation of Dlg is supported by the presence of nuclear proteins *S. rosetta* Dlg complex isolated by co-IP. The finding that also Homer, another postsynaptic scaffolding protein, was shown to localise to the nucleus of choanoflagellates and rat astrocytes (Burkhardt et al., 2014), further suggests that scaffolding proteins might have functions in the nucleus. Furthermore, splice variants of hDlg (human SAP97) have been found in the nucleus as well, and it was suggested that Dlg variants are relocated to the nucleus upon release of intramolecular SH3-GuK interaction, due to an otherwise blocked nuclear translocation signal in the HOOK region (Kohu et al., 2002). This nuclear translocation signal, consisting of four consecutive basic amino acids (KRKK) was identified in a splice variant of SAP97 that was sufficient for nuclear translocation, when the SH3-GuK interaction was broken (Kohu et al., 2002). In wild type neurons, SAP97 is localised partially to the membrane and partially to the nucleus (Kohu et al., 2002). Mammalian PSD-95 does not contain a nuclear translocation sequence and it is more consistently localised to the membrane (Kohu et al., 2002). We could identify four consecutive basic amino acids in *S. rosetta* Dlg (KKKK). This potential nuclear translocation signal further supports a nuclear localisation of *S. rosetta* Dlg, although it remains to be tested if the signal is potent. The putative nuclear translocation signal of *S. rosetta* Dlg (KKKK) lies in the SH3 domain instead of in the HOOK region as in rat SAP97 (Kohu et al., 2002). However, a computational structure prediction is always an approximation and it is possible that structures aligning to the rat SH3 domain are part of the HOOK region. Other proteins, identified in the co-IP experiment as putative

members of a Dlg complex in *S. rosetta*, are membrane proteins, but it needs to be investigated, if they are localised to the nuclear membrane, the plasma membrane, another cellular membrane, or various membranes.

Immunostaining data suggest a localisation of the Dlg scaffold to the nuclear membrane in choanoflagellates. MAGUKs including ZO-1, CASK and Dlg proteins, were found to localise to the nucleus and CASK has been shown to have some regulatory function (Hsueh et al., 2000; Sherman and Brophy, 2000; Kohu et al., 2002). Our data suggest the possibility that they could also be involved in a nuclear scaffold. Nuclear scaffolds have been described such as the NIPP1 nuclear scaffold (Van Eynde et al., 2004). Furthermore, it has been described that the cytoskeleton acts in the nucleus in similar ways as in the cytoplasma and that different nuclear compartments (nuclear bodies) are dynamic and interconnected, coordinating processes such as splicing, nuclear import and export, DNA replication and transcription as well as controlling expression (Zimber et al., 2004). Studying nuclear Dlg complexes could as well lead to the discovery of an unknown function of Dlg in animals. Another postsynaptic scaffolding protein was identified in the nucleus in choanoflagellates and was subsequently found in nuclei of rat astrocytes (Burkhardt et al., 2014). It seems worthwhile to consider the function of these scaffolding proteins in the nuclei of animal and choanoflagellate cells, considering that the nuclear scaffold could be more ancestral and could teach us more about the type of complexes postsynaptic signalling machineries originated from. Antibody staining to the basal side of the cell membrane of some cells suggests that under certain circumstances there might be a MAGUK regulated complex at this membrane. This would be particularly interesting, considering that at the basal pole cells attach to the substratum and have cell-cell contacts in colonies (Dayel et al., 2011).

4.4.6 Possible implications for the evolution of a postsynaptic scaffold in animals

Dlg proteins form a variety of protein complexes in bilaterian animals (illustrated in Figure 4-13). In vertebrate glutamatergic synapses PSD-95 forms distinct complexes at AMPA receptors, NMDA receptors, kainate receptors and Shaker potassium channels (Kim et al., 1995; Niethammer et al., 1996; Garcia et

al., 1998; Chen et al., 2000; Husi et al., 2000; Contractor et al., 2011; Kegel et al., 2013). Other Dlg homologs can rescue some of the PSD-95 functions (Xu, 2011). The recruitment of Dlg homologs into complexes is strictly regulated by alternative transcription (Schlüter et al., 2006), posttranslational modifications such as phosphorylation (Xu, 2011) and intra-protein activity regulations (Wu et al., 2000). The single Dlg homolog in invertebrates Dlg-1 is involved in complexes at the neuromuscular junction, as well as at adherens junctions and septate junctions (as shown for *C. elegans* and *D. melanogaster*, respectively) with implications for cell-cell adhesion and cell polarity establishment (Firestein and Rongo, 2001; Roh et al., 2002; Wang et al., 2014). In addition to its synaptic functions the vertebrate homolog to Dlg-1 (SAP97) also regulates the function of epithelial tight junctions and is involved in cell polarity establishment as well (Roh et al., 2002; Humbert et al., 2008). Interactions common to all these complexes are homo- and hetero-multimerisation processes with Dlg proteins and other membrane associated guanylate kinases (such as MAGUK p55 homologs) (Bachmann et al., 2004; Nakagawa et al., 2004; Stucke et al., 2007; Rademacher et al., 2016; Ghosh et al., 2018). These interactions explain the scaffolding functions of Dlg proteins. Our data point to a conservation of this kind of interaction in choanoflagellates, suggesting the presence of a Dlg/MAGUK p55 scaffold in the last common ancestor of animals and choanoflagellates.

We hypothesised that the *S. rosetta* Dlg homolog has scaffolding function and forms a scaffold similar to the one found in the postsynapse. Data presented in this chapter support this hypothesis. The Dlg/PSD-95 scaffolding function seems to be conserved in choanoflagellates, including the potential for homo- and heteromultimerisation with other membrane associated guanylate kinases. Moreover, a *S. rosetta* Dlg putatively interacts with a MAGUK p55 protein, an interaction that is important for protein scaffolds at cellular junctions at epithelia and synapses (Stucke et al., 2007; Rademacher et al., 2016). This suggests that a MAGUK organised scaffold predates the origin of first animals and was present in the last common ancestor of choanoflagellates and animals. However, our data do not support the presence of signalling machineries organised by Dlg in choanoflagellates that are acting in animal epithelia and/or glutamatergic synapses. This proposes that signalling machineries specialised to functions at

animal epithelia and synapses evolved in the animal lineage but might have been built upon a putatively ancestral structural framework.

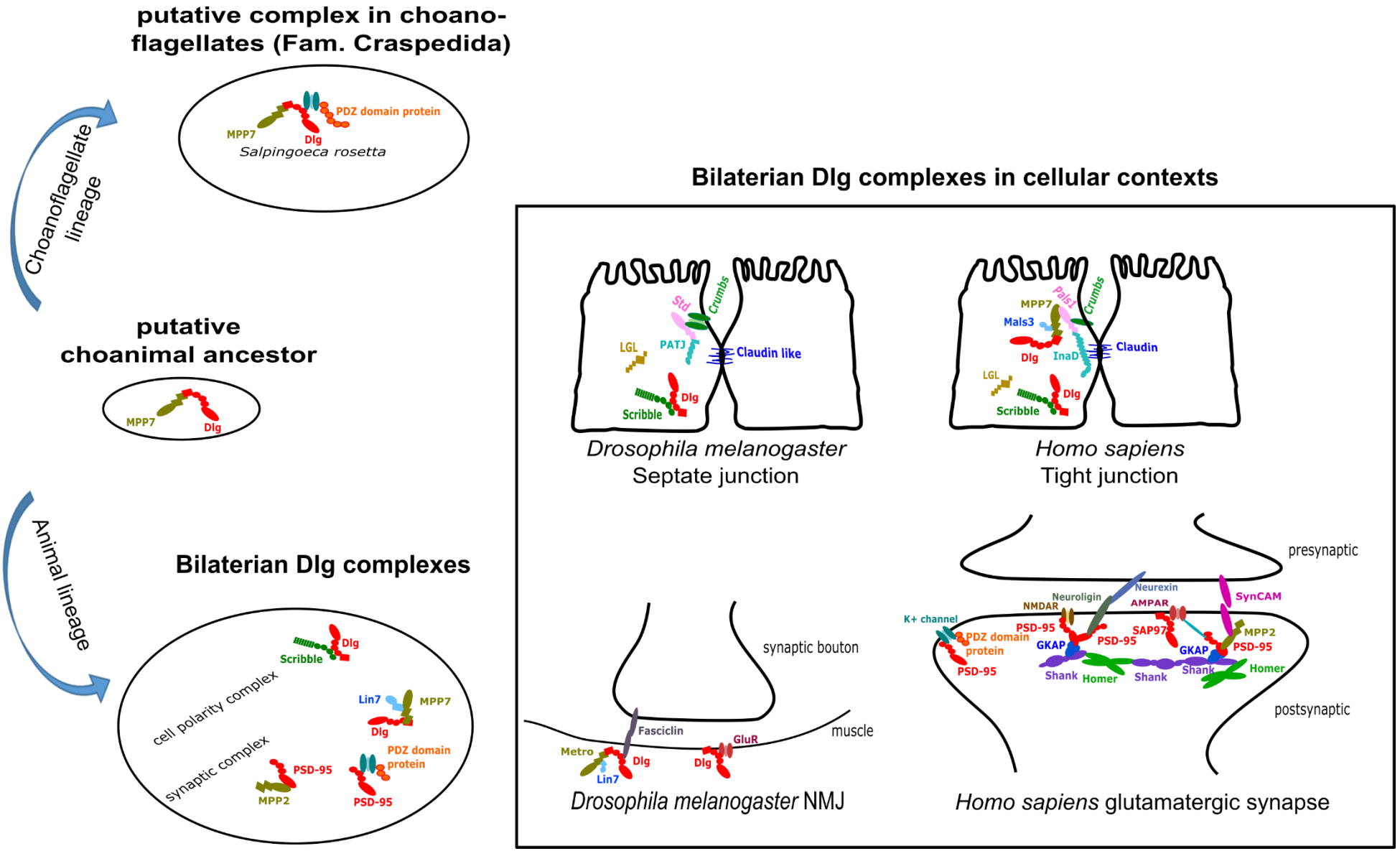

putative complex in choano-flagellates (Fam. Craspedida)
PDZ domain protein
MPP7
Dlg
Salpingoeca rosetta
Choanoflagellate lineage
putative choanimal ancestor
MPP7
Dlg
Animal lineage
Bilaterian Dlg complexes
Dlg
Scribble
cell polarity complex
Lin7
MPP7
Dlg
synaptic complex
PSD-95
PDZ domain protein
MPP2
PSD-95
Bilaterian Dlg complexes in cellular contexts
Std
Crumbs
LGL
PATJ
Claudin like
Dlg
Scribble
Drosophila melanogaster
Septate junction
MPP7
Pals1
Crumbs
Mals3
Dlg
InaD
Claudin
LGL
Dlg
Scribble
Homo sapiens
Tight junction
synaptic bouton
Fasciclin
muscle
Metro
Dlg
GluR
Dlg
Lin7
Drosophila melanogaster NMJ
presynaptic
Neurexin
Neuroligin
NMDAR
AMPAR
SynCAM
K+ channel
PSD-95
PDZ domain protein
PSD-95
GKAP
PSD-95
SAP97
MPP2
GKAP
PSD-95
Shank
Homer
Shank
Shank
Homer
postsynaptic
Homo sapiens glutamatergic synapse

Figure 4-13: **Hypothetical evolutionary scenario for the elaboration of an ancestral Dlg complex to modern complexes at bilaterian cellular junctions.** Left: choanoflagellate-animal ancestor (choanimal ancestor) as inferred from the comparison of Dlg complex components in *Salpingoeca rosetta* and animals. Hypothetical *S. rosetta* complex suggested by binding partners found in this study. The receptor/transmembrane protein involved and the specific interactions between the proteins remain to be investigated. Three hypothetical bilaterian ancestral complexes as inferred from the presence in both protostomes and deuterostomes, building on more ancestral complexes. Only a selection of bilaterian Dlg homolog complexes are shown. Complexes shown based on data/ graphs from: Roh et al., 2002 (std/PALS1/MPP5-crumbs-PATJ), Mathew et al., 2002 (Scribble-Dlg *Drosophila* via GUKholder (not shown)), Humbert et al. 2008 (Scribble-Dlg vertebrates via NHS (not shown)), Bohl et al., 2007 (Dlg-MPP7-Mals3), Poliak et al., 2002 (PATJ/InaD-Claudin), Bachmann et al., 2010 (Dlg-Metro-Lin7), Wang et al., 2014 (Neuromuscular junction), Kegel et al., 2013 (AMPAR complex), Shiraishi-Yamaguchi and Furuichi, 2007 (Shank, Homer), Blackstone and Sheng, 2002; Feng and Zhang, 2009 (other synaptic Dlg interactions), Kim et al., 1995; Poliak et al., 2003 (K+ channel complex). Abbreviations: Std: Stardust (*Drosophila* homolog of PALS1/MPP5); NMJ: Neuromuscular junction.

Interestingly, some choanoflagellate species express proteins with highest similarity to ionotropic glutamate receptors (section 2.3.4). It is therefore possible that a connection of the MAGUK scaffold and ionotropic glutamate receptors (potentially even at the plasma membrane) is conserved in some choanoflagellates, which would suggest that this interaction might be ancestral. This needs to be investigated in order to learn about the ancestry of these complexes. All choanoflagellate species investigated in section 2 are in culture (Richter et al., 2018). Protocols for protein extraction would need to be established for the species of interest. It is particularly interesting that putative glutamate receptors occur in *Choanoeca perplexa*, a species that is closely related to the newly discovered *Choanoeca flexa*, which occurs in colonies that contract when stimulated with light (Brunet et al., 2019). Although this response is collective, theoretically, it does not necessarily require coordination between the cells, as all cells are equally stimulated by the light (Brunet et al., 2019). Investigating the function of these receptors in this species could give new insights on the question, if choanoflagellate cells can communicate with each other, and if the MAGUK-iGluR interaction might have preceded animal origins.

Electron micrographs of cnidarian and ctenophore synapses show electron-dense regions near the postsynaptic membrane, which appear similar to the postsynaptic density observed at the bilaterian postsynapse (Hernandez-Nicaise, 1973; Westfall, 1996). We did not find records of biochemical investigations of

these structures in those phyla. Investigating if non-bilaterian animals form PSD-like scaffolds using bilaterian PSD scaffolding proteins is another important step to unravel the evolution of these complexes. It is known that these proteins occur in all animals (Sakarya et al., 2007; Alié and Manuel, 2010), but to our knowledge they have not been studied on protein level in sponges, cnidarians and ctenophores to this date.

4.5 Outlook

The co-IP dataset in combination with isothermal titration calorimetry measurements and immunostainings performed with a custom-made *S. rosetta* Dlg antibody presented in this chapter suggest that the Dlg and MAGUK p55 homologs of the choanoflagellate *Salpingoeca rosetta* interact. This supports the hypothesis that part of the Dlg complex found at synaptic junctions in bilaterian animals preceded the origin of first animals.

For more confidence on the co-IP data, replicates for this experiment created with the same antibody remain to be analysed. Replication of the analysis with another antibody would increase confidence in the data, as interaction partners that could be identified with both antibodies are unlikely to be proteins the antibodies cross-react with, if antibodies are raised against different antigens of the same protein. The three *S. rosetta* peptide antibodies tested, did not pull-down *S. rosetta* Dlg in co-IP experiments. The fourth created peptide antibody is an interesting candidate to test, because it shows membrane staining in immunostaining (section 6.3.1; Figure 6-11). Other approaches could be used to complement the co-IP and back up the data. One option would be a pull-down of proteins from *S. rosetta* lysate with recombinant *S. rosetta* Dlg bound to a column. This approach might show interactions that do not occur in the cell under native conditions, where they might occur at different cellular localisations. Additionally, some interactions that do occur in the cell might not be found with this approach, as the native protein might have posttranslational modifications or differences in folding with importance for binding. Nevertheless, this approach has the advantage that it circumvents cross-reactivity with an antibody. With the option to transfect *S. rosetta* (Booth et al., 2018) and to perform genome editing in this

species (Booth and King, 2020), it is possible that in the future it will be feasible to also integrate Dlg protein with different tags into the *S. rosetta* genome, which could be used for tandem affinity purification. This approach allows the isolation of protein complexes through two specific affinity purification steps, resulting in low levels of false positives and false negatives (Puig et al., 2001).

Furthermore, it would be interesting to quantify on protein level, if the Dlg protein is indeed more abundant in colonies. Generally, co-IP experiments will be performed on colony lysate as well, in order to look for variation in the *S. rosetta* Dlg complex in different life cycle stages. A repetition of immunostainings using an antibody targeting the nuclear pore complex could specify the nuclear Dlg position. Immunostainings on colonies will also be interesting to compare the localisation in colonies to the one in single cells.

The MPP7-Dlg interaction in *S. rosetta* is suggested by our data, but will require stronger data. We will express and purify the L27 tandem of MPP7 and perform ITC experiments, titrating the Dlg L27 domain to this region. Using only these shorter constructs will allow to increase concentrations, which will hopefully increase the signal.

It would be beneficial to rule out that the utilised antibody recognises MPP7 or PSD-95 alpha. Testing if the antibody can bind to the recombinant proteins would be a first indication showing if the proteins are in the co-IP complex due to nonspecific binding of the antibody or due to interaction with Dlg. This can be tested in different ways. One option would be to incubate beads with bound antibody with recombinant MPP7 or PSD-95 alpha, followed by washing steps and Western blot analysis. Another option would be to try if the antibody could be blocked with MPP7 or PSD-95 alpha PDZ domains in the same way as with Dlg PDZ domains (Figure 4-2 D). These two experiments could indicate whether the Dlg antibody has cross-reactivity with MPP7 and/or PSD-95 alpha. Nevertheless, if both experiments do not show binding of the antibody to these proteins, there is still the possibility, that the antibody does bind the native protein, which could be folded differently than the recombinant protein, or have posttranslational modifications required for binding.

Lastly, we are planning to transfect *S. rosetta* with plasmids encoding fluorescently tagged Dlg using the method described by (Booth et al., 2018) in

order to confirm localisations of this protein suggested by immunostaining experiments.

5 Main Discussion

5.1 Insights into the origin of postsynaptic signalling machineries

The major aim of this thesis was to improve our understanding of the evolutionary origin of postsynaptic signalling machineries. To do this, I studied the choanoflagellates, the closest known protistan relatives of animals (Carr et al., 2008; King et al., 2008; Ruiz-Trillo et al., 2008), and used my data in order to make predictions about the last common ancestor of choanoflagellates and animals. I made use of a variety of approaches, including transcriptome surveys, ancestral protein reconstruction combined with functional binding assays as well as co-immunoprecipitation and immunostaining. My results suggest which components of vertebrate-like glutamatergic postsynaptic signalling machineries preceded animal origins and were therefore prerequisites for the evolution of synapses. This further implies which important protein components and protein interactions were likely innovations concordant or within the animal radiation.

5.2 Key proteins for the formation of postsynaptic signalling machineries putatively preceded synapse evolution

I surveyed 19 new choanoflagellate transcriptomes (Richter et al., 2018) for the presence of proteins with statistically significant sequence similarity to key proteins of postsynaptic complexes at vertebrate glutamatergic synapses. I hypothesised that this survey would provide more insights into which postsynaptic proteins are putatively ancestral. Indeed, I made two major findings that could be of interest for future research.

First, I found that scaffolding proteins with important functions at vertebrate glutamatergic synapses are expressed in choanoflagellates that branch throughout the phylogenetic radiation of this group. Potential differences in the types of membrane associated guanylate kinases (MAGUKs) expressed between species of the two choanoflagellate families were observed. Whereas a majority of choanoflagellate species expressed putatively choanoflagellate-unique MAGUKs, only species of the family Craspedida expressed MAGUKs with domain architectures of animal-like Dlg and MAGUK p55 proteins. Only craspedidan species secrete organic cell covers, and include species that form colonies with cell-cell contact via incomplete cytokinesis (Leadbeater and Morton,

1974; Carr et al., 2008; Dayel et al., 2011; Nitsche et al., 2011; Stoupin et al., 2012; Leadbeater, 2015). Second, I found that the last common ancestor of choanoflagellates and animals probably had proteins with homology to animal ionotropic glutamate receptors, Shaker/Shal-like voltage gated potassium channels and nitric oxide synthase, as I identified proteins with statistically significant sequence similarity to these proteins in a subset of the choanoflagellate transcriptomes.

It was suggested that the main animal innovations that were necessary for the evolution of synapses were receptors and specialised signalling proteins that were integrated into new signalling machineries organised by previously existing protein scaffolds (Alié and Manuel, 2010). Indeed, choanoflagellates lack important synaptic adhesion molecules (such as Neuroligin and SynCAM) as well as many of the receptors required for synapse function (Alié and Manuel, 2010; Burkhardt et al., 2014). The identification of a putative ionotropic receptor and a voltage-gated potassium channel in a subset of choanoflagellate species suggests that the last common ancestor of animals could likely react to glutamate or other molecules and could make use of voltage-mediated signalling. This is supported by the conservation of ionotropic glutamate receptors between plants and animals (Chiu et al., 1999), and previous studies showing voltage-mediated signalling in other eukaryotes. Mechanical stimuli lead to changes in membrane polarisation in the ciliate *Paramecium caudatum*, which alters its swimming behaviour (Machemer and Ogura, 1979; Schlaepfer and Wessel, 2015). Furthermore, it has been elucidated that a cGMP-gated potassium channel plays a role in the phototactic response of the flagellated zoospores of the aquatic fungus *Blastocladiella emersonii* (Avelar et al., 2015). The presence of a putative nitric oxide synthase further suggests that cells in the last common ancestor of animals and choanoflagellates had simple means of communication. Putative implications of these findings are further discussed in section 5.7.

Given that the source of my findings are transcriptomes, these results are preliminary. Even though all transcriptomes are of good quality, expressing a majority of core eukaryotic genes (Richter et al., 2018), the genes regarded here could be facultative and might therefore not be expressed under the experimental conditions chosen for the analysis or in the life cycle stage the choanoflagellates were in prior to the procedure. Therefore, the observation of putative differences

between the expressed MAGUKs of craspedid and acanthoecid choanoflagellate species mentioned above has to be considered with caution. In order to test if this observation displays the real situation, genome sequencing of acanthoecid species would be necessary. The two choanoflagellate genomes sequenced to date are both of craspedid species (King et al., 2008; Fairclough et al., 2013). One further limitation of my data is that a survey alone is insufficient to resolve the identity of the proteins identified. In future experiments, the phylogenetic relationships of the choanoflagellate proteins identified and their putative animal counterparts will need to be established. Finally, only functional analysis of these proteins in the future will reveal their cellular function in choanoflagellates as well as their putative function in the last common ancestor of choanoflagellates and animals. Nevertheless, this survey highlights differences between choanoflagellate species that need to be understood in order to draw conclusions about their last common ancestor with animals and suggests that additional key genes required for synaptic function might have preceded animal origins.

5.3 The prerequisites for the formation of a Homer-Shank platform likely preceded animal origins

At vertebrate glutamatergic synapses, the interaction between the two scaffolding proteins Homer and Shank is of major importance as they form a binding platform that interconnects receptor-complexes with signalling machineries involved in actin cytoskeleton remodelling and calcium signalling (Naisbitt et al., 1999; Sala et al., 2001). I reconstructed putative ancestral sequences of the Homer EVH1 domain corresponding to the animal and the choanoflagellate ancestor, using maximum likelihood approaches. I then produced recombinantly expressed proteins of these putative ancestral sequences, as well as of sequences from selected animal and choanoflagellate species. These proteins were tested for binding towards the Shank PPxxF motif region. I hypothesised that the capacity of the Homer EVH1 domain to bind PPxxF motifs is conserved and was present in the last common ancestor of choanoflagellates and animals.

Sequence, structural and domain architecture comparisons of the Homer and Shank proteins in choanoflagellates revealed that the Homer binding site is conserved. PPxxF motifs are found in many but not all identified choanoflagellate

Shank proteins. Many of the Homer proteins contain coiled coil domains, which enable them to tetramerise (Hayashi et al., 2006, 2009), a feature which has been shown to be conserved in the choanoflagellate species *Salpingoeca rosetta* (Burkhardt et al., 2014). Likewise, many of the Shank proteins contain a SAM domain, that enables vertebrate Shank to multimerise (Naisbitt et al., 1999). In addition, all tested ancestral and candidate Homer EVH1 domains bound a rat Shank construct including the PPxxF motif. The capacity of *S. rosetta* Homer to bind the PPxxF motif of *S. rosetta* Shank was also demonstrated. My results support the hypothesis that the Homer and Shank binding capacity preceded the evolutionary origin of animals.

Even though I could demonstrate that *S. rosetta* Homer and Shank have the capacity to bind, Shank was not co-precipitated with Homer from colonial *S. rosetta* cell lysate (Burkhardt et al., 2014). This allows two alternative interpretations. Potentially the binding capacity of Homer is ancestral and Shank was later recruited into an interaction with Homer by molecular exploitation as described by Anderson et al. (2016). Another option is that Homer and Shank do bind in *S. rosetta*, but that the binding was simply not detected in the co-immunoprecipitation or that they only bind in another life cycle stage. Co-immunoprecipitation experiments on *S. rosetta* cell lysate from cells enriched in other life cycle stages might reveal binding of the two proteins in non-colonial cells.

A combination of Homer and Shank binding capacity and the capacity to tetramerise and multimerise in many choanoflagellate species, suggest that the molecular prerequisites for the formation of a Homer-Shank platform were already in place before the emergence of synapses in the animal lineage.

5.4 An ancestral scaffold probably served as basis for epithelial occluding junctions and synaptic junctions

The MAGUK PSD-95 is known as the key regulator of the postsynaptic density at vertebrate glutamatergic synapses, as it is involved in the formation and organisation of protein complexes at glutamate receptors (Kim and Sheng, 2004; Chen et al., 2011). One important feature of PSD-95 and other vertebrate Dlg paralogs is the formation of a scaffold via protein-protein interactions (Verpelli

et al., 2012). I hypothesised that this scaffold forming function is conserved in the *S. rosetta* Dlg homolog. In order to test this hypothesis, I first identified potential *S. rosetta* Dlg interaction partners via co-immunoprecipitation followed by mass spectrometry analysis.

My results indicate that the interactions of *S. rosetta* Dlg are mainly distinct from known synaptic PSD-95 interactions. However, I did indeed identify other scaffolding proteins including one MAGUK of the p55 family. Initial binding assays suggest that *S. rosetta* Dlg and MAGUK p55 proteins might interact via their L27 domains. Although there are two lines of evidence both suggesting binding between the two proteins, further experiments are required to validate these findings. Increasing the number of co-immunoprecipitation replicates and fine-tuning isothermal titration calorimetry experiments will likely provide conclusive results.

My findings indicate that the last common ancestor of choanoflagellates and animals likely had an ancestral Dlg-MAGUK p55 scaffold. This supports previous suggestions of an ancestral scaffold (Sakarya et al., 2007; Alié and Manuel, 2010). Harden et al. (2016) proposed that synapses and occluding junctions (tight junctions and septate junctions) have a common origin in pleated septate junctions. Consistent with the hypothesis of Harden et al. (2016), a Dlg-MAGUK p55 interaction occurs at both tight junctions and at glutamatergic postsynapses (Stucke et al., 2007; Rademacher et al., 2016). By showing that this interaction putatively preceded first animals, my data support a common origin of the complexes. At both junctions, MAGUKs have been shown to anchor junctional proteins and link them to the cytoskeleton and other signalling machineries. This linkage has been suggested to facilitate the plasticity of the junction (Harden et al., 2016).

5.5 Hypothetical scenario for the elaboration of ancestral scaffolds for postsynaptic signalling machineries

A combination of previous research and my new data suggest that many components required to build vertebrate-like postsynaptic signalling machineries at glutamatergic synapses preceded animal origins. This includes the structural scaffold formed by MAGUKs, and the Homer-Shank scaffold, as well as a putative

ancestral protein that might have given rise to animal ionotropic glutamate receptors. Animal evolutionary innovations were supposedly the establishment of a link between the two scaffolds and the integration of receptors into this scaffold (Alié and Manuel, 2010). This probably occurred via the evolution of binding sites as well as the recruiting of adapter proteins such as GKAP that links PSD-95 and Shank at vertebrate glutamatergic synapses (Kim et al., 1997; Naisbitt et al., 1999). Finally, signalling machineries specialised on the quick transfer of information evolved and were recruited into the complex, potentially replacing or adding on ancient signalling machineries that might have been built on the same structural scaffold.

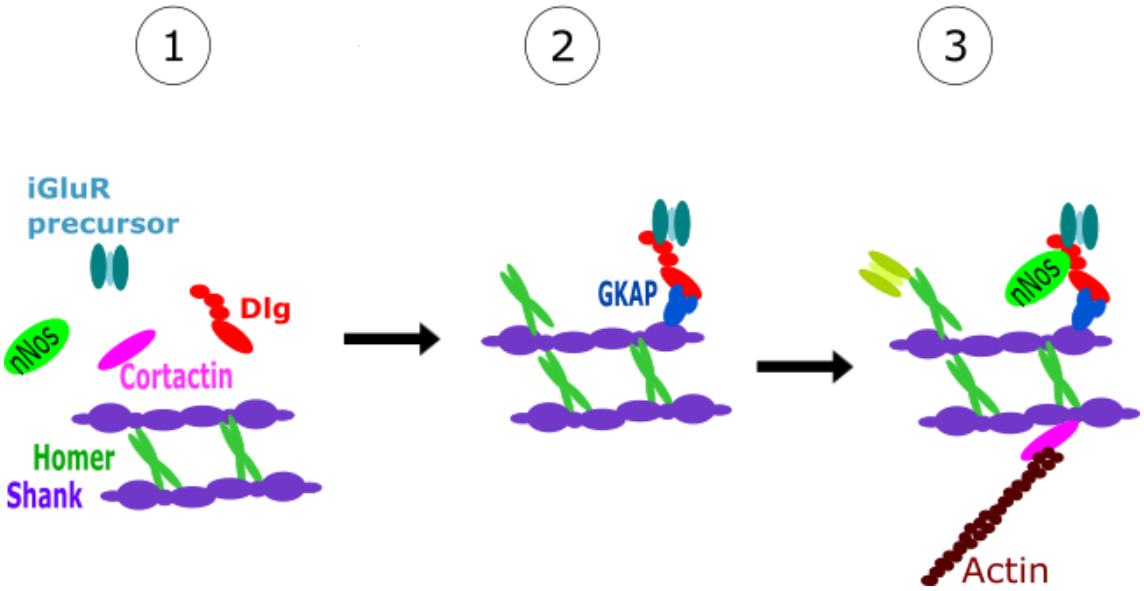


Figure 5-1: Hypothetical evolution of glutamatergic signalling machineries. 1) Many proteins that were recruited into vertebrate-like glutamatergic signalling machineries were present before the evolution of animals. 2) Important interactions between the receptors and scaffolds were formed. 3) More receptors and signalling molecules evolved or were recruited to the complex.

5.6 Putative ancestral functions of scaffolding proteins

As outlined in section 1.1.5, scaffolding proteins that are functionally important at postsynapses are pleiotropic and have a variety of functions in other cell types depending on the interaction partners present in the physiological context of this cell. Based on presence or absence of phylogenetically confirmed

homologs to synaptic binding partners of these scaffolding proteins in the genome of the choanoflagellate species *Monosiga brevicollis*, Alié and Manuel (2010) speculated about their putative ancestral functions. Alié and Manuel (2010) hypothesised that the Homer and Shank scaffold is ancestral and served as a link between the intracellular calcium signalling and the regulation of the actin cytoskeleton. My data support, that Homer and Shank might have formed a scaffold in the last common ancestor of animals and choanoflagellates. Although I showed that *S. rosetta* Homer has the capacity to bind Shank, Shank was not co-precipitated with Homer from colonial cells (Burkhardt et al., 2014), leaving it an unresolved question, if the two proteins are *in vivo* interaction partners in choanoflagellates. Instead, Burkhardt et al. (2014) made the discovery, that *S. rosetta* Homer interacts with flotillins in the nucleus of the choanoflagellate. Flotillins are proteins that have been shown to regulate cell proliferation and mitosis in human cell lines (Santamaria et al., 2005; Gómez et al., 2010). Burkhardt et al. (2014) subsequently revealed that both flotillin interaction and nuclear localisation of Homer is conserved in primary rat hippocampal astrocytes (specialised glia cells) but not primary rat hippocampal neurons. This suggests that nuclear translocation of Homer might be dependent on the cell type or a stimulus. In line with the hypothesis by Alié and Manuel (2010), Burkhardt et al. (2014) further identified an IP3 receptor among putative Homer binding proteins, which supports that Homer's calcium signalling function might in fact be ancestral. Evidence for the regulation of the actin cytoskeleton through Homer in choanoflagellates has so far not been found (Burkhardt et al., 2014), but might be detected through the investigation of Shank binding partners.

Immunostaining experiments suggested that Shank localises to the nucleus of *S. rosetta* as well (unpublished data by Dr Pawel Burkhardt). In cultured rat hippocampal neurons, Shank3 was found to translocate to the nucleus in an activity dependent manner (Grabrucker et al., 2014). In the nucleus Shank3 was found in a complex that is distinct from its PSD complex and included hnRNPs (heterogenous nuclear ribonucleoproteins) (Grabrucker et al., 2014). It was suggested that nuclear Shank3 can alter gene expression via assembly or modification of an hnRNP complex (Grabrucker et al., 2014). The expression of a constitutively nuclear Shank mutant led to a reduced synaptic density and reduced dendritic branching (Grabrucker et al., 2014). On a molecular level, it

was shown that the mutant displayed decreased transcription of Shank and increased transcription of several target genes including synaptotagmin-1 and leucine-rich repeat transmembrane neuronal protein 1 (LRRTM1) (Grabrucker et al., 2014). Synaptotagmin-1 is a protein at synaptic vesicles that is important for the calcium dependent fusion of vesicles with the plasma membrane (Pang et al., 2006)) and LRRTM1 is implicated in pre- and postsynaptic excitatory differentiation (Linhoff et al., 2009)).

These findings suggest that nuclear translocation by Shank might regulate pre- and postsynaptic processes. It is unknown what Homer and Shank might regulate in *S. rosetta*. Burkhardt et al. (2014) hypothesised that flotillin import into the nucleus (which they observed in all tested single cells but only in a few cells within rosette colonies) and its subsequent interaction with Homer regulates the differentiation of individual cells within the colony. One feature that has been suggested to allow cell differentiation in animals and their close relatives, is the controlled formation of actin-based projections, such as filopodia and microvilli (Sebé-Pedrós et al., 2013a). Filopodia-like structures (or pseudopodia) were morphologically described in a range of eukaryotic phyla (including excavates, stramenopiles, rhizarians, amoebozoans, apusozoans, nucleariids (the sister group of fungi), as well as holozoan phyla (filastereans, choanoflagellates, and animals)) (Leadbeater and Morton, 1974; Mikrjukov and Mylnikov, 2001; Zettler et al., 2001; Cavalier-Smith and Chao, 2003; Preston and King, 2003; Pawlowski, 2008; Cavalier-Smith and Chao, 2010; Dayel et al., 2011; Ota et al., 2012; Cavalier-Smith, 2013; Sebé-Pedrós et al., 2013a). Sebé-Pedrós et al. (2013) investigated the presence of animal filopodia-specific proteins across eukaryotes and revealed that there are some common proteins, such as core-actin linking proteins and members of a putative ancestral filopodia formation mechanism including the Arp2/3 complex, Diaphanous-related formins (DRFs) (actin remodelling proteins), as well as WASP (a nucleation promoting factor). The ancient mechanism was probably employed in both amoebozoans and animals with the evolution of independent signalling mechanisms for the control of filopodia formation (Sebé-Pedrós et al., 2013a). The authors found that key components of the animal molecular machinery for filopodia formation are encoded in the genomes of the choanoflagellates *Salpingoeca rosetta* and *Monosiga brevicollis* as well as the filasterean *Capsaspora owczarzaki*. They

further showed that one of these proteins, fascin (an actin-bundling protein), localises to filopodia and microvilli in *S. rosetta* (Sebé-Pedrós et al., 2013a). In addition, putative filopodial gene homologs are upregulated in *C. owczarzaki* cell types with filopodia in contrast to cell types without (Sebé-Pedrós et al., 2013a). They concluded that filopodia and microvilli of holozoans organisms have a common origin, building on more ancient eukaryotic mechanisms (Sebé-Pedrós et al., 2013a).

The choanoflagellate collar consists of microvilli (Leadbeater, 2015). In *S. rosetta* these microvilli are retracted in the transition to fast swimmers (Dayel et al., 2011). *S. rosetta* filopodia are important for the transition between attached cells and fast swimmers (Dayel et al., 2011). Fast swimmers use filopodia to attach to the substratum prior to the secretion of stalk and theca (Dayel et al., 2011). Filopodia are also used to lift and release a cell that is transitioning into a fast swimmer from its theca (Dayel et al., 2011). Furthermore, filopodia link cells in rosette colonies in addition to intercellular bridges and extracellular matrix (ECM) (Dayel et al., 2011). *C. owczarzaki* forms filopodiated amoebae that can form colonies through aggregation (Sebé-Pedrós et al., 2013b).

Even though, there is currently no evidence for the direct regulation of the actin cytoskeleton via Homer and Shank in choanoflagellates, it is hypothetically possible that this regulation occurs indirectly through the clustering of nuclear complexes. As elucidated above, vertebrate Shank interacts with hnRNPs in the nucleus, which reduces dendritic branching, a process that requires the maturation of filopodia into stable branches (Grabrucker et al., 2014; Leondaritis and Eickholt, 2015). Homer and Shank at the PSD are required for dendrite maturation via spine head growth (Sala et al., 2001). Synaptic maturation reduces synaptic filopodia that are formed during synapse formation as they facilitate the contact between two neurons (Shen and Cowan, 2010; Proepper et al., 2011). At synapses, Homer and Shank influence the actin cytoskeleton directly (as reviewed in section 1.1.5.2) and it is not clear, if their nuclear translocation has impact on filopodia formation in animals. There is however another protein at the postsynaptic density (Abelson-interacting protein 1 – Abi-1) that has a dual function at the PSD and in the nucleus (Innocenti et al., 2005; Proepper et al., 2007, 2011). At PSDs it is part of a complex that regulates the actin cytoskeleton (Innocenti et al., 2005). Synaptic stimulation leads to its translocation to the

nucleus (Proepper et al., 2007), where it interacts with an hnRNP, synergistically regulating the balance between filopodia formation and synaptic maturation (Proepper et al., 2011). Considering all these examples from animals together with the findings that Homer and Shank localise to the nucleus in choanoflagellates and animals (Burkhardt et al., 2014 and unpublished data by Dr Pawel Burkhardt; Grabrucker et al., 2014), it is possible that an ancient function of Homer and Shank was the clustering of nuclear complexes regulating cellular processes such as the formation of filopodia. Both the presence of postsynaptic scaffolding proteins as well as the animal-like mechanism of filopodia formation are restricted to holozoans (Alié and Manuel, 2010; Sebé-Pedrós et al., 2013a; Burkhardt et al., 2014; Burkhardt and Sprecher, 2017). Holozoans, such as choanoflagellates, filastereans, and animals are further capable of at least temporal cell differentiation, the establishment or maintenance of cell-cell contact in colonies (that are formed through either aggregation – as in *C. owczarzaki* – or incomplete cell division – as in *S. rosetta*) or in multicellular context in animals (Dayel et al., 2011; Sebé-Pedrós et al., 2013b, 2017). These features might be supported through the capacity to regulate the formation of filopodia in an animal-like fashion. The co-occurrence of this mechanism with the presence of postsynaptic scaffolding proteins suggests that these proteins could be involved in this mechanism, either through regulation in the nucleus, or directly at the plasma membrane, or both, but this requires future functional investigations.

Alié and Manuel (2010) further hypothesised that, ancestrally, Dlg formed a scaffold that could recruit and link cation channels in the plasma membrane as well as regulate the microtubule cytoskeleton (Alié and Manuel, 2010). In my data set of putative *S. rosetta* Dlg interaction partners I could not identify ion channels, but I did identify one V-type ATPase and many transmembrane proteins of unknown function. My immunostaining suggested, that just like the other scaffolding proteins investigated, *S. rosetta* Dlg localises to the nucleus. I did not find any evidence that Dlg interacts with Shank in the nucleus, as I identified neither Shank nor the known adapter protein GKAP in the *S. rosetta* Dlg complex. In vertebrates, MAGUKS (including MAGI-1, ZO-1, CASK and Dlg proteins) were found to shuttle between the cytoplasm and the nucleus (Gottardi et al., 1996; Dobrosotskaya et al., 1997; Hsueh et al., 2000; Kohu et al., 2002). The

localisation of ZO-1 was shown to differ depending on the state of cell contact in epithelial cells (Gottardi et al., 1996). Nuclear CASK interacts with the transcription factor Tbr-1, a T-box transcription factor, suggesting regulatory function of this MAGUK in the nucleus (Hsueh et al., 2000). Its nuclear localisation is reduced in presence of its synaptic binding partner syndecan-3 (Hsueh et al., 2000). My data suggest that the regulating function of MAGUKs is ancestral. The occasional staining of the plasma membrane at the basal side of *S. rosetta* cells with the Dlg antibody in immunoprecipitation experiments, suggests that there might be a Dlg complex closely associated with this region. This region is the place of cell-cell contact in rosette colonies as well as the location of theca secretion (Dayel et al., 2011). Harden et al. (2016) reviewed the link between junctional proteins and the cytoskeleton via MAGUKs. If the MAGUK-microtubule interaction would be ancestral, this would support the hypothesis of Alié and Manuel (2010). In my Dlg co-immunoprecipitation experiment, I detected actin, as well as a dynein motor protein (known to move along microtubules in inward direction towards the nucleus (Schroer and Steuer, 1989). Both proteins occurred in low amounts also in the negative control, and actin is a known contaminant in co-IP experiments (Norberg et al., 1982). Therefore, at this stage, it is difficult to say, if *S. rosetta* Dlg really interacts with cytoskeleton elements. Interestingly, in my survey, I did not detect MAGUKs with canonical Dlg and p55 domain architecture in the transcriptomes of acanthoecid species. Colonies with cell-cell contact and the production of an organic theca are features that have only been described in craspedid species (Carr et al., 2008; Nitsche et al., 2011; Leadbeater, 2015). As mentioned above, missing data in transcriptomes does not necessarily mean that the gene is not present in these species. Nevertheless, taking this observation and the putative localisation of Dlg to the plasma membrane on the basal side of the cell together, I hypothesise that these MAGUKs could be involved in the establishment or maintenance of contacts within colonies or to the substratum.

The hypotheses concerning putative ancestral functions of PSD scaffolding proteins are based on their localisation, functions described in vertebrate cells and the comparison of cellular features between animals and choanoflagellates. Other putative functions of these proteins are possible and this will have to be subjected to experimental testing in the future. Another limitation

of my data is that even though the used antibody evidently binds Dlg, I cannot exclude that it also binds another PDZ domain containing protein with similar region to the antigen (PSD-95 alpha) that was identified among the precipitated proteins. There are ongoing efforts to produce another *S. rosetta* Dlg antibody with specificity to another antigen. Binding of the current antibody to PSD-95 alpha and MAGUK p55 will be tested. Moreover, I will attempt to transform *S. rosetta* with a plasmid encoding fluorescently-labelled Dlg using an approach that has been shown to reveal the localisation of proteins in this species (Booth et al., 2018; Wetzel et al., 2018).

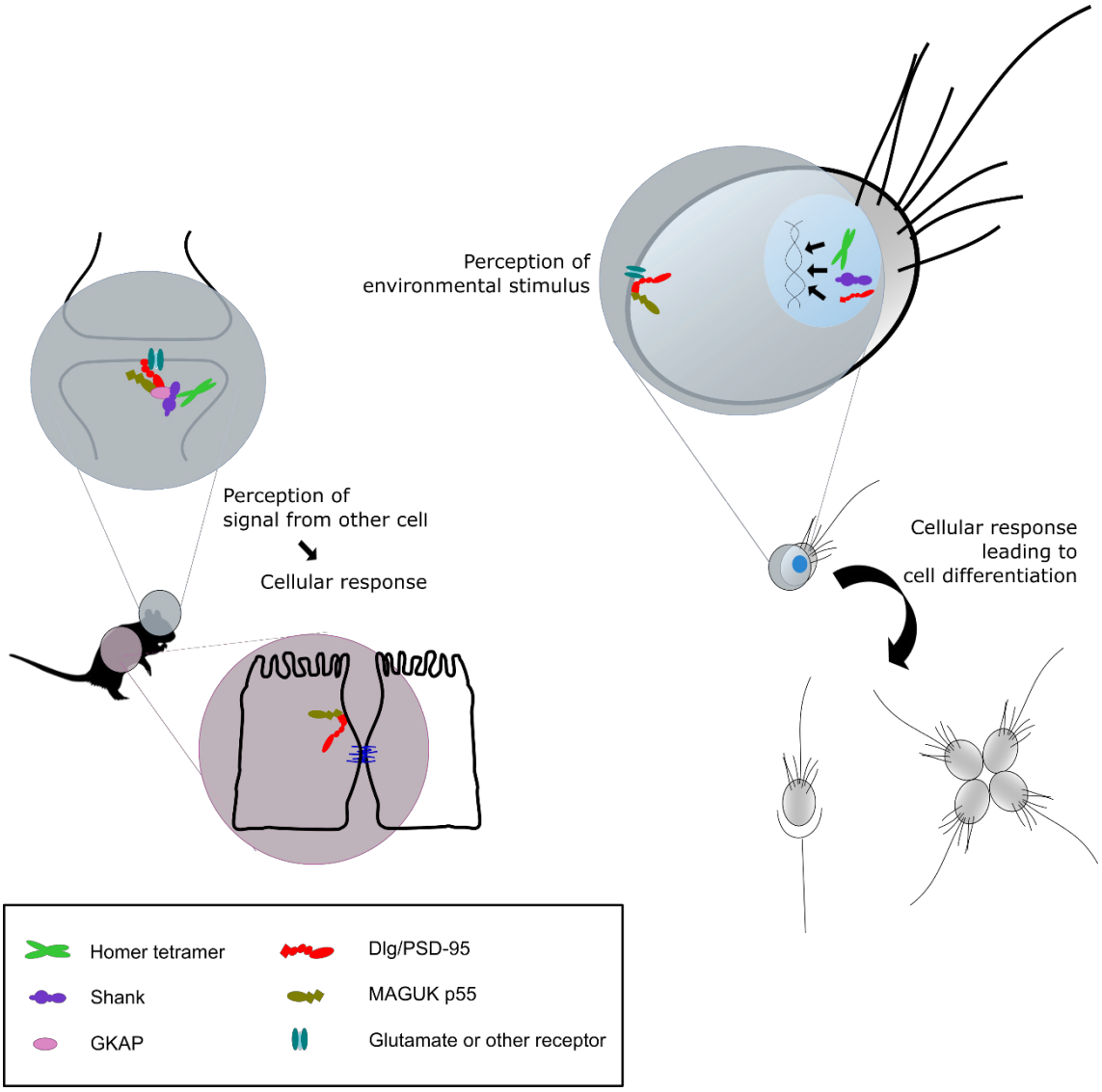


Figure 5-2: Functionality of scaffolding proteins in animals and their hypothetical functions in choanoflagellates. On the left: Scaffolding proteins organise signalling complexes at vertebrate synapses and epithelial tight junctions. They enable cells to perceive signals from other cells and response appropriately. On the right: Putative cellular context of scaffolding proteins in choanoflagellates that might be involved in the cellular response to environmental stimuli leading to cell differentiation. The organisation of scaffolding proteins in regulatory complexes in the nucleus could be partially conserved in animals as well as the Dlg-MAGUK p55 scaffold and the Homer-Shank interaction. Rat picture from phylopic.org.

5.7 Emergence of neuronal cell types and synapses

Choanoflagellates and other close relatives to animals have complex life cycles that are differentially regulated in their distinct life cycle stages (Sebé-Pedrós et al., 2017). This implies that the life cycle stages of these organisms correspond to distinct cell types that are capable of temporal cell differentiation (Sebé-Pedrós et al., 2017). This differentiation is induced by environmental stimuli and cells are not fixed in their cell type (Sebé-Pedrós et al., 2017). Sebé-Pedrós et al. (2017) propose a model (based on previous suggestions by Zakhvatkin, 1949 and Mikhailov et al., 2009) in which the last common ancestor of animals already had distinct gene regulatory programmes for the differentiation into cell types that were activated in this ancestor upon environmental stimulus. Ultrastructural studies of *S. rosetta* rosette colonies showed that some cells are very different in their phenotype than other cells (Naumann and Burkhardt, 2019). This suggests that *S. rosetta* might be capable of such a simple form of spatial cell differentiation (Naumann and Burkhardt, 2019).

My survey revealed the presence of putative ionotropic glutamate receptors, Shaker/Shal-like potassium channels and nitric oxide synthase in some choanoflagellates. Proteins with similarity to animal ionotropic glutamate receptors have been identified in plants, where they react to a broader range of amino acids (Forde and Roberts, 2014). In choanoflagellates and the putative animal ancestor, reception could have helped to react to environmental stimuli. Choanoflagellates require bacteria as their food source (Richter and Nitsche, 2017). Amino acid reception could be a cue for the presence of bacterial metabolites, potentially inducing attachment of choanoflagellates. Previous studies revealed that the choanoflagellate species *Salpingoeca rosetta* (which does not encode a putative ionotropic glutamate receptor) responds to specific species of environmental bacteria, with some species inducing rosette colony formation and other species inducing sexual recombination (Alegado et al., 2012; Woznica et al., 2017). It has not been established how the chemical cues that choanoflagellates react to are received.

Hitherto, there was no evidence that choanoflagellate cells can communicate with each other. Flagella beating in colonies is not synchronised (Kirkegaard et al., 2016) and even a form of contractility that was observed in colonial sheets of cells from the species *Choanoeca flexa*, seems to be induced

by an environmental stimulus (light) that acts on every single cell (Brunet et al., 2019). Cell-cell signalling is implicated in the internally controlled cell differentiation in animals (Perrimon et al., 2012; Sebé-Pedrós et al., 2017). Interactions between cells are furthermore indispensable for the formation of cellular junctions and the emergence of functional epithelial as well as neuronal cell types (Shen and Cowan, 2010; Hoelzle and Svitkina, 2012). I found a putatively ancestral complex in choanoflagellates that is important at cellular junctions. I further identified a putative nitric oxide synthase in some choanoflagellate species, which might be used as a messenger between choanoflagellate cells. Nitric oxide (NO) was suggested to be one of the first biological signalling molecules (Feelisch and Martin, 1995). Nitric oxide synthases (NOS) are found in eukaryotes, prokaryotes and archaea (Santana et al., 2017). In cyanobacteria and plants, stressors (such as reactive oxygen species and UV radiation) affect NO signalling, initiating cellular responses such as chlorophyll protection (Beligni and Lamattina, 1999; Chen et al., 2003). NO also seems to be important for plant-bacteria interactions (Santana et al., 2017). At vertebrate postsynapses, NOS produces NO, which acts on neighbouring synaptic terminals (Prast and Philippu, 2001).

Neuronal cell types and synapses presumably emerged in the animal lineage. As established by prior studies prerequisites for this evolutionary event that might have preceded the origin of animals are the capability to differentiate into distinct cell types (Sebé-Pedrós et al., 2017) (not including neuronal cell types or functional epithelial cell types), the presence of many genes with functional importance at synapses (Alié and Manuel, 2010; Burkhardt et al., 2014), as well as conserved protein interactions for secretion (Burkhardt et al., 2011). I here suggest that in addition to that, a complex existed that became important for the formation of different cellular junctions. It remains to be investigated if this complex was already involved in the linkage of cells or if it was integrated into this functional context in the animal lineage. Assuming that both contractility as well as mechanisms for secretion preceded the origin of animals (Burkhardt et al., 2011; Brunet et al., 2019), both hypothetical scenarios of the emergence of neurons (from myoepithelia and from secretory cells in the epithelium – as introduced in section 1.2.2) or even a combination of both scenarios are possible.

5.8 Common molecular foundations could be exploited in nervous system origin event(s)

In my studies of postsynaptic signalling machineries, I focussed on complexes with importance at glutamatergic synapses. This was done because this type of synapse is extensively studied (Reiner and Levitz, 2018). Furthermore, it was suggested that glutamate receptors are present in all animal lineages with nervous system (Kass-Simon and Pierobon, 2007; Moroz et al., 2014; Burkhardt and Sprecher, 2017). Even sponges have glutamate receptors and show a reaction to glutamate (Nickel, 2010; Riesgo et al., 2014). There are other types of synapses that were suggested to be ancestral (such as peptidergic synapses due to their high prevalence in Cnidaria and Ctenophora) (Grimmelikhuijzen and Westfall, 1995; Moroz and Kohn, 2016; Kristan, Jr., 2016) and require further investigation in order to study their ancestry.

Focussing here only on glutamatergic synapses, I could identify more proteins with statistically significant sequence similarity to genes with functional importance at this type of synapses. In line with other surveys (Alié and Manuel, 2010; Burkhardt et al., 2014), this suggests that many of the components required for the emergence of this type of synapse preceded animal origins. On a molecular level, this argues for a single origin of nervous systems, although the bias created by the comparison to the well-studied bilaterian signalling machineries, needs to be appreciated (Dunn et al., 2015; Liebeskind et al., 2016). This approach assumes that synapses generally work as they do in Bilateria, which is probably not accurate, as the bilaterian complexes are also derived from the ancestral complex. More protein-level studies in non-bilaterian animals are of major importance to get a more comprehensive view on synaptic signalling machineries.

In principle, it would be possible that there are multiple origins of nervous systems, even if many proteins utilised for synapses in general turn out to be conserved. This depends on the interpretation of the concept of the nervous system origin. Considered on the level of protein complex evolution and the addition of key proteins for synaptic functions as well as the evolution of cell differentiation programmes and the interaction of cells within multicellular organisms, this remains a possibility.

5.9 Future recommendation

In addition to the experiments suggested in the individual sections to strengthen the evidence to my findings, I recommend future research to focus on the following tasks in order to get a more comprehensive view on the origin of postsynaptic signalling machineries:

- Investigations of postsynaptic protein complexes in non-bilaterian animals. This will help to avoid a bias assuming that all synapses are bilaterian-like.
- More co-immunoprecipitation experiments on a) other life cycle stages of *S. rosetta* and b) other proteins with statistically significant sequence similarity to PSD scaffolding proteins (such as Shank) could reveal more conserved interactions.
- Finally, functional studies are necessary in order to test hypotheses surrounding putative functions of scaffolding proteins in close relatives to animals. This might elucidate shared functions of these proteins between animals and choanoflagellates that might have ancestral origin. The generation and analysis of *S. rosetta* mutants with defects in filopodia formation, cell-cell contact and theca formation might reveal if scaffolding proteins are involved in these processes. As genetic manipulation was recently established for *S. rosetta* (Booth and King, 2020), protein function could also be tested in a more direct way, knocking out *S. rosetta* Dlg and observing phenotypic changes orchestrated by that. The function of choanoflagellate-unique MAGUKs could additionally be investigated in the acanthoecid species *Diaphanoeca grandis*, where the gene could be silenced through a method established for this species (Li et al., 2018 preprint).

5.10 Conclusion

In this thesis, I collected evidence illustrating the ancestral origin of postsynaptic proteins and some of their key interactions. Using a variety of approaches helped me to get a better understanding for the evolutionary origin of these interactions. I found proteins with statistically significant sequence similarity to postsynaptic scaffolding proteins in all sequenced choanoflagellate transcriptomes. I established that the capacity of Homer to bind Shank presumably preceded the evolution of animals and choanoflagellates. Moreover, I found evidence for an ancestral scaffold between Dlg and a MAGUK p55 protein as this interaction seems to be conserved in choanoflagellates. This interaction occurs at animal tight junctions and at the postsynapse, demonstrating that the structural scaffold for junctional signalling machineries might have its origin in a complex that preceded first animals. This leads me to conclude:

Not only were many postsynaptic proteins present before the evolution of animals, it seems like many interactions between these proteins are also conserved. Ancestral functions remain to be resolved, but the foundations for the emergence of synapses seem to have been laid long before the evolution of neurons.

6 Appendix

6.1 Supplementary material for chapter 2

Table 6-1: Querys for BLASTp searches in choanoflagellate protein databases.

Protein	Query sequences
Shank	*Homo sapiens* Shank 2 (NP_036441.2)
	Monosiga brevicollis Shank (XP_001748613.1) full-length (aa 1-1438) or only SH3-SAM region (aa 1220 – 1438)
Homer	*Homo sapiens* Homer 1 (NP_004263.1)
MAGUKs	*Homo sapiens* PSD-95 (NP_001308004.1)
	Trichoplax adhaerens Dlg (XP_002108800.1)
	Capsaspora owczarzaki Dlg (CAOG_01229)
	Salpingoeca rosetta Dlg (cDNA translated sequence shown in supplement Fig 6.13)
mGluR	*Homo sapiens* mGluR1 (AAB05337.1)
	Homo sapiens mGluR4 (NP_000832.1)
	Drosophila melanogaster mGluR (NP_001259076.1)
iGluR	*Homo sapiens* NMDAR (NP_000823.1)
	Drosophila melanogaster NMDAR (NP_730940.1)
	Nematostella vectensis NMDAR (XP_001627645.1)

Table 6-2: Accession numbers of all proteins with statistically significant sequence similarity and with comparable domain architecture to postsynaptic proteins identified. NCBI accession numbers were provided for every protein accessible there. Other accession numbers provided were numbers used in the OrthoMCL analysis by Richter et al. (2018) and correspond to accession numbers of different databases. 19 choanoflagellate transcriptome based protein databases: https://dx.doi.org/10.6084/m9.figshare.5686984.v2. Dataset 2. *Oscarella pearsei* peptide translated transcriptome: compagen.org. *Mnemiopsis leidyi* peptide translated genome https://research.nhgri.nih.gov/mnemiopsis/genes/genewiki.cgi. *Nematostella vectensis* and *Trichoplax adhaerens* peptide translated genomes were accessed at JGI (https://mycocosm.jgi.doe.gov/). The Drosophila melanogaster accession number refers to the flybase.org. Approximated gene families according to OrthoMCL clusters by Richter et al. (2018) for the different proteins are listed as well. For *Stephanoeca diplocostata*, if not specified

differently, the French isolate was investigated as this isolate was included in the OrthoMCL clustering.

Protein	Species	Gene family	Accession number
Dlg/PSD-95	*Capsaspora owkzarzaki*	670	CAOG_01229
	Salpingoeca rosetta	670	PTSG_01141
	Microstomoeca roanoka	670	m.56385
	Hartaetosiga balthica	670	m.15905
	Hartaetosiga gracilis	670	m.62679
	Salpingoeca infusionum	670	m.27203
	Monosiga brevicollis	670	209 (XP_001743865.1)
	Choanoeca perplexa	670	m.132977
	Salpingoeca kvevrii	670	m.76739
	Salpingoeca urceolata	670	m.333787; m.333860
	Salpingoeca macrocollata	670	m.98347
	Salpingoeca punica	670	m.201747
	Salpingoeca helianthica	670	m.61369
	Mylnosiga fluctuans	670	m.242855
	Codosiga hollandica	670	m.548917
	Didymoeca costata	670	m.342288
	Oscarella pearsei	670	m.306928
	Mnemiopsis leidyi	670	ML02777
	Trichoplax adhaerens	670	63520 (XP_002108800.1)
	Drosophila melanogaster	670	FBpp0089351 (NP_996406.1)
	Mus musculus	670	6681195 (NP_031890.1)
MAGUK p55	*Capsaspora owkzarzaki*	6285	CAOG_08723
	Salpingoeca rosetta	6285	PTSG_09863
	Microstomoeca roanoka	6285	m.41930
	Hartaetosiga balthica	6285	m.37757
	Hartaetosiga gracilis	6285	m.221252; m.221265
	Salpingoeca infusionum	7142	m.269432
	Monosiga brevicollis	6285	32262 (XP_001745620.1)
	Choanoeca perplexa	6285	m.253058
	Salpingoeca kvevrii	6285	m.299993
	Salpingoeca urceolata	6285	m.82242
	Salpingoeca macrocollata	6285	m.157751
	Salpingoeca punica	6285	m.123077
	Salpingoeca helianthica	6285	m.87043
	Mylnosiga fluctuans	6285	m.17003
	Codosiga hollandica	6285	m.89692; m.89717
	Salpingoeca dolichothecata	6285	m.72552
	Oscarella pearsei	6285	m.5868
		7087	m.307869
		7142	m.5379
	Mnemiopsis leidyi	6285	ML096814
		7087	ML120734; ML35887
	Trichoplax adhaerens	6285	54413 (XP_002110426.1)
		7087	63748 (XP_002110949.1)
		7142	52734 (XP_002109073.1)
	Nematostella vectensis	7087	116416 (XP_032233802.1)
		7142	173637 (EDO33577.1)
	Drosophila melanogaster	6285	FBpp0087292
		7078	(NP_610642.2)
		7142	FBpp0088886 (AAN11089.3)
			FBpp0305870 (AAF46351.2)
	Mus musculus	6285	239051602
		7078	(NP_001155092.1)
		7142	7710062 (NP_057904)
			9625023 (NP_001074756.2)
Choanoflagellate	*Salpingoeca rosetta*	22778	PTSG_02659

MAGUK		7659	PTSG_12450
	Microstomoeca roanoka	7659	m.244150
		22778	m.236465
	Hartaetosiga balthica	7659	m.121362
	Hartaetosiga gracilis	7659	m.220164
	Salpingoeca infusionum	7659	m.263278
		22778	m.103781
	Monosiga brevicollis	22778	35505 (NCBI XP_001742640.1)
	Choanoeca perplexa	7659	m.261367
		22778	268729
	Salpingoeca kvevrii	7659	m.163969
	Salpingoeca urceolata	7659	m.66462
	Salpingoeca macrocollata	7659	m.160592
	Salpingoeca punica	7659	m.44694
	Codosiga hollandica	7659	m.912948
	Helgoeca nana	47837	m.268630
	Savillea parva	47837	m.17495
	Diaphanoeca grandis	7659	m.230852
	Didymoeca costata	73131	m.58669
	*Stephanoeca diplocostata*_AU		m.1350202
Homer	*Capsaspora owkzarzaki*	2756	CAOG_02708
	Salpingoeca rosetta	2756	PTSG_00432
	Microstomoeca roanoka	2756	m.208464; m.208450; m.208456
	Hartaetosiga balthica	2756	m.137955
	Hartaetosiga gracilis	2756	m.12338
	Salpingoeca infusionum	2756	m.8158
	Monosiga brevicollis	2756	14345 (EDQ91787.1)
	Choanoeca perplexa	2756	m.166648
	Salpingoeca kvevrii	2756	m.41574; m.41579
	Salpingoeca urceolata	2756	m.50012
	Salpingoeca macrocollata	2756	m.48677; m.48686
	Salpingoeca punica	2756	m.234757; m.30343
	Salpingoeca helianthica	2756	m.10913
	Mylnosiga fluctuans	2756	m.51570; m.51617
	Codosiga hollandica	2756	m.142293; m.142300
	Salpingoeca dolichothecata	2756	m.23193; m.23192; m.8487
	Acanthoeca spectabilis	2756	m.56065
	Helgoeca nana	2756	m.414893
	Savillea parva	2756	m.195488
	Diaphanoeca grandis	2756	m.67103; m.54127
	Didymoeca costata	2756	m.340273
	Stephanoeca diplocostata	2756	m.58182
	Oscarella pearsei	2756	m.8451
	Mnemiopsis leidyi	2756	ML06361
	Trichoplax adhaerens	2756	22944 (XP_002111045.1)
	Nematostella vectensis	2756	91920 (EDO45622.1)
	Drosophila melanogaster	2756	FBpp0302927 (NP_477396.1)
	Mus musculus	2756	6754224 (NP_036112.1)
Shank	*Capsaspora owkzarzaki*		XP_004364328.1
	Salpingoeca rosetta		XP_004998206.1
	Microstomoeca roanoka		m.36562
	Hartaetosiga balthica		m.42726
	Hartaetosiga gracilis		m.79922; m.79930
	Salpingoeca infusionum		m.305885
	Monosiga brevicollis		XP_001748613.1
	Choanoeca perplexa		m.158209
	Salpingoeca kvevrii		m.158234
	Salpingoeca urceolata		m.143926

	Salpingoeca macrocollata		m.185115; m.72510
	Salpingoeca punica		m.24131
	Salpingoeca helianthica		m.84573
	Mylnosiga fluctuans		m.145719; m.145732
	Codosiga hollandica		m.654371; m.654456
	Salpingoeca dolichothecata		m.150520; m.8517
	Acanthoeca spectabilis		m.44830
	Helgoeca nana		m.181527
	Savillea parva		m.228055; m.234857; m.194431
	Diaphanoeca grandis		m.159568; m.159556
	Didymoeca costata		m.38810; m.164364
	Stephanoeca diplocostata		m.962348
	Oscarella pearsei		m.126548
	Nematostella vectensis		XP_032239996.1
	Drosophila melanogaster		NP_610925.3
	Mus musculus		NP_001029287.1
GKAP (DLGAP1-4)	*Capsaspora owkzarzaki*	10111	CAOG_07783
	Salpingoeca kvevrii	28096	m.53536
	Salpingoeca urceolata	10111	m.21322
		16408	m.295828
	Salpingoeca macrocollata	22194	m.295183; m.295214
		10111	m.46887
	Salpingoeca punica	10111	m.131774
		22194	m.29793
	Mylnosiga fluctuans	10111	m.36351; m.36358
		22194	m.5018
	Codosiga hollandica	10111	m.80382
	Salpingoeca dolichothecata	22194	m.295784
		10111	m.91392
	Oscarella pearsei	16408	m.18178
	Mnemiopsis leidyi	10111	ML03326
	Trichoplax adhaerens	10111	57926
	Nematostella vectensis	16408	241923
	Drosophila melanogaster	16408	FBpp0085375 (NP001163056.1)
	Mus musculus	28096	109891942 (NP_001035953.1)
GKAP (DLGAP5)	*Salpingoeca rosetta*	109910	PTSG_05865
	Microstomoeca roanoka	109910	m.99149
	Hartaetosiga balthica	36869	m.138658
	Hartaetosiga gracilis	36869	m.195645
	Monosiga brevicollis	15861	32557 (EDQ89126.1)
	Choanoeca perplexa	12123	m.296378
	Salpingoeca kvevrii	15861	m.3571
	Salpingoeca urceolata	15861	m.63161
	Salpingoeca helianthica	22194	m.5662
	Codosiga hollandica	15861	m.386583
	Acanthoeca spectabilis	15861	m.452231
	Helgoeca nana	15861	m.65558
	Savillea parva	15861	m.10891
	Diaphanoeca grandis	15861	m.112158
	Didymoeca costata	15861	m.340703
	Stephanoeca diplocostata	15861	m.793754
	Oscarella pearsei	12123	m.309954
	Mnemiopsis leidyi	None	ML143015; ML35061
	Trichoplax adhaerens	12123	60595
	Nematostella vectensis	12123	214316
	Drosophila melanogaster	73778	FBpp0086788 (NP_01163142.1)
	Mus musculus	12123	225543150 (NP_653136.2)
Lin-7	*Capsaspora owkzarzaki*	7369	CAOG_07087

	Oscarella pearsei	7369	m.14113
	Mnemiopsis leidyi	7369	ML05292
	Trichoplax adhaerens	7369	61629 (XP_002117602.1)
	Nematostella vectensis	7369	160186 (XP_001639785.1)
	Drosophila melanogaster	7369	FBpp0084162 (NP_651330.1)
	Mus musculus	7369	6755973 (NP_035829.1)
iGluR	*Choanoeca perplexa*	9908	m.166147
		95	m.99980
	Salpingoeca urceolata	95	m.271231
	Diaphanoeca grandis	None	m.117958
		95	m.4410
		32739	m.164754
	Didymoeca costata	19327	m.132649
		39844	m.301599
	Stephanoeca diplocostata	9908	m.1633213
	Oscarella pearsei	95	m.1834
	Mnemiopsis leidyi	95	ML016317
	Trichoplax adhaerens	95	55165 (XP_002111317.1)
	Nematostella vectensis	95	239847 (EDO46633.1)
		9908	171792 (EDO35545.1)
	Drosophila melanogaster	9908	FBpp0078410
		95	(NP_730940.1) FBpp0076691 (NP_476855.1)
	Mus musculus	95	124487364
		9908	(NP_001074566.1) 294997257 (NP_001171128.1)
mGluR	*Oscarella pearsei*		m.309622
	Mnemiopsis leidyi		ML04045
	Trichoplax adaherens		XP_002112665.1
	Nematostella vectensis		XP_032221914.1
	Drosophila melanogaster		NP_524639.2
	Mus musculus		XP_006512612.1
IP3R	*Capsaspora owkzarzaki*	552	CAOG_04510
		591	CAOG_04826
	Salpingoeca rosetta	552	PTSG_06996
		591	PTSG_07655
	Microstomoeca roanoka	11602	m.196625
		552	m.213666; m.220769
	Hartaetosiga balthica	591	m.129598
		552	m.53014
	Hartaetosiga gracilis	591	m.110530
		552	m.153526
	Salpingoeca infusionum	552	m.193039; m.357042
		591	m.231617
	Monosiga brevicollis	552	31670 (XP_001747685.1)
		591	37956 (XP_001747685.1)
	Choanoeca perplexa	591	m.178213
		552	m.267230; m.276491
	Salpingoeca kvevrii	552	m.139103; m.171596; m.196040; m.196116;
		591	m.196227 m.172928
	Salpingoeca urceolata	552	m.20572
	Salpingoeca macrocollata	552	m.113029; m.236300
	Salpingoeca punica	552	m.168492
	Salpingoeca helianthica	552	m.72766
	Mylnosiga fluctuans	552	m.20810; m.77250
		591	m.252205
	Codosiga hollandica	552	m.710944; m.885791

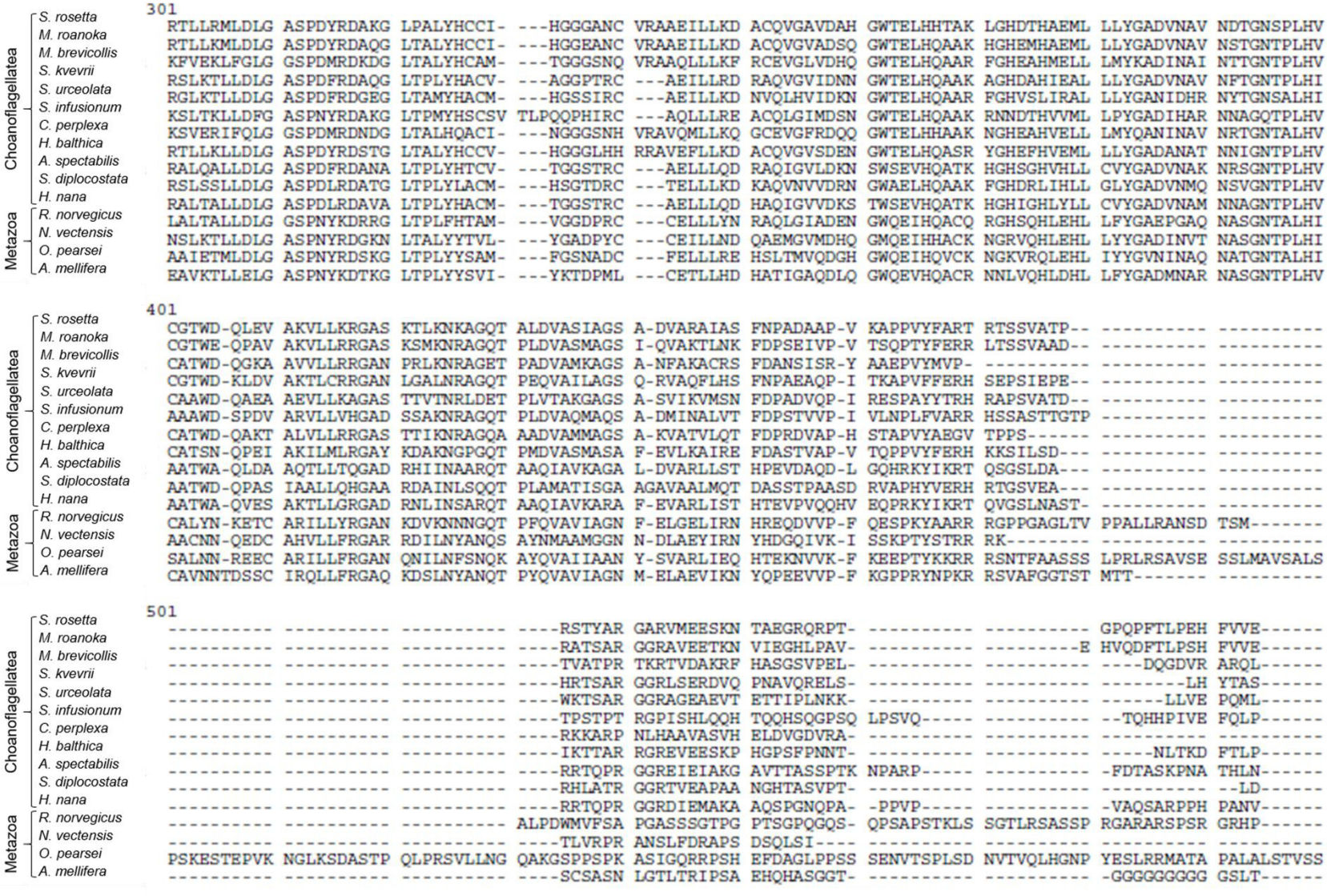
	Salpingoeca dolichothecata	552	m.136903; m.161045; m.193054
	Acanthoeca spectabilis	552	m.244930; m.449323;
		591	m.49160
			m.87044
	Helgoeca nana	591	m.275934
		552	m.30504; m.357148
	Diaphanoeca grandis	552	m.22601; m.247043; m.247080; m.247181; m.247298; m.247338; m.247377
	Didymoeca costata	552	m.161015; m.56376
	Stephanoeca diplocostata	591	m.990990
	Oscarella pearsei	552	m.114956
		591	m.117174
	Trichoplax adhaerens	552	55241 (XP_002111348.1)
		591	56365 (XP_002112826.1)
	Nematostella vectensis	591	244707 (EDO37876.1)
	Drosophila melanogaster	591	FBpp0078335 (NP_001287180.1)
	Mus musculus	591	291327470 (NP_034715.3)
K^+ Shaker/Shal	*Salpingoeca helianthica*	3751	m.62049
	Mylnosiga fluctuans	5877	m.105653; m.105697; m.121176; m.121184
		3496	m.31412
	Salpingoeca dolichothecata	5877	m.79705
		3751	m.89349; m.89357
	Mnemiopsis leidyi	5877	ML022314
	Trichoplax adhaerens	3496	1922 (XP_002107932.1)
		3751	3098 (XP_002114266.1)
	Nematostella vectensis	3496	21646 (XP_001634787.2)
		5877	135889 (EDO31803.1)
		3751	85267 (XP_001639934.2)
	Drosophila melanogaster	5877	FBpp0074741
		3496	(NP_524159.1)
		3751	FBpp0088331 (NP_728783.1) FBpp0293731 (NP_476721.1)
	Mus musculus	3496	22122333 (NP_666034.1)
		3751	31560819 (NP_032446.2)
		5877	6680526 (NP_032449.1)
CAMKII	*Capsaspora owkzarzaki*	41	CAOG_07915
	Salpingoeca rosetta	41	PTSG_10090
	Microstomoeca roanoka	41	m.56031
	Hartaetosiga balthica	41	m.12778
	Hartaetosiga gracilis	41	m.240565
	Salpingoeca infusionum	41	m.355872
	Monosiga brevicollis	41	32484 (XP_001745863.1)
	Choanoeca perplexa	41	m.25242
	Salpingoeca kvevrii	41	m.186408
	Salpingoeca urceolata	41	m.174574; m.174582
	Salpingoeca macrocollata	41	m.39937
	Salpingoeca punica	41	m.65597
	Salpingoeca helianthica	41	m.227575
	Mylnosiga fluctuans	41	m.232189
	Codosiga hollandica	41	m.920109
	Salpingoeca dolichothecata	41	m.267256
	Acanthoeca spectabilis	41	m.133898
	Helgoeca nana	41	m.26748
	Savillea parva	41	m.63778
	Diaphanoeca grandis	41	m.105205

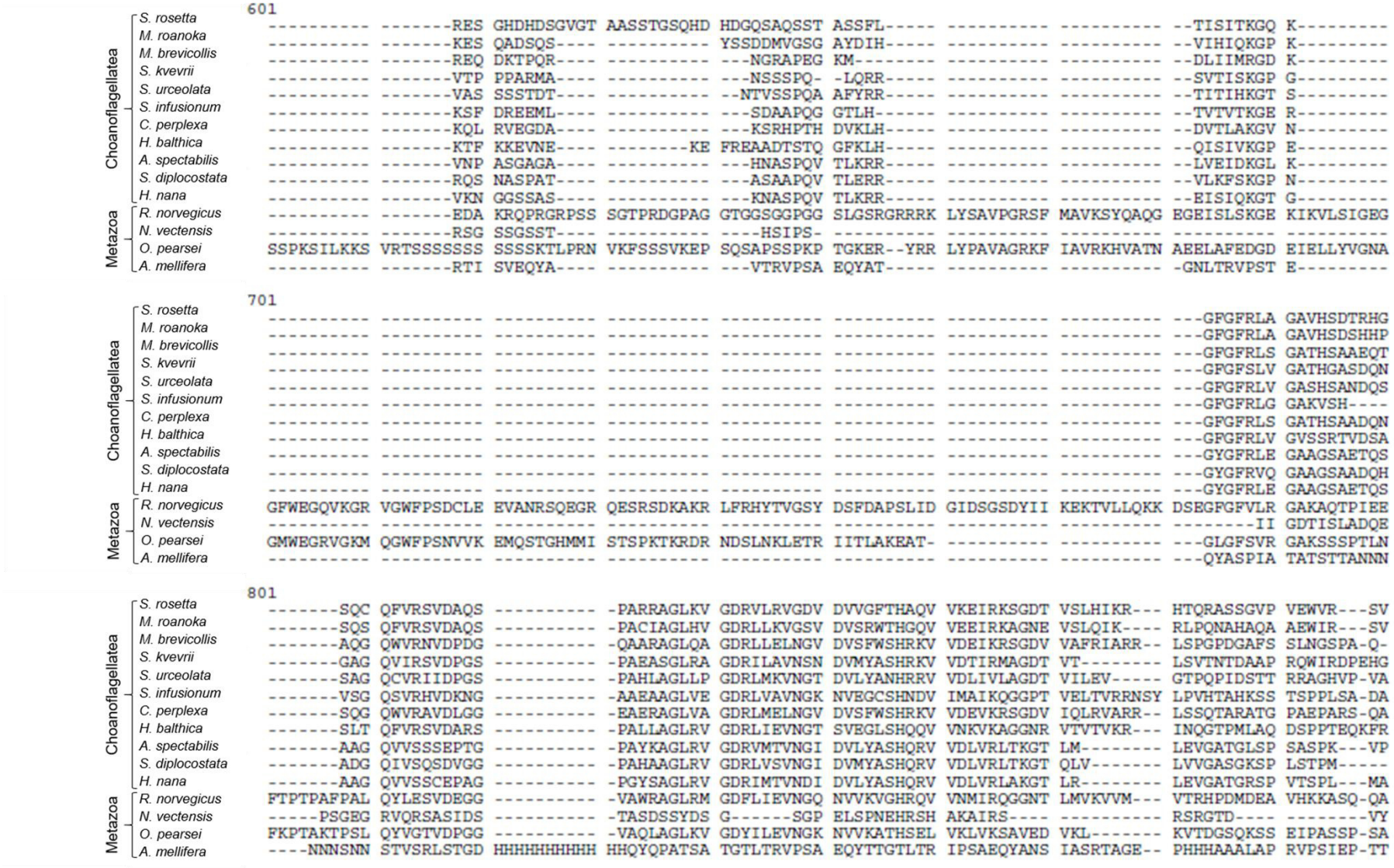
	Stephanoeca diplocostata	41	m.669015
	Oscarella pearsei	41	m.66586
	Mnemiopsis leidyi	41	ML03781; ML35309
	Trichoplax adhaerens	41	50756 (XP_002115697.1)
	Nematostella vectensis	41	157808 (XP_001641961.1)
	Drosophila melanogaster	41	FBpp0289606 (NP_726633.2)
	Mus musculus	41	28916677 (NP_803126.1)
NOS	*Salpingoeca infusionum*	4786	m.15126
	Choanoeca perplexa	4786	m.32057
	Salpingoeca urceolata	4786	m.116692
	Helgoeca nana	4786	m.69061
	Oscarella pearsei	4768	m.6626
	Mnemiopsis leidyi	4768	ML074215
	Trichoplax adhaerens	4768	18893 (XP_002108333.1)
	Nematostella vectensis	4768	110599 (XP_002108333.1)
	Drosophila melanogaster	4768	FBpp0079777 (NP_523541.2)
	Mus musculus	4768	6724321 (NP_032738.1)

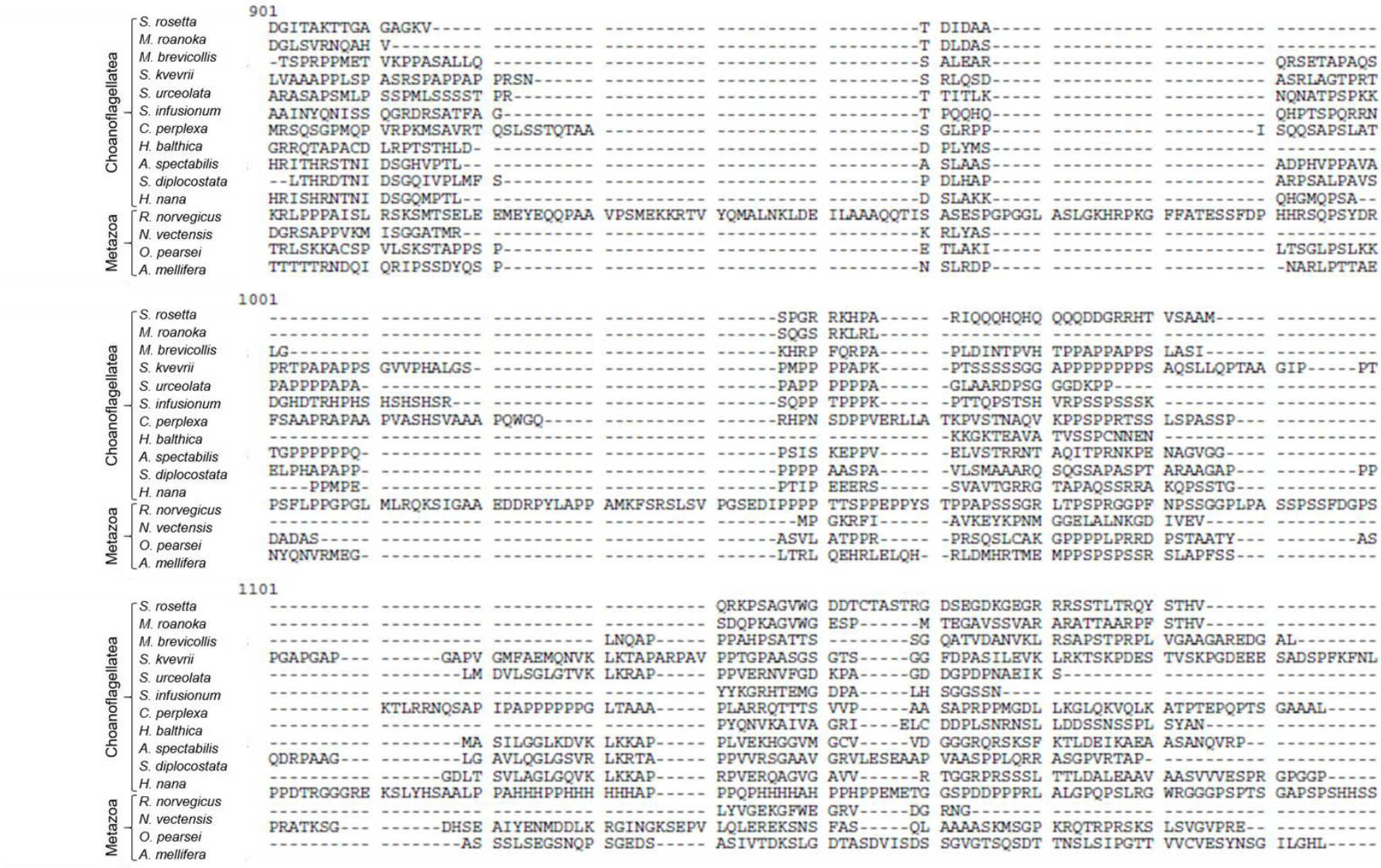
6.2 Supplementary material for chapter 3

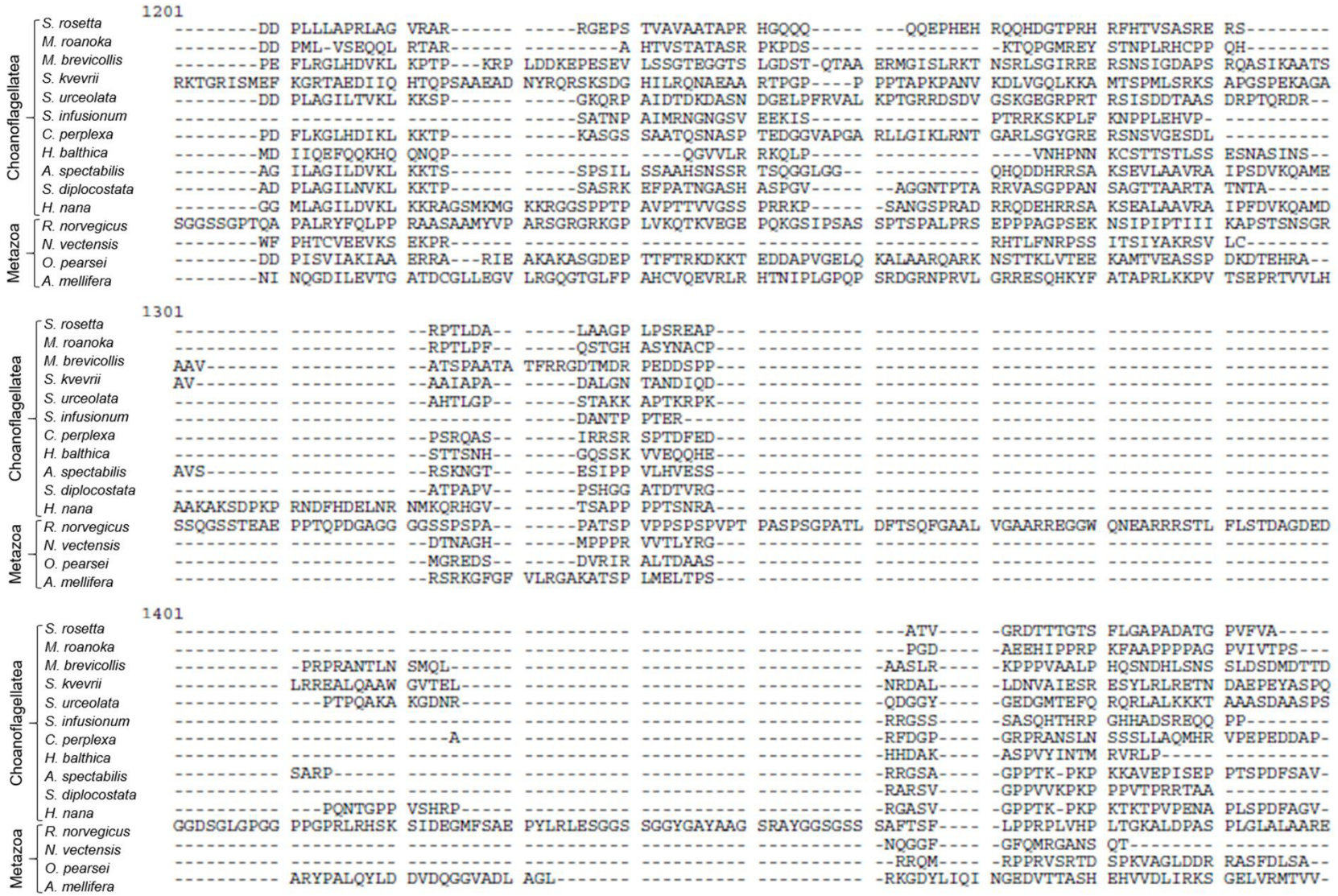
6.2.1 Supplementary figures

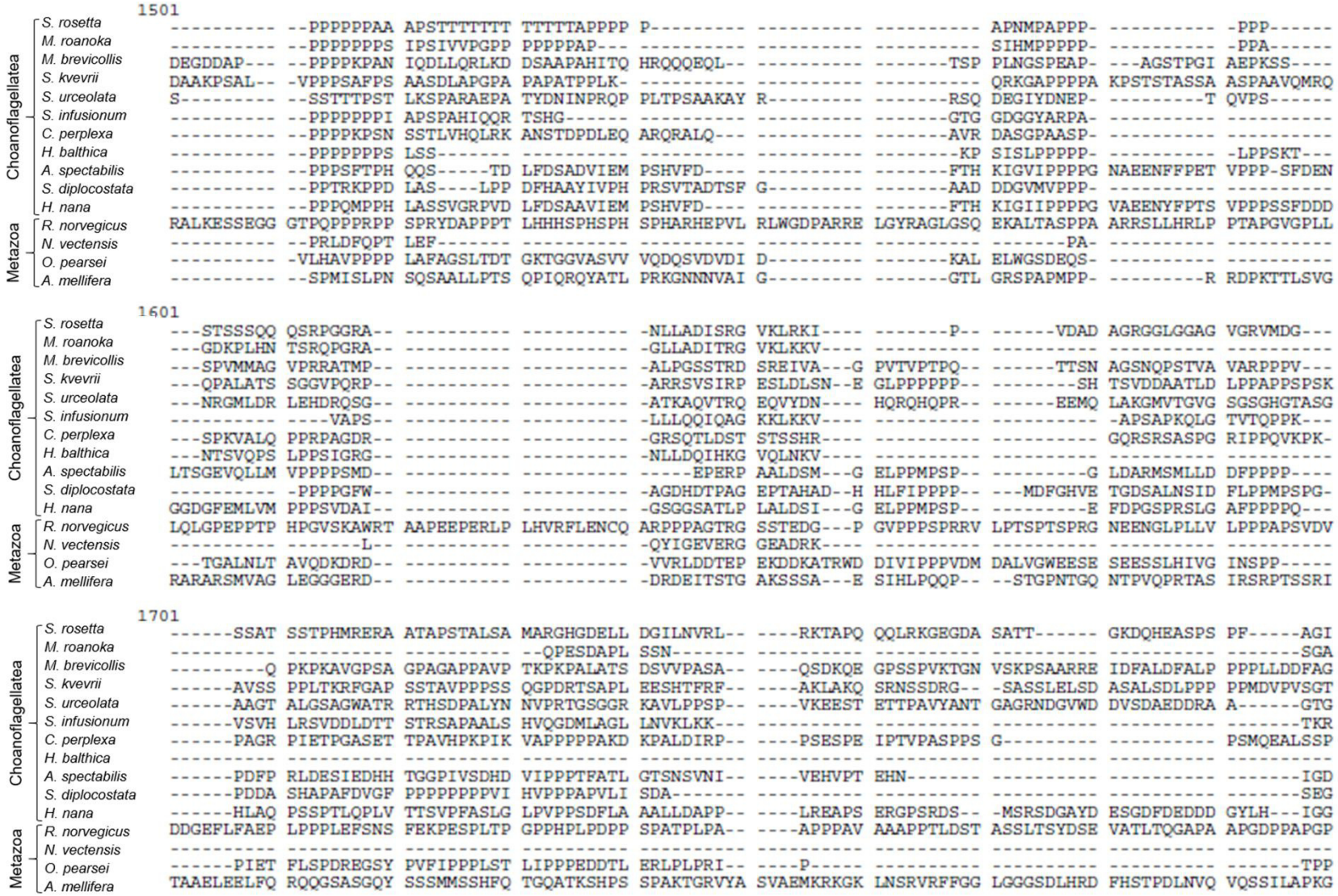

1
S. rosetta ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
M. roanoka ---------- ---------- ---FFLSFCF LFCLPIAVVV LTNMSRKSQP DGGYTGDHMN DLRARSQSYS -QPRSQTDDV TSSPLRPTAA SSSSILVRVS
M. brevicollis ---------- ---------- ---------- ---------- ---------- -------MAQ PQTSSATGSP MRLADAIPQA SASAVAAALT NDDTILCRVS
S. kvevrii ---------- ---------- ---------- ---------- ---------- ---------- ---------- -----SRMSD DSSSVLVRVS
S. urceolata ---------- ---------- ---------- ---------- ---------- ---------- MADPHAMLAR PPFKMHAPAG SGAAAGQEEP ASASILVRVA
S. infusionum ---------- ---------- ---------- ---------- ---------- ----MAHPID MASPRGRAST ISTASPATAS KMSPAG-NAA PDNTLLVRVT
C. perplexa ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---MAQRRAT TATSAAPKGS GGDSVLIRIT
H. balthica ---------- ---------- --FFFFQFFA PSSFPTSIYI TAVVVVTSTF RNRIMSNNTD KIRRNTVASG IRSTKPKEGD EQQPCPMKEK PESSILVRIS
A. spectabilis ---------- ---------- ---------- ---------- ---------- ---------- ----MAAAHF DEADGGVMGF ADDTVDVEPD GADTILTRVC
S. diplocostata ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----MEASAA HQDTILVRVS
H. nana ---------- ---------- ---------- ---------- ---------- ---------- ----MAATQW HDEDDGGIMG FSDTVDAEPQ GADTILTRVC
R. norvegicus ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----MVFRIG
N. vectensis ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
O. pearsei ---------- ---------- ---------- ---------- -------GVP DGRGSRSLPS ISSFHSIQSE NTNPDSSMAS PPALDATGAL RSGSILVRIS
A. mellifera MEAGPKQQQQ QQQQQQQQQQ QQQQQQQQQT QQQQQAQQNS SCGASSGQAA SGQATSPQSG GQPAQGQPQA QQNQDAAAVA ACTTAEAGPV EEGVLLARVH
101
S. rosetta ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- -----MHLVQ FKYRSRVDTA
M. roanoka VVDQDLAKCL LFDKRDTVWT AKQQVLNRLA K--------- ------DLDD AINYGLYLPP SHGRAGKYLD EERILGDYA- -LTGPVAELE FKYRARVDNA
M. brevicollis IPEQELQKCL VFDKTETVWT AKQLVLNKLA KVPFACLSHV PPPRANDIPG NINYGLYLPP ANGRAGKYLS EERLLGDYG- -LAGPVSELQ FKLKRRIYLS
S. kvevrii VVSQNLQKCL MFDQRQTVWT AKQKVLESLA K--------- ------DLLD AINYGLYLPP SQGRAGKFLE EERILSDYA- -LSGPVAELE FKYKQRIYTD
S. urceolata VPEQNLQKML LFDKRQTVWT AKQRVLERLA K--------- ------NLVD AVNYGLYLPP AGGRAGKFLN EERMLGDYA- -LTGPVGYLE FKYKRRVYLT
S. infusionum VVEQSLSKCI MFDTDRTVWE AKQLVLDRLA K--------- ------GLKD AINYGLYQPP KNGRAGKFLD ENRLIGQYG- -LAGPVAELE FKLKRRVYKD
C. perplexa VPDQDLQKCL LFNKKETVWH AKQSVLDRLA Q--------- ------TVPN NINYGLYLPP VNGQAGRYLS EERMLGDYR- -LPGPVAQLQ FKYKRRVYTT
H. balthica VADQNLSKCL LLDQRDTVWT AKQEVLKRLA K--------- ------DLKH AMNYGLYLPA QGGRAGKFLA EERVLGDYA- -LNGPVAELE FKYRRRVYVS
A. spectabilis VSEQGIQKCL LFDQNSTVWT AKQIVLDRLA R--------- ------DLDD KINYGLYLPA SGGRAGKFLA EERLIGDYS- -LQGPVATLE FKYKRRVYLA
S. diplocostata IQEQNVQKCL LFDRTATVWT AKQLVLDKLA K--------- ------DLTD AINYGLYIPP TNGRAGKFLA EERLLGDYA- -LAGPVASLD FKYKRRVYPT
H. nana VSEQGLQKCL LFDKNSTVWT AKQIVLDRLA R--------- ------DLED KINYGLYQPA SGGRAGKFLA EERLIGDYS- -LQGPVATLE FKYKRRVYLA
R. norvegicus IPDLHQTKCL RFNPDATIWT AKQQVLCALS E--------- ------SLQD VLNYGLFQPA TSGRDANFLE EERLLREYPQ SFEKGVPYLE FRYKTRVYKQ
N. vectensis -------KLF LFDDQERIWL AKLQVISELA K--------- ------GVKD ALNYGFYEPP CNGRSGKFLD EERTFREYP- -QRVKPAQLE FKYKRRVYKS
O. pearsei VPEQDLQKCL QFELNESVWT AKENVLTRLA K--------- ------DVGD AMNYGLYLPP ANGRSGKFLE EERRLGEYS- -LKEPIA-LE FKRKKRDYTG
A. mellifera VPELYVSKCL QFPKDQLVWD VKQQCLASLP K--------- ------ELKE SFNYGLFCPP VNGKAGKFLD EERRLGDYP- -FNGPVGYLE LKYKRRVYKM
201
S. rosetta -LSMPEKVHA KLHTKANKRK FFEAVVKNNL NKVEDLLKKG FDPNFHDDDS GETPLTRAAA H--DQ-QDMV VLLFSNGAHF DYRSKDGFTP IHKAAVPGHH
M. roanoka LSGVSKKAHT KLHTKANYKR FYDAVGKGNK ERVEDMLTKG FDPNFQDEES GETPLTRAAL N--DD-LEMA MLLFSNGAHF DYRSKDSFTP IHRAAVAGHH
M. brevicollis -LPTTSKSLA KLHSRANVRK FFDAVVKHNY ERVTNMLVKG IDPNMVDDES GETPLTIAAK G--DD-VDML VILYNGGAHL DFRSRDGLTP LHKAATRGMF
S. kvevrii -APTIEKTRS KLHTRSNLKR FLEYCTKGSL LKVKEQLARG LDPNFHDEET GETPMTFAAA S--DN-LSLI TLLVEGGGHM DFRSRDGLTP IHKAAVASHH
S. urceolata -VPGTEKTRS KLHTRASLRK FFDLVSKEKL SQVKELLARG TDPNFHDESS GETPLTLATA S--EN-VSVL IALVEGGAHM DFRARSTLTP VHTATLNGNH
S. infusionum LVETSTKARA KLHTRANLKK FFDHVVKNNV DRVREQLDKG FDPNMHCNDT GATPLTHAAA M--DQ-VDMV VALVGGGALL DYRSKEGLTA VHTATQNGCL
C. perplexa -PPASEKTMS KLHSRAGLKK FFDAVAKRNY NKVKDMLERG LDPNFQDEES GEAPLTLSAK V--DD-ADML RMLYEGGAHL DFRAKDGATA LHKATVRGGF
H. balthica STSLTPKALT KLHTQSNYKK FFDAVCKGNE QKVKDMLDKG FDPNFQDEES GDTPLTRAAF V--DN-TELA VLLFSHGAHF DFRAKDGFTA IHRAAINGCH
A. spectabilis -PNFTEKMYA KLHSKSALKR FLEQCAKRNV QRVTELIAKG TDPNFHDEDN GETPLTLAVM Q--DH-VELI TVLVEGGGHT DFRAKDGMTP VHKAAIHGMH
S. diplocostata -TLPVDKARV KLHGKSSLKR VCDLISKGGV AKIADMIGKG LDPNFQDEGS GETPLTIAVA H--DR-TEMI TTLVEGGAHM DFRAKDGLTP LHKAAVTGMH
H. nana -PTFTEKTYA KLHTKSSLKR FFELCAKRNV PRVTELIGKG IDPNFHDDDS GETPLTLAVM Q--DH-VELI TVLVEGGGHT DFRAKDGLTP VHKAAVHGMH
R. norvegicus -TNLDEKQLA KLHTKTGLKK FLEYVQLGTS DKVARLLDKG LDPNYHDSDS GETPLTLAAQ T--EGSVEVI RTLCLGGAHI DFRARDGMTA LHKAACARHC
N. vectensis -LQYDAKALQ KVHTKGNLKK LLDAVSLGSI PQVTKLTDKG LDPNFQEDKT GETPLTIAAM G--DQGMDLI MSLVNGGAHL DYRAGDGLTS VHKAAIKGKV
O. pearsei -QNFDEKQLM KINSKVNQKK VMDHLANVNA DKLRKILVKG LDPNFFDEDT GITPLALAIM APEKHLKEII LTLVENGAYL DFRLKDGQTA MHKAATHGKT
A. mellifera -LHLDEKQLK AMHTRTNLRR LLEYVANSQV EKIAKMCSKG LDPNFHCQET GETPLTLATT L--KKPSKVI IALVNGGALL DYRTKEGLTA MHRAVERNSL
Choanoflagellatea
Metazoa

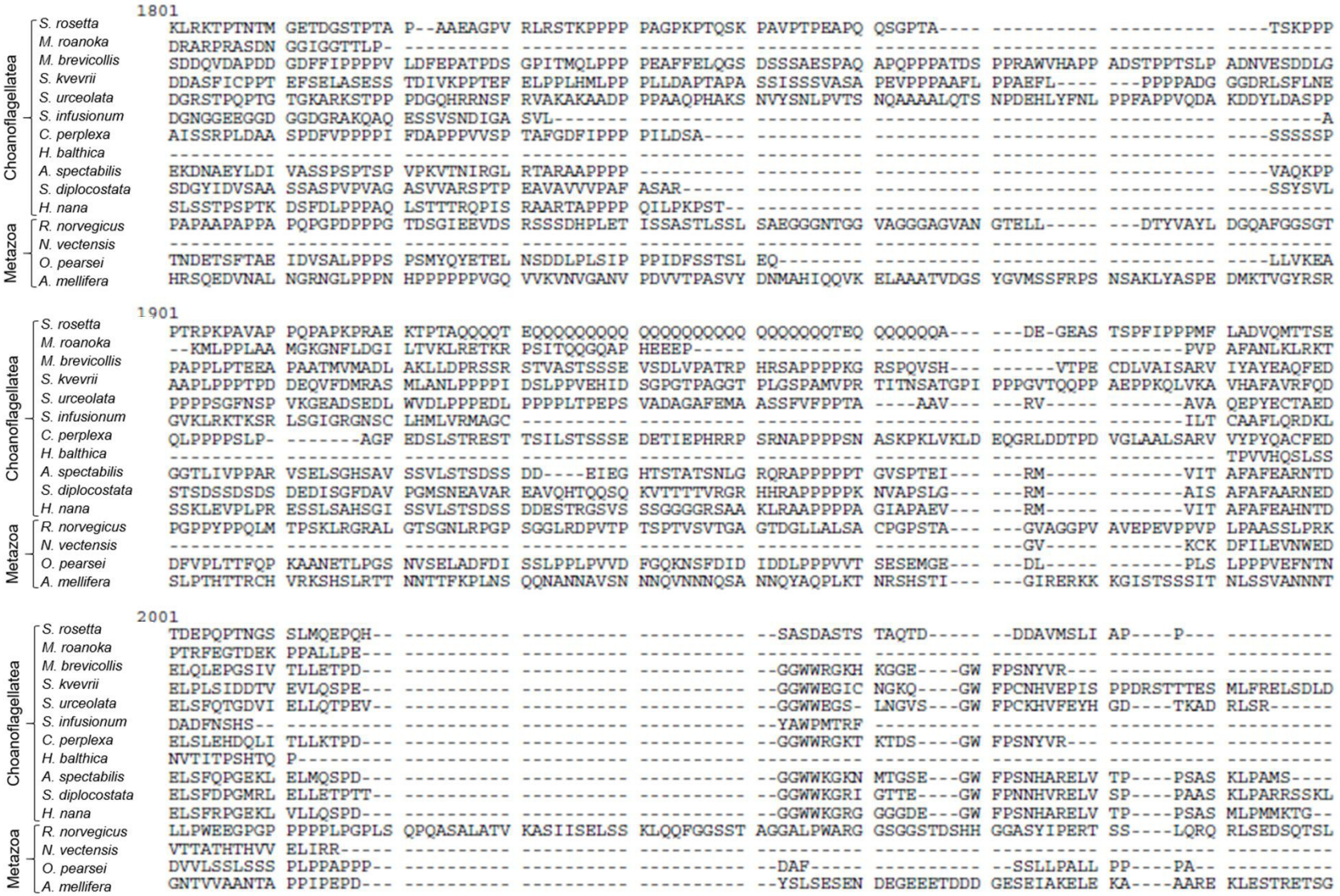

Choanoflagellatea
Metazoa
S. rosetta
M. roanoka
M. brevicollis
S. kvevrii
S. urceolata
S. infusionum
C. perplexa
H. balthica
A. spectabilis
S. diplocostata
H. nana
R. norvegicus
N. vectensis
O. pearsei
A. mellifera
301
RTLLRMLDLG ASPDYRDAKG LPALYHCCI- ---HGGGANC VRAAEILLKD ACQVGAVDAH GWTELHHTAK LGHDTHAEML LLYGADVNAV NDTGNSPLHV
RTLLKMLDLG ASPDYRDAQG LTALYHCCI- ---HGGEANC VRAAEILLKD ACQVGVADSQ GWTELHQAAK HGHEMHAEML LLYGADVNAV NSTGNTPLHV
KFVEKLFGLG GSPDMRDKDG LTALYHCAM- ---TGGGSNQ VRAAQLLLKF RCEVGLVDHQ GWTELHQAAR FGHEAHMELL LMYKADINAI NTTGNTPLHV
RSLKTLLDLG ASPDFRDAQG LTPLYHACV- ---AGGPTRC ---AEILLRD RAQVGVIDNN GWTELHQAAK AGHDAHIEAL LLYGADVNAV NFTGNTPLHI
RGLKTLLDLG ASPDFRDGEG LTAMYHACM- ---HGSSIRC ---AEILLKD NVQLHVIDKN GWTELHQAAR FGHVSLIRAL LLYGANIDHR NYTGNSALHI
KSLTKLLDFG ASPNYRDAKG LTPMYHSCSV TLPQQPHIRC ---AQLLLRE ACQLGIMDSN GWTELHQAAK RNNDTHVVML LPYGADIHAR NNAGQTPLHV
KSVERIFQLG GSPDMRDNDG LTALHQACI- ---NGGGSNH VRAVQMLLKQ GCEVGFRDQQ GWTELHHAAK NGHEAHVELL LMYQANINAV NRTGNTALHV
RTLLKLLDLG ASPDYRDSTG LTALYHCCV- ---HGGGLHH RRAVEFLLKD ACQVGVSDEN GWTELHQASR YGHEFHVEML LLYGADANAT NNIGNTPLHV
RALQALLDLG ASPDFRDANA LTPLYHTCV- ---TGGSTRC ---AELLLQD RAQIGVLDKN SWSEVHQATK HGHSGHVHLL CVYGADVNAK NRSGNTPLHV
RSLSSLLDLG ASPDLRDATG LTPLYLACM- ---HSGTDRC ---TELLLKD KAQVNVVDRN GWAELHQAAK FGHDRLIHLL GLYGADVNMQ NSVGNTPLHV
RALTALLDLG ASPDLRDAVA LTPLYHACM- ---TGGSTRC ---AELLLQD HAQIGVVDKS TWSEVHQATK HGHIGHLYLL CVYGADVNAM NNAGNTPLHV
LALTALLDLG GSPNYKDRRG LTPLFHTAM- ---VGGDPRC ---CELLLYN RAQLGIADEN GWQEIHQACQ RGHSQHLEHL LFYGAEPGAQ NASGNTALHI
NSLKTLLDLG ASPNYRDGKN LTALYYTVL- ---YGADPYC ---CEILLND QAEMGVMDHQ GMQEIHHACK NGRVQHLEHL LYYGADINVT NASGNTPLHI
AAIETMLDLG ASPNYRDSKG LTPLYYSAM- ---FGSNADC ---FELLLRE HSLTMVQDGH GWQEIHQVCK NGKVRQLEHL IYYGVNINAQ NATGNTALHI
EAVKTLLELG ASPNYKDTKG LTPLYYSVI- ---YKTDPML ---CETLLHD HATIGAQDLQ GWQEVHQACR NNLVQHLDHL LFYGADMNAR NASGNTPLHV
401
CGTWD-QLEV AKVLLKRGAS KTLKNKAGQT ALDVASIAGS A-DVARAIAS FNPADAAP-V KAPPVYFART RTSSVATP-- ---------- ----------
CGTWE-QPAV AKVLLRRGAS KSMKNRAGQT PLDVASMAGS I-QVAKTLNK FDPSEIVP-V TSQPTYFERR LTSSVAAD-- ---------- ----------
CATWD-QGKA AVVLLRRGAN PRLKNRAGET PADVAMKAGS A-NFAKACRS FDANSISR-Y AAEPVYMVP- ---------- ----------
CGTWD-KLDV AKTLCRRGAN LGALNRAGQT PEQVAILAGS Q-RVAQFLHS FNPAEAQP-I TKAPVFFERH SEPSIEPE-- ---------- ----------
CAAWD-QAEA AEVLLKAGAS TTVTNRLDET PLVTAKGAGS A-SVIKVMSN FDPADVQP-I RESPAYYTRH RAPSVATD-- ---------- ----------
AAAWD-SPDV ARVLLVHGAD SSAKNRAGQT PLDVAQMAQS A-DMINALVT FDPSTVVP-I VLNPLFVARR HSSASTTGTP ---------- ----------
CATWD-QAKT ALVLLRRGAS TTIKNRAGQA AADVAMMAGS A-KVATVLQT FDPRDVAP-H STAPVYAEGV TPPS------ ---------- ----------
CATSN-QPEI AKILMLRGAY KDAKNGPGQT PMDVASMASA F-EVLKAIRE FDASTVAP-V TQPPVYFERH KKSILSD--- ---------- ----------
AATWA-QLDA AQTLLTQGAD RHIINAARQT AAQIAVKAGA L-DVARLLST HPEVDAQD-L GQHRKYIKRT QSGSLDA--- ---------- ----------
AATWD-QPAS IAALLQHGAA RDAINLSQQT PLAMATISGA AGAVAALMQT DASSTPAASD RVAPHYVERH RTGSVEA--- ---------- ----------
AATWA-QVES AKTLLGRGAD RNLINSARQT AAQIAVKARA F-EVARLIST HTEVPVQQHV EQPRKYIKRT QVGSLNAST- ---------- ----------
CALYN-KETC ARILLYRGAN KDVKNNNGQT PFQVAVIAGN F-ELGELIRN HREQDVVP-F QESPKYAARR RGPPGAGLTV PPALLRANSD TSM-------
AACNN-QEDC AHVLLFRGAR RDILNYANQS AYNMAAMGGN N-DLAEYIRN YHDGQIVK-I SSKPTYSTRR RK-------- ---------- ----------
SALNN-REEC ARILLFRGAN QNILNFSNQK AYQVAIIAAN Y-SVARLIEQ HTEKNVVK-F KEEPTYKKRR RSNTFAASSS LPRLRSAVSE SSLMAVSALS
CAVNNTDSSC IRQLLFRGAQ KDSLNYANQT PYQVAVIAGN M-ELAEVIKN YQPEEVVP-F KGPPRYNPKR RSVAFGGTST MTT------- ----------
501
---------- ---------- ---------- ----RSTYAR GARVMEESKN TAEGRQRPT- ---------- ---------- GPQPFTLPEH FVVE------
---------- ---------- ---------- ----RATSAR GGRAVEETKN VIEGHLPAV- ---------- ----------E HVQDFTLPSH FVVE------
---------- ---------- ---------- ----TVATPR TKRTVDAKRF HASGSVPEL- ---------- ---------- ----DQGDVR ARQL------
---------- ---------- ---------- ----HRTSAR GGRLSERDVQ PNAVQRELS- ---------- ---------- --------LH YTAS------
---------- ---------- ---------- ----WKTSAR GGRAGEAEVT ETTIPLNKK- ---------- ---------- ------LLVE PQML------
---------- ---------- ---------- ----TPSTPT RGPISHLQQH TQQHSQGPSQ LPSVQ----- ---------- --TQHHPIVE FQLP------
---------- ---------- ---------- ----RKKARP NLHAAVASVH ELDVGDVRA- ---------- ---------- ---------- ----------
---------- ---------- ---------- ----IKTTAR RGREVEESKP HGPSFPNNT- ---------- ---------- -----NLTKD FTLP------
---------- ---------- ---------- ----RRTQPR GGREIEIAKG AVTTASSPTK NPARP----- ---------- -FDTASKPNA THLN------
---------- ---------- ---------- ----RHLATR GGRTVEAPAA NGHTASVPT- ---------- ---------- ---------- --LD------
---------- ---------- ---------- ----RRTQPR GGRDIEMAKA AQSPGNQPA- -PPVP----- ---------- -VAQSARPPH PANV------
---------- ---------- ---------- ALPDWMVFSA PGASSSGTPG PTSGPQGQS- QPSAPSTKLS SGTLRSASSP RGARARSPSR GRHP------
---------- ---------- ---------- ----TLVRPR ANSLFDRAPS DSQLSI---- ---------- ---------- ---------- ----------
PSKESTEPVK NGLKSDASTP QLPRSVLLNG QAKGSPPSPK ASIGQRRPSH EFDAGLPPSS SENVTSPLSD NVTVQLHGNP YESLRRMATA PALALSTVSS
---------- ---------- ---------- ----SCSASN LGTLTRIPSA EHQHASGGT- ---------- ---------- -GGGGGGGGG GSLT------


601
Choanoflagellatea
S. rosetta
M. roanoka
M. brevicollis
S. kvevrii
S. urceolata
S. infusionum
C. perplexa
H. balthica
A. spectabilis
S. diplocostata
H. nana
Metazoa
R. norvegicus
N. vectensis
O. pearsei
A. mellifera
701
801


901
1001
1101
Choanoflagellatea
Metazoa
S. rosetta
M. roanoka
M. brevicollis
S. kvevrii
S. urceolata
S. infusionum
C. perplexa
H. balthica
A. spectabilis
S. diplocostata
H. nana
R. norvegicus
N. vectensis
O. pearsei
A. mellifera


1201
Choanoflagellatea
S. rosetta --------DD PLLLAPRLAG VRAR------ -----RGEPS TVAVAATAPR HGQQQ----- ---QQEPHEH RQQHDGTPRH RFHTVSASRE RS--------
M. roanoka --------DD PML-VSEQQL RTAR------ ---------A HTVSTATASR PKPDS----- ---------- -KTQPGMREY STNPLRHCPP QH--------
M. brevicollis --------PE FLRGLHDVKL KPTP---KRP LDDKEPESEV LSSGTEGGTS LGDST-QTAA ERMGISLRKT NSRLSGIRRE RSNSIGDAPS RQASIKAATS
S. kvevrii RKTGRISMEF KGRTAEDIIQ HTQPSAAEAD NYRQRSKSDG HILRQNAEAA RTPGP----P PPTAPKPANV KDLVGQLKKA MTSPMLSRKS APGSPEKAGA
S. urceolata --------DD PLAGILTVKL KKSP------ -----GKQRP AIDTDKDASN DGELPFRVAL KPTGRRDSDV GSKGEGRPRT RSISDDTAAS DRPTQRDR--
S. infusionum ---------- ---------- ---------- -----SATNP AIMRNGNGSV EEKIS----- ---------- PTRRKSKPLF KNPPLEHVP- ----------
C. perplexa --------PD FLKGLHDIKL KKTP------ -----KASGS SAATQSNASP TEDGGVAPGA RLLGIKLRNT GARLSGYGRE RSNSVGESDL ----------
H. balthica --------MD IIQEFQQKHQ QNQP------ ---------- ----QGVVLR RKQLP----- ---------- ----VNHPNN KCSTTSTLSS ESNASINS--
A. spectabilis --------AG ILAGILDVKL KKTS------ -----SPSIL SSAAHSNSSR TSQGGLGG-- ---------- QHQDDHRRSA KSEVLAAVRA IPSDVKQAME
S. diplocostata --------AD PLAGILNVKL KKTP------ -----SASRK EFPATNGASH ASPGV----- --AGGNTPTA RRVASGPPAN SAGTTAARTA TNTA------
H. nana --------GG MLAGILDVKL KKRAGSMKMG KKRGGSPPTP AVPTTVVGSS PRRKP----- -SANGSPRAD RRQDEHRRSA KSEALAAVRA IPFDVKQAMD
Metazoa
R. norvegicus SGGSSGPTQA PALRYFQLPP RAASAAMYVP ARSGRGRKGP LVKQTKVEGE PQKGSIPSAS SPTSPALPRS EPPPAGPSEK NSIPIPTIII KAPSTSNSGR
N. vectensis --------WF PHTCVEEVKS EKPR------ ---------- ---------- ---------- RHTLFNRPSS ITSIYAKRSV LC--------
O. pearsei --------DD PISVIAKIAA ERRA---RIE AKAKASGDEP TTFTRKDKKT EDDAPVGELQ KALAARQARK NSTTKLVTEE KAMTVEASSP DKDTEHRA--
A. mellifera --------NI NQGDILEVTG ATDCGLLEGV LRGQGTGLFP AHCVQEVRLR HTNIPLGPQP SRDGRNPRVL GRRESQHKYF ATAPRLKKPV TSEPRTVVLH
1301
Choanoflagellatea
S. rosetta ---------- ---------- --RPTLDA-- -----LAAGP LPSREAP--- ---------- ---------- ---------- ---------- ----------
M. roanoka ---------- ---------- --RPTLPF-- -----QSTGH ASYNACP--- ---------- ---------- ---------- ---------- ----------
M. brevicollis AAV------- ---------- --ATSPAATA TFRRGDTMDR PEDDSPP--- ---------- ---------- ---------- ---------- ----------
S. kvevrii AV-------- ---------- --AAIAPA-- -----DALGN TANDIQD--- ---------- ---------- ---------- ---------- ----------
S. urceolata ---------- ---------- --AHTLGP-- -----STAKK APTKRPK--- ---------- ---------- ---------- ---------- ----------
S. infusionum ---------- ---------- ---------- -----DANTP PTER------ ---------- ---------- ---------- ---------- ----------
C. perplexa ---------- ---------- --PSRQAS-- -----IRRSR SPTDFED--- ---------- ---------- ---------- ---------- ----------
H. balthica ---------- ---------- --STTSNH-- -----GQSSK VVEQQHE--- ---------- ---------- ---------- ---------- ----------
A. spectabilis AVS------- ---------- --RSKNGT-- -----ESIPP VLHVESS--- ---------- ---------- ---------- ---------- ----------
S. diplocostata ---------- ---------- --ATPAPV-- -----PSHGG ATDTVRG--- ---------- ---------- ---------- ---------- ----------
H. nana AAKAKSDPKP RNDFHDELNR NMKQRHGV-- -----TSAPP PPTSNRA--- ---------- ---------- ---------- ---------- ----------
Metazoa
R. norvegicus SSQGSSTEAE PPTQPDGAGG GGSSPSPA-- -----PATSP VPPSPSPVPT PASPSGPATL DFTSQFGAAL VGAARREGGW QNEARRRSTL FLSTDAGDED
N. vectensis ---------- ---------- --DTNAGH-- -----MPPPR VVTLYRG--- ---------- ---------- ---------- ---------- ----------
O. pearsei ---------- ---------- --MGREDS-- -----DVRIR ALTDAAS--- ---------- ---------- ---------- ---------- ----------
A. mellifera ---------- ---------- --RSRKGFGF VLRGAKATSP LMELTPS--- ---------- ---------- ---------- ---------- ----------
1401
Choanoflagellatea
S. rosetta ---------- ---------- ---------- ---------- ---------- ---------- ---ATV---- -GRDTTTGTS FLGAPADATG PVFVA-----
M. roanoka ---------- ---------- ---------- ---------- ---------- ---------- ---PGD---- -AEEHIPPRP KFAAPPPPAG PVIVTPS---
M. brevicollis ---------- -PRPRANTLN SMQL------ ---------- ---------- ---------- -AASLR---- -KPPPVAALP HQSNDHLSNS SLDSDMDTTD
S. kvevrii ---------- LRREALQAAW GVTEL----- ---------- ---------- ---------- -NRDAL---- -LDNVAIESR ESYLRLRETN DAEPEYASPQ
S. urceolata ---------- ---PTPQAKA KGDNR----- ---------- ---------- ---------- -QDGGY---- -GEDGMTEFQ RQRLALKKKT AAASDAASPS
S. infusionum ---------- ---------- ---------- ---------- ---------- ---------- -RRGSS---- -SASQHTHRP GHHADSREQQ PP--------
C. perplexa ---------- ---------- ----A----- ---------- ---------- ---------- -RFDGP---- -GRPRANSLN SSSLLAQMHR VPEPEDDAP-
H. balthica ---------- ---------- ---------- ---------- ---------- ---------- -HHDAK---- -ASPVYINTM RVRLP----- ----------
A. spectabilis ---------- SARP------ ---------- ---------- ---------- ---------- -RRGSA---- -GPPTK-PKP KKAVEPISEP PTSPDFSAV-
S. diplocostata ---------- ---------- ---------- ---------- ---------- ---------- -RARSV---- -GPPVVKPKP PPVTPRRTAA ----------
H. nana ---------- ---PQNTGPP VSHRP----- ---------- ---------- ---------- -RGASV---- -GPPTK-PKP KTKTPVPENA PLSPDFAGV-
Metazoa
R. norvegicus GGDSGLGPGG PPGPRLRHSK SIDEGMFSAE PYLRLESGGS SGGYGAYAAG SRAYGGSGSS SAFTSF---- -LPPRPLVHP LTGKALDPAS PLGLALAARE
N. vectensis ---------- ---------- ---------- ---------- ---------- ---------- -NQGGF---- -GFQMRGANS QT-------- ----------
O. pearsei ---------- ---------- ---------- ---------- ---------- ---------- --RRQM---- -RPPRVSRTD SPKVAGLDDR RASFDLSA--
A. mellifera ---------- ARYPALQYLD DVDQGGVADL AGL------- ---------- ---------- -RKGDYLIQI NGEDVTTASH EHVVDLIRKS GELVRMTVV-


1501
Choanoflagellatea
S. rosetta ---------- --PPPPPPAA APSTTTTTTT TTTTTAPPPP P--------- ---------- ---------- APNMPAPPP- ---------- -PPP------
M. roanoka ---------- --PPPPPPPS IPSIVVPGPP PPPPPAP--- ---------- ---------- ---------- SIHMPPPPP- ---------- -PPA------
M. brevicollis DEGDDAP--- --PPPPKPAN IQDLLQRLKD DSAAPAHITQ HRQQQEQL-- ---------- -------TSP PLNGSPEAP- ---AGSTPGI AEPKSS----
S. kvevrii DAAKPSAL-- -VPPPSAFPS AASDLAPGPA PAPATPPLK- ---------- ---------- ---------- QRKGAPPPPA KPSTSTASSA ASPAAVQMRQ
S. urceolata S--------- --SSTTTPST LKSPARAEPA TYDNINPRQP PLTPSAAKAY R--------- -------RSQ DEGIYDNEP- ---------T QVPS------
S. infusionum ---------- --PPPPPPPI APSPAHIQQR TSHG------ ---------- ---------- -------GTG GDGGYARPA- ---------- ----------
C. perplexa ---------- --PPPPKPSN SSTLVHQLRK ANSTDPDLEQ ARQRALQ--- ---------- -------AVR DASGPAASP- ---------- ----------
H. balthica ---------- --PPPPPPPS LSS------- ---------- ---------- ---------- --------KP SISLPPPPP- ---------- -LPPSKT---
A. spectabilis ---------- --PPPSFTPH QQS-----TD LFDSADVIEM PSHVFD---- ---------- -------FTH KIGVIPPPPG NAEENFFPET VPPP-SFDEN
S. diplocostata ---------- --PPTRKPPD LAS----LPP DFHAAYIVPH PRSVTADTSF G--------- -------AAD DDGVMVPPP- ---------- ----------
H. nana ---------- --PPPQMPPH LASSVGRPVD LFDSAAVIEM PSHVFD---- ---------- -------FTH KIGIIPPPPG VAEENYFPTS VPPPSSFDDD
Metazoa
R. norvegicus RALKESSEGG GTPQPPPRPP SPRYDAPPPT LHHHSPHSPH SPHARHEPVL RLWGDPARRE LGYRAGLGSQ EKALTASPPA ARRSLLHRLP PTAPGVGPLL
N. vectensis ---------- --PRLDFQPT LEF------- ---------- ---------- ---------- ---------- -------PA- ---------- ----------
O. pearsei ---------- -VLHAVPPPP LAFAGSLTDT GKTGGVASVV VQDQSVDVDI D--------- -------KAL ELWGSDEQS- ---------- ----------
A. mellifera ---------- --SPMISLPN SQSAALLPTS QPIQRQYATL PRKGNNNVAI G--------- -------GTL GRSPAPMPP- ---------R RDPKTTLSVG
1601
Choanoflagellatea
S. rosetta ---STSSSQQ QSRPGGRA-- ---------- ---------- -NLLADISRG VKLRKI---- -------P-- ------VDAD AGRGGLGGAG VGRVMDG---
M. roanoka ---GDKPLHN TSRQPGRA-- ---------- ---------- -GLLADITRG VKLKKV---- ---------- ---------- ---------- ----------
M. brevicollis ---SPVMMAG VPRRATMP-- ---------- ---------- -ALPGSSTRD SREIVA---G PVTVPTPQ-- ------TTSN AGSNQPSTVA VARPPPV---
S. kvevrii ---QPALATS SGGVPQRP-- ---------- ---------- -ARRSVSIRP ESLDLSN--E GLPPPPPP-- --------SH TSVDDAATLD LPPAPPSPSK
S. urceolata ---NRGMLDR LEHDRQSG-- ---------- ---------- -ATKAQVTRQ EQVYDN---- HQRQHQPR-- ------EEMQ LAKGMVTGVG SGSGHGTASG
S. infusionum ---------- ----VAPS-- ---------- ---------- -LLLQQIQAG KKLKKV---- ---------- ---------- -APSAPKQLG TVTQPPK---
C. perplexa ---SPKVALQ PPRPAGDR-- ---------- ---------- -GRSQTLDST STSSHR---- ---------- ---------- GQRSRSASPG RIPPQVKPK-
H. balthica ---NTSVQPS LPPSIGRG-- ---------- ---------- -NLLDQIHKG VQLNKV---- ---------- ---------- ---------- ----------
A. spectabilis LTSGEVQLLM VPPPPSMD-- ---------- ---------- -----EPERP AALDSM---G ELPPMPSP-- ---------G LDARMSMLLD DFPPPP----
S. diplocostata ---------- -PPPPGFW-- ---------- ---------- -AGDHDTPAG EPTAHAD--H HLFIPPPP-- ---MDFGHVE TGDSALNSID FLPPMPSPG-
H. nana GGDGFEMLVM PPPSVDAI-- ---------- ---------- -GSGGSATLP LALDSI---G ELPPMPSP-- ---------E FDPGSPRSLG AFPPPPQ---
Metazoa
R. norvegicus LQLGPEPPTP HPGVSKAWRT AAPEEPERLP LHVRFLENCQ ARPPPAGTRG SSTEDG---P GVPPPSPRRV LPTSPTSPRG NEENGLPLLV LPPPAPSVDV
N. vectensis ---------- -------L-- ---------- ---------- -QYIGEVERG GEADRK---- ---------- ---------- ---------- ----------
O. pearsei ---TGALNLT AVQDKDRD-- ---------- ---------- -VVRLDDTEP EKDDKATRWD DIVIPPPVDM DALVGWEESE SEESSLHIVG INSPP-----
A. mellifera RARARSMVAG LEGGGERD-- ---------- ---------- -DRDEITSTG AKSSSA---E SIHLPQQP-- --STGPNTGQ NTPVQPRTAS IRSRPTSSRI
1701
Choanoflagellatea
S. rosetta ------SSAT SSTPHMRERA ATAPSTALSA MARGHGDELL DGILNVRL-- ----RKTAPQ QQLRKGEGDA SATT------ GKDQHEASPS PF-----AGI
M. roanoka ---------- ---------- ---------- --QPESDAPL SSN------- ---------- ---------- ---------- ---------- ------SGA
M. brevicollis ---------Q PKPKAVGPSA GPAGAPPAVP TKPKPALATS DSVVPASA-- ----QSDKQE GPSSPVKTGN VSKPSAARRE IDFALDFALP PPPLLDDFAG
S. kvevrii ------AVSS PPLTKRFGAP SSTAVPPPSS QGPDRTSAPL EESHTFRF-- ----AKLAKQ SRNSSDRG-- -SASSLELSD ASALSDLPPP PPMDVPVSGT
S. urceolata ------AAGT ALGSAGWATR RTHSDPALYN NVPRTGSGGR KAVLPPSP-- ----VKEEST ETTPAVYANT GAGRNDGVWD DVSDAEDDRA A------GTG
S. infusionum ------VSVH LRSVDDLDTT STRSAPAALS HVQGDMLAGL LNVKLKK--- ---------- ---------- ---------- ---------- -------TKR
C. perplexa ------PAGR PIETPGASET TPAVHPKPIK VAPPPPPAKD KPALDIRP-- ----PSESPE IPTVPASPPS G--------- ---------- PSMQEALSSP
H. balthica ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
A. spectabilis ------PDFP RLDESIEDHH TGGPIVSDHD VIPPPTFATL GTSNSVNI-- ----VEHVPT EHN------- ---------- ---------- -------IGD
S. diplocostata ------PDDA SHAPAFDVGF PPPPPPPPVI HVPPPAPVLI SDA------- ---------- ---------- ---------- ---------- -------SEG
H. nana ------HLAQ PSSPTLQPLV TTSVPFASLG LPVPPSDFLA AALLDAPP-- ----LREAPS ERGPSRDS-- -MSRSDGAYD ESGDFDEDDD GYLH---IGG
Metazoa
R. norvegicus DDGEFLFAEP LPPPLEFSNS FEKPESPLTP GPPHPLPDPP SPATPLPA-- ----APPPAV AAAPPTLDST ASSLTSYDSE VATLTQGAPA APGDPPAPGP
N. vectensis ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
O. pearsei ------PIET FLSPDREGSY PVFIPPPLST LIPPPEDDTL ERLPLPRI-- ----P----- ---------- ---------- ---------- -------TPP
A. mellifera TAAELEELFQ RQQGSASGQY SSSMMSSHFQ TGQATKSHPS SPAKTGRVYA SVAEMKRKGK LNSRVRFFGG LGGGSDLHRD FHSTPDLNVQ VQSSILAPKG


1801
S. rosetta KLRKTPTNTM GETDGSTPTA P--AAEAGPV RLRSTKPPPP PAGPKPTQSK PAVPTPEAPQ QSGPTA---- ---------- ---------- ----TSKPPP
M. roanoka DRARPRASDN GGIGGTTLP- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
M. brevicollis SDDQVDAPDD GDFFIPPPPV LDFEPATPDS GPITMQLPPP PEAFFELQGS DSSSAESPAQ APQPPPATDS PPRAWVHAPP ADSTPPTSLP ADNVESDDLG
S. kvevrii DDASFICPPT EFSELASESS TDIVKPPTEF ELPPLHMLPP PLLDAPTAPA SSISSSVASA PEVPPPAAFL PPAEFL---- ---PPPPADG GGDRLSFLNE
S. urceolata DGRSTPQPTG TGKARKSTPP PDGQHRRNSF RVAKAKAADP PPAAQPHAKS NVYSNLPVTS NQAAAALQTS NPDEHLYFNL PPFAPPVQDA KDDYLDASPP
S. infusionum DGNGGEEGGD GGDGRAKQAQ ESSVSNDIGA SVL------- ---------- ---------- ---------- ---------- ---------- ---------A
C. perplexa AISSRPLDAA SPDFVPPPPI FDAPPPVVSP TAFGDFIPPP PILDSA---- ---------- ---------- ---------- ---------- ----SSSSSP
H. balthica ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
A. spectabilis EKDNAEYLDI VASSPSPTSP VPKVTNIRGL RTARAAPPPP ---------- ---------- ---------- ---------- ---------- ----VAQKPP
S. diplocostata SDGYIDVSAA SSASPVPVAG ASVVARSPTP EAVAVVVPAF ASAR------ ---------- ---------- ---------- ---------- ----SSYSVL
H. nana SLSSTPSPTK DSFDLPPPAQ LSTTTRQPIS RAARTAPPPP QILPKPST-- ---------- ---------- ---------- ---------- ----------
R. norvegicus PAPAAPAPPA PQPGPDPPPG TDSGIEEVDS RSSSDHPLET ISSASTLSSL SAEGGGNTGG VAGGGAGVAN GTELL----- ---DTYVAYL DGQAFGGSGT
N. vectensis ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
O. pearsei TNDETSFTAE IDVSALPPPS PSMYQYETEL NSDDLPLSIP PPIDFSSTSL EQ-------- ---------- ---------- ---------- ----LLVKEA
A. mellifera HRSQEDVNAL NGRNGLPPPN HPPPPPPVGQ VVKVNVGANV PDVVTPASVY DNMAHIQQVK ELAAATVDGS YGVMSSFRPS NSAKLYASPE DMKTVGYRSR
1901
S. rosetta PTRPKPAVAP PQPAPKPRAE KTPTAQQQQT EQQQQQQQQQ QQQQQQQQQQ QQQQQQQTEQ QQQQQQA--- ---DE-GEAS TSPFIPPPMF LADVQMTTSE
M. roanoka --KMLPPLAA MGKGNFLDGI LTVKLRETKR PSITQQGQAP HEEEP----- ---------- ---------- ---------- -------PVP AFANLKLRKT
M. brevicollis PAPPLPTEEA PAATMVMADL AKLLDPRSSR STVASTSSSE VSDLVPATRP HRSAPPPPKG RSPQVSH--- ------VTPE CDLVAISARV IYAYEAQFED
S. kvevrii AAPLPPPTPD DEQVFDMRAS MLANLPPPPI DSLPPVEHID SGPGTPAGGT PLGSPAMVPR TITNSATGPI PPPGVTQQPP AEPPKQLVKA VHAFAVRFQD
S. urceolata PPPPSGFNSP VKGEADSEDL WVDLPPPEDL PPPPLTPEPS VADAGAFEMA ASSFVFPPTA ----AAV--- ---RV----- -------AVA QEPYECTAED
S. infusionum GVKLRKTKSR LSGIGRGNSC LHMLVRMAGC ---------- ---------- ---------- ---------- -------ILT CAAFLQRDKL
C. perplexa QLPPPPSLP- -------AGF EDSLSTREST TSILSTSSSE DETIEPHRRP SRNAPPPPSN ASKPKLVKLD EQGRLDDTPD VGLAALSARV VYPYQACFED
H. balthica ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- TPVVHQSLSS
A. spectabilis GGTLIVPPAR VSELSGHSAV SSVLSTSDSS DD----EIEG HTSTATSNLG RQRAPPPPPT GVSPTEI--- ---RM----- -------VIT AFAFEARNTD
S. diplocostata STSDSSDSDS DEDISGFDAV PGMSNEAVAR EAVQHTQQSQ KVTTTTVRGR HHRAPPPPPK NVAPSLG--- ---RM----- -------AIS AFAFAARNED
H. nana SSKLEVPLPR ESSLSAHSGI SSVLSTSDSS DDESTRGSVS SSGGGGRSAA KLRAAPPPPA GIAPAEV--- ---RM----- -------VIT AFAFEAHNTD
R. norvegicus PGPPYPPQLM TPSKLRGRAL GTSGNLRPGP SGGLRDPVTP TSPTVSVTGA GTDGLLALSA CPGPSTA--- ---GVAGGPV AVEPEVPPVP LPAASSLPRK
N. vectensis ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---GV----- -------KCK DFILEVNWED
O. pearsei DFVPLTTFQP KAANETLPGS NVSELADFDI SSLPPLPVVD FGQKNSFDID IDDLPPPVVT SESEMGE--- ---DL----- -------PLS LPPPVEFNTN
A. mellifera SLPTHTTRCH VRKSHSLRTT NNTTFKPLNS QQNANNAVSN NNQVNNNQSA NNQYAQPLKT NRSHSTI--- ---GIRERKK KGISTSSSIT NLSSVANNNT
2001
S. rosetta TDEPQPTNGS SLMQEPQH-- ---------- ---------- ---------- --SASDASTS TAQTD----- --DDAVMSLI AP----P--- ----------
M. roanoka PTRFEGTDEK PPALLPE--- ---------- ---------- ---------- ---------- ---------- ---------- ----------
M. brevicollis ELQLEPGSIV TLLETPD--- ---------- ---------- ---------- --GGWWRGKH KGGE----GW FPSNYVR--- ---------- ----------
S. kvevrii ELPLSIDDTV EVLQSPE--- ---------- ---------- ---------- --GGWWEGIC NGKQ----GW FPCNHVEPIS PPDRSTTTES MLFRELSDLD
S. urceolata ELSFQTGDVI ELLQTPEV-- ---------- ---------- ---------- --GGWWEGS- LNGVS---GW FPCKHVFEYH GD----TKAD RLSR------
S. infusionum DADFNSHS-- ---------- ---------- ---------- ---------- --YAWPMTRF ---------- ---------- ---------- ----------
C. perplexa ELSLEHDQLI TLLKTPD--- ---------- ---------- ---------- --GGWWRGKT KTDS----GW FPSNYVR--- ---------- ----------
H. balthica NVTITPSHTQ P--------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
A. spectabilis ELSFQPGEKL ELMQSPD--- ---------- ---------- ---------- --GGWWKGKN MTGSE---GW FPSNHARELV TP----PSAS KLPAMS----
S. diplocostata ELSFDPGMRL ELLETPTT-- ---------- ---------- ---------- --GGWWKGRI GTTE----GW FPNNHVRELV SP----PAAS KLPARRSSKL
H. nana ELSFRPGEKL VLLQSPD--- ---------- ---------- ---------- --GGWWKGRG GGGDE---GW FPSNHARELV TP----PSAS MLPMMKTG--
R. norvegicus LLPWEEGPGP PPPPLPGPLS QPQASALATV KASIISELSS KLQQFGGSST AGGALPWARG GSGGSTDSHH GGASYIPERT SS----LQRQ RLSEDSQTSL
N. vectensis VTTATHTHVV ELIRR----- ---------- ---------- ---------- ---------- ---------- ---------- ----------
O. pearsei DVVLSSLSSS PLPPAPPP-- ---------- ---------- ---------- --DAF----- ---------- --SSLLPALL PP----PA-- ----------
A. mellifera GNTVVAANTA PPIPEPD--- ---------- ---------- ---------- --YSLSESEN DEGEEETDDD GESEIAKELE KA----AARE KLESTRETSG
Choanoflagellatea
Metazoa

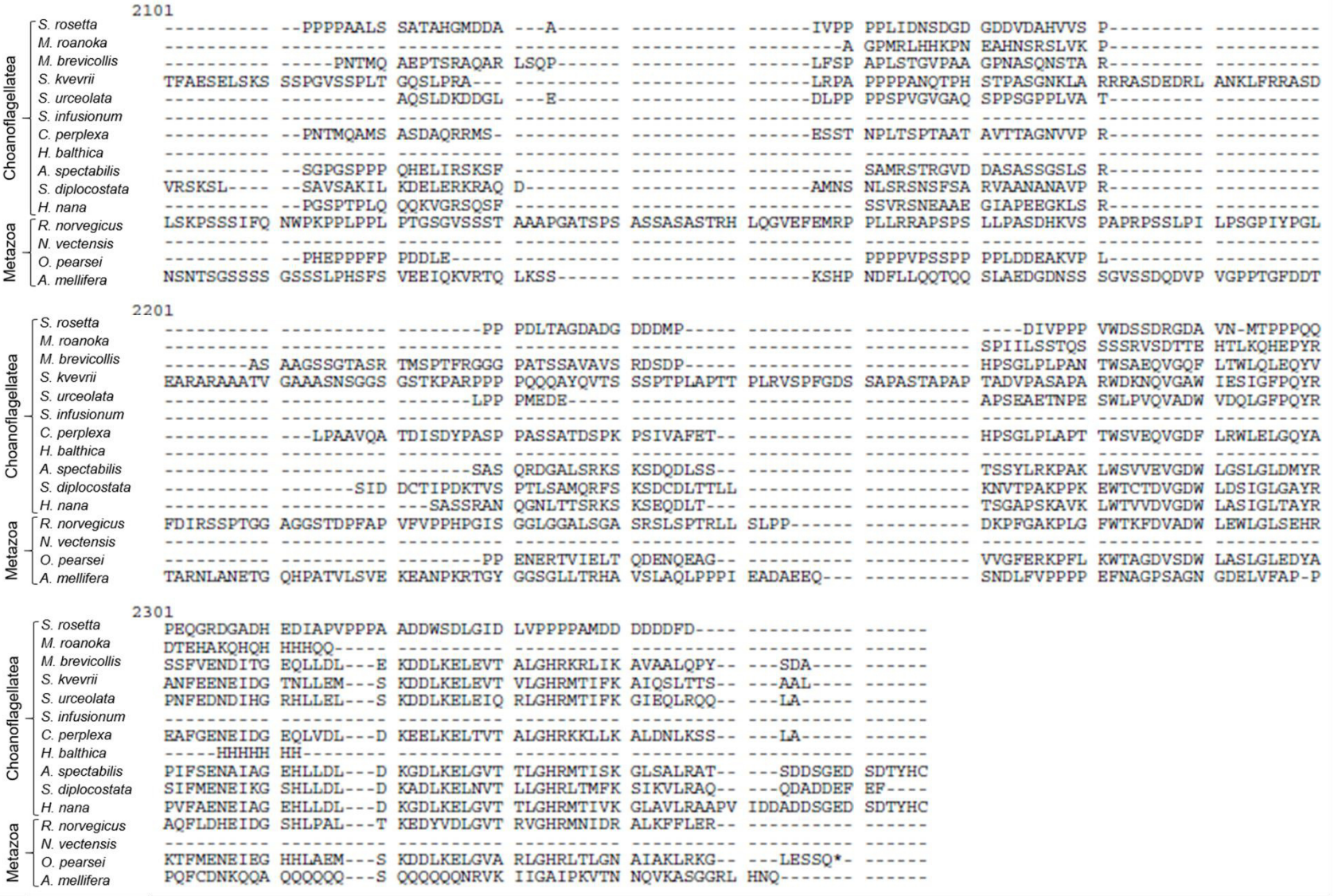
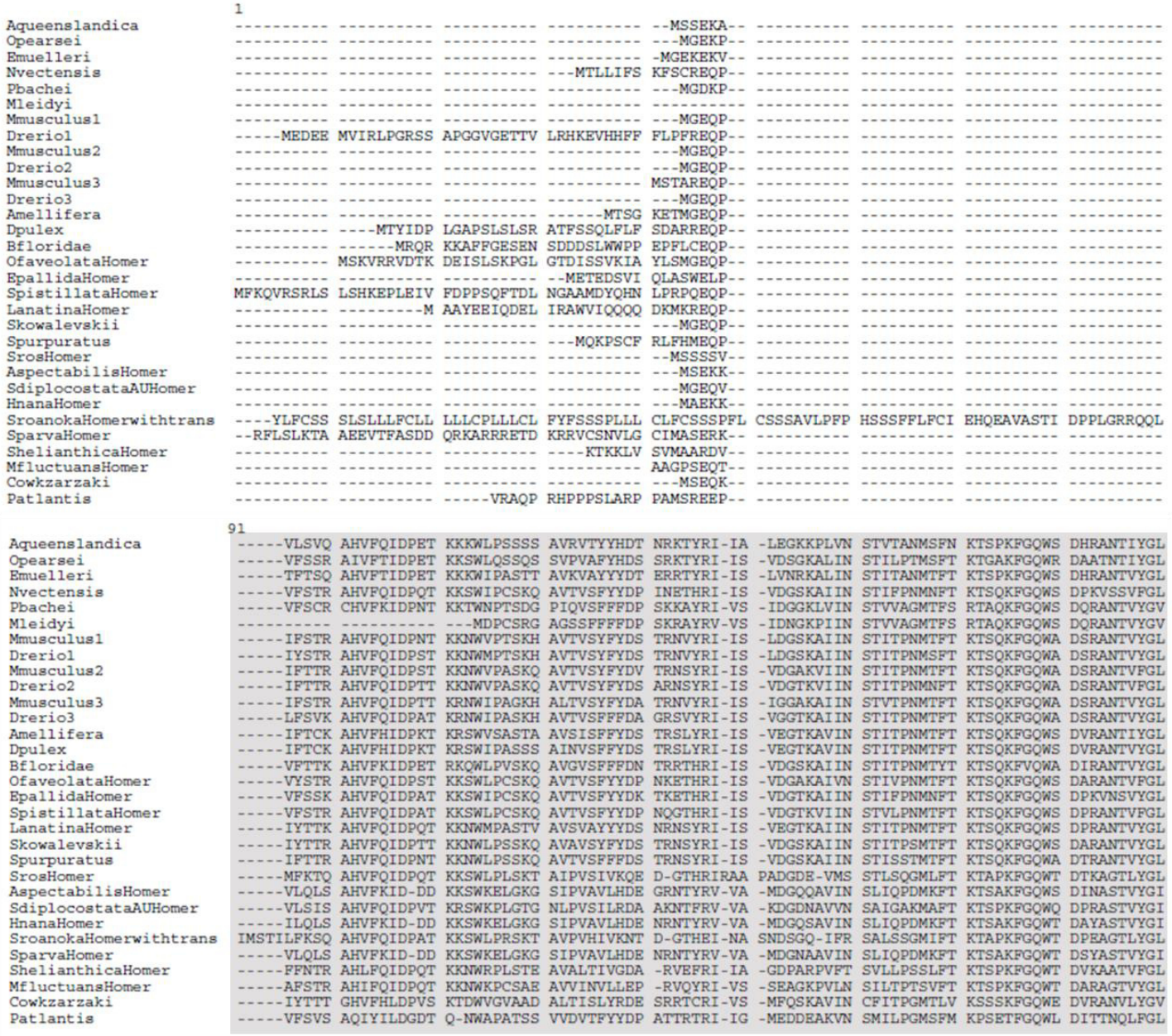
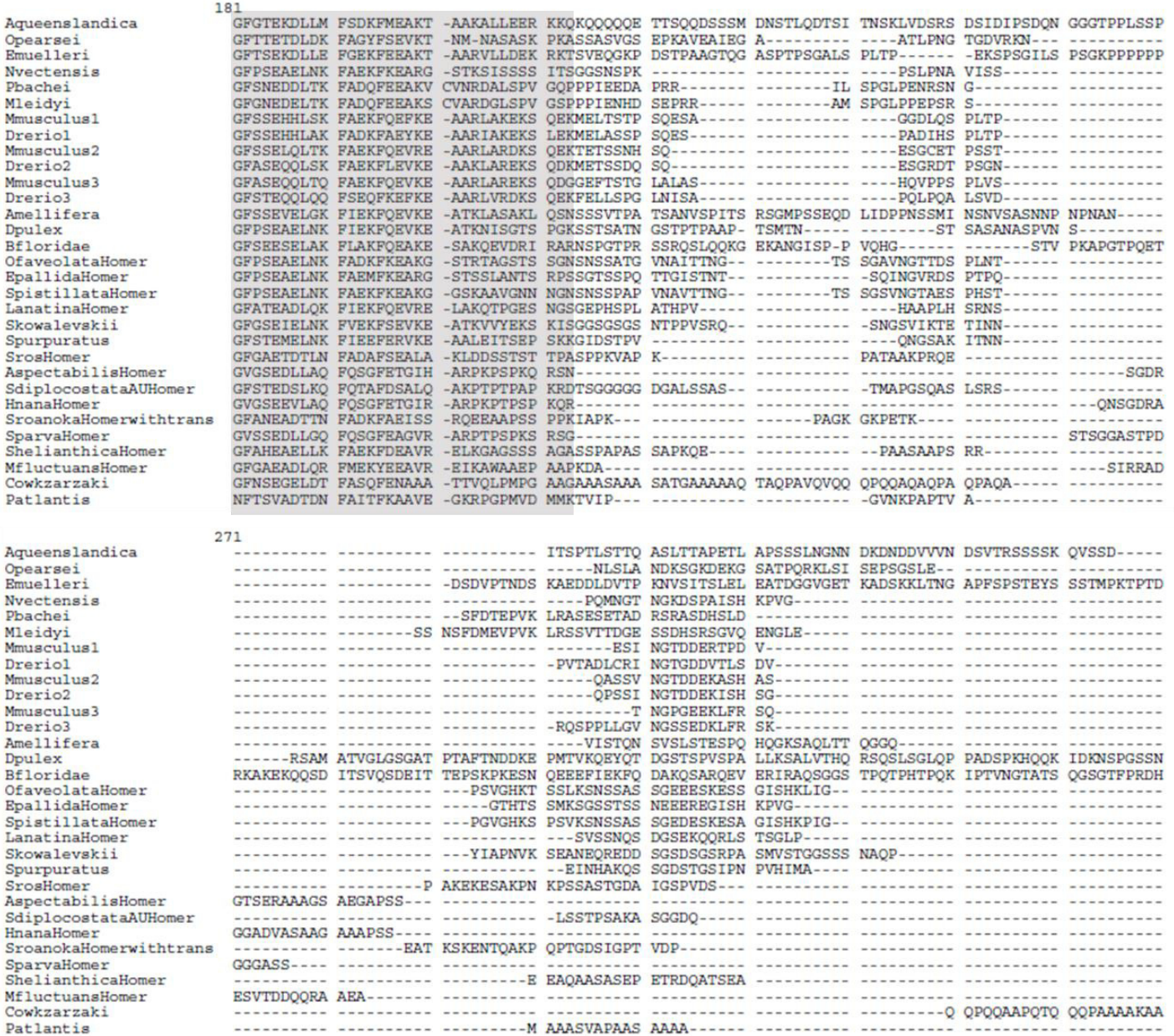
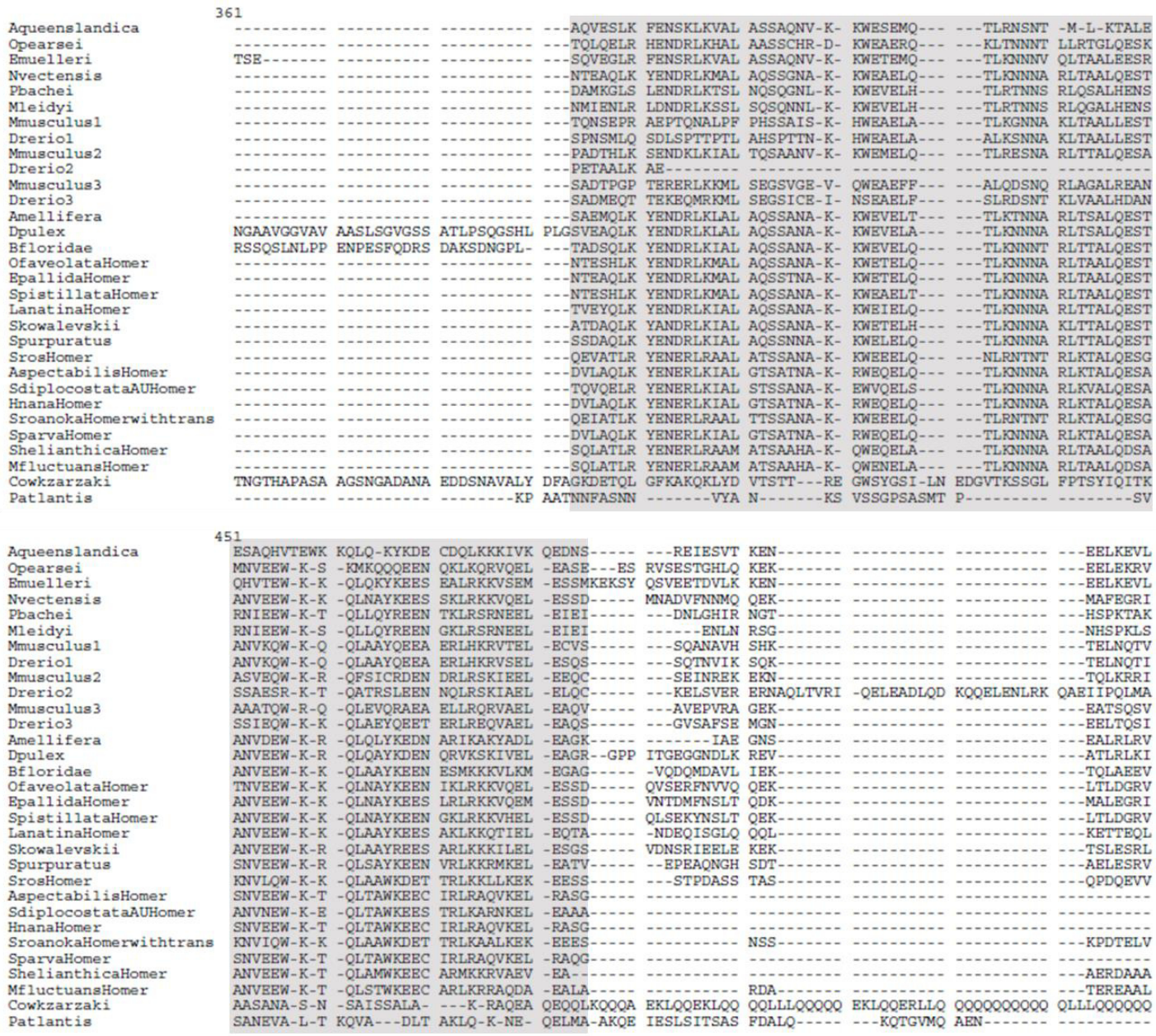
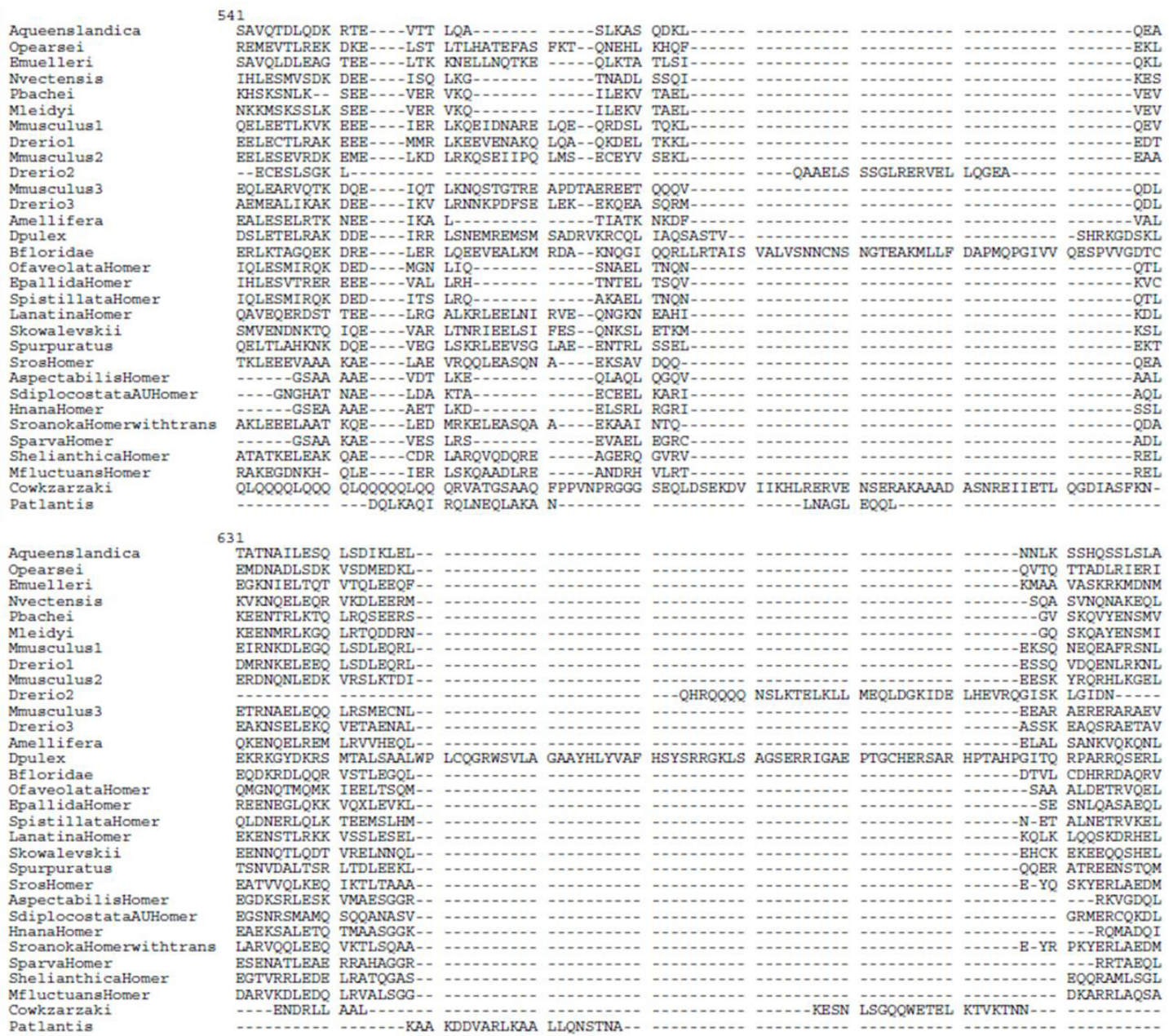


Figure 6-1: Alignment of choanoflagellate and animal Shank sequences.


1
Aqueenslandica ---------- ---------- ---------- ---------- --MSSEKA-- ---------- ---------- ---------- ----------
Opearsei ---------- ---------- ---------- ---------- ---MGEKP-- ---------- ---------- ---------- ----------
Emuelleri ---------- ---------- ---------- ---------- -MGEKEKV-- ---------- ---------- ---------- ----------
Nvectensis ---------- ---------- ---------- ---MTLLIFS KFSCREQP-- ---------- ---------- ---------- ----------
Pbachei ---------- ---------- ---------- ---------- ---MGDKP-- ---------- ---------- ---------- ----------
Mleidyi ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Mmusculus1 ---------- ---------- ---------- ---------- ---MGEQP-- ---------- ---------- ---------- ----------
Dreriol -----MEDEE MVIRLPGRSS APGGVGETTV LRHKEVHHFF FLPFREQP-- ---------- ---------- ---------- ----------
Mmusculus2 ---------- ---------- ---------- ---------- ---MGEQP-- ---------- ---------- ---------- ----------
Drerio2 ---------- ---------- ---------- ---------- ---MGEQP-- ---------- ---------- ---------- ----------
Mmusculus3 ---------- ---------- ---------- ---------- MSTAREQP-- ---------- ---------- ---------- ----------
Drerio3 ---------- ---------- ---------- ---------- ---MGEQP-- ---------- ---------- ---------- ----------
Amellifera ---------- ---------- ---------- ------MTSG KETMGEQP-- ---------- ---------- ---------- ----------
Dpulex ---------- ----MTYIDP LGAPSLSLSR ATFSSQLFLF SDARREQP-- ---------- ---------- ---------- ----------
Bfloridae ---------- ------MRQR KKAFFGESEN SDDDSLWWPP EPFLCEQP-- ---------- ---------- ---------- ----------
OfaveolataHomer ---------- MSKVRRVDTK DEISLSKPGL GTDISSVKIA YLSMGEQP-- ---------- ---------- ---------- ----------
EpallidaHomer ---------- ---------- ---------- --METEDSVI QLASWELP-- ---------- ---------- ---------- ----------
SpistillataHomer MFKQVRSRLS LSHKEPLEIV FDPPSQFTDL NGAAMDYQHN LPRPQEQP-- ---------- ---------- ---------- ----------
LanatinaHomer ---------- ---------M AAYEEIQDEL IRAWVIQQQQ DKMKREQP-- ---------- ---------- ---------- ----------
Skowalevskii ---------- ---------- ---------- ---------- ---MGEQP-- ---------- ---------- ---------- ----------
Spurpuratus ---------- ---------- ---------- ---MQKPSCF RLFHMEQP-- ---------- ---------- ---------- ----------
SrosHomer ---------- ---------- ---------- ---------- --MSSSSV-- ---------- ---------- ---------- ----------
AspectabilisHomer ---------- ---------- ---------- ---------- ---MSEKK-- ---------- ---------- ---------- ----------
SdiplocostataAUHomer ---------- ---------- ---------- ---------- ---MGEQV-- ---------- ---------- ---------- ----------
HnanaHomer ---------- ---------- ---------- ---------- ---MAEKK-- ---------- ---------- ---------- ----------
SroanokaHomerwithtrans ----YLFCSS SLSLLLFCLL LLLCPLLLCL FYFSSSPLLL CLFCSSSPFL CSSSAVLPFP HSSSFFLFCI EHQEAVASTI DPPLGRRQQL
SparvaHomer --RFLSLKTA AEEVTFASDD QRKARRRETD KRRVCSNVLG CIMASERK-- ---------- ---------- ---------- ----------
ShelianthicaHomer ---------- ---------- ---------- ----KTKKLV SVMAARDV-- ---------- ---------- ---------- ----------
MfluctuansHomer ---------- ---------- ---------- ---------- AAGPSEQT-- ---------- ---------- ---------- ----------
Cowkzarzaki ---------- ---------- ---------- ---------- ---MSEQK-- ---------- ---------- ---------- ----------
Patlantis ---------- ---------- -----VRAQP RHPPPSLARP PAMSREEP-- ---------- ---------- ---------- ----------
91
Aqueenslandica -----VLSVQ AHVFQIDPET KKKWLPSSSS AVRVTYYHDT NRKTYRI-IA -LEGKKPLVN STVTANMSFN KTSPKFGQWS DHRANTIYGL
Opearsei -----VFSSR AIVFTIDPET KKSWLQSSQS SVPVAFYHDS SRKTYRI-IS -VDSGKALIN STILPTMSFT KTGAKFGQWR DAATNTIYGL
Emuelleri -----TFTSQ AHVFTIDPET KKKWIPASTT AVKVAYYYDT ERRTYRI-IS -LVNRKALIN STITANMTFT KTSPKFGQWS DHRANTVYGL
Nvectensis -----VFSTR AHVFQIDPQT KKSWIPCSKQ AVTVSFYYDP INETHRI-IS -VDGSKAIIN STIFPNMNFT KTSQKFGQWS DPKVSSVFGL
Pbachei -----VFSCR CHVFKIDPNT KKTWNPTSDG PIQVSFFFDP SKKAYRI-VS -IDGGKLVIN STVVAGMTFS RTAQKFGQWS DQRANTVYGV
Mleidyi ---------- ---------- ---MDPCSRG AGSSFFFFDP SKRAYRV-VS -IDNGKPIIN STVVAGMTFS RTAQKFGQWS DQRANTVYGV
Mmusculus1 -----IFSTR AHVFQIDPNT KKNWVPTSKH AVTVSYFYDS TRNVYRI-IS -LDGSKAIIN STITPNMTFT KTSQKFGQWA DSRANTVYGL
Dreriol -----IYSTR AHVFQIDPST KKNWMPTSKH AVTVSYFYDS TRNVYRI-IS -LDGSKAIIN STITPNMSFT KTSQKFGQWA DSRANTVYGL
Mmusculus2 -----IFTTR AHVFQIDPST KKNWVPASKQ AVTVSYFYDV TRNSYRI-IS -VDGAKVIIN STITPNMTFT KTSQKFGQWA DSRANTVFGL
Drerio2 -----IFTTR AHVFQIDPTT KKNWVPASKQ AVTVSYFYDS ARNSYRI-IS -VDGTKVIIN STITPNMNFT KTSQKFGQWA DSRANTVFGL
Mmusculus3 -----IFSTR AHVFQIDPTT KRNWIPAGKH ALTVSYFYDA TRNVYRI-IS -IGGAKAIIN STVTPNMTFT KTSQKFGQWA DSRANTVYGL
Drerio3 -----LFSVK AHVFQIDPAT KRNWIPASKH AVTVSFFFDA GRSVYRI-IS -VGGTKAIIN STITPNMTFT KTSQKFGQWA DSRANTVYGL
Amellifera -----IFTCK AHVFHIDPKT KRSWVSASTA AVSISFFYDS TRSLYRI-IS -VEGTKAVIN STITPNMTFT KTSQKFGQWS DVRANTIYGL
Dpulex -----IFTCK AHVFHIDPKT KRSWIPASSS AINVSFFYDS TRSLYRI-IS -VEGTKAVIN STITPNMTFT KTSQKFGQWS DVRANTVYGL
Bfloridae -----VFTTK AHVFKIDPET RKQWLPVSKQ AVGVSFFFDN TRRTHRI-IS -VDGSKAIIN STITPNMTYT KTSQKFVQWA DIRANTVYGL
OfaveolataHomer -----VYSTR AHVFQIDPST KKSWLPCSKQ AVTVSFYYDP NKETHRI-IS -VDGAKAIVN STIVPNMTFT KTSQKFGQWS DARANTVFGL
EpallidaHomer -----VFSSK AHVFQIDPAT KKSWIPCSKQ AVTVSFYYDK TKETHRI-IS -VDGTKAIIN STIFPNMNFT KTSQKFGQWS DPKVNSVYGL
SpistillataHomer -----VFSTR AHVFQIDPAT KKSWLPCSKQ AVTVSFYYDP NQGTHRI-IS -VDGTKVIIN STVLPNMTFT KTSQKFGQWS DPRANTVFGL
LanatinaHomer -----IYTTK AHVFQIDPQT KKNWMPASTV AVSVAYYYDS NRNSYRI-IS -VEGTKAIIN STITPNMTFT KTSQKFGQWS DPRANTVYGL
Skowalevskii -----IYTTR AHVFQIDPTT KKNWLPSSKQ AVAVSYFYDS TRNSYRI-IS -VDGSKAIIN STITPSMTFT KTSQKFGQWS DARANTVYGL
Spurpuratus -----IFTTR AHVFQIDPNT KKNWLPSSKQ AVTVSFFFDS TRNSYRI-IS -VDGSKAIIN STISSTMTFT KTSQKFGQWA DTRANTVYGL
SrosHomer -----MFKTQ AHVFQIDPQT KKSWLPLSKT AIPVSIVKQE D-GTHRIRAA PADGDE-VMS STLSQGMLFT KTAPKFGQWT DTKAGTLYGL
AspectabilisHomer -----VLQLS AHVFKID-DD KKSWKELGKG SIPVAVLHDE GRNTYRV-VA -MDGQQAVIN SLIQPDMKFT KTSAKFGQWS DINASTVYGI
SdiplocostataAUHomer -----VLSIS AHVFQIDPVT KRSWKPLGTG NLPVSILRDA AKNTFRV-VA -KDGDNAVVN SAIGAKMAFT KTSPKFGQWQ DPRASTVYGI
HnanaHomer -----ILQLS AHVFKID-DD KKSWKELGKG SIPVAVLHDE NRNTYRV-VA -MDGQSAVIN SLIQPDMKFT KTSAKFGQWT DAYASTVYGI
SroanokaHomerwithtrans IMSTILFKSQ AHVFQIDPAT KKSWLPRSKT AVPVHIVKNT D-GTHEI-NA SNDSGQ-IFR SALSSGMIFT KTAPKFGQWT DPEAGTLYGL
SparvaHomer -----VLQLS AHVFKID-DD KKSWKELGKG SIPVAVLHDE NRNTYRV-VA -MDGNAAVIN SLIQPDMKFT KTSAKFGQWT DSYASTVYGI
ShelianthicaHomer -----FFNTR AHLFQIDPQT KKNWRPLSTE AVALTIVGDA -RVEFRI-IA -GDPARPVFT SVLLPSSLFT KTSPKFGQWT DVKAATVFGL
MfluctuansHomer -----AFSTR AHIFQIDPQT KKNWKPCSAE AVVINVLLEP -RVQYRI-VS -SEAGKPVLN SILTPTSVFT KTSPKFGQWT DARAGTVYGL
Cowkzarzaki -----IYTTT GHVFHLDPVS KTDWVGVAAD ALTISLYRDE SRRTCRI-VS -MFQSKAVIN CFITPGMTLV KSSSKFGQWE DVRANVLYGV
Patlantis -----VFSVS AQIYILDGDT Q-NWAPATSS VVDVTFYYDP ATTRTRI-IG -MEDDEAKVN SMILPGMSFM KPSETFGQWL DITTNQLFGL


181
Aqueenslandica GFGTEKDLLM FSDKFMEAKT -AAKALLEER KKQKQQQQQE TTSQQDSSSM DNSTLQDTSI TNSKLVDSRS DSIDIPSDQN GGGTPPLSSP
Opearsei GFTTETDLDK FAGYFSEVKT -NM-NASASK PKASSASVGS EPKAVEAIEG A--------- ----ATLPNG TGDVRKN--- ----------
Emuelleri GFTSEKDLLE FGEKFEEAKT -AARVLLDEK RKTSVEQGKP DSTPAAGTQG ASPTPSGALS PLTP------ -EKSPSGILS PSGKPPPPPP
Nvectensis GFPSEAELNK FAEKFKEARG -STKSISSSS ITSGGSNSPK ---------- ---------- ----PSLPNA VISS------ ----------
Pbachei GFSNEDDLTK FADQFEEAKV CVNRDALSPV GQPPPIEEDA PRR------- --------IL SPGLPENRSN G--------- ----------
Mleidyi GFGNEDELTK FADQFEEAKS CVARDGLSPV GSPPPIENHD SEPRR----- --------AM SPGLPPEPSR S--------- ----------
Mmusculus1 GFSSEHHLSK FAEKFQEFKE -AARLAKEKS QEKMELTSTP SQESA----- ---------- ----GGDLQS PLTP------ ----------
Drerio1 GFSSEHHLAK FADKFAEYKE -AARIAKEKS LEKMELASSP SQES------ ---------- ----PADIHS PLTP------ ----------
Mmusculus2 GFSSELQLTK FAEKFQEVRE -AARLARDKS QEKTETSSNH SQ-------- ---------- ----ESGCET PSST------ ----------
Drerio2 GFASEQQLSK FAEKFLEVKE -AAKLAREKS QDKMETSSDQ SQ-------- ---------- ----ESGRDT PSGN------ ----------
Mmusculus3 GFASEQQLTQ FAEKFQEVKE -AARLAREKS QDGGEFTSTG LALAS----- ---------- ----HQVPPS PLVS------ ----------
Drerio3 GFSTEQQLQQ FSEQFKEFKE -AARLVRDKS QEKFELLSPG LNISA----- ---------- ----PQLPQA LSVD------ ----------
Amellifera GFSSEVELGK FIEKFQEVKE -ATKLASAKL QSNSSSVTPA TSANVSPITS RSGMPSSEQD LIDPPNSSMI NSNVSASNNP NPNAN-----
Dpulex GFPSEAELNK FIEKFQEVKE -ATKNISGTS PGKSSTSATN GSTPTPAAP- TSMTN----- --------ST SASANASPVN S---------
Bfloridae GFSEESELAK FLAKFQEAKE -SAKQEVDRI RARNSPGTPR SSRQSLQQKG EKANGISP-P VQHG------ -------STV PKAPGTPQET
OfaveolataHomer GFPSEAELNK FADKFKEAKG -STRTAGSTS SGNSNSSATG VNAITTNG-- --------TS SGAVNGTTDS PLNT------ ----------
EpallidaHomer GFPSEAELNK FAEMFKEARG -STSSLANTS RPSSGTSSPQ TTGISTNT-- ---------- -SQINGVRDS PTPQ------ ----------
SpistillataHomer GFPSEAELNK FAEKFKEAKG -GSKAAVGNN NGNSNSSPAP VNAVTTNG-- --------TS SGSVNGTAES PHST------ ----------
LanatinaHomer GFATEADLQK FIEKFQEVRE -LAKQTPGES NGSGEPHSPL ATHPV----- ---------- ----HAAPLH SRNS------ ----------
Skowalevskii GFGSEIELNK FVEKFSEVKE -ATKVVYEKS KISGGSGSGS NTPPVSRQ-- ---------- -SNGSVIKTE TINN------ ----------
Spurpuratus GFSTEMELNK FIEEFERVKE -AALEITSEP SKKGIDSTPV ---------- ---------- ----QNGSAK ITNN------ ----------
SrosHomer GFGAETDTLN FADAFSEALA -KLDDSSTST TPASPPKVAP K--------- ---------- PATAAKPRQE ---------- ----------
AspectabilisHomer GVGSEDLLAQ FQSGFETGIH -ARPKPSPKQ RSN------- ---------- ---------- ---------- ---------- ------SGDR
SdiplocostataAUHomer GFSTEDSLKQ FQTAFDSALQ -AKPTPTPAP KRDTSGGGGG DGALSSAS-- ---------- -TMAPGSQAS LSRS------ ----------
HnanaHomer GVGSEEVLAQ FQSGFETGIR -ARPKPTPSP KQR------- ---------- ---------- ---------- ---------- ---QNSGDRA
SroanokaHomerwithtrans GFANEADTTN FADKFAEISS -RQEEAAPSS PPKIAPK--- ---------- ------PAGK GKPETK---- ---------- ----------
SparvaHomer GVSSEDLLGQ FQSGFEAGVR -ARPTPSPKS RSG------- ---------- ---------- ---------- ---------- STSGGASTPD
ShelianthicaHomer GFAHEAELLK FAEKFDEAVR -ELKGAGSSS AGASSPAPAS SAPKQE---- ---------- --PAASAAPS RR-------- ----------
MfluctuansHomer GFGAEADLQR FMEKYEEAVR -EIKAWAAEP AAPKDA---- ---------- ---------- ---------- ---------- ----SIRRAD
Cowkzarzaki GFNSEGELDT FASQFENAAA -TTVQLPMPG AAGAAASAAA SATGAAAAAQ TAQPAVQVQQ QPQQAQAQPA QPAQA----- ----------
Patlantis NFTSVADTDN FAITFKAAVE -GKRPGPMVD MMKTVIP--- ---------- ---------- -GVNKPAPTV A--------- ----------
271
Aqueenslandica ---------- ---------- ---------- ITSPTLSTTQ ASLTTAPETL APSSSLNGNN DKDNDDVVVN DSVTRSSSSK QVSSD-----
Opearsei ---------- ---------- ---------- -----NLSLA NDKSGKDEKG SATPQRKLSI SEPSGSLE-- ---------- ----------
Emuelleri ---------- ---------- -DSDVPTNDS KAEDDLDVTP KNVSITSLEL EATDGGVGET KADSKKLTNG APFSPSTEYS SSTMPKTPTD
Nvectensis ---------- ---------- ---------- ----PQMNGT NGKDSPAISH KPVG------ ---------- ---------- ----------
Pbachei ---------- ---------- --SFDTEPVK LRASESETAD RSRASDHSLD ---------- ---------- ---------- ----------
Mleidyi ---------- --------SS NSFDMEVPVK LRSSVTTDGE SSDHSRSGVQ ENGLE----- ---------- ---------- ----------
Mmusculus1 ---------- ---------- ---------- -------ESI NGTDDERTPD V--------- ---------- ---------- ----------
Drerio1 ---------- ---------- ---------- -PVTADLCRI NGTGDDVTLS DV-------- ---------- ---------- ----------
Mmusculus2 ---------- ---------- ---------- -----QASSV NGTDDEKASH AS-------- ---------- ---------- ----------
Drerio2 ---------- ---------- ---------- -----QPSSI NGTDDEKISH SG-------- ---------- ---------- ----------
Mmusculus3 ---------- ---------- ---------- ---------T NGPGEEKLFR SQ-------- ---------- ---------- ----------
Drerio3 ---------- ---------- ---------- -RQSPPLLGV NGSSEDKLFR SK-------- ---------- ---------- ----------
Amellifera ---------- ---------- ---------- ----VISTQN SVSLSTESPQ HQGKSAQLTT QGGQ------ ---------- ----------
Dpulex ------RSAM ATVGLGSGAT PTAFTNDDKE PMTVKQEYQT DGSTSPVSPA LLKSALVTHQ RSQSLSGLQP PADSPKHQQK IDKNSPGSSN
Bfloridae RKAKEKQQSD ITSVQSDEIT TEPSKPKESN QEEEFIEKFQ DAKQSARQEV ERIRAQSGGS TPQTPHTPQK IPTVNGTATS QGSGTFPRDH
OfaveolataHomer ---------- ---------- ---PSVGHKT SSLKSNSSAS SGEEESKESS GISHKLIG-- ---------- ---------- ----------
EpallidaHomer ---------- ---------- -----GTHTS SMKSGSSTSS NEEEREGISH KPVG------ ---------- ---------- ----------
SpistillataHomer ---------- ---------- ---PGVGHKS PSVKSNSSAS SGEDESKESA GISHKPIG-- ---------- ---------- ----------
LanatinaHomer ---------- ---------- ---------- ---SVSSNQS DGSEKQQRLS TSGLP----- ---------- ---------- ----------
Skowalevskii ---------- ---------- ---YIAPNVK SEANEQREDD SGSDSGSRPA SMVSTGGSSS NAQP------ ---------- ----------
Spurpuratus ---------- ---------- ---------- --EINHAKQS SGDSTGSIPN PVHIMA---- ---------- ---------- ----------
SrosHomer ---------- ---------P AKEKESAKPN KPSSASTGDA IGSPVDS--- ---------- ---------- ---------- ----------
AspectabilisHomer GTSERAAAGS AEGAPSS--- ---------- ---------- ---------- ---------- ---------- ---------- ----------
SdiplocostataAUHomer ---------- ---------- ---------- -LSSTPSAKA SGGDQ----- ---------- ---------- ---------- ----------
HnanaHomer GGADVASAAG AAAPSS---- ---------- ---------- ---------- ---------- ---------- ---------- ----------
SroanokaHomerwithtrans ---------- -------EAT KSKENTQAKP QPTGDSIGPT VDP------- ---------- ---------- ---------- ----------
SparvaHomer GGGASS---- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
ShelianthicaHomer ---------- ---------- ---------E EAQAASASEP ETRDQATSEA ---------- ---------- ---------- ----------
MfluctuansHomer ESVTDDQQRA AEA------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Cowkzarzaki ---------- ---------- ---------- ---------- ---------- ---------- ---------Q QPQQAAPQTQ QQPAAAAKAA
Patlantis ---------- ---------- ---------M AAASVAPAAS AAAA------ ---------- ---------- ---------- ----------


361
Aqueenslandica ---------- ---------- ---------- ---AQVESLK FENSKLKVAL ASSAQNV-K- KWESEMQ--- ---TLRNSNT -M-L-KTALE
Opearsei ---------- ---------- ---------- ---TQLQELR HENDRLKHAL AASSCHR-D- KWEAERQ--- ---KLTNNNT LLRTGLQESK
Emuelleri TSE------- ---------- ---------- ---SQVEGLR FENSRLKVAL ASSAQNV-K- KWETEMQ--- ---TLKNNNV QLTAALEESR
Nvectensis ---------- ---------- ---------- ---NTEAQLK YENDRLKMAL AQSSGNA-K- KWEAELQ--- ---TLKNNNA RLTAALQEST
Pbachei ---------- ---------- ---------- ---DAMKGLS LENDRLKTSL NQSQGNL-K- KWEVELH--- ---TLRTNNS RLQSALHENS
Mleidyi ---------- ---------- ---------- ---NMIENLR LDNDRLKSSL SQSQNNL-K- KWEVELH--- ---TLRTNNS RLQGALHENS
Mmusculus1 ---------- ---------- ---------- ---TQNSEPR AEPTQNALPF PHSSAIS-K- HWEAELA--- ---TLKGNNA KLTAALLEST
Dreriol ---------- ---------- ---------- ---SPNSMLQ SDLSPTTPTL AHSPTTN-K- HWEAELA--- ---ALKSNNA KLTAALLEST
Mmusculus2 ---------- ---------- ---------- ---PADTHLK SENDKLKIAL TQSAANV-K- KWEMELQ--- ---TLRESNA RLTTALQESA
Drerio2 ---------- ---------- ---------- ---PETAALK AE-------- ---------- ---------- ---------- ----------
Mmusculus3 ---------- ---------- ---------- ---SADTPGP TERERLKKML SEGSVGE-V- QWEAEFF--- ---ALQDSNQ RLAGALREAN
Drerio3 ---------- ---------- ---------- ---SADMEQT TEKEQMRKML SEGSICE-I- NSEAELF--- ---SLRDSNT KLVAALHDAN
Amellifera ---------- ---------- ---------- ---SAEMQLK YENDRLKLAL AQSSANA-K- KWEVELT--- ---TLKTNNA RLTSALQEST
Dpulex NGAAVGGVAV AASLSGVGSS ATLPSQGSHL PLGSVEAQLK YENDRLKLAL AQSSANA-K- KWEVELA--- ---TLKNNNA RLTSALQEST
Bfloridae RSSQSLNLPP ENPESFQDRS DAKSDNGPL- ---TADSQLK YENDRLKIAL AQSSANA-K- KWEVELQ--- ---TLKNNNT RLTTALQEST
OfaveolataHomer ---------- ---------- ---------- ---NTESHLK YENDRLKMAL AQSSANA-K- KWETELQ--- ---TLKNNNA RLTAALQEST
EpallidaHomer ---------- ---------- ---------- ---NTEAQLK YENDRLKMAL AQSSTNA-K- KWETELQ--- ---TLKNNNA RLTAALQEST
SpistillataHomer ---------- ---------- ---------- ---NTESHLK YENDRLKMAL AQSSANA-K- KWEAELT--- ---TLKNNNA RLTAALQEST
LanatinaHomer ---------- ---------- ---------- ---TVEYQLK YENDRLKIAL AQSSANA-K- KWEIELQ--- ---TLKNNNA RLTTALQEST
Skowalevskii ---------- ---------- ---------- ---ATDAQLK YANDRLKIAL AQSSANA-K- KWETELH--- ---TLKNNNA KLTTALQEST
Spurpuratus ---------- ---------- ---------- ---SSDAQLK YENDRLKIAL AQSSNNA-K- KWELELQ--- ---TLKNNNA RLTTALQEST
SrosHomer ---------- ---------- ---------- ---QEVATLR YENERLRAAL ATSSANA-K- KWEEELQ--- ---NLRNTNT RLKTALQESG
AspectabilisHomer ---------- ---------- ---------- ---DVLAQLK YENERLKIAL GTSATNA-K- RWEQELQ--- ---TLKNNNA RLKTALQESA
SdiplocostataAUHomer ---------- ---------- ---------- ---TQVQELR YENERLKIAL STSSANA-K- EWVQELS--- ---TLKNNNA RLKVALQESA
HnanaHomer ---------- ---------- ---------- ---DVLAQLK YENERLKIAL GTSATNA-K- RWEQELQ--- ---TLKNNNA RLKTALQESA
SroanokaHomerwithtrans ---------- ---------- ---------- ---QEIATLK YENERLRAAL TTSSANA-K- KWEEELQ--- ---TLRNTNT RLKTALQESG
SparvaHomer ---------- ---------- ---------- ---DVLAQLK YENERLKIAL GTSATNA-K- RWEQELQ--- ---TLKNNNA RLKTALQESA
ShelianthicaHomer ---------- ---------- ---------- ---SQLATLR YENERLRAAM ATSAAHA-K- QWEQELA--- ---TLKNNNA RLTAALQDSA
MfluctuansHomer ---------- ---------- ---------- ---SQLATLR YENERLRAAM ATSAAHA-K- QWENELA--- ---TLKNNNA RLTAALQDSA
Cowkzarzaki TNGTHAPASA AGSNGADANA EDDSNAVALY DFAGKDETQL GFKAKQKLYD VTSTT---RE GWSYGSI-LN EDGVTKSSGL FPTSYIQITK
Patlantis ---------- ---------- --------KP AATNNFASNN -------VYA N-------KS VSSGPSASMT P--------- --------SV
451
Aqueenslandica ESAQHVTEWK KQLQ-KYKDE CDQLKKKIVK QEDNS----- ---REIESVT KEN------- ---------- ---------- ---EELKEVL
Opearsei MNVEEW-K-S -KMKQQQEEN QKLKQRVQEL -EASE---ES RVSESTGHLQ KEK------- ---------- ---------- ---EELEKRV
Emuelleri QHVTEW-K-K -QLQKYKEES EALRKKVSEM -ESSMKEKSY QSVEETDVLK KEN------- ---------- ---------- ---EELKEVL
Nvectensis ANVEEW-K-K -QLNAYKEES SKLRKKVQEL -ESSD----- MNADVFNNMQ QEK------- ---------- ---------- ---MAFEGRI
Pbachei RNIEEW-K-T -QLLQYREEN TKLRSRNEEL -EIEI----- ---DNLGHIR NGT------- ---------- ---------- ---HSPKTAK
Mleidyi RNIEEW-K-S -QLLQYREEN GKLRSRNEEL -EIEI----- ------ENLN RSG------- ---------- ---------- ---NHSPKLS
Mmusculus1 ANVKQW-K-Q -QLAAYQEEA ERLHKRVTEL -ECVS----- ---SQANAVH SHK------- ---------- ---------- ---TELNQTV
Dreriol ANVKQW-K-Q -QLAAYQEEA ERLHKRVSEL -ESQS----- ---SQTNVIK SQK------- ---------- ---------- ---TELNQTI
Mmusculus2 ASVEQW-K-R -QFSICRDEN DRLRSKIEEL -EEQC----- ---SEINREK EKN------- ---------- ---------- ---TQLKRRI
Drerio2 SSAESR-K-T -QATRSLEEN NQLRSKIAEL -ELQC----- ---KELSVER ERNAQLTVRI -QELEADLQD KQQELENLRK QAEIIPQLMA
Mmusculus3 AAATQW-R-Q -QLEVQRAEA ELLRQRVAEL -EAQV----- ---AVEPVRA GEK------- ---------- ---------- ---EATSQSV
Drerio3 SSIEQW-K-K -QLAEYQEET ERLREQVAEL -EAQS----- ---GVSAFSE MGN------- ---------- ---------- ---EELTQSI
Amellifera ANVDEW-K-R -QLQLYKEDN ARIKAKYADL -EAGK----- -------IAE GNS------- ---------- ---------- ---EALRLRV
Dpulex ANVEEW-K-R -QLQAYKDEN QRVKSKIVEL -EAGR--GPP ITGEGGNDLK REV------- ---------- ---------- ---ATLRLKI
Bfloridae ANVEEW-K-K -QLAAYKEEN ESMKKKVLKM -EGAG----- -VQDQMDAVL IEK------- ---------- ---------- ---TQLAEEV
OfaveolataHomer TNVEEW-K-K -QLNAYKEEN IKLRKKVQEL -ESSD----- QVSERFNVVQ QEK------- ---------- ---------- ---LTLDGRV
EpallidaHomer ANVEEW-K-K -QLNAYKEES LRLRKKVQEM -ESSD----- VNTDMFNSLT QDK------- ---------- ---------- ---MALEGRI
SpistillataHomer ANVEEW-K-K -QLNAYKEEN GKLRKKVHEL -ESSD----- QLSEKYNSLT QEK------- ---------- ---------- ---LTLDGRV
LanatinaHomer ANVEEW-K-K -QLAAYKEES AKLKKQTIEL -EQTA----- -NDEQISGLQ QQL------- ---------- ---------- ---KETTEQL
Skowalevskii ANVEEW-K-R -QLAAYREES ARLKKKILEL -ESGS----- VDNSRIEELE KEK------- ---------- ---------- ---TSLESRL
Spurpuratus SNVEEW-K-R -QLSAYKEEN VRLKKRMKEL -EATV----- --EPEAQNGH SDT------- ---------- ---------- ---AELESRV
SrosHomer KNVLQW-K-K -QLAAWKDET TRLKKLLKEK -EESS----- ---STPDASS TAS------- ---------- ---------- ---QPDQEVV
AspectabilisHomer SNVEEW-K-T -QLTAWKEEC IRLRAQVKEL -RASG----- ---------- ---------- ---------- ---------- ----------
SdiplocostataAUHomer ANVNEW-K-E -QLTAWKEES TRLKARNKEL -EAAA----- ---------- ---------- ---------- ---------- ----------
HnanaHomer SNVEEW-K-T -QLTAWKEEC IRLRAQVKEL -RASG----- ---------- ---------- ---------- ---------- ----------
SroanokaHomerwithtrans KNVIQW-K-K -QLAAWKDET TRLKAALKEK -EEES----- ---------- NSS------- ---------- ---------- ---KPDTELV
SparvaHomer SNVEEW-K-T -QLTAWKEEC IRLRAQVKEL -RAQG----- ---------- ---------- ---------- ---------- ----------
ShelianthicaHomer ANVEEW-K-T -QLAMWKEEC ARMKKRVAEV -EA------ ---------- ---------- ---------- ---------- ---AERDAAA
MfluctuansHomer ANVEEW-K-T -QLSTWKEEC ARLKRRAQDA -EALA----- ---------- RDA------- ---------- ---------- ---TEREAAL
Cowkzarzaki AASANA-S-N -SAISSALA- ---K-RAQEA QEQQLKQQQA EKLQQEKLQQ QQLLLQQQQQ EKLQQERLLQ QQQQQQQQQQ QLLLQQQQQQ
Patlantis SANEVA-L-T KQVA---DLT AKLQ-K-NE- QELMA-AKQE IESLSITSAS FDALQ----- ---KQTGVMQ AEN------- ----------


541
Aqueenslandica SAVQTDLQDK RTE----VTT LQA------- -----SLKAS QDKL------ ---------- ---------- ---------- -------QEA
Opearsei REMEVTLREK DKE----LST LTLHATEFAS FKT--QNEHL KHQF------ ---------- ---------- ---------- -------EKL
Emuelleri SAVQLDLEAG TEE----LTK KNELLNQTKE -----QLKTA TLSI------ ---------- ---------- ---------- -------QKL
Nvectensis IHLESMVSDK DEE----ISQ LKG------- -----TNADL SSQI------ ---------- ---------- ---------- -------KES
Pbachei KHSKSNLK-- SEE----VER VKQ------- -----ILEKV TAEL------ ---------- ---------- ---------- -------VEV
Mleidyi NKKMSKSSLK SEE----VER VKQ------- -----ILEKV TAEL------ ---------- ---------- ---------- -------VEV
Mmusculus1 QELEETLKVK EEE----IER LKQEIDNARE LQE--QRDSL TQKL------ ---------- ---------- ---------- -------QEV
Drerio1 EELECTLRAK EEE----MMR LKEEVENAKQ LQA--QKDEL TKKL------ ---------- ---------- ---------- -------EDT
Mmusculus2 EELESEVRDK EME----LKD LRKQSEIIPQ LMS--ECEYV SEKL------ ---------- ---------- ---------- -------EAA
Drerio2 --ECESLSGK L--------- ---------- ---------- ---------- ----QAAELS SSGLRERVEL LQGEA----- ----------
Mmusculus3 EQLEARVQTK DQE----IQT LKNQSTGTRE APDTAEREET QQQV------ ---------- ---------- ---------- -------QDL
Drerio3 AEMEALIKAK DEE----IKV LRNNKPDFSE LEK--EKQEA SQRM------ ---------- ---------- ---------- -------QDL
Amellifera EALESELRTK NEE----IKA L--------- -----TIATK NKDF------ ---------- ---------- ---------- -------VAL
Dpulex DSLETELRAK DDE----IRR LSNEMREMSM SADRVKRCQL IAQSASTV-- ---------- ---------- ---------- -SHRKGDSKL
Bfloridae ERLKTAGQEK DRE----LER LQEEVEALKM RDA--KNQGI QQRLLRTAIS VALVSNNCNS NGTEAKMLLF DAPMQPGIVV QESPVVGDTC
OfaveolataHomer IQLESMIRQK DED----MGN LIQ------- -----SNAEL TNQN------ ---------- ---------- ---------- -------QTL
EpallidaHomer IHLESVTRER EEE----VAL LRH------- -----TNTEL TSQV------ ---------- ---------- ---------- -------KVC
SpistillataHomer IQLESMIRQK DED----ITS LRQ------- -----AKAEL TNQN------ ---------- ---------- ---------- -------QTL
LanatinaHomer QAVEQERDST TEE----LRG ALKRLEELNI RVE--QNGKN EAHI------ ---------- ---------- ---------- -------KDL
Skowalevskii SMVENDNKTQ IQE----VAR LTNRIEELSI FES--QNKSL ETKM------ ---------- ---------- ---------- -------KSL
Spurpuratus QELTLAHKNK DQE----VEG LSKRLEEVSG LAE--ENTRL SSEL------ ---------- ---------- ---------- -------EKT
SrosHomer TKLEEEVAAA KAE----LAE VRQQLEASQN A----EKSAV DQQ------- ---------- ---------- ---------- -------QEA
AspectabilisHomer ------GSAA AAE----VDT LKE------- -----QLAQL QGQV------ ---------- ---------- ---------- -------AAL
SdiplocostataAUHomer ----GNGHAT NAE----LDA KTA------- -----ECEEL KARI------ ---------- ---------- ---------- -------AQL
HnanaHomer ------GSEA AAE----AET LKD------- -----ELSRL RGRI------ ---------- ---------- ---------- -------SSL
SroanokaHomerwithtrans AKLEEELAAT KQE----LED MRKELEASQA A----EKAAI NTQ------- ---------- ---------- ---------- -------QDA
SparvaHomer ------GSAA KAE----VES LRS------- -----EVAEL EGRC------ ---------- ---------- ---------- -------ADL
ShelianthicaHomer ATATKELEAK QAE----CDR LARQVQDQRE -----AGERQ GVRV------ ---------- ---------- ---------- -------REL
MfluctuansHomer RAKEGDNKH- QLE----IER LSKQAADLRE -----ANDRH VLRT------ ---------- ---------- ---------- -------REL
Cowkzarzaki QLQQQQLQQQ QLQQQQQLQQ QRVATGSAAQ FPPVNPRGGG SEQLDSEKDV IIKHLRERVE NSERAKAAAD ASNREIIETL QGDIASFKN-
Patlantis ---------- ---DQLKAQI RQLNEQLAKA N--------- ---------- -----LNAGL EQQL------ ---------- ----------
631
Aqueenslandica TATNAILESQ LSDIKLEL-- ---------- ---------- ---------- ---------- ---------- ------NNLK SSHQSSLSLA
Opearsei EMDNADLSDK VSDMEDKL-- ---------- ---------- ---------- ---------- ---------- ------QVTQ TTADLRIERI
Emuelleri EGKNIELTQT VTQLEEQF-- ---------- ---------- ---------- ---------- ---------- ------KMAA VASKRKMDNM
Nvectensis KVKNQELEQR VKDLEERM-- ---------- ---------- ---------- ---------- ---------- -------SQA SVNQNAKEQL
Pbachei KEENTRLKTQ LRQSEERS-- ---------- ---------- ---------- ---------- ---------- --------GV SKQVYENSMV
Mleidyi KEENMRLKGQ LRTQDDRN-- ---------- ---------- ---------- ---------- ---------- --------GQ SKQAYENSMI
Mmusculus1 EIRNKDLEGQ LSDLEQRL-- ---------- ---------- ---------- ---------- ---------- ------EKSQ NEQEAFRSNL
Drerio1 DMRNKELEEQ LSDLEQRL-- ---------- ---------- ---------- ---------- ---------- ------ESSQ VDQENLRKNL
Mmusculus2 ERDNQNLEDK VRSLKTDI-- ---------- ---------- ---------- ---------- ---------- ------EESK YRQRHLKGEL
Drerio2 ---------- ---------- ---------- ---------- ---QHRQQQQ NSLKTELKLL MEQLDGKIDE LHEVRQGISK LGIDN-----
Mmusculus3 ETRNAELEQQ LRSMECNL-- ---------- ---------- ---------- ---------- ---------- ------EEAR AERERARAEV
Drerio3 EAKNSELEKQ VETAENAL-- ---------- ---------- ---------- ---------- ---------- ------ASSK EAQSRAETAV
Amellifera QKENQELREM LRVVHEQL-- ---------- ---------- ---------- ---------- ---------- ------ELAL SANKVQKQNL
Dpulex EKRKGYDKRS MTALSAALWP LCQGRWSVLA GAAYHLYVAF HSYSRRGKLS AGSERRIGAE PTGCHERSAR HPTAHPGITQ RPARRQSERL
Bfloridae EQDKRDLQQR VSTLEGQL-- ---------- ---------- ---------- ---------- ---------- ------DTVL CDHRRDAQRV
OfaveolataHomer QMGNQTMQMK IEELTSQM-- ---------- ---------- ---------- ---------- ---------- -------SAA ALDETRVQEL
EpallidaHomer REENEGLQKK VQXLEVKL-- ---------- ---------- ---------- ---------- ---------- --------SE SNLQASAEQL
SpistillataHomer QLDNERLQLK TEEMSLHM-- ---------- ---------- ---------- ---------- ---------- ------N-ET ALNETRVKEL
LanatinaHomer EKENSTLRKK VSSLESEL-- ---------- ---------- ---------- ---------- ---------- ------KQLK LQQSKDRHEL
Skowalevskii EENNQTLQDT VRELNNQL-- ---------- ---------- ---------- ---------- ---------- ------EHCK EKEEQQSHEL
Spurpuratus TSNVDALTSR LTDLEEKL-- ---------- ---------- ---------- ---------- ---------- ------QQER ATREENSTQM
SrosHomer EATVVQLKEQ IKTLTAAA-- ---------- ---------- ---------- ---------- ---------- ------E-YQ SKYERLAEDM
AspectabilisHomer EGDKSRLESK VMAESGGR-- ---------- ---------- ---------- ---------- ---------- ---------- ---RKVGDQL
SdiplocostataAUHomer EGSNRSMAMQ SQQANASV-- ---------- ---------- ---------- ---------- ---------- ---------- GRMERCQKDL
HnanaHomer EAEKSALETQ TMAASGGK-- ---------- ---------- ---------- ---------- ---------- ---------- ---RQMADQI
SroanokaHomerwithtrans LARVQQLEEQ VKTLSQAA-- ---------- ---------- ---------- ---------- ---------- ------E-YR PKYERLAEDM
SparvaHomer ESENATLEAE RRAHAGGR-- ---------- ---------- ---------- ---------- ---------- ---------- ---RRTAEQL
ShelianthicaHomer EGTVRRLEDE LRATQGAS-- ---------- ---------- ---------- ---------- ---------- ---------- EQQRAMLSGL
MfluctuansHomer DARVKDLEDQ LRVALSGG-- ---------- ---------- ---------- ---------- ---------- ---------- DKARRLAQSA
Cowkzarzaki ----ENDRLL AAL------- ---------- ---------- ---------- ------KESN LSGQQWETEL KTVKTNN--- ----------
Patlantis ---------- -------KAA KDDVARLKAA LLQNSTNA-- ---------- ---------- ---------- ---------- ----------

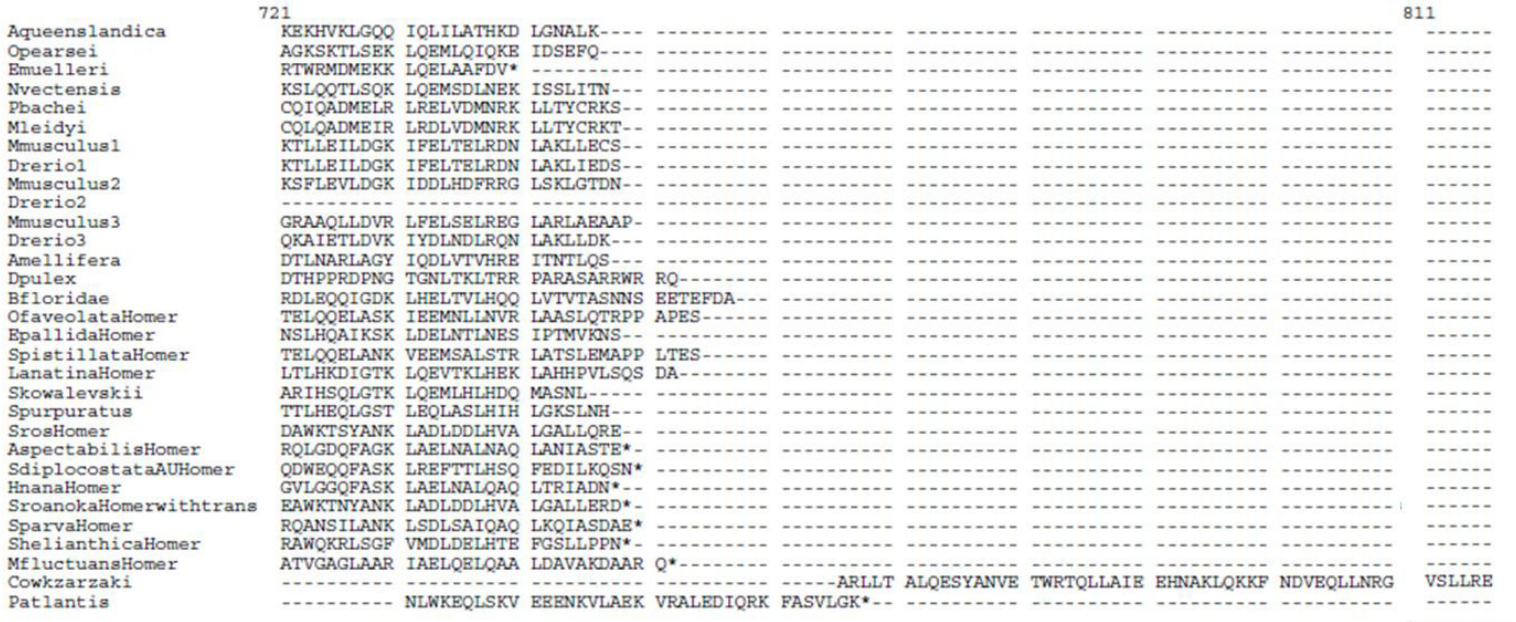


Figure 6-2: Full-length Homer Alignment. In grey conserved areas that were used for phylogeny reconstruction (alignment main text Figure 3-2). For the alignment in Figure 3-2 rat Homer sequences were subsequently added to the alignment (because rat and not mouse Homer was used as a positive control), and ctenophore sequences, as well as *B. floridae*, *L. anatina*, *S. kowalevskii*, *C. owkzarzaki* and *P. atlantis* were taken out. The sequences were re-ordered but the alignment was not edited further. The alignment was also base for the alternative phylogenies (section 6.2.4), altering only the length of the alignment used and which sequences were included for phylogeny reconstruction. *Sroanoka = Mroanoka*

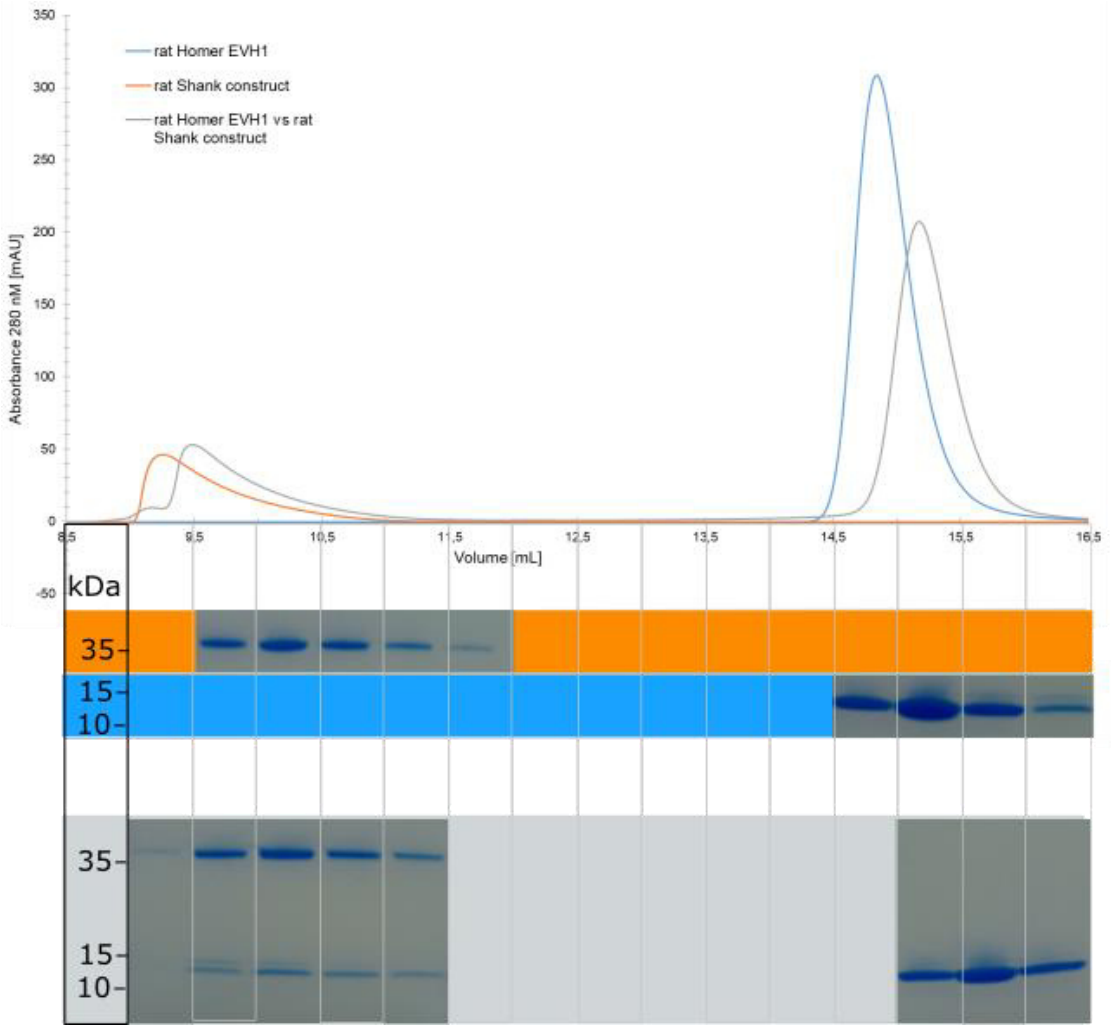


Figure 6-3: Size exclusion chromatography shows weak binding between rat Homer EVH1 and a rat Shank construct. Chromatographic elution profiles for rat Homer EVH1 (individually, orange curve), rat Shank construct (individually, blue curve) and a combination of both proteins (grey curve). A combination of Homer and Shank elutes in the same way as the individual proteins, with a very small additional peak (the combined curve is slightly shifted due to a small volume difference, but they seem congruent). Elution fractions were collected and analysed by SDS PAGE (shown below the elution graph). Colours surrounding the gel photographs are corresponding to curves with orange showing the gel for the individual Shank construct, blue showing the gel for individual Homer EVH1, and grey showing the gel for the Homer-Shank combination. The fractions are shown corresponding to volumes of the shifted grey curve to make them comparable. The Homer and Shank mixture (pre-incubated at room temperature for 1 hour) shows a co-migration of Homer EVH1 with rat Shank (as shown on the gel), and indication for their capacity to bind.

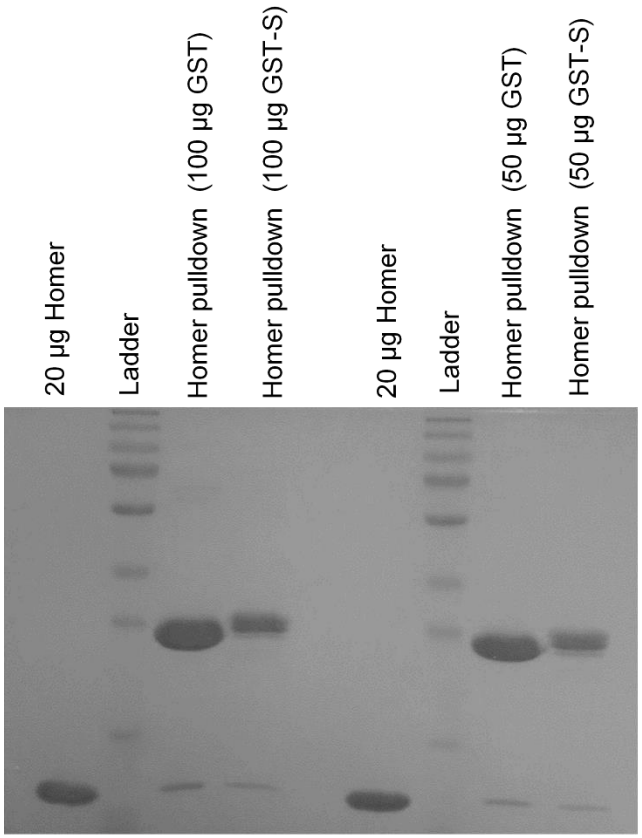


Figure 6-4: GST pull-down experiment with *S. rosetta* Homer EVH1 and *S. rosetta* Shank peptide. GST (negative control) vs. GST-S (*S. rosetta* Shank peptide attached to GST). Homer EVH1 is pulled down from both GST and GST-S, indicating unspecific binding of Homer EVH1 to GST-beads or some contamination of the pull-down sample with beads. It seems that *S. rosetta* Homer-Shank interaction cannot be shown under these conditions. Ladder = protein standard.

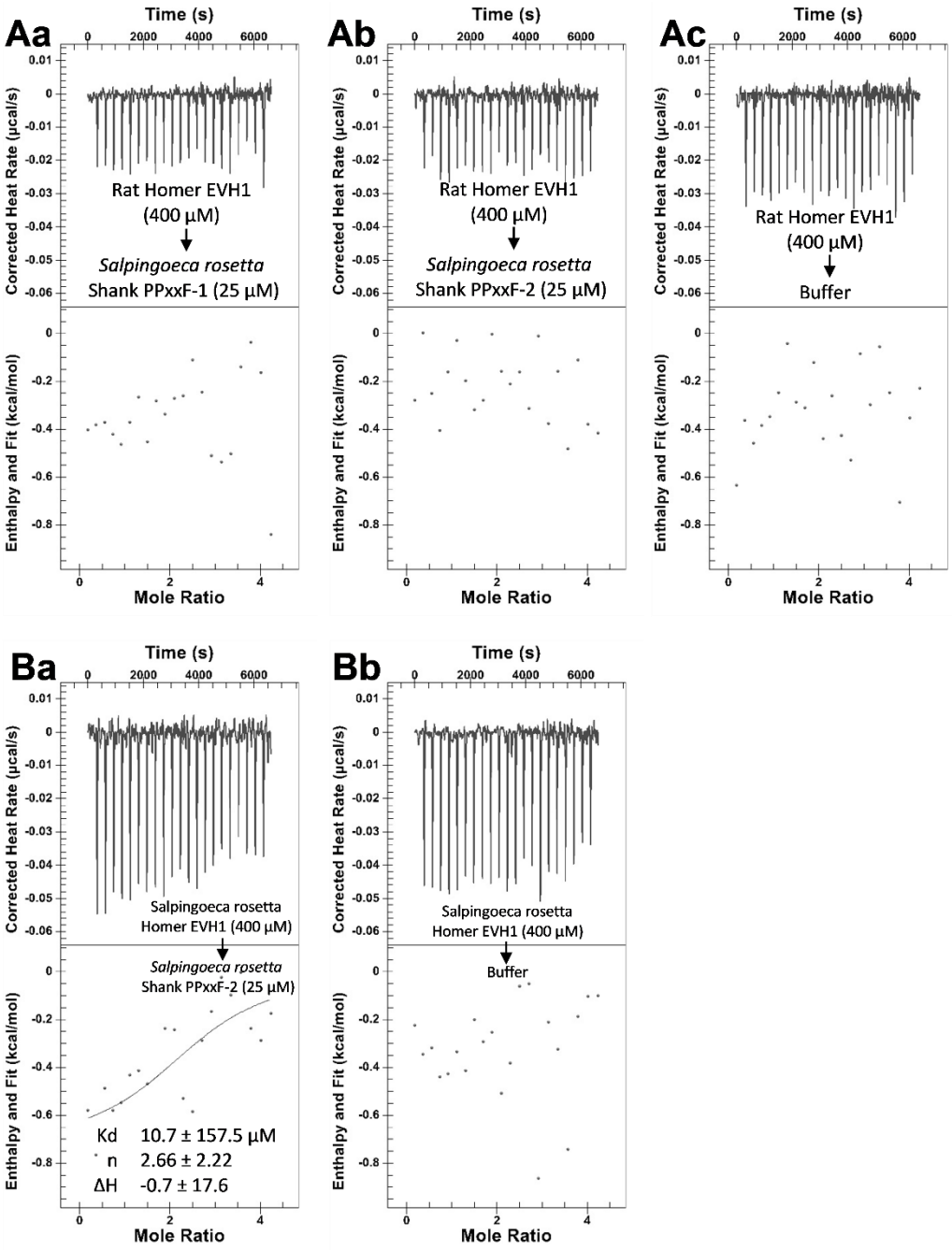


Figure 6-5: Rat Homer EVH1 does not bind to *Salpingoeca rosetta* Shank PPXXF motifs. All diagrams show calorimetric titrations; the upper panel of the diagram describes the corrected heat rate in µcal/s over the time course of the incremental titration. The lower panel describes the integrated areas normalised to the amount of Shank vs the molar ratio of Homer to Shank. The solid line represents the best fit to the data using a single site model. A) Rat Homer EVH1 titrations to different *S. rosetta* peptides with PPXXF motif vs buffer. Aa) Titration of rat Homer EVH1 into S. rosetta Shank peptide I. Ab) Titration of rat Homer EVH1 into *S. rosetta* Shank peptide II. Ac) Control titration of rat Homer EVH1 into experimental buffer. B) *S. rosetta* Homer EVH1 titration at the same concentration (concentrations too low, but at least some tendency of binding can be shown). Ba) Titration of *S. rosetta* Homer EVH1 into *S. rosetta* Shank peptide II. Bb) Control titration of *S. rosetta* Homer EVH1 into experimental buffer. All titrations: Homer EVH1 concentration 400 µM; peptide concentration 25 µM; 22 injections, first injection 0.5 µl (not included in fit), 21 injections of 1.9 µl. Kd = equilibrium dissociation constant describing binding affinity (the smaller the value, the higher the affinity); n = stoichiometry of binding with n=1 describing a 1:1 interaction; ΔH = free enthalpy difference in kcal/mol. The upper graphs show the corrected heat peaks following each titration; the lower graphs show the enthalpy for each titration (filled circles) and the fit to the independent model (one binding site).

6.2.2 Supplementary Table

Table 6-3: Sequence references for Homer and Shank proteins used in alignments.

Species	Protein	Sequence reference
Amphimedon queenslandica	Homer	NCBI: XP_019851701.1
Apis mellifera	Homer	NCBI: XP_006559185.1
Apis mellifera	Shank	NCBI: XP_026294872.1
Branchiostoma floridae	Homer	NCBI: XP_002609962.1
Capsaspora owczarzaki	Homer	NCBI: XP_004349458.1
Danio rerio	Homer1	NCBI: NP_001002496.3
Danio rerio	Homer2	NCBI: NP_001018470.1
Danio rerio	Homer3	NCBI: NP_957407.1
Daphnia pulex	Homer	NCBI: EFX90289.1
Ephydatia muelleri	Homer	Compagen.org: EMUE_TPEP_130911.fa (m.51385)
Exaiptasia pallida	Homer	NCBI: KXJ16109.1
Lingula anatina	Homer	NCBI: XP_013385142.1
Monosiga brevicollis	Shank	NCBI: XP_001748613.1
Mus musculus	Homer1b	NCBI: NP_001271118.1
Mus musculus	Homer2	NCBI: NP_036113.1
Mus musculus	Homer3	NCBI: NP_001139625.1
Nematostella vectensis	Homer	NCBI: XP_001637685.1
Nematostella vectensis	Shank	NCBI: XP_001634679.1
Orbicella faveolata	Homer	NCBI: XP_020620008.1
Oscarella pearsei	Homer	Compagen.org: OCAR_TPEP_130911.fa (m.8451)
Oscarella pearsei	Shank	Compagen.org: OCAR_TPEP_130911.fa (m.126548)
Parvularia atlantis	Homer	Multicellgenome.com: Nuclearia_a-Unigene.fa.transdecoder.pep (m.6481)
Rattus norvegicus	Homer1	NCBI: NP_113895.1
Rattus norvegicus	Homer2	NCBI: NP_445761.1
Rattus norvegicus	Homer3	NCBI: NP_445762.1
Rattus norvegicus	Shank1a	NCBI: AAD29417.1
Saccoglossus kowalevskii	Homer	NCBI: NP_001161569.1
Salpingoeca rosetta	Homer	NCBI XP_004998981.1
Salpingoeca rosetta	Shank	NBCI: XP_004998206.1
Strongylocentrotus purpuratus	Homer	NCBI XP_011679722.1 (annotation release 101)
Stylophora pistillata	Homer	NCBI: XP_022803393.1

6.2.3 R code

```
###Script to extract data from rst file###

setwd('/path/to/working/directory')

library(stringr)


filename <- "file1.rst"

conn <- file(file1.rst, open="r")

linn <- readLines(conn)
```

```
REQnode <- 47 # node number requires - user input - the number of the node of interest can be identified by copying the tree data in the rst file into a new file and opening it with the figtree programme. Every node will be depicted a certain number and the numbers of interest are ancestral nodes.

aalength <- 263 # length of protein - user input


matchStatement <- paste("Prob distribution at node", REQnode)


ctr1 <- 0

matched <- NA

bad <- TRUE

while(bad){

  ctr1 <- ctr1+1

  matched <- pmatch(matchStatement, linn[ctr1])

  if(!is.na(matched)) bad <- FALSE


}


nodedatastart <- ctr1 + 4


pat <- ": A"


probvalues <- matrix(NA, nrow=aalength, ncol=20) # container for output - later written to csv

colnames(probvalues) <- c("A", "R", "N", "D", "C", "Q", "E", "G", "H", "I", "L", "K", "M", "F", "P", "S", "T", "W", "Y", "V")

datapos <- nodedatastart


for(i in 1:aalength){

  st1 <- str_locate_all(linn[datapos], pat)[[1]][2]+2

  for(j in 1:20){

    probvalues[i,j] <- substr(linn[datapos], st1, st1+4)
```

```
    st1 <- st1 + 9
  }
  datapos <- datapos+1
}
```

```
write.csv(probvalues, file="file1.csv")
```

##Manually before proceeding with next script: edit csv file by adding a column at second position, which is called "Removed". Depict Y for every amino acid that was probably not present in ancestral sequence (indels that occur only in one single species), and N for all other amino acids. Save the file and proceed.

###Script to extract ancestral sequence###

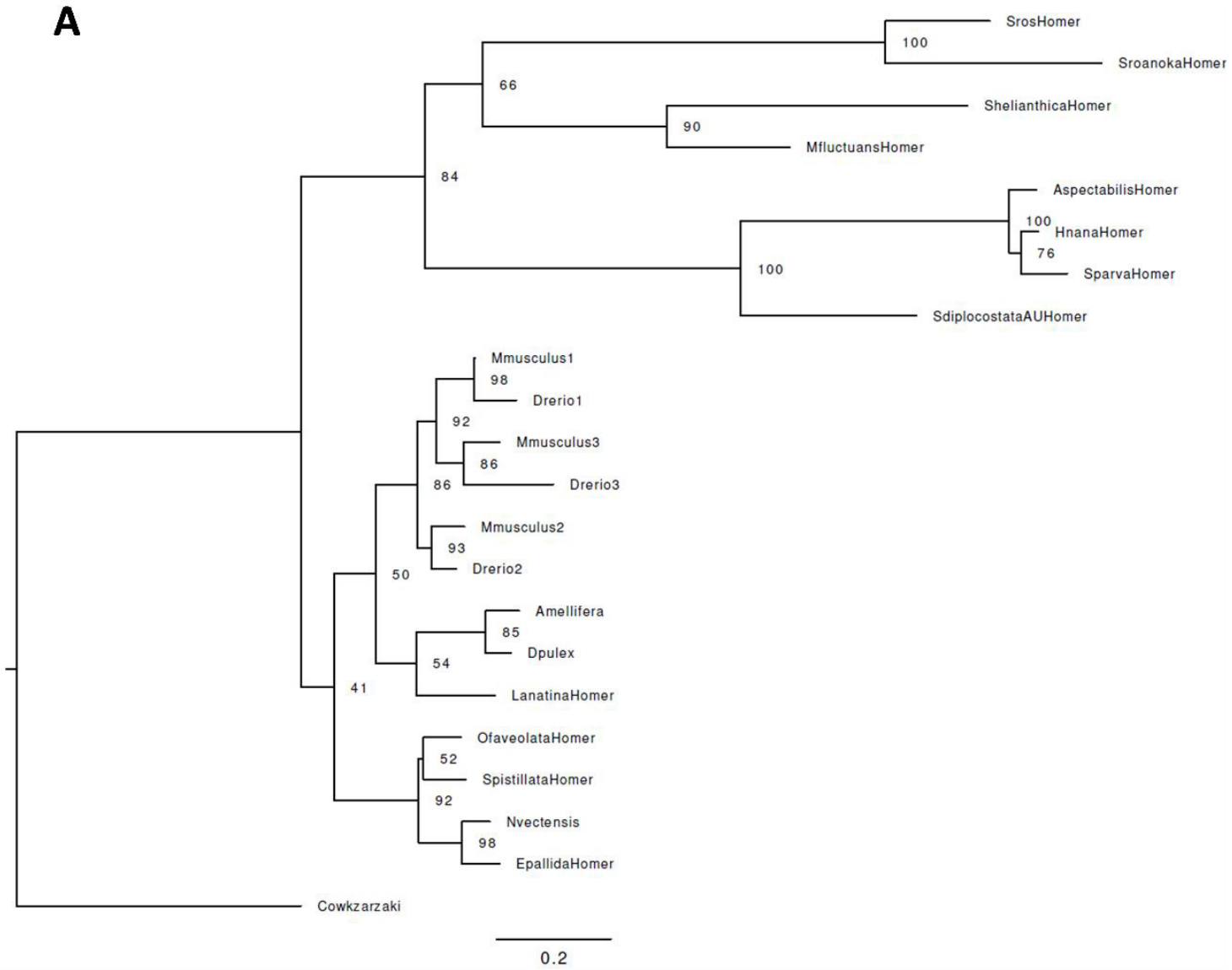
```
setwd('/path/to/working/directory')
library(stringr)

PP <- read.csv("file1.csv")
#attach(PP)
names(PP)
str(PP)

## remove sites inferred to be insertions

PP_edited <- subset(PP, Removed=="N") #retain everything where Remove equals N
PP_edited_aa <- data.frame(PP_edited[,3:22])

dim(PP_edited_aa)
which.max(PP_edited_aa[1,])

## create empty structure to store results
```

```
maxPP <- matrix(ncol=2,nrow=188) # check dim(PP) ## requires manual input, nrow number depends on number of amino acids in ancestral sequence excluding indels removed before ## also change row dimension in for loop below


## extract maximum value for each row and corresponding amino acid


for (i in 1:188) {


  maxPP[i,1] <- names(which.max(PP_edited_aa[i,]))

  maxPP[i,2] <- max(PP_edited_aa[i,])


}


## mean PP for ancestral sequence: Will give a value between 0 and 1, closer to one means better ancestral reconstruction (quality test)


mean(as.numeric(maxPP[,2]))


ASR <- maxPP[,1]


writeChar(ASR, "file1.txt") #This file will contain the most probable ancestral sequence.


rm(PP) # remove dataset to not confuse R with future dataframes




###Script to learn which sites have more than one plausible state and which one###


PP_47 <- read.csv("file1.csv") #number referring to node number that was looked at in the beginning

names(PP_47)
```

```
str(PP_47)

## remove sites inferred to be insertions

PP_edited_47 <- subset(PP_47, Removed=="N")
PP_edited_aa_47 <- data.frame(PP_edited_47[,3:22])

more_than_1_47 <- rep(0, times=dim(PP_edited_aa_47)[1]) # when there is more than one possible site, when it is identifying plausible alternative states (Thornton papers)
#which_aas <- rep(NA, times = dim(PP_edited_aa_47)[1])
for(i in 1:(dim(PP_edited_aa_47)[1])){
  tempa <- which(PP_edited_aa_47[i,] >= 0.2) # all plausible alternative states larger equals 0.2
  tempb <- which(PP_edited_aa_47[i,] == max(PP_edited_aa_47)) #best one that were already identified
  tempc <- tempa[!(tempa %in% tempb)] # plausible ones without best ones
  if(length(tempc>0)) more_than_1_47[i] <- length(tempc)
}
more_than_1_47 # file shows how many alternative states, in my case 0 (meaning only best one), 1 (meaning best one and one plausible alternative) or 2 (best one and two plausible alternatives)

which(more_than_1_47 > 1) #shows me all sites that have a problem (more than one plausible state)

PP_edited_aa_47[108,] #shows me posterior probabilities for this site. This can be done for every site with more than one plausible state in order to figure out which are plausible alternative amino acids at this site and to infer plausible alternative sequence including all plausible alternative amino acids.
```

6.2.4 Alternative Homer phylogenies

The ancestral sequences calculated depend on the topology of the phylogeny used as base for the calculation. The phylogeny utilised for tested ancestral sequences resulted in the most robust bootstrap values and it matched with the known phylogenetic relationships of the included species. In order to show how ancestral sequences change depending on the inclusion of certain groups of species, more phylogenies were calculated. All phylogenies were based on the full-length Homer alignment (Figure 6-2), using either only the Homer EVH1 domain (as marked in Figure 3-2) or the longer alignment with the EVH1 domain in combination with other conserved areas (grey areas in Figure 6-2) (long). For this comparison a fast bootstrap analysis was used (same procedure as described in section 3.2.1 with the difference that the fast bootstrap instead of the long bootstrap method was chosen. It is less accurate according to the bootstrap values.). Only alignments that resulted in reasonable phylogenies (e. g. separating choanoflagellates from animals; allowing rooting on outgroup or (if not included) on choanoflagellates) were used to calculate ancestral sequences. Including the sequence of the nuclearian *P. atlantis* in combination with the sequences of the filasterean *C. owczarzaki* never resulted in reasonable phylogenies, as the two species did not cluster together, making rooting on both of them impossible. Including *C. owczarzaki* as an outgroup only worked when sponges were taken out of the alignment, because it otherwise resulted in the clustering of choanoflagellates within animals. Sponges seemed to add valuable information to the ancestral sequence. Additionally, the *C. owczarzaki* sequence resulted in a very long branch, therefore altering the phylogeny. Therefore, for tests, sponge sequences were included to calculate ancestral sequences. However, for comparison reasons two of the alternative phylogenies (Figure 6-6) are presented here. The resulting ancestral sequences are shown in Figure 6-7. Figure 6-7 also includes ancestral sequences used for ITC tests, as well as their plausible alternative sequences (with all amino acids having other plausible alternatives of at least 20 % probability exchanged to these alternatives).


A
SrosHomer
100
SroanokaHomer
66
ShelianthicaHomer
90
MfluctuansHomer
84
AspectabilisHomer
100
HnanaHomer
76
SparvaHomer
100
SdiplocostataAUHomer
Mmusculus1
98
Drerio1
92
Mmusculus3
86
Drerio3
86
Mmusculus2
93
Drerio2
50
Amellifera
85
Dpulex
54
LanatinaHomer
41
OfaveolataHomer
52
SpistillataHomer
92
Nvectensis
98
EpallidaHomer
Cowkzarzaki
0.2

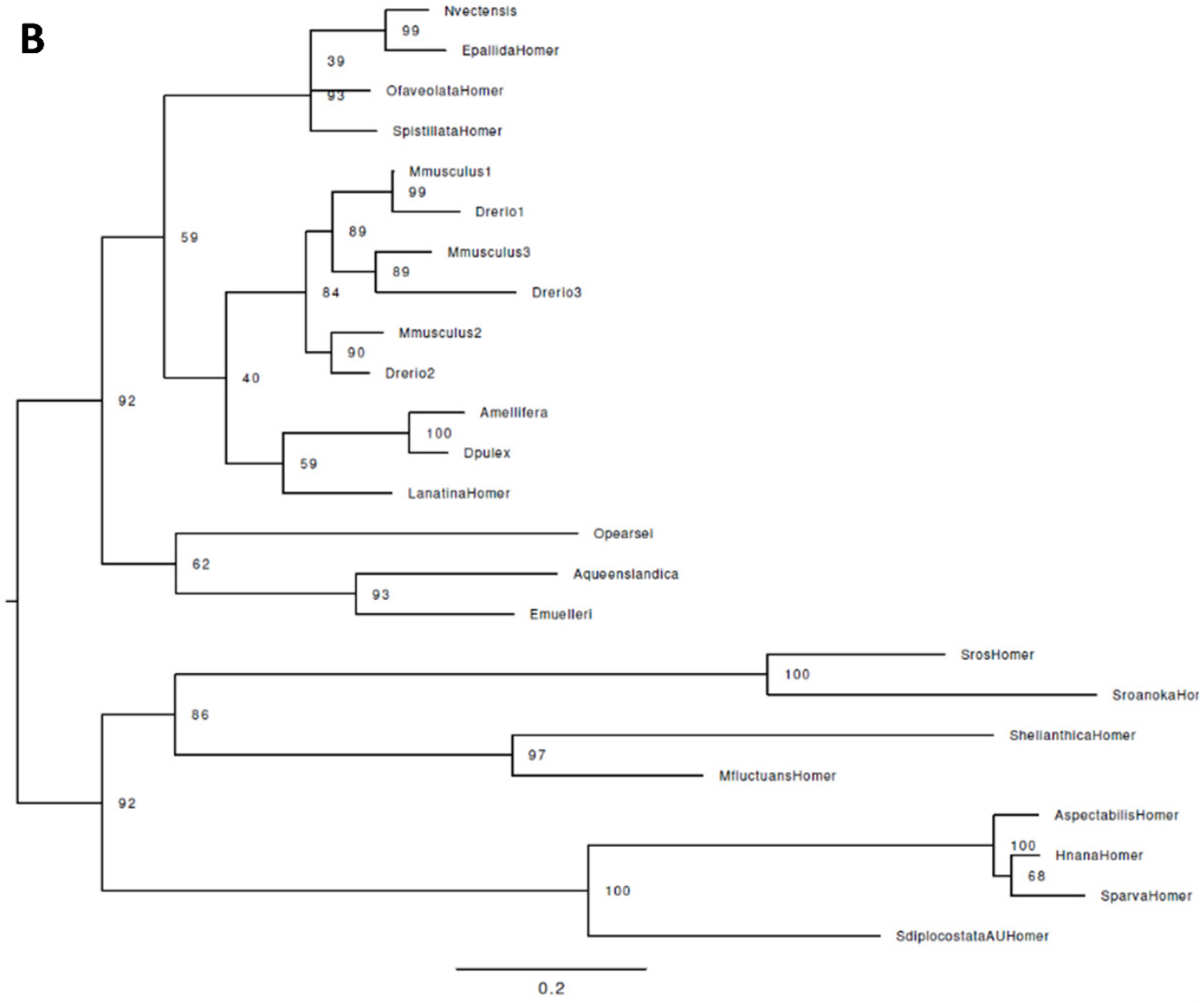

B
Nvectensis
99
EpallidaHomer
39
93
OfaveolataHomer
SpistillataHomer
Mmusculus1
99
Drerio1
89
Mmusculus3
89
Drerio3
84
Mmusculus2
90
Drerio2
59
40
92
Amellifera
100
Dpulex
59
LanatinaHomer
Opearsei
62
Aqueenslandica
93
Emuelleri
SrosHomer
100
SroanokaHor
86
ShellanthicaHomer
97
MfluctuansHomer
92
AspectabilisHomer
100
HnanaHomer
68
SparvaHomer
100
SdiplocostataAUHomer
0.2

Figure 6-6: Alternative Homer phylogenies. The phylogenies shown here resulted from the short alignment (Homer EVH1 only). A) Excluding sponges and including *C. owczarzaki* as outgroup. Maximum likelihood phylogenies were calculated with IQ tree for 108 amino acid characters and 22 taxa included. B) Including sponges and excluding *C. owczarzaki* as outgroup (rooting on animal/choanoflagellates division). Maximum likelihood phylogenies were calculated with IQ tree for 108 amino acid characters and 24 taxa included. In the phylogeny used finally, a version of B was used, however the long alignment was used and some other species were included (*R. norvegicus, S. purpuratus*) and excluded (*L. anatina*), as this improved bootstrap values. Numbers on nodes give the bootstrap support (using 1000 bootstraps). Branch lengths are proportionate to the number of nucleotide substitutions as indicated by the scale bar.

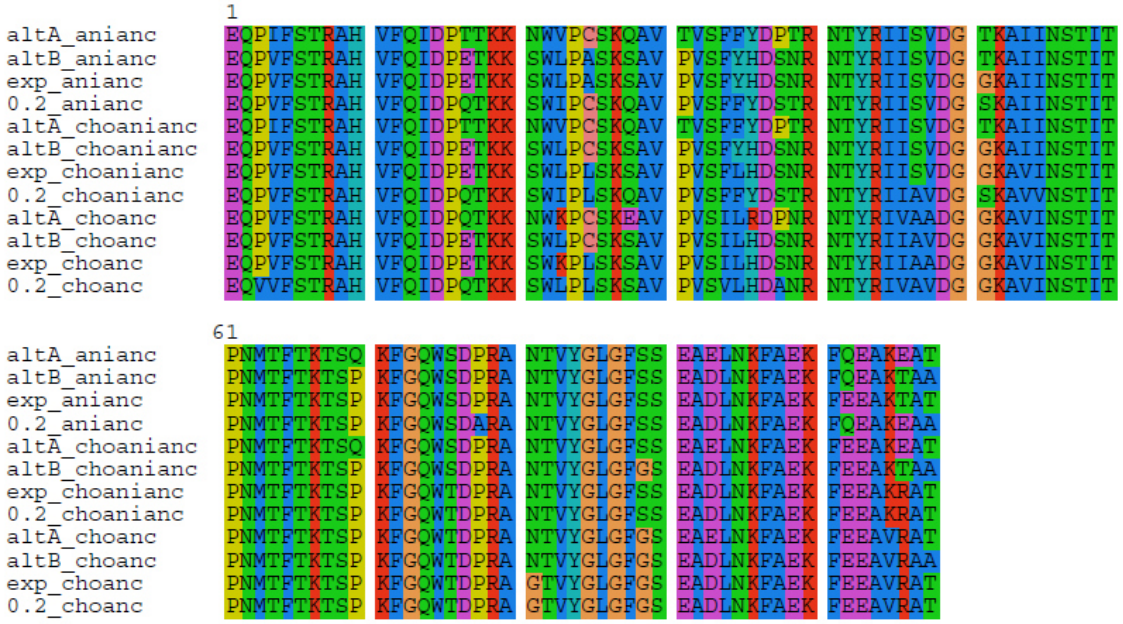


Figure 6-7: **Alternative ancestral sequences.** altA – alternative phylogeny A; altB – alternative phylogeny B; exp – sequence used in experiments; 0.2 – sequence used in experiments with all amino acids with a plausible alternatives of a probability of at least 20 % changed to this alternative; anianc = animal ancestor; choanianc = choanimal ancestor; choanc = choanoflagellates ancestor.

The alternative ancestral sequences presented here demonstrate that they are dependent on the sequences included and on the phylogeny provided for the calculation. Plausible alternative sequences already incorporate many of the changes introduced by alternative phylogenies. Using those sequences to test if binding of ancestral Homer sequences to rat Shank is still possible will therefore be a good control.

6.3 Supplementary material for chapter 4

6.3.1 Description of the behaviour of all tested *S. rosetta* Dlg antibodies on Western blots and in immunostaining experiments

Initially, twelve *S. rosetta* Dlg antibodies were tested. The antibodies were first tested in Western blots (Figure 6-8) for binding to their respective antigen (Figure 6-8 A-D). Subsequently, the antibodies were tested on full-length recombinant *S. rosetta* Dlg (Figure 6-8 E-G). Finally, they were tested on *S. rosetta* cell lysate (Figure 6-8 H-I). The two antibodies raised against the L27 domain seemed to have the lowest specificity, recognising additional bands on a Western blot, even when the recombinant antigen was applied (Figure 6-8C) and not recognising a band at the expected height of *S. rosetta* Dlg when applied to *S. rosetta* single cell lysate (Figure 6-8 I). Antibodies raised against the first two PDZ domains seemed to be generally more specific to their target antigen. Five of six of these antibodies specifically bound to the target antigen, as well as to full-length recombinant *S. rosetta* Dlg (Figure 6-8 A, B, E, F). As described in the main text they also bind to native *S. rosetta* Dlg in single cell lysate, however, specificity and affinity of the antibodies varied (Figure 6-8H).

The antibody CA5681 was shown to have high specificity towards its antigen and affinity in comparison to other antibodies tested and was therefore used for subsequent experiments described in the main text. In immunostaining experiments, this antibody consistently showed nucleus staining in replicated experiments, but we also observed other staining patterns (Figure 6-9). Occasionally, cells with staining in the nucleus showed additional staining of the plasma membrane at the basal side of the cell (implications explained in the main text). In a single experiment with this antibody, we observed a distinct staining pattern. Some staining was visible in the nucleus, whereas the most intensive staining was punctate around the microtubules in the choanoflagellate flagellum (Figure 6-9, 3rd row). It is a possibility that under certain conditions, Dlg is localised inside the flagellum. Dlg proteins have been shown to be mobile within the cell and were suggested to move along microtubules via binding to kinesin motor proteins (Yamada et al., 2007). We did however not identify a protein of the kinesin family among the *S. rosetta* Dlg interaction partners. The staining could not be replicated and it might simply be an artefact. We observed similar staining patterns with some of the other antibodies (antibody CA5684, another

antibody with the PDZ1-2 antigen (Figure 6-10, 2nd row) and antibody 1742005, an antibody raised against another antigen as described below (Figure 6-11, 3rd row). However, all these stainings have in common that they show much background staining. Antibody CA5682 also stains the nucleus, replicating this staining with another antibody that on the Western blot shows high affinity to the band corresponding to *S. rosetta* Dlg (Figure 6-8 H), but seems to bind an additional protein. This protein might be of bacterial origin, as the antibody also stains bacteria. Antibodies CA5683 and CA5684 seem to be either in low concentration or have low affinity to their antigen. Antibody CA5683, in particular, seems to show stronger binding to another protein. This antibody stains a structure surrounding the nucleus, which might correspond to mitochondria, according to vital dye stainings by Laundon et al. (2019). Antibody CA5684 stains bacteria and shows punctate stainings within the cell (which might correspond to bacteria in food vacuoles or potentially endoplasmatic reticulum, which has been shown to be disconnected and spread over the whole cell in *S. rosetta* single cells (Laundon et al., 2019)) and around the microtubules within the flagellum (described above). The antibody CA5685 did neither bind recombinant nor native *S. rosetta* Dlg on Western blots. In immunostainings the antibody stained bacteria and the *S. rosetta* cytosol. Antibody CA5686 binds recombinant and native *S. rosetta* Dlg on Western blots, but also recognises another protein from *S. rosetta* cell lysate. The antibody stains bacteria and puncta within the cells, which might as well correspond to bacteria in food vacuoles or endoplasmatic reticulum.

Due to the discovery of other PDZ domain containing proteins in the isolated protein via Dlg co-IP with antibody CA5681, we attempted to generate further antibodies that were raised against other regions of the protein. We recombinantly expressed and purified poly-histidine-tagged peptides with 49 and 51 amino acids, respectively. These peptides were found in regions distinct from conserved domains in order to increase specificity. All four antibodies showed binding to the antigen (Figure 6-8 D), which was weaker with antibody 1742008. Antibodies 1742005 and 1742007 (same antigen – pep2) also showed binding to full-length recombinant *S. rosetta* Dlg (***Figure 6-8*** G), whereas the antibodies 1742004 and 1742008 (same antigen – pep1) did not, or the binding was too weak for detection. On *S. rosetta* single cell lysate, a band that might correspond to Dlg was only detected for antibody 172004.

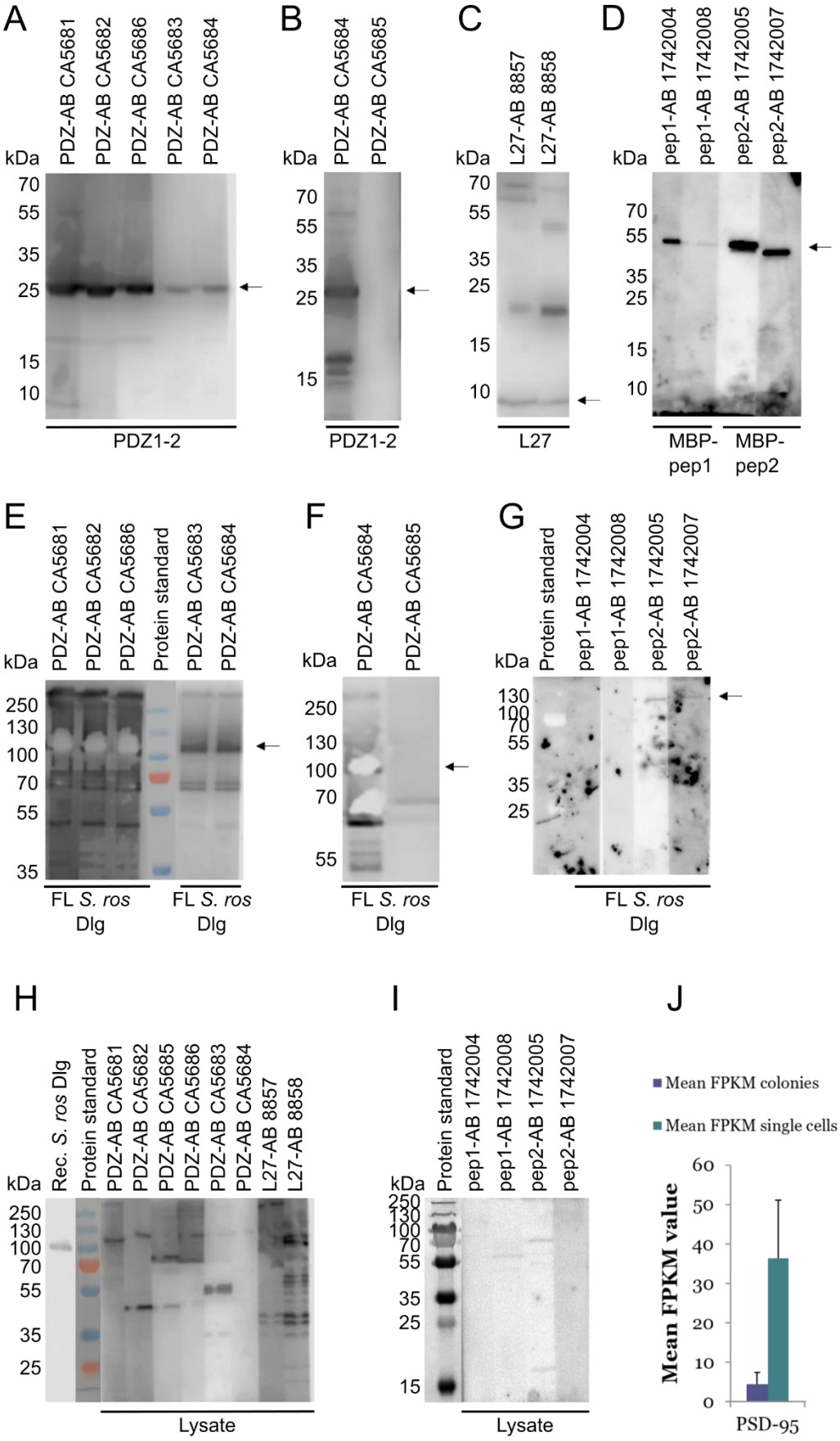

A
PDZ-AB CA5681
PDZ-AB CA5682
PDZ-AB CA5686
PDZ-AB CA5683
PDZ-AB CA5684
kDa
70
55
35
25
15
10
PDZ1-2
B
PDZ-AB CA5684
PDZ-AB CA5685
kDa
70
55
35
25
15
PDZ1-2
C
L27-AB 8857
L27-AB 8858
kDa
70
55
35
25
15
10
L27
D
pep1-AB 1742004
pep1-AB 1742008
pep2-AB 1742005
pep2-AB 1742007
kDa
70
55
35
25
15
10
MBP-pep1
MBP-pep2
E
PDZ-AB CA5681
PDZ-AB CA5682
PDZ-AB CA5686
Protein standard
PDZ-AB CA5683
PDZ-AB CA5684
kDa
250
130
100
70
55
35
FL S. ros Dlg
FL S. ros Dlg
F
PDZ-AB CA5684
PDZ-AB CA5685
kDa
250
130
100
70
55
FL S. ros Dlg
G
Protein standard
pep1-AB 1742004
pep1-AB 1742008
pep2-AB 1742005
pep2-AB 1742007
kDa
130
100
70
55
35
25
FL S. ros Dlg
H
Rec. S. ros Dlg
Protein standard
PDZ-AB CA5681
PDZ-AB CA5682
PDZ-AB CA5685
PDZ-AB CA5686
PDZ-AB CA5683
PDZ-AB CA5684
L27-AB 8857
L27-AB 8858
kDa
250
130
100
70
55
35
25
Lysate
I
Protein standard
pep1-AB 1742004
pep1-AB 1742008
pep2-AB 1742005
pep2-AB 1742007
kDa
250
130
100
70
55
35
25
15
Lysate
J
Mean FPKM colonies
Mean FPKM single cells
Mean FPKM value
60
50
40
30
20
10
0
PSD-95

Figure 6-8: Validation of 12 custom made *Salpingoeca rosetta* Dlg/PSD-95 antibodies. A-I: Western blots showing the specificity of the antibodies on recombinant antigens (A-D), on full-length recombinant *S. rosetta* (*S. ros*) Dlg (E-G) and on *S. rosetta* single cell lysate (H-I). A) Five PDZ1-2 antibodies (PDZ-ABs) bind to recombinant *S. rosetta* PDZ1-2. B) One PDZ-AB does not show any immunoreactivity towards the PDZ1-2 antigen (shown in comparison to the already overexposed blot stained with one of the other PDZ-ABs). C) The L27 antibodies (L27-Abs) show immunoreactivity towards the antigen they were raised against, but are unspecific (other bands on blot). D) The four peptide antibodies bind to their respective antigens (the antibodies were raised against the peptides attached to a poly-histidine tag; in order to exclude that immunoreactivity was towards this tag the recombinant protein used for these Western blots was expressed with a maltose binding protein tag). Bands at approx. 50 kDa correspond to the peptides (5-5.7 kDa plus the maltose binding protein (runs at 45 kDa alone). The peptide 1 antibody raised in rabbit 1742008 shows less immunoreactivity. The peptide 2 antibodies show stronger immunoreactivity. The last lane was not aligned properly to the other lanes resulting in a band that seems slightly lower than the other bands. E) Five PDZ-ABs bind to recombinant *S. rosetta* Full length Dlg. F) The antibody without immunoreactivity towards the PDZ1-2 antigen, also does not show immunoreactivity towards full length recombinant *S. rosetta* Dlg/PSD-95 (shown in comparison to the already overexposed blot stained with one of the other PDZ-ABs). G) The peptide 2 antibodies show immunoreactivity to full-length recombinant *S. rosetta* Dlg, whereas no immunoreactivity towards this protein can be observed for the peptide 1 antibodies. H) Six PDZ-Abs and two L27-Abs were incubated on blot of separated *S. rosetta* single-cell lysate proteins. First lane: CA5684 antibody reveals MW of recombinant full-length (FL) *S. rosetta* Dlg I) Four peptide antibodies (pep1 from two different rabbits and pep2 from two different rabbits) were incubated on blot of separated *S. rosetta* single-cell lysate proteins. J) Graph showing mean FPKM (Fragments per kilobase of transcript per million mapped reads) values (describing differential expression of genes) and standard deviation for *S. rosetta* Dlg in colonial versus attached single cells (data from Fairclough et al. (2013). Lysate was won from single *S. rosetta* cells in culture (col- culture). Some Western blots were cropped, and the orientation of some Western blots was changed. Original Western blots are shown in Fig. 6.12.

Notably, we tested the antibodies 1742004, 1742008 and 1742005 in co-immunoprecipitation experiments, but could not show binding of either of these antibodies to Dlg. It is possible that the antibodies do not bind the native protein (potentially because the regions outside of the domains are not accessible in the folded protein). On the other hand, binding of antibody and antigen could be of low affinity, rendering the antibodies useless in immunostaining and co-immunoprecipitation experiments. All three antibodies stain bacteria in immunostainings. Antibody 1742007 staining can be observed at the choanoflagellate cell membrane and microvilli.

Antibody specificities and corresponding stainings are quite variable and one can assume that many structures observed in stainings could correspond to other proteins. Purifying antibody CA5682 might allow us to replicate experiments with an independent antibody, although it has been raised against the same antigen. Furthermore, we plan to test antibody 1742007 in co-immunoprecipitation in order to resolve if the antibody can bind native Dlg.

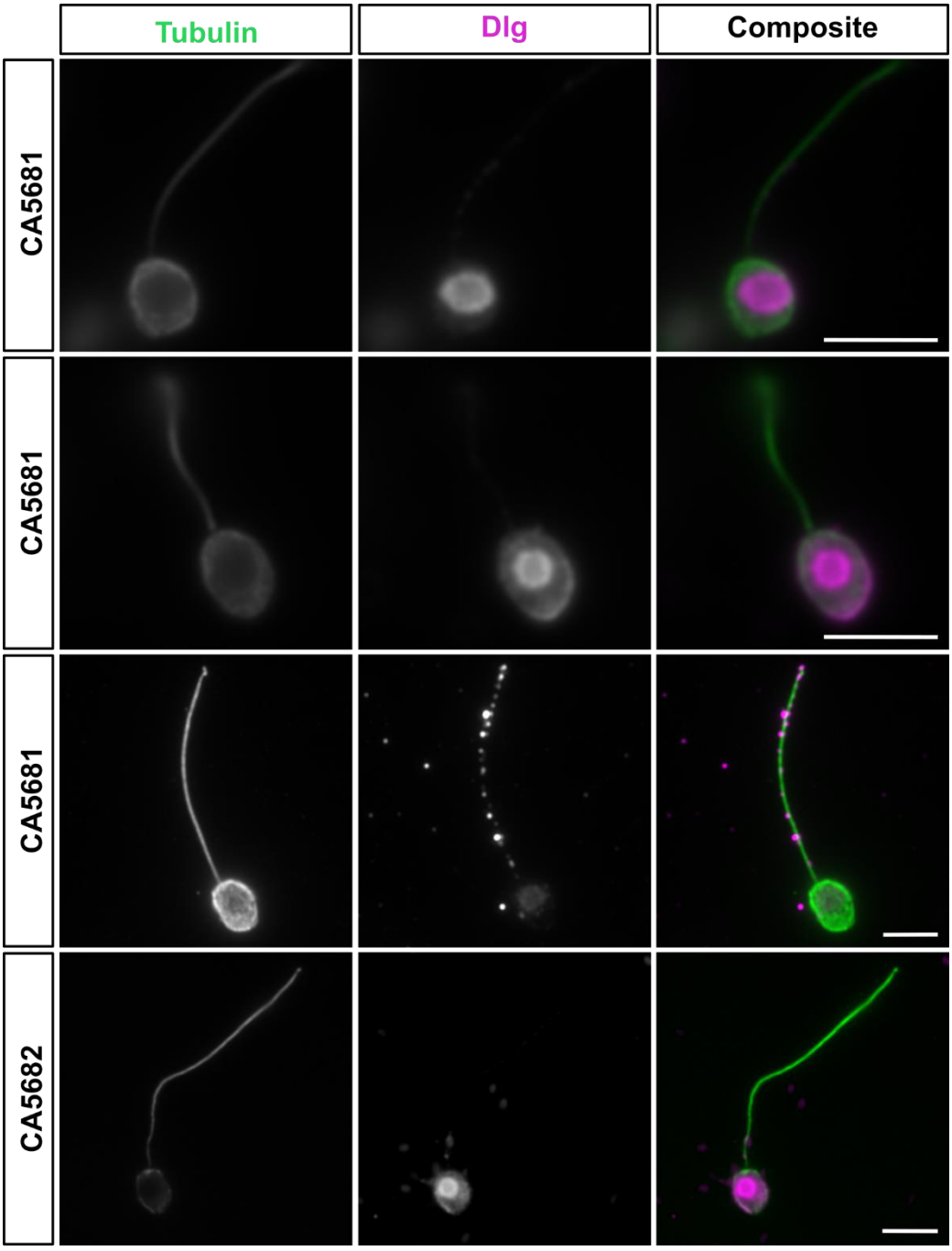


Figure 6-9: Immunostaining patterns observed with *S. rosetta* Dlg PDZ1-2 antibodies CA5681 and CA5682. The tubulin antibody (E7 antibody, green) highlights microtubules in the cell body and the flagellum. Rows 1-2 show again the nucleus staining and plasma membrane staining at the basal side of the cell described for CA5681 in the main text. Row 3 shows an alternative staining pattern (punctate staining around flagella microtubules) observed with the same antibody. This staining could not be replicated. Row 4 shows also nuclear staining pattern with antibody CA5682. Additionally, this antibody stained bacteria. Choanoflagellates were imaged with a Leica DMi8 fluorescence microscope with a 100x magnification oil objective. Scale bars: 5 µm.

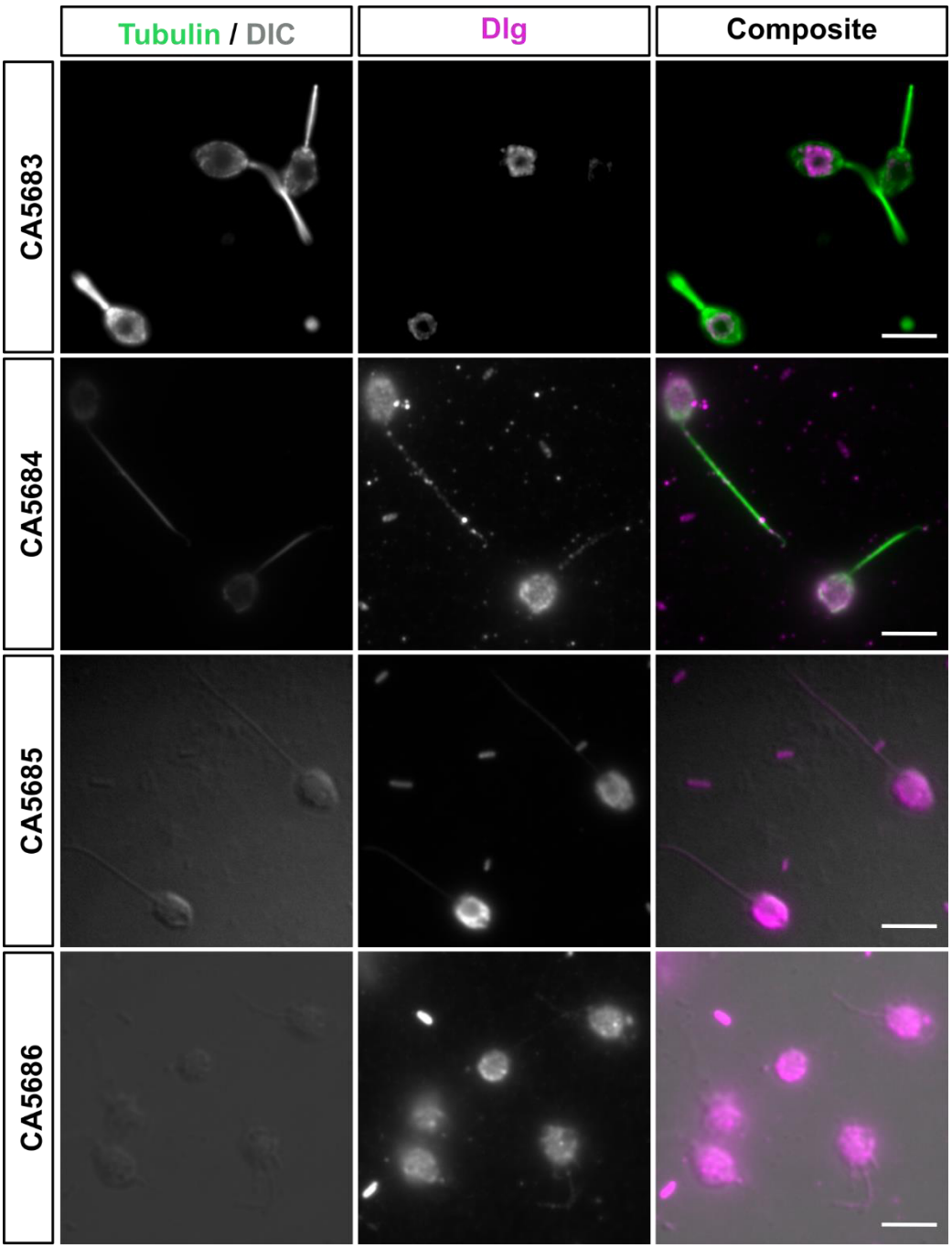


Figure 6-10: Immunostaining patterns observed with *S. rosetta* Dlg PDZ1-2 antibodies CA5683, CA5684, CA5685, and CA5686. The tubulin antibody (E7 antibody, green) highlights microtubules in the cell body and flagellum for antibodies CA5683 and CA5684. Secondary antibodies against mouse tubulin antibodies were cross-reactive to antibodies CA5685 and CA6886, which were raised in rats. Therefore, for these treatments differential interference contrast (DIC) light microscopic pictures were taken as reference to show cell bodies and flagella. Antibody CA5683 stains a structure surrounding the nucleus. Antibodies CA5684, CA5685 and CA5686 stain bacteria and structures in the cell. Choanoflagellates were imaged with a Leica DMi8 fluorescence microscope with a 100x magnification oil objective. Scale bars: 5 µm.

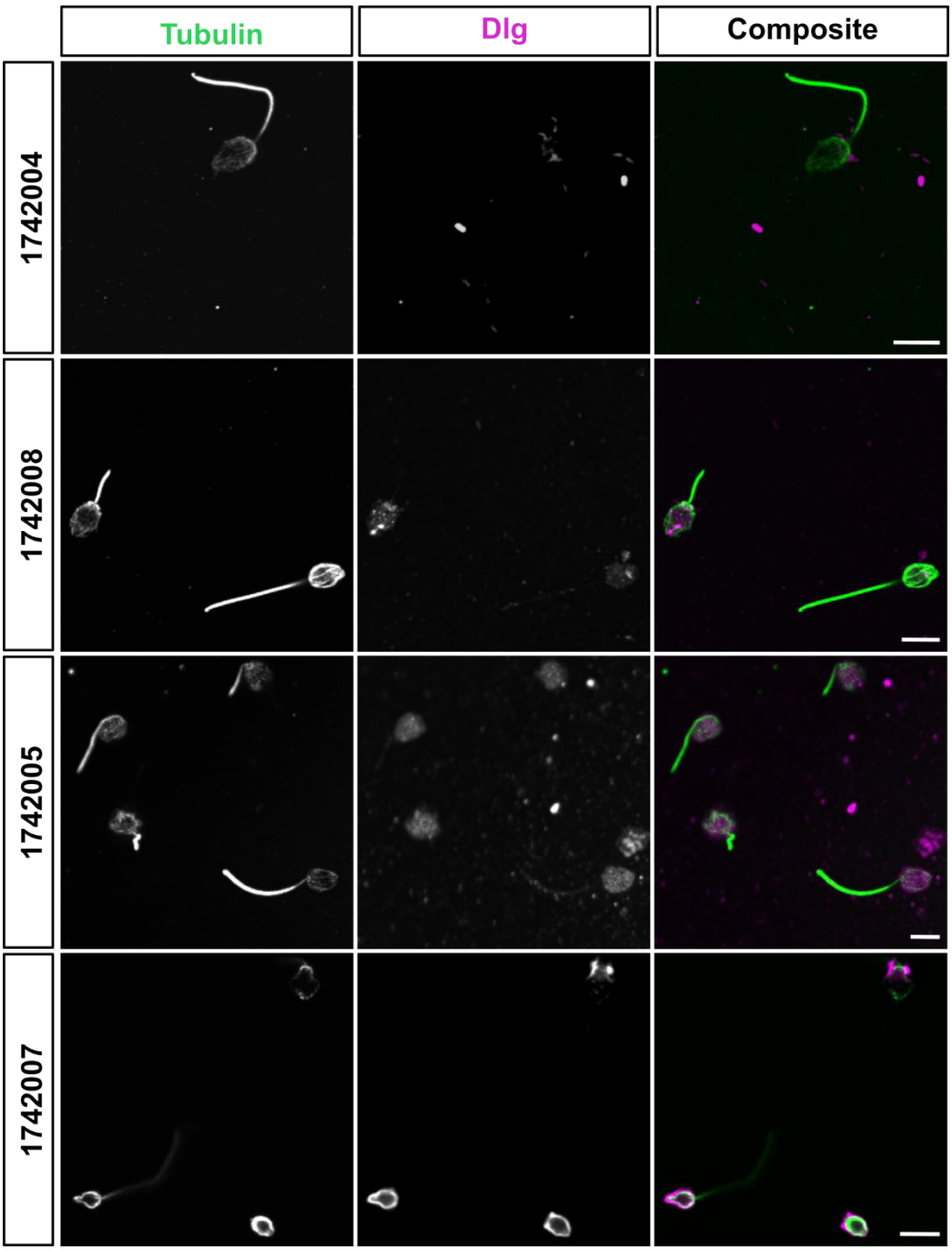


Figure 6-11: Immunostaining patterns observed with *S. rosetta* Dlg peptide antibodies 1742004, 1742008, 1742005, and 1742007. The tubulin antibody (E7 antibody, green) highlights microtubules in the cell body and the flagellum. Antibodies 1742004, 1742008, and 1742005 stain bacteria. The antibody 1742007 stains the cell membrane and microvilli (surrounding the flagellum).

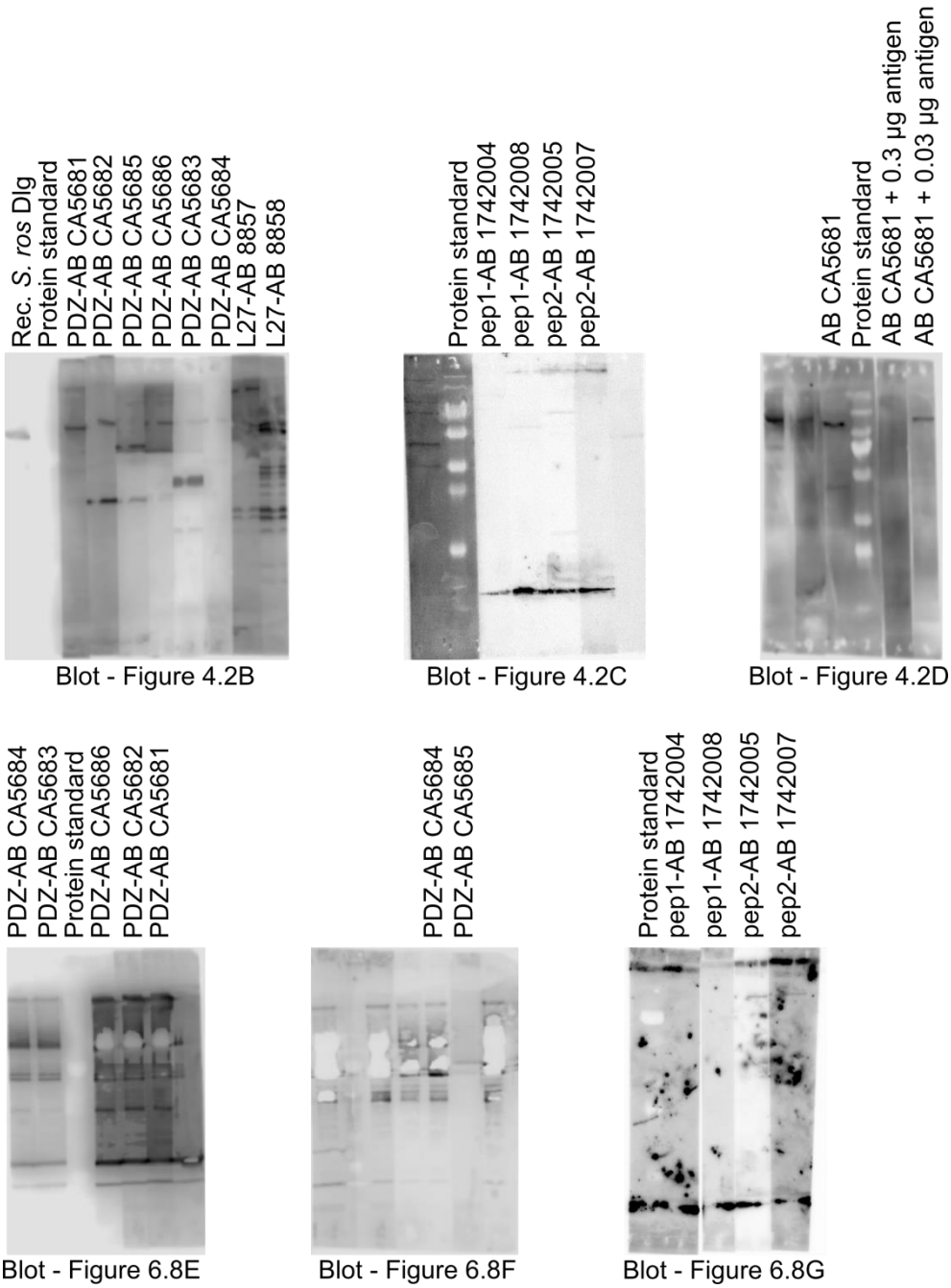


Figure 6-12: Original pictures of Western blots used for antibody validations – uncropped and in original orientation and order. Lanes that were shown in Figures 4.2 and 6.8 (as depicted below the blots) are labelled.

6.3.2 Supplementary Figures

Figure 6-13: *S. rosetta* Dlg/PSD-95 sequence showing domains and expression constructs. On top: poly-histidine-tag from expression vector (the underlined region remains in protein construct after thrombin cleavage; last serine only in PSD-95 SH3-HOOK-GuK and PSD-95 Full length). Below: PSD-95 sequence with the domains in coloured writing: Grey-blue: L27 domain; orange: PDZ domains; blue: SH3 domain; green: GuK domain. Boxes show expression constructs. Coloured boxes are labelled; grey boxes: 1. L27 aa 1-116 (aa 1-79 underlined); 2. Peptide 2; 3. Peptide 1.

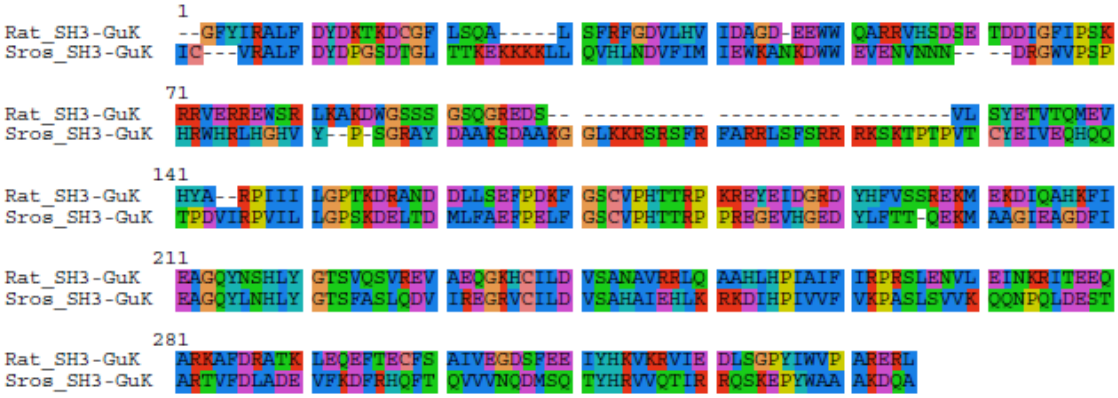


Figure 6-14: Alignment SH3-HOOK-GuK module. Alignment corresponding to structure homology-modelling between the rat and the *S. rosetta* module.

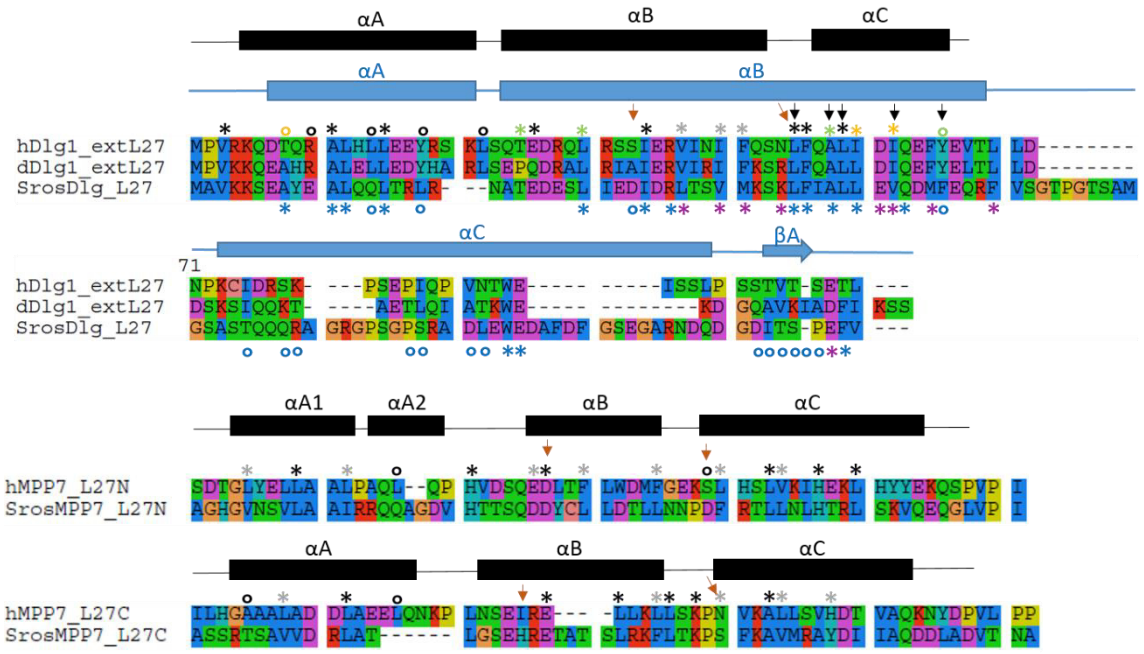


Figure 6-15: Alignment of *Salpingoeca rosetta* L27 domains to *H. sapiens* and *D. melanogaster* L27 domains including comparison of amino acid sites with functional importance for homo- or heterodimerisation in human, rat and *Drosophila melanogaster* proteins. On top alignment of human (h), *Drosophila melanogaster* (d) and *S. rosetta* (Sros) Dlg-1. The *D. melanogaster* Dlg L27 domain is extended and the alignment is shown also for the extended region. Below are alignments of the MPP7 N-terminal L27 domains (L27N) and C-terminal L27 domains (L27C) of human and *S. rosetta*. Secondary structure elements are displayed above the alignment. In black: human L27 secondary structure elements; in blue: *D. melanogaster* L27 secondary structure elements. Residues important for L27 homo- and heterodimerisation are indicated. Above the sequences such residues identified in human with importance for heterodimerisation are shown (black stars: conservation in *S. rosetta*; grey stars: amino acid with similar properties in *S. rosetta* sequence; black circle: no conservation in *S. rosetta*); complemented are amino acids that have been shown to be important for homodimerisation in rat, excluding the ones that are also important for heterodimerisation in human (green stars: conservation in *S. rosetta*; yellow stars: amino acid with similar properties in *S. rosetta* sequence; green circles: no conservation in *S. rosetta*; black arrows indicate amino acids that have been shown to be important for the interaction between two dimers). Below the sequence residues identified in *Drosophila melanogaster* with importance for homodimerisation are shown (blue stars: conservation in *S. rosetta*; purple stars: amino acid with similar properties in *S. rosetta* sequence; blue circle: no conservation in *S. rosetta*). Brown arrows indicate key residues described to favour heterodimerisation over homodimerisation in most L27 domains except for the Dlg L27 domain. The first arrow in the Dlg sequence corresponds to a site that in other L27 domains of Type 1 is positively charged and therefore cannot bind with other Type 1 L27 domains, but only with Type 2 L27 domains, with a negative charge at this position (as shown for MPP7 L27N below, but not for MPP7 L27C, which is neutral). The second brown arrow indicates another key residue that is usually hydrophic in Type A molecules and polar in Type B molecules as shown for human Dlg and MPP7 L27 domains.

6.3.3 Supplementary Tables

Table 6-4: Biochemical properties and purification conditions of recombinant proteins. His = Histidine tag; MBP = maltose binding protein tag; pI = isoelectrical point; Ext coeff = extinction coefficient.

S. rosetta expressed proteins	Biochemical properties	Purification conditions
Dlg full-length (His)	92972.52 Da (cleaved His) pI 6.18; Ext coeff: 66350	Dialysis into 200 mM NaCl, 20 mM Tris pH 7.4 and 1 mM DTT; Ion exchange on HiTrap™ SP HP
Dlg SH3-HOOK-GuK (His)	38442.58 Da (cleaved His) pI 8.75; Ext coeff: 46410	Dialysis into 200 mM NaCl, 20 mM Tris pH 7.4 and 1 mM DTT; Ion exchange on HiTrap™ SP HP
Dlg PDZ1-2 (His)	25145.13 Da (cleaved His) pI 5.11; Ext coeff: 8480	Dialysis into 100 mM NaCl, 20 mM Tris pH 7.4 and 1 mM DTT; Ion exchange on HiTrap™ Q HP
Dlg PDZ1-3 (His)	42874.1 Da (cleaved His) pI 5.9; Ext coeff: 12950	Dialysis into 100 mM NaCl, 20 mM Tris pH 7.4 and 1 mM DTT; Ion exchange on HiTrap™ Q HP
Dlg L27 (His)	9223.45 Da (aa 1-79, cleaved His) pI 5.77; Ext coeff: 1490	Dialysis into 100 mM NaCl, 20 mM Tris pH 7.4 and 1 mM DTT; Ion exchange on HiTrap™ Q HP
Dlg peptide1 (His)	8043.08 Da (incl His) pI 12.01; Ext coeff: 2980	Dialysis into 500 mM NaCl, 20 mM Tris pH7.4; Ion exchange on HiTrap™ SP HP
Dlg peptide2 (His)	7332.17 Da (incl His) pI 9.70; Ext coeff: 1490	Dialysis into 100 mM NaCl, 20 mM Tris pH 7.4; Ion exchange on HiTrap™ SP HP
Dlg peptide1 (MBP)	48689.07 Da (incl. MBP) pI 8.27; Ext coeff: 69330	Proceeded in column buffer with reduced salinity (dilution) 100 mM NaCl, 20 mM Tris pH 7.4, 1 mM EDTA, ion exchange on HiTrap™ SP HP
Dlg peptide2 (MBP)	47978.16 Da (incl. MBP) pI 5.2; Ext coeff: 67840	Proceeded in column buffer 200 mM NaCl, 20 mM Tris pH 7.4, 1 mM EDTA; no ion exchange
MPP7 full-length (His)	61798.54 Da (cleaved His) pI 6.83, Ext coeff: 42400	Dialysis into 500 mM NaCl, 20 mM Tris pH 7.4 2 mM DTT; salt concentration brought to 250 mM NaCl before ion exchange on HiTrap™ SP HP

Table 6-5: **Unique proteins in the Dlg co-IP mass spectrometry data vs the control co-IP mass spectrometry data.**

Unique from peptide counts and LFQ (yellow) and unique from normalised LFQ value only.
LFQ = label free quantification

NCBI identifier	**Annotation** Available annotation (bold) Own research via BLASTp searches and SMART domain architecture observations (not bold)	**LFQ value**
PTSG_01553	**non-muscle actin**	4422600000
PTSG_00410	**vacuolar iron family transporter**	1254100000
PTSG_03306	**PSD-95 alpha**	888740000
PTSG_01141	**Dlg/PSD-95**	342660000
PTSG_09863	MAGUK p55	251350000
PTSG_07478	Ankyrin-repeat containing protein	106980000
PTSG_08367	Protein with cytochrome-c oxidase domain Cox4	77703000
PTSG_11918	Hypothetical protein	63945000
PTSG_11624	**serine/threonine protein kinase**	68719000
PTSG_00416	Dynein light chain Tctex1 domain	45738000
PTSG_03018	**Rieske iron sulfur protein** + ubiquinone cytochrome-c reductase domain	43166000
PTSG_13055	Phosphoinositide specific phospholipase C (smart)/ 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta subunit (BLASTp search)	40778000
PTSG_08202	katanin p60 ATPase containing subunit A-like 2	41120000
PTSG_10302	globin	44572000
PTSG_00943	**polyadenylate binding protein**	30934000
PTSG_03120		27479000
PTSG_04752	**DERA (Desoxyribose-phosphate aldolase) protein**	25805000
PTSG_08036		18412000
PTSG_03062	**ubiquinol-cytochrome C reductase**	25773000
PTSG_02512	Nucleoprotein TPR	22978000
PTSG_02182	**alpha-galactosidase**	23890000
PTSG_02188		34350000
PTSG_07832		24067000
PTSG_00698	**nicotinate-nucleotide pyrophosphorylase**	18465000
PTSG_00708	**pirin family protein**	20093000
PTSG_09945	**splicing factor**	25463000
PTSG_07129		20810000
PTSG_00667		19907000
PTSG_08081		19896000
PTSG_05972		19604000
PTSG_09554		16760000
PTSG_12676		20839000
PTSG_12696	**actin**	15045000

PTSG_04141	**growth factor receptor-bound protein 2**	16150000
PTSG_01568	**methylmalonyl-CoA epimerase**	13903000
PTSG_04145		15920000
PTSG_06707	**formaldehyde dehydrogenase**	14217000
PTSG_05945		14100000
PTSG_10400		15148000
PTSG_04085	**ribosomal protein L35a**	12250000
PTSG_02022		12602000
PTSG_02579		12414000
PTSG_11636		13473000
PTSG_06685	**Rho1 GTPase**	11810000
PTSG_04266	**aminopeptidase**	13602000
PTSG_05576		11296000
PTSG_10796		11264000
PTSG_05798		12207000
PTSG_12877	universal stress protein	10399000
PTSG_09495	**ribosomal protein rpl36**	11999000
PTSG_05423	**uracil phosphoribosyltransferase**	13599000
PTSG_01619	**glutathione reductase**	8912600
PTSG_05953		12406000
PTSG_05473	**ribose-5-phosphate isomerase**	10681000
:PTSG_10379	**saccharopine dehydrogenase**	7547400
PTSG_06812	**arginine kinase**	10569000
PTSG_07208		8339100
PTSG_09913		9478900
PTSG_00307	**endophilin**	10428000
PTSG_11831		8776000
PTSG_10977	**40s ribosomal protein S27**	7551600
PTSG_04414	**tubulin beta chain**	8495000
PTSG_05372		8642100
PTSG_03338	**UDP-n-acteylglucosamine pyrophosphorylase**	9621700
PTSG_03104	**villin-1**	8620400
PTSG_04745	**ribosomal protein rpl17**	7409900
PTSG_04054	**myo-inositol-1-phosphate synthase**	7394900
PTSG_04388		7262700
PTSG_04898		11385000
PTSG_02200	**3-hydroxyisobutyrate dehydrogenase**	7117700
PTSG_02301	**thioredoxin reductase 1**	13345000
PTSG_04441	**MEMO1 protein**	8261000
PTSG_12806		7976100
PTSG_05410	**crk-like protein**	8808700
PTSG_12511		7153900
PTSG_09540	**spectrin**	6363600
PTSG_06718	**ubiquitin carboxyl-terminal hydrolase**	8152200
PTSG_06665	**proliferation-associated protein 2G4**	7538100
PTSG_01461	**dynein light chain roadblock-type 1**	8215700
PTSG_05774	**3-hydroxyanthranilate 3,4-dioxygenase**	7738200

PTSG_08551		7370400
PTSG_04531		5181900
PTSG_02905	**acyl-CoA hydrolase**	6861400
PTSG_05675	**transmembrane 9 superfamily member 2**	6475700
PTSG_12376	**NADH dehydrogenase flavoprotein 1**	5744700
PTSG_10880	**frequenin-1**	8037400
PTSG_06071	**arginase**	5459500
PTSG_01048		5729900
PTSG_07903		6635500
PTSG_02147	**MYG1 protein**	5199900
PTSG_10903	Hypothetical protein	5147100
PTSG_10137	**ribosomal protein S17**	7394000
PTSG_09167		7345600
PTSG_11774		5906800
PTSG_12122	**proline iminopeptidase**	4931000
PTSG_09593	**rps15A**	6208600
PTSG_11722		5858000
PTSG_10299	**voltage-gated potassium channel subunit beta-1**	6350700
PTSG_00836	**gamma-glutamyl hydrolase**	6069300
PTSG_05393		4669300
PTSG_09317	**phosphoglucomutase 3**	5554500
PTSG_03314	**lysozyme**	8658600
PTSG_00406	**polyhydroxybutyrate depolymerase**	5973200
PTSG_06336	**NADH-ubiquinone oxido-reductase 23 kDa subunit**	4522800
PTSG_04718	**thymidine phosphorylase**	4487400
PTSG_02259	**arbp-prov protein**	5036100
PTSG_04588		4907900
PTSG_00211		4429300
PTSG_07800		5067500
PTSG_06187		4148500
PTSG_01865		6150400
PTSG_04201	**AGC/PKA protein kinase**	5663000
PTSG_04797		4526900
PTSG_12178	**histone deacetylase superfamily protein**	4492400
PTSG_06763	ankyrin repeat containing	3640700
PTSG_08640		3596400
PTSG_02258	**acyl-CoA binding protein**	3491100
PTSG_08625		3881900
PTSG_11067	**argininosuccinate synthetase 1**	4319900
PTSG_05755		3344100
PTSG_06297		3344000
PTSG_04215		4023100
PTSG_13198		3776800
PTSG_12887		4128900
PTSG_04187		4205300
PTSG_11749	**DUS4L protein**	3073000

PTSG_06495		3954700
PTSG_02961		2956100
PTSG_08194		1919800
PTSG_04442	**P-glycoprotein**	3680400
PTSG_11120	**p75 neurotrophin receptor b**	3471500
PTSG_02376	**GDP-D-mannose pyrophosphorylase**	5295800
PTSG_07295		2813700
PTSG_10902	**RNA-binding protein 8A**	3211200
PTSG_01490		3312700
PTSG_01213	**quemao protein**	3518000
PTSG_13234		2737300
PTSG_04665		5078200
PTSG_08200	**fumarate hydratase class I**	2665000
PTSG_10929	**Ppp2r2a protein (Serine/threonine protein phosphatase)**	2623400
PTSG_09483	**coatomer subunit delta**	3121500
PTSG_03025		3477500
PTSG_04177	**coatomer subunit beta**	3376600
PTSG_09659		2476200
PTSG_13117		3415000
PTSG_00568	**ATP-dependent RNA helicase eIF4A**	2387700
PTSG_11901	Pyrroline-5-carboxylate reductase	2304200
PTSG_08641		3038400
PTSG_11672		2706900
PTSG_08850		2953800
PTSG_12699		2131000
PTSG_11553		2109100
PTSG_04212		2055300
PTSG_02696		2823800
PTSG_05481		2751100
PTSG_10219	**26S proteasome non-ATPase regulatory subunit 11**	3066100
PTSG_12259	**CAMK/CAMKL protein kinase**	2012300
PTSG_01640	**eIF4AIII-PA**	2008800
PTSG_11682	**vesicle-fusing ATPase**	1921800
PTSG_04369	**hypothetical protein**	1892800
PTSG_05386		2466200
PTSG_03447		1823900
PTSG_00332	putative Glycosylphosphatidyl inositol (GPI) anchor	1819900
PTSG_04123		1754800
PTSG_08773	**cytochrome c oxidase copper chaperone**	3227400
PTSG_09566		1524300
PTSG_04583	**hydroxysteroid dehydrogenase 4**	1491200
PTSG_00960		2183200
PTSG_07498	**beta-glucosidase**	1403100

PTSG_03456	**C-type MBL-2 protein**	1399000
PTSG_12177		1392000
PTSG_01284	**tyrosine-protein phosphatase 9**	1338700
PTSG_02496	**heterotrimeric G protein beta subunit 1**	1202600
PTSG_12637		1201600
PTSG_12744		1200700
PTSG_10860		1172200
PTSG_03830	**phospholipase C gamma**	1156100
PTSG_05764		1140400
PTSG_04900	**LZIC protein**	1136600
PTSG_05994	**Atp6v0a1 protein**	1114700
PTSG_00770	**cathepsin** (protease)	1109700
PTSG_01852		1072100
PTSG_11258	L-lactate dehydrogenase	1061700
PTSG_01007	Glutathione S-Transferase (GST)	1018000
PTSG_10087	Enoyl-CoA hydratase or similar enzyme	1011600
PTSG_08705	**V-type proton ATPase subunit F**	1004100
PTSG_06947		1243200
PTSG_06411		940700
PTSG_01337		924810
PTSG_03929	**deoxyuridine 5'-triphosphate nucleotidohydrolase**	893170
PTSG_01302	**acyloxyacyl hydrolase**	862940
PTSG_03142	**acyl-CoA dehydrogenase**	853240
PTSG_11403		843160
PTSG_07828	pyruvate water dikinase	1567100
PTSG_07272	ef hand family protein	774080
PTSG_13141	RUN, coiled coil and FYVE domain containing protein	757410
PTSG_11360	**alpha-glucosidase II alpha subunit, family GH31**	1378100
PTSG_05032	**phosphofructokinase**	708840
PTSG_11840	Protein with immunoglobulin- like fold	697160
PTSG_11241	Protein with 4x EF hand	673390
PTSG_05324	**AMP-dependent synthetase and ligase**	649420
PTSG_0954	**ThiF family protein**	640670
PTSG_12256	**thioredoxin**	636270
PTSG_04272	Pleckstrin homology domain containing protein	615140
PTSG_12410		1059300
PTSG_02297		556330
PTSG_05063	**prolyl endopeptidase**	551040
PTSG_03363	protein with RNA recognition motif RRM	1019400
PTSG_09727		792230
PTSG_07105	PID domain	729020

7 Scientific output

Publication:

Hoffmeyer, T.T., Burkhardt, P., 2016. Choanoflagellate Models – *Monosiga brevicollis* and *Salpingoeca rosetta*. Current Opinion in Genetic & Development 39:42-47. (Review)

Publications in preparation:

Hoffmeyer, T. T., Grønborg, M., Walter, T. S., Aricescu, A. R., Burkhardt, P. A protein scaffold mediated by a PSD-95 homolog in choanoflagellates reveals insights into the origin of postsynaptic signalling machineries. (in preparation, based on chapter 4)

Hoffmeyer, T. T., Savory, F. R., Richards, T. A., Burkhardt, P. Homer and Shank, two proteins organising signalling machineries in the postsynapse, were putatively ancestral binding partners before the evolution of animals. (in preparation, based on chapter 3)

8 References

Abe, T., Sugihara, H., Nawa, H., et al. (1992) Molecular characterization of a novel metabotropic glutamate receptor mGluR5 coupled to inositol phosphate/Ca^{2+} signal transduction. *Journal of Biological Chemistry*, 267 (19): 13361–13368.

Abedin, M. and King, N. (2008) The premetazoan ancestry of cadherins. *Science*, 319 (5865): 946–948. doi:10.1126/science.1151084.

Achim, K. and Arendt, D. (2014) Structural evolution of cell types by step-wise assembly of cellular modules. *Current Opinion in Genetics and Development*, 27: 102–108. doi:10.1016/j.gde.2014.05.001.

Ahmmed, G.U., Mehta, D., Vogel, S., et al. (2004) Protein kinase Cα phosphorylates the TRPC1 channel and regulates store-operated Ca^{2+} entry in endothelial cells. *The Journal of Biological Chemistry*, 279 (20): 20941–20949. doi:10.1074/jbc.M313975200.

Alberts, B., Johnson, A., Lewis, J., et al. (2002) *Molecular biology of the cell*. New York: Garland Science.

Alegado, R.A., Brown, L.W., Cao, S., et al. (2012) A bacterial sulfonolipid triggers multicellular development in the closest living relatives of animals. *eLife*, 1: e00013. doi:10.7554/eLife.00013.

Alié, A. and Manuel, M. (2010) The backbone of the post-synaptic density originated in a unicellular ancestor of choanoflagellates and metazoans. *BMC Evolutionary Biology*, 10: 34. doi:10.1186/1471-2148-10-34.

Altschul, S.F., Gish, W., Myers, E.W., et al. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, 215: 403–410.

Anctil, M. (1985) Cholinergic and monoaminergic mechanisms associated with control of bioluminescence in the ctenophore Mnemiopsis leidyi. *Journal of Experimental Biology*, 119: 225–238.

Anderson, D.P., Whitney, D.S., Hanson-Smith, V., et al. (2016) Evolution of an ancient protein function involved in organized multicellularity in animals. *eLife*, 5: e10147. doi:10.7554/eLife.10147.

Anderson, J.M. (1996) Cell signalling: MAGUK magic. *Current Biology*, 6 (4): 382–384. doi:10.1016/S0960-9822(02)00501-8.

Anderson, P.A. V (1985) Physiology of a bidirectional, excitatory, chemical synapse. *Journal of Neurophysiology*, 53 (3): 821–835.

Aramori, I. and Nakanishi, S. (1992) Signal transduction and pharmacological characteristics of a metabotropic glutamate receptor, mGluRI, in transfected CHO cells. *Neuron*, 8 (4): 757–765. doi:10.1016/0896-6273(92)90096-V.

Arendt, D. (2008) The evolution of cell types in animals: emerging principles from molecular studies. *Nature Reviews Genetics*, 9: 868–882. doi:10.1038/nrg2416.

Arendt, D., Musser, J.M., Baker, C.V.H., et al. (2016a) The origin and evolution of cell types. *Nature Reviews Genetics*, 17 (12): 744–757. doi:10.1038/nrg.2016.127.

Arendt, D., Tosches, M.A. and Marlow, H. (2016b) From nerve net to nerve ring, nerve cord and brain-evolution of the nervous system. *Nature Reviews Neuroscience*. 17 (1) pp. 61–72. doi:10.1038/nrn.2015.15.

Armon, S., Bull, M.S., Aranda-Diaz, A., et al. (2018) Ultrafast epithelial contractions provide insights into contraction speed limits and tissue integrity. *Proceedings of the National Academy of Sciences of the United States of America*, 115 (44): E10333–E10341. doi:10.1073/pnas.1802934115.

Augustine, G.J., Burns, M.E., DeBello, W.M., et al. (1999) Proteins involved in synaptic vesicle trafficking. *Journal of Physiology*, 520 (1): 33–41.

Ausubel, F.M. (2005) Are innate immune signaling pathways in plants and animals conserved? *Nature Immunology*, 6 (10): 973–979. doi:10.1038/ni1253.

Avelar, G.M., Glaser, T., Leonard, G., et al. (2015) A cyclic GMP-dependent K^+ channel in the blastocladiomycete fungus *Blastocladiella emersonii*. *Eukaryotic Cell*, 14 (9): 958–963. doi:10.1128/EC.00087-15.

Babonis, L.S. and Martindale, M.Q. (2017) Phylogenetic evidence for the modular evolution of metazoan signalling pathways. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372 (1713): 20150477. doi:10.1098/rstb.2015.0477.

Bachmann, A., Kobler, O., Kittel, R.J., et al. (2010) A perisynaptic ménage à trois between Dlg, *D*Lin-7, and Metro controls proper organization of *Drosophila* synaptic junctions. *The Journal of Neuroscience*, 30 (17): 5811–5824. doi:10.1523/JNEUROSCI.0778-10.2010.

Bachmann, A., Timmer, M., Sierralta, J., et al. (2004) Cell type-specific recruitment of *Drosophila* Lin-7 to distinct MAGUK-based protein complexes defines novel roles for Sdt and Dlg-S97. *Journal of Cell Science*, 117 (10): 1899–1909. doi:10.1242/jcs.01029.

Baguñà, J., Martinez, P., Paps, J., et al. (2008) Back in time: A new systematic proposal for the Bilateria. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363 (1496): 1481–1491. doi:10.1098/rstb.2007.2238.

Baluška, F., Šamaj, J., Wojtaszek, P., et al. (2003) Cytoskeleton-plasma membrane-cell wall continuum in plants . Emerging links revisited. *Plant Physiology*, 133 (October): 482–491. doi:10.1104/pp.103.027250.482.

Banner, D.W., D'Arcy, A., Janes, W., et al. (1993) Crystal structure of the soluble human 55 kd TNF receptor-human TNFβ complex: Implications for TNF receptor activation. *Cell*, 73 (3): 431–445. doi:10.1016/0092-8674(93)90132-A.

Baran, Y., Bercovich, A., Sebe-Pedros, A., et al. (2019) MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. *Genome Biology*, 20: 206. doi:10.1186/s13059-019-1812-2.

Baron, M.K., Boeckers, T.M., Vaida, B., et al. (2006) An architectural framework that may lie at the core of the postsynaptic density. *Science*, 311: 531–535. doi:10.1126/science.1117715.

Barzik, M., Carl, U.D., Schubert, W.D., et al. (2001) The N-terminal domain of Homer/Vesl is a new class II EVH1 domain. *Journal of molecular biology*, 309 (1): 155–69. doi:10.1006/jmbi.2001.4640.

Bassell, G.J. and Warren, S.T. (2008) Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function. *Neuron*, 60: 201–214. doi:10.1016/j.neuron.2008.10.004.

Beligni, M.V. and Lamattina, L. (1999) Nitric oxide counteracts cytotoxic processes mediated by reactive oxygen species in plant tissues. *Planta*, 208 (3): 337–344. doi:10.1007/s004250050567.

Bement, W.M., Mandato, C.A. and Kirsch, M.N. (1999) Wound-induced assembly and closure of an actomyosin purse string in *Xenopus* oocytes. *Current Biology*, 9 (11): 579–587. doi:10.1016/S0960-9822(99)80261-9.

Beneken, J., Tu, J.C., Xiao, B., et al. (2000) Structure of the Homer EVH1 domain-peptide complex reveals a new twist in polyproline recognition. *Neuron*, 26 (1): 143–154. doi:10.1016/S0896-6273(00)81145-9.

Bergstralh, D.T. and St Johnston, D. (2012) Epithelial cell polarity: what flies can teach us about cancer. *Essays In Biochemistry*, 53: 129–140. doi:10.1042/bse0530129.

Bilder, D. (2001) PDZ proteins and polarity: functions from the fly. *Trends in Genetics*, 17 (9): 511–519. doi:10.1016/S0168-9525(01)02407-6.

Blackstone, C. and Sheng, M. (2002) Postsynaptic calcium signaling microdomains in neurons. *Frontiers in Bioscience*, 7: d872-885. doi:10.2741/blacksto.

Boeckers, T.M. (2006) The postsynaptic density. *Cell and tissue research*, 326 (2): 409–422. doi:10.1007/s00441-006-0274-5.

Boeckers, T.M., Bockmann, J., Kreutz, M.R., et al. (2002) ProSAP/Shank proteins - a family of higher order organizing molecules of the postsynaptic density with an emerging role in human neurological disease. *Journal of Neurochemistry*, 81 (5): 903–910. doi:10.1046/j.1471-4159.2002.00931.x.

Bohl, J., Brimer, N., Lyons, C., et al. (2007) The Stardust family protein MPP7 forms a tripartite complex with LIN7 and DLG1 that regulates the stability and localization of DLG1 to cell junctions. *Journal of Biological Chemistry*, 282 (13): 9392–9400. doi:10.1074/jbc.M610002200.

Bonaglia, M.C., Giorda, R., Borgatti, R., et al. (2001) Disruption of the ProSAP2 gene in a t(12;22)(q24.1;q13.3) is associated with the 22q13.3 deletion syndrome. *The American Journal of Human Genetics*, 69 (2): 261–268. doi:10.1086/321293.

Booth, D.S. and King, N. (2020) Genome editing enables reverse genetics of

multicellular development in the choanoflagellate Salpingoeca rosetta. *eLife*, 9: e56193. doi:10.1101/2020.02.18.948406.

Booth, D.S., Szmidt-Middleton, H. and King, N. (2018) Choanoflagellate transfection illuminates their cell biology and the ancestry of animal septins. *Molecular Biology of the Cell*, 29: 3026–3038. doi:10.1091/mbc.E18-08-0514.

Bork, P., Doerks, T., Springer, T.A., et al. (1999) Domains in plexins: links to integrins and transcription factors. *Trends in Biochemical Sciences*, 24 (7): 261–263. doi:10.1016/S0968-0004(99)01416-4.

Bortolotto, Z.A., Fitzjohn, S.M. and Collingridge, G.L. (1999) Roles of metabotropic glutamate receptors in LTP and LTD in the hippocampus. *Current Opinion in Neurobiology*. doi:10.1016/S0959-4388(99)80044-0.

Boucher, J.L., Moali, C. and Tenu, J.P. (1999) Nitric oxide biosynthesis, nitric oxide synthase inhibitors and arginase competition for L-arginine utilization. *Cellular and Molecular Life Sciences*. 55 (8–9) pp. 1015–1028. doi:10.1007/s000180050352.

Brenman, J.E., Chao, D.S., Gee, S.H., et al. (1996) Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and α1-Syntrophin mediated by PDZ domains. *Cell*, 84: 757–767. doi:10.1016/s0092-8674(00)81053-3.

Brodal, P. (2004) *The central nervous system : structure and function*. 3rd ed. Brodal, P. (ed.). Oxford University Press.

Brunet, T. and Arendt, D. (2016) From damage response to action potentials: early evolution of neural and contractile modules in stem eukaryotes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 371 (1685): 20150043. doi:10.1098/rstb.2015.0043.

Brunet, T., Larson, B.T., Linden, T.A., et al. (2019) Light-regulated collective contractility in a multicellular choanoflagellate. *Science*, 366 (6463): 326–334. doi:10.1126/science.aay2346.

Bucher, D. and Anderson, P.A.V. (2015) Evolution of the first nervous systems - What can we surmise? *Journal of Experimental Biology*, 218 (4): 501–503. doi:10.1242/jeb.111799.

Burkhardt, P. (2015) The origin and evolution of synaptic proteins - choanoflagellates lead the way. *The Journal of Experimental Biology*, 218: 506–514. doi:10.1242/jeb.110247.

Burkhardt, P., Gronborg, M., McDonald, K., et al. (2014) Evolutionary insights into premetazoan functions of the neuronal protein Homer. *Molecular Biology and Evolution*, 31 (9): 2342–2355. doi:10.1093/molbev/msu178.

Burkhardt, P. and Sprecher, S.G. (2017) Evolutionary origin of synapses and neurons - Bridging the gap. *BioEssays*, 39: 1700024. doi:10.1002/bies.201700024.

Burkhardt, P., Stegmann, C.M., Cooper, B., et al. (2011) Primordial

neurosecretory apparatus identified in the choanoflagellate Monosiga brevicollis. *Proceedings of the National Academy of Sciences*, 108 (37): 15264–15269. doi:10. 1073/pnas.1106189108.

Busetta, B. and Barrans, Y. (1984) The prediction of protein domains. *Biochimica et Biophysica Acta*, 790 (2): 117–124. doi:10.1016/0167-4838(84)90214-0.

Cajal, S.R. (1888) Estructura de los centros nerviosos de las aves. *Revista Trimestral de Histología Normal y Patológica*, 1: 1–10.

Camacho, C., Coulouris, G., Avagyan, V., et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, 10: 421. doi:10.1186/1471-2105-10-421.

Caroni, P., Donato, F. and Muller, D. (2012) Structural plasticity upon learning: regulation and functions. *Nature Reviews Neuroscience*, 13 (7): 478–490. doi:10.1038/nrn3258.

Carr, M., Leadbeater, C., Hassan, R., et al. (2008) Molecular phylogeny of choanoflagellates, the sister group to Metazoa. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (43): 16641–16646.

Carr, M., Richter, D.J., Fozouni, P., et al. (2017) A six-gene phylogeny provides new insights into choanoflagellate evolution. *Molecular Phylogenetics and Evolution*, 107: 166–178. doi:10.1016/j.ympev.2016.10.011.

Del Castillo, J. and Katz, B. (1954) Quantal components of the end-plate potential. *J. Physiol. (I954)*, 124: 560–573.

Cavalier-Smith, T. (2013) Early evolution of eukaryote feeding modes, cell structural diversity, and classification of the protozoan phyla Loukozoa, Sulcozoa, and Choanozoa. *European Journal of Protistology*, 49 (2): 115–178. doi:10.1016/j.ejop.2012.06.001.

Cavalier-Smith, T. and Chao, E.E.-Y. (2003) Phylogeny of Choanozoa, Apusozoa, and other Protozoa and early eukaryote megaevolution. *Journal of Molecular Evolution*, 56: 540–563. doi:10.1007/s00239-002-2424-z.

Cavalier-Smith, T. and Chao, E.E. (2010) Phylogeny and evolution of Apusomonadida (Protozoa: Apusozoa): new genera and species. *Protist*, 161 (4): 549–576. doi:10.1016/j.protis.2010.04.002.

Chapman, J.A., Kirkness, E.F., Simakov, O., et al. (2010) The dynamic genome of *Hydra*. *Nature*, 464 (7288): 592–596. doi:10.1038/nature08830.

Chen, K., Feng, H., Zhang, M., et al. (2003) Nitric oxide alleviates oxidative damage in the green alga *Chlorella pyrenoidosa* caused by UV-B radiation. *Folia Microbiologica*, 48 (3): 389–393. doi:10.1007/BF02931372.

Chen, L., Chetkovich, D.M., Petraliak, R.S., et al. (2000) Stargazin regulates synaptic targeting of AMPA receptors by two distinct mechanisms. *Nature*, 408: 936–943.

Chen, X., Levy, J.M., Hou, A., et al. (2015) PSD-95 family MAGUKs are

essential for anchoring AMPA and NMDA receptor complexes at the postsynaptic density. *Proceedings of the National Academy of Sciences of the United States of America*, 112 (50): E6983–E6992. doi:10.1073/pnas.1517045112.

Chen, X., Nelson, C.D., Li, X., et al. (2011) PSD-95 is required to sustain the molecular organization of the postsynaptic density. *The Journal of Neuroscience*, 31 (17): 6329–6338. doi:10.1523/JNEUROSCI.5968-10.2011.

Cheng, D., Hoogenraad, C.C., Rush, J., et al. (2006) Relative and absolute quantification of postsynaptic density proteome isolated from rat forebrain and cerebellum. *Molecular & Cellular Proteomics*, 5 (6): 1158–1170. doi:10.1074/mcp.D500009-MCP200.

Chishti, A.H., Kim, A.C., Marfatia, S.M., et al. (1998) The FERM domain: A unique module involved in the linkage of cytoplasmic proteins to the membrane. *Trends in Biochemical Sciences*, 23 (8): 281–282. doi:10.1016/S0968-0004(98)01237-7.

Chiu, J., Desalle, R., Lam, H.-M., et al. (1999) Molecular evolution of glutamate receptors: a primitive signaling mechanism that existed before plants and animals diverged. *Molecular Biology and Evolution*, 16 (6): 826–838.

Cho, K.-O., Hunt, C.A. and Kennedy, M.B. (1992) The rat brain postsynaptic density fraction contains a homolog of the *Drosophila* discs-large tumor suppressor protein. *Neuron*, 9 (5): 929–942. doi:10.1016/0896-6273(92)90245-9.

Clapham, D.E. (2007) Calcium Signaling. *Cell*, 131 (6): 1047–1058. doi:10.1016/j.cell.2007.11.028.

Cohen, N.A. and Brenman, J.E. (1996) Binding of the inward rectifier K+ channel Kir 2.3 to PSD-95iIs regulated by protein kinase A phosphorylation. *Neuron*, 17: 759–767.

Cohen, S.L., Chait, B.T., Ferré-D'Amaré, A.R., et al. (1995) Probing the solution structure of the DNA-binding protein Max by a combination of proteolysis and mass spectrometry. *Protein Science*, 4 (6): 1088–1099. doi:10.1002/pro.5560040607.

Collins, M.O., Husi, H., Yu, L., et al. (2006) Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. *Journal of Neurochemistry*, 97: 16–23. doi:10.1111/j.1471-4159.2005.03507.x.

Contractor, A., Mulle, C. and Swanson, G.T. (2011) Kainate receptors coming of age: milestones of two decades of research. *Trends in Neurosciences*, 34 (3): 154–163. doi:10.1016/j.tins.2010.12.002.

Cordes, V.C., Reidenbach, S., Rackwitz, H.-R., et al. (1997) Identification of protein p270/Tpr as a constitutive component of the nuclear pore complex–attached intranuclear filaments. *The Journal of Cell Biology J. Cell Biol*, 136 (127): 515–529.

Craven, S.E., El-Husseini, A.E. and Bredt, D.S. (1999) Synaptic targeting of the postsynaptic density protein PSD-95 mediated by a tyrosine-based trafficking signal. *Journal of Biological Chemistry*, 22: 497–509. doi:10.1074/jbc.M910153199.

Davidson, E.H. (2010) Emerging properties of animal gene regulatory networks. *Nature*. 468 (7326) pp. 911–920. doi:10.1038/nature09645.

Dayel, M.J., Alegado, R.A., Fairclough, S.R., et al. (2011) Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Developmental Biology*, 357: 73–82. doi:10.1016/j.ydbio.2011.06.003.

Degnan, B.M., Adamska, M., Richards, G.S., et al. (2015) "Porifera." In Wanninger, A. (ed.) *Evolutionary developmental biology of invertebrates*. Vienna, Austria: Springer. pp. 65–106.

Dobrosotskaya, I., Guy, R.K. and James, G.L. (1997) MAGI-1, a membrane-associated guanylate kinase with a unique arrangement of protein-protein interaction domains. *Journal of Biological Chemistry*, 272 (50): 31589–31597. doi:10.1074/jbc.272.50.31589.

Dong, A., Xu, X., Edwards, A.M., et al. (2007) In situ proteolysis for protein crystallization and structure determination. *Nature Methods*, 4 (12): 1019–1021. doi:10.1038/nmeth1118.

Dosemeci, A., Weinberg, R.J., Reese, T.S., et al. (2016) The postsynaptic density: there is more than meets the eye. *Frontiers in Synaptic Neuroscience*, 8 (23). doi:10.3389/fnsyn.2016.00023.

Dunn, C.W., Hejnol, A., Matus, D.Q., et al. (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature*, 452: 745–750. doi:10.1038/nature06614.

Dunn, C.W., Leys, S.P. and Haddock, S.H.D. (2015) The hidden biology of sponges and ctenophores. *Trends in Ecology and Evolution*. 30 (5) pp. 282–291. doi:10.1016/j.tree.2015.03.003.

Ebnet, K. (2008) Organization of multiprotein complexes at cell–cell junctions. *Histochemistry and Cell Biology*, 130: 1–20. doi:10.1007/s00418-008-0418-7.

Edgar, R.C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32 (5): 1792–1797. doi:10.1093/nar/gkh340.

Eick, G.N., Bridgham, J.T., Anderson, D.P., et al. (2017) Robustness of reconstructed ancestral protein functions to statistical uncertainty. *Molecular Biology and Evolution*, 34 (2): 247–261. doi:10.1093/molbev/msw223.

Elliott, G.R.D. and Leys, S.P. (2007) Coordinated contractions effectively expel water from the aquiferous system of a freshwater sponge. *Journal of Experimental Biology*, 210 (21): 3736–3748. doi:10.1242/jeb.003392.

Ellwanger, K. and Nickel, M. (2006) Neuroactive substances specifically modulate rhythmic body contractions in the nerveless metazoon *Tethya*

wilhelma (Demospongiae, Porifera). *Frontiers in Zoology*, 3 (1): 7. doi:10.1186/1742-9994-3-7.

Van Eynde, A., Nuytten, M., Dewerchin, M., et al. (2004) The nuclear scaffold protein NIPP1 is essential for early embryonic development and cell proliferation. *Molecular and Cellular Biology*, 24 (13): 5863–5874. doi:10.1128/mcb.24.13.5863-5874.2004.

Fairclough, S.R. (2015) "Choanoflagellates: perspectives on the origin of multicellularity." In Ruiz-Trillo, I. and Nedelcu, A.M. (eds.) *Evolutionary transitions to multicellular life*. Advances in Marine Genomics. Dordrecht: Springer Netherlands. pp. 99–116. doi:10.1007/978-94-017-9642-2.

Fairclough, S.R., Chen, Z., Kramer, E., et al. (2013) Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biology*, 14 (2): R15.

Fatt, P. and Katz, B. (1951) An analysis of the end-plate potential recorded with an intra-cellular electrode. *The Journal of Physiology*, 115: 320–370.

Fatt, P. and Katz, B. (1952) Spontaneous subthreshold activity at motor nerve endings. *The Journal of Physiology*, 117: 109–128.

Feelisch, M. and Martin, J.F. (1995) The early role of nitric oxide in evolution. *Trends in Ecology & Evolution*, 10 (12): 496–499. doi:10.1016/S0169-5347(00)89206-X.

Feng, W., Long, J.F., Fan, J.S., et al. (2004) The tetrameric L27 domain complex as an organization platform for supramolecular assemblies. *Nature Structural and Molecular Biology*, 11 (5): 475–480. doi:10.1038/nsmb751.

Feng, W. and Zhang, M. (2009) Organization and dynamics of PDZ-domain-related supramodules in the postsynaptic density. *Nature Reviews Neuroscience*, 10 (2): 87–99. doi:10.1038/nrn2540.

Fernandez, D., Bonilla, E., Mirza, N., et al. (2006) Rapamycin reduces disease activity and normalizes T cell activation-induced calcium fluxing in patients with systemic lupus erythematosus. *Arthritis and Rheumatism*, 54 (9): 2983–2988. doi:10.1002/art.22085.

Fernández, E., Collins, M.O., Uren, R.T., et al. (2009) Targeted tandem affinity purification of PSD-95 recovers core postsynaptic complexes and schizophrenia susceptibility proteins. *Molecular Systems Biology*, 5: 269. doi:10.1038/msb.2009.27.

Firestein, B.L. and Rongo, C. (2001) DLG-1 is a MAGUK similar to SAP97 and is required for adherens junction formation. *Molecular biology of the cell*, 12 (November): 3465–3475.

Forde, B.G. and Roberts, M.R. (2014) Glutamate receptor-like channels in plants: A role as amino acid sensors in plant defence? *F1000Prime Reports*, 6: 37. doi:10.12703/P6-37.

Franz, A. and Riechmann, V. (2010) Stepwise polarisation of the *Drosophila*

follicular epithelium. *Developmental Biology*, 338 (2): 136–147. doi:10.1016/j.ydbio.2009.11.027.

Fujisawa, T. (2008) *Hydra* peptide project 1993-2007. *Development, Growth & Differentiation*, 50: S257–S268. doi:10.1111/j.1440-169X.2008.00997.x.

Funke, L., Dakoji, S. and Bredt, D.S. (2005) Membrane-Associated Guanylate Kinases Regulate Adhesion and Plasticity At Cell Junctions. *Annual Review of Biochemistry*, 74 (1): 219–245. doi:10.1146/annurev.biochem.74.082803.133339.

Furness, J.B. and Stebbing, M.J. (2018) The first brain: species comparisons and evolutionary implications for the enteric and central nervous systems. *Neurogastroenterology and Motility*, 30: e13234. doi:10.1111/nmo.13234.

Gale, J.E., Marcotti, W., Kennedy, H.J., et al. (2001) FM1-43 dye behaves as a permeant blocker of the hair-cell mechanotransducer channel. *Journal of Neuroscience*, 21 (18): 7013–7025. doi:10.1523/jneurosci.21-18-07013.2001.

Garcia, E.P., Mehta, S., Blair, L.A.C., et al. (1998) SAP90 binds and clusters kainate receptors causing incomplete desensitization. *Neuron*, 21: 727–739.

Gardiol, D., Kühne, C., Glaunsinger, B., et al. (1999) Oncogenic human papillomavirus E6 proteins target the discs large tumour suppressor for proteasome mediated degradation. *Oncogene*, 18 (40): 5487–5496. doi:10.1038/sj.onc.1202920.

Gardner, L.A., Naren, A.P. and Bahouth, S.W. (2007) Assembly of an SAP97-AKAP79-cAMP-dependent protein kinase scaffold at the type 1 PSD-95/DLG/ZO1 motif of the human β1-adrenergic receptor generates a receptosome involved in receptor recycling and networking. *Journal of Biological Chemistry*, 282 (7): 5085–5099. doi:10.1074/jbc.M608871200.

Garner, C.C. and Kindler, S. (1996) Synaptic proteins and the assembly of synaptic junctions. *Trends in Cell Biology*, 6 (11): 429–433. doi:10.1016/S0962-8924(96)10036-2.

Garner, C.C., Nash, J. and Huganir, R.L. (2000) PDZ domains in synapse assembly and signalling. *Trends in Cell Biology*, 10 (7): 274–280. doi:10.1016/S0962-8924(00)01783-9.

Gasic, G.P. and Hollmann, M. (1992) Molecular neurobiology of glutamate receptors. *Annual Review of Physiology*, 54: 507–536.

Gaspar, T., Kevers, C., Faivre-Rampant, O., et al. (2003) Changing concepts in plant hormone action. *In Vitro Cellular & Developmental Biology - Plant*, 39: 85–106. doi:10.1079/IVP2002393.

George, R.A. and Heringa, J. (2002) Protein domain identification and improved sequence similarity searching using PSI-BLAST. *Proteins: Structure, Function, and Genetics*, 48 (4): 672–681. doi:10.1002/prot.10175.

Ghosh, A., Ramagopal, U.A., Bonanno, J.B., et al. (2018) Structures of the L27 d of Disc Large homologue 1 protein illustrate a self-assembly module.

Biochemistry, 57 (8): 1293–1305. doi:10.1021/acs.biochem.7b01074.

Goddard, R.H. and La Claire II, J.W. (1991) Calmodulin and wound healing in the coenocytic green alga Ernodesmis verticillata (Kützing) Børgesen : ultrastructure of the cortical cytoskeleton and immunogold labeling. *Planta*, 186: 17–26.

Gómez, V., Sesé, M., Santamaría, A., et al. (2010) Regulation of Aurora B kinase by the lipid raft protein Flotillin-1. *Journal of Biological Chemistry*, 285 (27): 20683–20690. doi:10.1074/jbc.M110.130591.

Gottardi, C., Arpin, M., Fanningt, A.S., et al. (1996) The junction-associated protein, zonula occludens-1, localizes to the nucleus before the maturation and during the remodeling of cell-cell contacts. *Cell Biology*, 93: 10779–10784.

Gouy, M., Guindon, S. and Gascuel, O. (2010) Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, 27 (2): 221–224. doi:10.1093/molbev/msp259.

Grabrucker, S., Proepper, C., Mangus, K., et al. (2014) The PSD protein ProSAP2/Shank3 displays synapto-nuclear shuttling which is deregulated in a schizophrenia-associated mutation. *Experimental Neurology*, 253: 126–137. doi:10.1016/j.expneurol.2013.12.015.

Graham, T.A., Clements, W.K., Kimelman, D., et al. (2002) The crystal structure of the β-catenin/ICAT complex reveals the inhibitory mechanism of ICAT. *Molecular Cell*, 10: 563–571.

Grau-Bové, X., Torruella, G., Donachie, S., et al. (2017) Dynamics of genomic innovation in the unicellular ancestry of animals. *eLife*, 6. doi:10.7554/eLife.26036.

Grimmelikhuijzen, C.J. and Westfall, J.A. (1995) “The nervous systems of cnidarians.” In Breidbach, O. and Kutsch, W. (eds.) *The nervous system of invertebrates: an evolutionary and comparative approach*. Birkhäuser Basel. pp. 7–24. doi:10.1007/978-3-0348-9219-3_2.

Grossoehme, N.E., Spuches, A.M. and Wilcox, D.E. (2010) Application of isothermal titration calorimetry in bioinorganic chemistry. *Journal of Biological Inorganic Chemistry*, 15 (8): 1183–1191. doi:10.1007/s00775-010-0693-3.

Grundfest, H. (1959) “Evolution of conduction in the nervous system.” In Bass, A.D. and Brodie, B.B. (eds.) *Evolution of nervous control from primitive organisms to man.* Whitefish MT, USA: Literary Licensing LLC. pp. 43–86.

Haen Whitmer, K.M. (2018) “Model systems for exploring the evolutionary origins of the nervous system.” In Kloc, M. and Kubiak, J.Z. (eds.) *Marine organisms as model systems in Biology and Medicine*. Cham, Switzerland: Springer International Publishing AG. pp. 185–196.

Harden, N., Hau Wang, S.J. and Krieger, C. (2016) Making the connection - shared molecular machinery and evolutionary links underlie the formation and

plasticity of occluding junctions and synapses. *Journal of Cell Science*, 129 (16): 3067–3076. doi:10.1242/jcs.186627.

Hayashi, M.K., Ames, H.M. and Hayashi, Y. (2006) Tetrameric hub structure of postsynaptic scaffolding protein Homer. *The Journal of Neuroscience*, 26 (33): 8492–8501. doi:10.1523/JNEUROSCI.2731-06.2006.

Hayashi, M.K., Tang, C., Verpelli, C., et al. (2009) The postsynaptic density proteins Homer and Shank form a polymeric network structure. *Cell*, 137 (1): 159–171. doi:10.1016/j.cell.2009.01.050.

Hejnol, A. and Rentzsch, F. (2015) Neural nets. *Current Biology*, 25 (18): R782–R786. doi:10.1016/j.cub.2015.08.001.

Hernandez-Nicaise, M.L. (1973) The nervous system of ctenophores III. Ultrastructure of synapses. *Journal of Neurocytology*, 2: 249–263. doi:10.1007/BF01104029.

Hillier, B.J., Christopherson, K.S., Prehoda, K.E., et al. (1999) Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-Syntrophin complex. *Science*, 284 (5415): 812–815.

Hobert, O. (2016) "Terminal selectors of neuronal identity." In *Current Topics in Developmental Biology*. Academic Press Inc. pp. 455–475. doi:10.1016/bs.ctdb.2015.12.007.

Hodgkin, A.L. (1951) The ionic basis of electrical activity in nerve and muscle. *Biological Reviews*, 26 (4): 339–409. doi:10.1111/j.1469-185X.1951.tb01204.x.

Hoelzle, M.K. and Svitkina, T. (2012) The cytoskeletal mechanisms of cell-cell junction formation in endothelial cells. *Molecular Biology of the Cell*, 23: 310–323. doi:10.1091/mbc.E11-08-0719.

Hoffmeyer, T.T. and Burkhardt, P. (2016) Choanoflagellate models - *Monosiga brevicollis* and *Salpingoeca rosetta*. *Current Opinion in Genetics and Development*, 39: 42–47. doi:10.1016/j.gde.2016.05.016.

Horridge, B.G.A. and Mackay, B. (1964) Neurociliary synapses in Pleurobrachia (Ctenophora). *The Quarterly Journal of Microscopical Science*, 105 (2): 163–174.

Horridge, G.A. (1965) Relations between nerves and cilia in ctenophores. *American Zoologist*, 5: 357–375. doi:10.1093/icb/5.3.357.

Horridge, G.A. and Mackay, B. (1962) Naked axons and symmetrical synapses in coelenterates. *Quarterly Journal of Microscopical Science*, 103 (4): 531–541. doi:10.1038/193899a0.

Hsueh, Y., Wang, T., Yang, F., et al. (2000) Nuclear translocation and guanylate kinase CASK / LIN-2. *Nature*, 404 (March).

Huang, G.N., Huso, D.L., Bouyain, S., et al. (2008) NFAT binding and regulation of T cell activation by the cytoplasmic scaffolding Homer proteins. *Science*, 319 (5862): 476–481. doi:10.1126/science.1151227.

Humbert, P., Russell, S. and Richardson, H. (2003) Dlg, Scribble and Lgl in cell polarity, cell proliferation and cancer. *BioEssays*, 25 (6): 542–553. doi:10.1002/bies.10286.

Humbert, P.O., Grzeschik, N.A., Brumby, A.M., et al. (2008) Control of tumourigenesis by the Scribble/Dlg/Lgl polarity module. *Oncogene*, 27 (55): 6888–6907. doi:10.1038/onc.2008.341.

Husi, H., Ward, M.A., Choudhary, J.S., et al. (2000) Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. *Nature Neuroscience*, 3 (7): 661–669. doi:10.1038/76615.

Huxley, J.S. (1935) Chemical regulation and the hormone concept. *Biological Reviews*, 10 (4): 427–441.

Innocenti, M., Gerboth, S., Rottner, K., et al. (2005) Abi1 regulates the activity of N-WASP and WAVE in distinct actin-based processes. *Nature Cell Biology*, 7 (10): 969–976. doi:10.1038/ncb1304.

Irie, M., Hata, Y., Takeuchi, M., et al. (1997) Binding of Neuroligins to PSD-95. *Science*, 277: 1511–1515.

Jager, M., Chiori, R., Alié, A., et al. (2011) New insights on ctenophore neural anatomy: immunofluorescence study in *Pleurobrachia pileus* (Müller, 1776). *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 316B (3): 171–187. doi:10.1002/jez.b.21386.

Jameson, D.M. and Mocz, G. (2005) “Fluorescence polarization / anisotropy approaches to study protein-ligand interactions.” In Nienhaus, G.U. (ed.) *Protein-ligand interactions. Methods in molecular biology™*. vol 305. Totowa, NJ, USA: Humana Press. pp. 301–322.

Jékely, G., Keijzer, F. and Godfrey-Smith, P. (2015a) An option space for early neural evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370: 20150181. doi:10.1098/rstb.2015.0181.

Jékely, G., Paps, J. and Nielsen, C. (2015b) The phylogenetic position of ctenophores and the origin(s) of nervous systems. *EvoDevo*, 6: 1. doi:10.1186/2041-9139-6-1.

Johnston, C.A., Doe, C.Q. and Prehoda, K.E. (2012) Structure of an enzyme-derived phosphoprotein recognition domain. *PLoS ONE*, 7 (4): e36014. doi:10.1371/.

de Juan, D., Pazos, F. and Valencia, A. (2013) Emerging methods in protein co-evolution. *Nature Reviews Genetics*, 14: 249–261. doi:10.1038/nrg3414.

Kandel, E.R., Schwartz, J.H. and Jessell, T.M. (eds.) (2000) *Principles of neural science*. 4th ed. New York: McGraw-Hill Medical.

Kang, T., Zhang, K., Yin, C., et al. (2016) Binding affinity analysis of the interaction between Homer EVH domain and ryanodine receptor with biosensors based on imaging ellipsometry. *Analytical Methods*, 8 (14): 2936–2940. doi:10.1039/C6AY00089D.

Kass-Simon, G. and Pierobon, P. (2007) Cnidarian chemical neurotransmission, an updated overview. *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, 146 (1): 9–25. doi:10.1016/j.cbpa.2006.09.008.

Kegel, L., Aunin, E., Meijer, D., et al. (2013) LGI proteins in the nervous system. *ASN Neuro*, 5 (3): e00115. doi:10.1042/AN20120095doi:10.1042/AN20120095.

Kennedy, M.B. (2000) Signal-processing machines at the postsynaptic density. *Science*, 290 (5492): 750–754. doi:10.1126/science.290.5492.750.

Kennedy, M.B. (2016) Synaptic signaling in learning and memory. *Cold Spring Harbor Perspectives in Biology*, 8: a016824. doi:10.1101/cshperspect.a016824.

Kim, E., Naisbitt, S., Hsueh, Y.-P., et al. (1997) GKAP, a novel synaptic protein that interacts with the guanylate kinase-like domain of the PSD-95/SAP90 family of channel clustering molecules. *The Journal of Cell Biology*, 136 (3): 669–678.

Kim, E., Niethammer, M., Rothschild, A., et al. (1995) Clustering of Shaker-type K+ channels by interaction with a family of membrane-associated guanylate kinases. *Nature*, 378: 85–88.

Kim, E. and Sheng, M. (2004) PDZ domain proteins of synapses. *Nature Reviews Neuroscience*, 5 (10): 771–781. doi:10.1038/nrn1517.

Kim, J.H., Kim, J.H., Yang, E., et al. (2009) Shank 2 expression coincides with neuronal differentiation in the developing retina. *Experimental and Molecular Medicine*, 41 (4): 236–242. doi:10.3858/emm.2009.41.4.026.

King, N., Westbrook, M.J., Young, S.L., et al. (2008) The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. *Nature*, 451 (7180): 783–788.

Kirillova, S., Kumar, S. and Carugo, O. (2009) Protein domain boundary predictions: a structural Biology perspective. *The Open Biochemistry Journal*, 3: 1–8. doi:10.2174/1874091x00903010001.

Kirkegaard, J.B., Marron, A.O. and Goldstein, R.E. (2016) Motility of colonial choanoflagellates and the statistics of aggregate random walkers. *Physical Review Letters*, 116: 038102. doi:10.1103/PhysRevLett.116.038102.

Kistner, U., Garner, C.C. and Linial, M. (1995) Nucleotide bindind by the synapse associated protein SAP90. *FEBS Letters*, 359 (2–3): 159–163. doi:10.1016/0014-5793(95)00030-D.

Klemmer, P., Smit, A.B. and Li, K.W. (2009) Proteomics analysis of immuno-precipitated synaptic protein complexes. *Journal of Proteomics*, 72: 82–90. doi:10.1016/j.jprot.2008.10.005.

Kloepper, T.H., Kienle, C.N. and Fasshauer, D. (2007) An elaborate classification of SNARE proteins sheds light on the conservation of the eukaryotic endomembrane system. *Molecular Biology of the Cell*, 18 (9): 3463–3471. doi:10.1091/mbc.E07-03-0193.

Knöpfel, T., Kuhn, R. and Allgeier, H. (1995) Metabotropic glutamate receptors: novel targets for drug development. *Journal of Medicinal Chemistry*, 38 (9): 1417–1426. doi:10.1021/jm00009a001.

Kohu, K., Ogawa, F. and Akiyama, T. (2002) The SH3, HOOK and guanylate kinase-like domains of hDLG are important for its cytoplasmic localization. *Genes to Cells*, 7: 707–715. doi:10.1046/j.1365-2443.2002.00555.x.

Koonin, E. V (2005) Orthologs, paralogs, and evolutionary genomics. *The Annual Review of Genetics*, 39: 309–38. doi:10.1146/annurev.genet.39.073003.114725.

Kristan, Jr., W.B. (2016) Early evolution of neurons. *Current Biology*, 26 (20): R949–R954. doi:10.1016/j.cub.2016.05.030.

Krugmann, S., Jordens, I., Gevaert, K., et al. (2001) Cdc42 induces filopodia by promoting the formation of an IRSp53:Mena complex. *Current Biology*, 11 (21): 1645–1655.

Kruijssen, D.L.H. and Wierenga, C.J. (2019) Single synapse LTP: a matter of context? *Frontiers in Cellular Neuroscience*, 13. doi:10.3389/fncel.2019.00496.

Kuffler, S.W. and Edwards, C. (1958) Mechanism of gamma aminobutyric acid (GABA) action and its relation to synaptic inhibition. *Journal of Neurophysiology*, 21 (6): 589–610.

Lahey, T., Gorczyca, M., Jia, X.-X., et al. (1994) The Drosophila tumor suppressor gene dlg is required for normal synaptic bouton structure. *Neuron*, 13 (4): 823–835.

Lam, H.M., Chiu, J., Hsieh, M.H., et al. (1998) Glutamate-receptor genes in plants. *Nature*, 396 (6707): 125–126. doi:10.1038/24066.

Lancaster, C.E., Ho, C.Y., Hipolito, V.E.B., et al. (2019) Phagocytosis: what's on the menu? *Biochemistry and Cell Biology*, 97: 21–29. doi:10.1139/bcb-2018-0008.

Lanner, J.T., Georgiou, D.K., Joshi, A.D., et al. (2010) Ryanodine receptors: structure, expression, molecular details, and function in calcium release. *Cold Spring Harbor perspectives in biology*, 2 (11): a003996. doi:10.1101/cshperspect.a003996.

Laundon, D., Larson, B.T., McDonald, K., et al. (2019) The architecture of cell differentiation in choanoflagellates and sponge choanocytes. *PLoS Biology*, 17 (4). doi:10.1371/journal.pbio.3000226.

Lea, W.A. and Simeonov, A. (2011) Fluorescence polarization assays in small molecule screening. *Expert Opinion on Drug Discovery*. 6 (1) pp. 17–32. doi:10.1517/17460441.2011.537322.

Leadbeater, B.S.C. (2015) *The choanoflagellates: evolution, biology, and ecology.* Cambridge, UK: Cambridge University Press.

Leadbeater, B.S.C. and Morton, C. (1974) A microscopical study of a marine

species of Codosiga James-Clark (Choanoflagellata) with special reference to the ingestion of bacteria. *Biological Journal of the Linnean Society*, 6 (4): 337–347.

Leondaritis, G. and Eickholt, B.J. (2015) Short lives with long-lasting effects: filopodia protrusions in neuronal branching morphogenesis. *PLOS Biology*, 13 (9): e1002241. doi:10.1371/journal.pbio.1002241.

Leonoudakis, D., Conti, L.R., Anderson, S., et al. (2004) Protein trafficking and anchoring complexes revealed by proteomic analysis of inward rectifier potassium channel (Kir2.x)-associated proteins. *The Journal of Biological Chemistry*, 279 (21): 22331–22346.

Letunic, I. and Bork, P. (2018) 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research*, 46 (D1): D493–D496. doi:10.1093/nar/gkx922.

Leuckart, R. (1848) *Ueber Morphologie und Verwandtschaftsverhaeltnisse der wirbellosen Tiere.* Braunschweig: Friedrich Vieweg und Sohn.

Levin, T.C., Greaney, A.J., Wetzel, L., et al. (2014) The Rosetteless gene controls development in the choanoflagellate *S. rosetta*. *eLife*, 3: e04070. doi:10.7554/eLife.04070.

Levin, T.C. and King, N. (2013) Evidence for sex and recombination in the choanoflagellate *Salpingoeca rosetta*. *Current biology : CB*, 23 (21): 2176–2180. doi:10.1016/j.cub.2013.08.061.

Li, J., Mahajan, A. and Tsai, M.D. (2006) Ankyrin repeat: A unique motif mediating protein-protein interactions. *Biochemistry*, 45 (51): 15168–15178. doi:10.1021/bi062188q.

Li, L., Stoeckert, C.J. and Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, 13: 2178–2189. doi:10.1101/gr.1224503.

Li, R., Neundorf, I. and Nitsche, F. (2018) First efficient transfection in choanoflagellates using cell-penetrating peptides. *bioRxiv*. doi:10.1101/260190.

Li, Y., Karnak, D., Demeler, B., et al. (2004) Structural basis for L27 domain-mediated assembly of signaling and cell polarity complexes. *EMBO Journal*, 23 (14): 2723–2733. doi:10.1038/sj.emboj.7600294.

Liebeskind, B.J., Hillis, D.M. and Zakon, H.H. (2011) Evolution of sodium channels predates the origin of nervous systems in animals. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (22): 9154–9159. doi:10.1073/pnas.1106363108.

Liebeskind, B.J., Hillis, D.M., Zakon, H.H., et al. (2016) Complex Homology and the Evolution of Nervous Systems. *Trends in Ecology and Evolution*, 31 (2): 127–135. doi:10.1016/j.tree.2015.12.005.

Lilja, J., Zacharchenko, T., Georgiadou, M., et al. (2017) SHANK proteins limit integrin activation by directly interacting with Rap1 and R-Ras. *Nature Cell*

Biology, 19 (4): 292–305. doi:10.1038/ncb3487.

Ling, E.M., Smith, T., Nguyen, X.D., et al. (2004) Relation of CD4+CD25+ regulatory T-cell suppression of allergen-driven T-cell activation to atopic status and expression of allergic disease. *Lancet*, 363 (9409): 608–615. doi:10.1016/S0140-6736(04)15592-X.

Linhoff, M.W., Laurén, J., Cassidy, R.M., et al. (2009) An unbiased expression screen for synaptogenic proteins identifies the LRRTM protein family as synaptic organizers. *Neuron*, 61: 734–749. doi:10.1016/j.neuron.2009.01.017.

López-Muñoz, F., Boya, J. and Alamo, C. (2006) Neuron theory, the cornerstone of neuroscience, on the centenary of the Nobel Prize award to Santiago Ramón y Cajal. *Brain Research Bulletin*, 70: 391–405. doi:10.1016/j.brainresbull.2006.07.010.

Lowe, J.S. and Anderson, P.G. (2015) "Epithelial Cells." In *Stevens Lowe's Human Histology*. 4th ed. Phyladelphia, PA, USA: Elsevier MOSBY. pp. 37–54. doi:10.1016/b978-0-7234-3502-0.00003-6.

Lu, M.S. and Johnston, C.A. (2013) Molecular pathways regulating mitotic spindle orientation in animals cells. *Development*, 140: 1843–1856. doi:10.1242/dev.087627.

Ludeman, D.A., Farrar, N., Riesgo, A., et al. (2014) Evolutionary origins of sensation in metazoans: functional evidence for a new sensory organ in sponges. *BMC Evolutionary Biology*, 14: 3. doi:10.1186/1471-2148-14-3.

Machemer, H. and Ogura, A. (1979) Ionic conductances of membranes in ciliated and deciliated *Paramecium*. *The Journal of Physiology*, 296 (1): 49–60. doi:10.1113/jphysiol.1979.sp012990.

Mackie, G.O. (1970) Neuroid conduction and the evolution of conducting tissues. *The Quarterly Review of Biology*, 45 (4): 319–332.

Mah, J.L., Christensen-Dalsgaard, K.K. and Leys, S.P. (2014) Choanoflagellate and choanocyte collar-flagellar systems and the assumption of homology. *Evolution & Development*, 16 (1): 25–37. doi:10.1111/ede.12060.

Malosio, M.L., Benfante, R., Racchetti, G., et al. (1999) Neurosecretory cells without neurosecretion: evidence of an independently regulated trait of the cell phenotype. *Journal of Physiology*, 520 (1): 43–52. doi:10.1111/j.1469-7793.1999.t01-1-00043.x.

Manning, G., Young, S.L., Miller, W.T., et al. (2008) The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (28): 9674–9679. doi:10.1073/pnas.0801314105.

Martín-Durán, J.M. and Hejnol, A. (2019) A developmental perspective on the evolution of the nervous system. *Developmental Biology*. doi:10.1016/j.ydbio.2019.10.003.

Masu, M., Tanabe, Y., Tsuchida, K., et al. (1991) Sequence and expression of a metabotropic glutamate receptor. *Nature*, 349 (6312): 760–765. doi:10.1038/349760a0.

Mathew, D., Gramates, L.S., Packard, M., et al. (2002) Recruitment of Scribble to the synaptic scaffolding complex requires GUK-holder, a novel DLG binding protein. *Current Biology*, 12 (7): 531–539. doi:10.1016/S0960-9822(02)00758-3.

Mauceri, D., Cattabeni, F., Di Luca, M., et al. (2004) Printed in calmodulin-dependent protein kinase II phosphorylation drives synapse-associated protein 97 into spines. *Journal of Biological Chemistry*, 279 (22): 23813–23821. doi:10.1074/jbc.M402796200.

Mayer, M.L., Westbrook, G.L. and Guthrie, P.B. (1984) Voltage-dependent block by Mg2+ of NMDA responses in spinal cord neurones. *Nature*, 309: 261–263.

Mayorova, T.D., Hammar, K., Winters, C.A., et al. (2019) The ventral epithelium of *Trichoplax adhaerens* deploys in distinct patterns cells that secrete digestive enzymes, mucus or diverse neuropeptides. *Biology Open*, 8: bio045674. doi:10.1242/bio.045674.

McGee, A.W., Dakoji, S.R., Olsen, O., et al. (2001) Structure of the SH3-guanylate kinase module from PSD-95 suggests a mechanism for regulated assembly of MAGUK scaffolding proteins. *Molecular Cell*, 8 (6): 1291–1301. doi:10.1016/S1097-2765(01)00411-7.

Meldrum, B.S. (2000) Glutamate as a Neurotransmitter in the Brain: Review of Physiology and Pathology. *The Journal of Nutrition*, 130 (4): 1007S-1015S. doi:10.1093/jn/130.4.1007s.

de Mendoza, A., Suga, H., Permanyer, J., et al. (2015) Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. *eLife*, 4: e08904. doi:10.7554/eLife.08904.

de Mendoza, A., Suga, H. and Ruiz-Trillo, I. (2010) Evolution of the MAGUK protein gene family in premetazoan lineages. *BMC Evolutionary Biology*, 10 (1): 93. doi:10.1186/1471-2148-10-93.

Mi, H., Dong, Q., Muruganujan, A., et al. (2010) PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Research*, 38 (Database issue): D204–D210. doi:10.1093/nar/gkp1019.

Mikhailov, K. V., Konstantinova, A. V., Nikitin, M.A., et al. (2009) The origin of Metazoa: a transition from temporal to spatial cell differentiation. *BioEssays*, 31 (7): 758–768. doi:10.1002/bies.200800214.

Mikrjukov, K.A. and Mylnikov, A.P. (2001) A study of the fine structure and the mitosis of a lamellicristate amoeba, *Micronuclearia podoventralis* gen. et sp. nov. (Nucleariidae, Rotosphaerida). *European Journal of Protistology*, 37 (1): 15–24. doi:10.1078/0932-4739-00783.

Moncada, S. (1999) Nitric oxide: discovery and impact on clinical medicine. *Journal of the Royal Society of Medicine*, 92 (4): 164–169. doi:10.1177/014107689909200402.

Montgomery, J.M., Zamorano, P.L. and Garner, C.C. (2004) MAGUKs in synapse assembly and function: an emerging view. *Cellular and Molecular Life Sciences (CMLS)*, 61 (7–8): 911–929. doi:10.1007/s00018-003-3364-5.

Moran, Y., Barzilai, M.G., Liebeskind, B.J., et al. (2015) Evolution of voltage-gated ion channels at the emergence of Metazoa. *The Journal of Experimental Biology*, 218 (4): 515–25. doi:10.1242/jeb.110270.

Moroz, L.L., Kocot, K.M., Citarella, M.R., et al. (2014) The ctenophore genome and the evolutionary origins of neural systems. *Nature*, 510 (7503): 109–114. doi:10.1038/nature13400.

Moroz, L.L. and Kohn, A.B. (2016) Independent origins of neurons and synapses: insights from ctenophores. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 371: 20150041.

Musser, J.M., Schippers, K.J., Nickel, M., et al. (2019) Profiling cellular diversity in sponges informs animal cell type and nervous system evolution. *bioRxiv*, p. 758276. doi:10.1101/758276.

de Nadal, E., Ammerer, G. and Posas, F. (2011) Controlling gene expression in response to stress. *Nature reviews. Genetics*, 12 (12): 833–845. doi:10.1038/nrg3055.

Naisbitt, S., Eunjoon, K., Tu, J.C., et al. (1999) Shank, a novel family of postsynaptic density proteins that binds to the NMDA receptor/PSD-95/GKAP complex and cortactin. *Neuron*, 23 (3): 569–582. doi:10.1016/S0896-6273(00)80809-0.

Nakagawa, T., Futai, K., Lashuel, H.A., et al. (2004) Quaternary structure, protein dynamics, and synaptic function of SAP97 controlled by L27 domain interactions. *Neuron*, 44 (3): 453–467. doi:10.1016/j.neuron.2004.10.012.

Naumann, B. and Burkhardt, P. (2019) Spatial cell disparity in the colonial choanoflagellate *Salpingoeca rosetta*. *Frontiers in Cell and Developmental Biology*, 7: 231. doi:10.3389/fcell.2019.00231.

Neurath, M.F., Weigmann, B., Finotto, S., et al. (2002) The transcription factor T-bet regulates mucosal T cell activation in experimental colitis and Crohn's disease. *Journal of Experimental Medicine*, 195 (9): 1129–1143. doi:10.1084/jem.20011956.

Nguyen, L., Schmidt, H.A., Haeseler, A. Von, et al. (2014) IQ-TREE : A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32 (1): 268–274. doi:10.1093/molbev/msu300.

Nickel, M. (2010) Evolutionary emergence of synaptic nervous systems: What can we learn from the non-synaptic, nerveless Porifera? *Invertebrate Biology*, 129 (1): 1–16. doi:10.1111/j.1744-7410.2010.00193.x.

Nielsen, C. (2008) Six major steps in animal evolution: Are we derived sponge larvae? *Evolution and Development*, 10 (2): 241–257. doi:10.1111/j.1525-142X.2008.00231.x.

Niethammer, M., Kim, E. and Sheng, M. (1996) Interaction between the C terminus of NMDA receptor subunits and multiple members of the PSD-95 family of membrane-associated guanylate kinases. *The Journal of Neuroscience*, 76 (7): 2157–2163.

Nitsche, F., Carr, M., Arndt, H., et al. (2011) Higher level taxonomy and molecular phylogenetics of the Choanoflagellatea. *Journal of Eukaryotic Microbiology*, 58 (5): 452–462.

Norberg, R., Thorstensson, R. and Utter, G. (1982) Non-specific precipitin reactions of IgG at low ionic strength. *Journal of Immunological Methods*, 49 (1): 113–116. doi:10.1016/0022-1759(82)90372-6.

Okabe, S. (2007) Molecular anatomy of the postsynaptic density. *Molecular and Cellular Neuroscience*. 34 (4) pp. 503–518. doi:10.1016/j.mcn.2007.01.006.

Ota, S., Eikrem, W. and Edvardsen, B. (2012) Ultrastructure and molecular phylogeny of thaumatomonads (Cercozoa) with emphasis on *Thaumatomastix salina* from Oslofjorden, Norway. *Protist*, 163 (4): 560–573. doi:10.1016/j.protis.2011.10.007.

Paarmann, I., Spangenberg, O., Lavie, A., et al. (2002) Formation of complexes between Ca^{2+}-calmodulin and the synapse-associated protein SAP97 requires the SH3 domain-guanylate kinase domain-connecting HOOK region. *Journal of Biological Chemistry*, 277 (43): 40832–40838. doi:10.1074/jbc.M205618200.

Paketurytė, V., Zubrienė, A., Chen, W.-Y., et al. (2019) “Inhibitor binding to carbonic anhydrases by isothermal titration calorimetry.” In Matulis, D. (ed.) *Carbonic anhydrase as drug target*. Cham, Switzerland: Springer International Publishing. pp. 79–95. doi:10.1007/978-3-030-12780-0_6.

Palay, S.L. (1956) Synapses in the central nervous system. *Journal of Biophysical and Biochemical Cytology*, 2 (4): 193–207.

Pang, Z.P., Shin, O.H., Meyer, A.C., et al. (2006) A gain-of-function mutation in synaptotagmin-1 reveals a critical role of Ca^{2+}-dependent soluble N-ethylmaleimide-sensitive factor attachment protein receptor complex binding in synaptic exocytosis. *Journal of Neuroscience*, 26 (48): 12556–12565. doi:10.1523/JNEUROSCI.3804-06.2006.

Parker, G.H. (1919) *The elementary nervous system*. Loeb, J., Morgan, T.H. and Osterhout, W.J. V. (eds.). Philadelphia, USA: J. B. Lippincott Company.

Passafaro, M., Sala, C., Niethammer, M., et al. (1999) Microtubule binding by CRIPT and its potential role in the synaptic clustering of PSD-95. *Nature Neuroscience*, 2 (12): 1063–1069.

Pawlowski, J. (2008) The twilight of Sarcodina: a molecular perspective on the polyphyletic origin of amoeboid protists. *Protistology*, 5 (4): 281–302.

Pawson, T. and Scott, J.D. (1997) Signaling through scaffold, anchoring, and adaptor proteins. *Science*, 278 (5346): 2075–2080.

Pearson, W.R. (2013) An introduction to sequence similarity ("homology") searching. *Current Protocols in Bioinformatics*, 42 (1): 3–1. doi:10.1002/0471250953.bi0301s42.

Perrimon, N., Pitsouli, C. and Shilo, B.Z. (2012) Signaling mechanisms controlling cell fate and embryonic patterning. *Cold Spring Harbor Perspectives in Biology*, 4 (8). doi:10.1101/cshperspect.a005975.

Pettitt, M.E., Orme, B. a a, Blake, J.R., et al. (2002) The hydrodynamics of filter feeding in choanoflagellates. *European Journal of Protistology*, 38: 313–332. doi:10.1078/0932-4739-00854.

Pike, L.J. (2006) Rafts defined: a report on the keystone symposium on lipid rafts and cell function. *The Journal of Lipid Research*, 47 (7): 1597–1598. doi:10.1194/jlr.E600002-JLR200.

Plickert, G. and Schneider, B. (2004) "Neuropeptides and photic behavior in Cnidaria." In *Hydrobiologia*. November 2004. pp. 49–57. doi:10.1007/s10750-004-2689-x.

Poliak, S., Matlis, S., Ullmer, C., et al. (2002) Distinct claudins and associated PDZ proteins form different autotypic tight junctions in myelinating Schwann cells. *Journal of Cell Biology*, 159 (2): 361–371. doi:10.1083/jcb.200207050.

Poliak, S., Salomon, D., Elhanany, H., et al. (2003) Juxtaparanodal clustering of Shaker-like K^+ channels in myelinated axons depends on Caspr2 and TAG-1. *Journal of Cell Biology*, 162 (6): 1149–1160. doi:10.1083/jcb.200305018.

Ponting, C.P. and Phillips, C. (1997) Identification of Homer as a homologue of the Wiskott-Aldrich syndrome protein suggests a receptor-binding function for WH1 domains. *Journal of Molecular Medicine*, 75: 769–771.

Pouliquin, P. and Dulhunty, A.F. (2009) Homer and the ryanodine receptor. *European Biophysical Societies*, 39: 91–102. doi:10.1007/s00249-009-0494-1.

Praetorius, H.A. and Spring, K.R. (2001) Bending the MDCK cell primary cilium increases intracellular calcium. *The Journal of Membrane Biology*, 184: 71–79. doi:10.1007/s00232-001-0075-4.

Prast, H. and Philippu, A. (2001) Nitric oxide as modulator of neuronal function. *Progress in Neurobiology*. 64 (1) pp. 51–68. doi:10.1016/S0301-0082(00)00044-7.

Preston, T.M. and King, C.A. (2003) Locomotion and phenotypic transformation of the amoeboflagellate *Naegleria gruberi* at the water-air interface. *The Journal of Eukaryotic Microbiology*, 50 (4): 245–251. doi:10.1111/j.1550-7408.2003.tb00128.x.

Proepper, C., Johannsen, S., Liebau, S., et al. (2007) Abelson interacting protein 1 (Abi-1) is essential for dendrite morphogenesis and synapse formation. *EMBO Journal*, 26 (5): 1397–1409. doi:10.1038/sj.emboj.7601569.

Proepper, C., Steinestel, K., Schmeisser, M.J., et al. (2011) Heterogeneous nuclear ribonucleoprotein K interacts with Abi-1 at postsynaptic sites and modulates dendritic spine morphology. *PLoS ONE*, 6 (11): e27045. doi:10.1371/journal.pone.0027045.

Puig, O., Caspary, F., Rigaut, G., et al. (2001) The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *METHODS*, 24: 218–229. doi:10.1006/meth.2001.1183.

Putnam, N.H., Srivastava, M., Hellsten, U., et al. (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science*, 317 (5834): 86–94. doi:10.1126/science.1139158.

Quina, L.A., Wang, S., Ng, L., et al. (2009) Brn3a and Nurr1 mediate a gene regulatory pathway for habenula development. *Journal of Neuroscience*, 29 (45): 14309–14322. doi:10.1523/JNEUROSCI.2430-09.2009.

R Core Team (2017) *R: a language and environment for statistical computing.* Available at: https://www.r-project.org/.

Rademacher, N., Schmerl, B., Lardong, J.A., et al. (2016) MPP2 is a postsynaptic MAGUK scaffold protein that links SynCAM1 cell adhesion molecules to core components of the postsynaptic density. *Scientific Reports*, 6 (1): 35283. doi:10.1038/srep35283.

Rajanala, K., Nandicoori, V.K. and Mata, J. (2012) Localization of nucleoporin Tpr to the nuclear pore complex is essential for Tpr mediated regulation of the export of unspliced RNA. *PLoS ONE*, 7 (1): e29921. doi:10.1371/journal.pone.0029921.

Ramos-Vicente, D., Ji, J., Gratacòs-Batlle, E., et al. (2018) Metazoan evolution of glutamate receptors reveals unreported phylogenetic groups and divergent lineage-specific events. *eLife*, 7. doi:10.7554/eLife.35774.

Reiner, A. and Levitz, J. (2018) Glutamatergic signaling in the central nervous system: ionotropic and metabotropic receptors in concert. *Neuron*, 98: 1080–1098. doi:10.1016/j.neuron.2018.05.018.

Rentzsch, F., Layden, M. and Manuel, M. (2017) The cellular and molecular basis of cnidarian neurogenesis. *Wiley Interdisciplinary Reviews: Developmental Biology*, 6: e257. doi:10.1002/wdev.257.

Richards, G.S. and Rentzsch, F. (2014) Transgenic analysis of a *SoxB* gene reveals neural progenitor cells in the cnidarian *Nematostella vectensis*. *Development (Cambridge)*, 141 (24): 4681–4689. doi:10.1242/dev.112029.

Richards, G.S. and Rentzsch, F. (2015) Regulation of *Nematostella* neural progenitors by *SoxB*, Notch and bHLH genes. *Development (Cambridge)*, 142 (19): 3332–3342. doi:10.1242/dev.123745.

Richter, D.J., Fozouni, P., Eisen, M.B., et al. (2018) Gene family innovation, conservation and loss on the animal stem lineage. *eLife*, 7: e34226. doi:10.7554/eLife.34226.

Richter, D.J. and King, N. (2013) The genomic and cellular foundations of animal origins. *Annual review of genetics*, 47: 509–537. doi:10.1146/annurev-genet-111212-133456.

Richter, D.J. and Nitsche, F. (2017) “Choanoflagellatea.” In Archibald, J.M., Simpson, A.G.B. and Slamovits, C.H. (eds.) *Handbook of the protists*. 2nd ed. Cham, Switzerland: Springer International Publishing AG. pp. 1479–1496. doi:10.1007/978-3-319-28149-0_5.

Riesgo, A., Farrar, N., Windsor, P.J., et al. (2014) The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. *Molecular Biology and Evolution*, 31 (5): 1102–1120. doi:10.1093/molbev/msu057.

Roh, M.H., Makarova, O., Liu, C.J., et al. (2002) The Maguk protein, Pals1, functions as an adapter, linking mammalian homologues of crumbs and discs lost. *Journal of Cell Biology*, 157 (1): 161–172. doi:10.1083/jcb.200109010.

Rstudio Team (2015) *Rstudio: integrated development for R*. Available at: http://www.rstudio.com.

Ruiz-Trillo, I., Roger, A.J., Burger, G., et al. (2008) A phylogenomic investigation into the origin of Metazoa. *Molecular Biology and Evolution*, 25 (4): 664–672. doi:10.1093/molbev/msn006.

Rumbaugh, G., Sia, G.M., Garner, C.C., et al. (2003) Synapse-associated protein-97 isoform-specific regulation of surface AMPA receptors and synaptic function in cultured neurons. *Journal of Neuroscience*, 23 (11): 4567–4576. doi:10.1523/jneurosci.23-11-04567.2003.

Ryan, J.F. and Chiodin, M. (2015) Where is my mind? How sponges and placozoans may have lost neural cell types. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 370: 20150059.

Ryan, J.F., Pang, K., Schnitzler, C.E., et al. (2013) The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science*, 342 (6164): 1242592. doi:10.1126/science.1242592.

Saadaoui, M., Machicoane, M., di Pietro, F., et al. (2014) Dlg1 controls planar spindle orientation in the neuroepithelium through direct interaction with LGN. *The Journal of Cell Biology*, 206 (6): 707–717. doi:10.1083/jcb.201405060.

Sachkova, M. and Burkhardt, P. (2019) “Exciting times to study the identity and evolution of cell types.” In *Development (Cambridge)*. 2019. Company of Biologists Ltd. p. dev178996. doi:10.1242/dev.178996.

Sakarya, O., Armstrong, K.A., Adamska, M., et al. (2007) A post-synaptic scaffold at the origin of the animal kingdom. *PLoS ONE*, 2 (6): e506. doi:10.1371/journal.pone.0000506.

Sala, C., Piëch, V., Wilson, N.R., et al. (2001) Regulation of dendritic spine morphology and synaptic function by Shank and Homer. *Neuron*, 31: 115–130. doi:10.1016/S0896-6273(01)00339-7.

Santamaria, A., Castellanos, E., Gomez, V., et al. (2005) PTOV1 enables the nuclear translocation and mitogenic activity of Flotillin-1, a major protein of lipid rafts. *Molecular and Cellular Biology*, 25 (5): 1900–1911. doi:10.1128/mcb.25.5.1900-1911.2005.

Santana, M.M., Gonzalez, J.M. and Cruz, C. (2017) Nitric oxide accumulation: The evolutionary trigger for phytopathogenesis. *Frontiers in Microbiology*, 8 (OCT): 1947. doi:10.3389/fmicb.2017.01947.

Satterlie, R.A. (2011) Do jellyfish have central nervous systems? *Journal of Experimental Biology*. 214 (8) pp. 1215–1223. doi:10.1242/jeb.043687.

Schierwater, B. and Eitel, M. (2015) "Placozoa." In Wanninger, A. (ed.) *Evolutionary developmental biology of invertebrates*. Vienna, Austria: Springer. pp. 107–114.

Schlaepfer, C.H. and Wessel, R. (2015) Excitable membranes and action potentials in paramecia: An analysis of the electrophysiology of ciliates. *Journal of Undergraduate Neuroscience Education*, 14 (1): A82–A86.

Schlüter, O.M., Xu, W. and Malenka, R.C. (2006) Alternative N-Terminal domains of PSD-95 and SAP97 govern activity-dependent regulation of synaptic AMPA receptor function. *Neuron*, 51 (1): 99–111. doi:10.1016/j.neuron.2006.05.016.

Schroer, T.A. and Steuer, E.R. (1989) Cytoplasmic dynein is a minus end-directed motor for membranous organelles. *Cell*, 56: 937–946.

Schultz, J., Milpetz, F., Bork, P., et al. (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences of the United States of America*, 95: 5857–5864. doi:10.1073/pnas.95.11.5857.

Sebé-Pedrós, A., Burkhardt, P., Sánchez-Pons, N., et al. (2013a) Insights into the origin of metazoan filopodia and microvilli. *Molecular Biology and Evolution*, 30 (9): 2013–2023. doi:10.1093/molbev/mst110.

Sebé-Pedrós, A., Chomsky, E., Pang, K., et al. (2018a) Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nature Ecology and Evolution*, 2 (7): 1176–1188. doi:10.1038/s41559-018-0575-6.

Sebé-Pedrós, A., Degnan, B.M. and Ruiz-Trillo, I. (2017) The origin of Metazoa: a unicellular perspective. *Nature Reviews Genetics*, 18: 498–512. doi:10.1038/nrg.2017.21.

Sebé-Pedrós, A., Irimia, M., del Campo, J., et al. (2013b) Regulated aggregative multicellularity in a close unicellular relative of metazoa. *eLife*, 2: e01287. doi:10.7554/eLife.01287.

Sebé-Pedrós, A., Saudemont, B., Chomsky, E., et al. (2018b) Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-Seq. *Cell*, 173 (6): 1520-1534.e20. doi:10.1016/j.cell.2018.05.019.

Senatore, A., Reese, T.S. and Smith, C.L. (2017) Neuropeptidergic integration

of behavior in *Trichoplax adhaerens*, an animal without synapses. *Journal of Experimental Biology*, 220 (18): 3381–3390. doi:10.1242/jeb.162396.

Serrano-Saiz, E., Leyva-Díaz, E., De La Cruz, E., et al. (2018) BRN3-type POU Homeobox genes maintain the identity of mature postmitotic neurons in nematodes and mice. *Current Biology*, 28 (17): 2813-2823.e2. doi:10.1016/j.cub.2018.06.045.

Seymour, G.B., Tucker, G. and Leach, L.A. (2004) Cell adhesion molecules in plants and animals. *Biotechnology and Genetic Engineering Reviews*, 21 (1): 123–132. doi:10.1080/02648725.2004.10648051.

Shen, K. and Cowan, C.W. (2010) Guidance molecules in synapse formation and plasticity. *Cold Spring Harbor perspectives in biology*. 2 (4). doi:10.1101/cshperspect.a001842.

Shen, X.-X., Hittinger, C.T. and Rokas, A. (2017) Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecology & Evolution*, 1 (5): 0126. doi:10.1038/s41559-017-0126.

Sheng, M. and Kim, E. (2011) The postsynaptic organization of synapses. *Cold Spring Harbor Perspectives in Biology*, 3: a005678. doi:10.1101/cshperspect.a005678.

Sherman, D.L. and Brophy, P.J. (2000) A tripartite nuclear localization signal in the PDZ-domain protein L- periaxin. *Journal of Biological Chemistry*, 275 (7): 4537–4840. doi:10.1074/jbc.275.7.4537.

Shin, J., Shin, Y., Oh, S.-M., et al. (2014) MiR-29b controls fetal mouse neurogenesis by regulating ICAT-mediated Wnt/β-catenin signaling. *Cell death & disease*, 5 (10): e1473. doi:10.1038/cddis.2014.439.

Shiraishi-Yamaguchi, Y. and Furuichi, T. (2007) The Homer family proteins. *Genome Biology*, 8 (2): 206. doi:10.1186/gb-2007-8-2-206.

Shiraishi, Y., Mizutani, A., Bito, H., et al. (1999) Cupidin, an isoform of Homer/Vesl, interacts with the actin cytoskeleton and activated rho family small GTPases and is expressed in developing mouse cerebellar granule cells. *The Journal of Neuroscience*, 19 (19): 8389–8400.

Siegrist, S.E. and Doe, C.Q. (2005) Microtubule-induced Pins/Gαi cortical polarity in *Drosophila* neuroblasts. *Cell*, 123 (7): 1323–1335. doi:10.1016/j.cell.2005.09.043.

Simion, P., Philippe, H., Baurain, D., et al. (2017) A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Current Biology*, 27 (7): 958–967. doi:10.1016/j.cub.2017.02.031.

Simmons, D.K. and Martindale, M.Q. (2016) “Ctenophora.” In Schmidt-Rhaesa, A., Harzsch, S. and Purschke, G. (eds.) *Structure and evolution of invertebrate nervous systems*. Oxford, UK: Oxford University Press.

Smith, C.L., Pivovarova, N. and Reese, T.S. (2015) Coordinated feeding behavior in *Trichoplax*, an animal without synapses. *PLoS ONE*, 10 (9): 1–15.

doi:10.1371/journal.pone.0136098.

Soltau, M., Berhorster, K., Kindler, S., et al. (2004) Insulin receptor substrate of 53kDa links postsynaptic shank to PSD-95. *Journal of Neurochemistry*, 90 (3): 659–665. doi:10.1111/j.1471-4159.2004.02523.x.

Squire, L.R., Bloom, F.E., Spitzer, N.C., et al. (eds.) (2008) *Fundamental Neuroscience*. 3rd ed. Academic Press/Elsevier.

Srivastava, M., Begovic, E., Chapman, J., et al. (2008) The *Trichoplax* genome and the nature of placozoans. *Nature*, 454 (7207): 955–960. doi:10.1038/nature07191.

Steinhardt, R.A., Bi, G. and Alderton, J.M. (1994) Cell membrane resealing by a vesicular mechanism similar to neurotransmitter release. *Science*, 263 (5145): 390–393. doi:10.1126/science.7904084.

Stiber, J.A., Tabatabaei, N., Hawkins, A.F., et al. (2005) Homer modulates NFAT-dependent signaling during muscle differentiation. *Developmental Biology*, 287 (2): 213–224. doi:10.1016/j.ydbio.2005.06.030.

Stolzer, M., Siewert, K., Lai, H., et al. (2015) Event inference in multidomain families with phylogenetic reconciliation. *BMC Bioinformatics*, 16 (Suppl 14): S8. doi:10.1186/1471-2105-16-S14-S8.

Stoupin, D., Kiss, A.K., Arndt, H., et al. (2012) Cryptic diversity within the choanoflagellate morphospecies complex *Codosiga botrytis* - phylogeny and morphology of ancient and modern isolates. *European Journal of Protistology*. doi:10.1016/j.ejop.2012.01.004.

Stucke, V.M., Timmerman, E., Vandekerckhove, J., et al. (2007) The MAGUK protein MPP7 binds to the polarity protein hDlg1 and facilitates epithelial tight junction formation. *Molecular Biology of the Cell*, 18: 1744–1755.

Su, W.H., Mruk, D.D., Wong, E.W.P., et al. (2013) Polarity protein complex scribble/lgl/dlg and epithelial cell barriers. *Advances in Experimental Medicine and Biology*, 763: 149–170. doi:10.1007/978-1-4614-4711-5_7.

Sugiyama, H., Ito, I. and Watanabe, M. (1989) Glutamate receptor subtypes may be classified into two major categories: A study on *Xenopus* oocytes injected with rat brain mRNA. *Neuron*, 3: 129–132. doi:10.1016/0896-6273(89)90121-9.

Suzuki, T., Zhang, J., Miyazawa, S., et al. (2011) Association of membrane rafts and postsynaptic density: proteomics, biochemical, and ultrastructural analyses. *Journal of Neurochemistry*, 119 (1): 64–77. doi:10.1111/j.1471-4159.2011.07404.x.Association.

Szafranski, P. and Goode, S. (2007) Basolateral junctions are sufficient to suppress epithelial invasion during *Drosophila* oogenesis. *Developmental Dynamics*, 236 (2): 364–373. doi:10.1002/dvdy.21020.

Szumlinski, K. and Woodward, J.J. (2014) “Glutamate signaling in alcohol abuse and dependence.” In Noronha, A.B.C., Cui, C., Harris, R.A., et al. (eds.)

Neurobiology of alcohol dependence. Elsevier. pp. 173–206.

Tago, K., Nakamura, T., Nishita, M., et al. (2000) Inhibition of Wnt signaling by ICAT, a novel beta-catenin-interacting protein. *Genes & development*, 14 (14): 1741–9.

Takahashi, T., Kobayakawa, Y., Muneoka, Y., et al. (2003) Identification of a new member of the GLWamide peptide family: physiological activity and cellular localization in cnidarian polyps. *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, 135 (2): 309–324. doi:10.1016/S1096-4959(03)00088-5.

Taniura, H., Sanada, N., Kuramoto, N., et al. (2006) A metabotropic glutamate receptor family gene in dictyostelium discoideum. *Journal of Biological Chemistry*, 281 (18): 12336–12343. doi:10.1074/jbc.M512723200.

Taylor, A.R. (2009) A fast Na^{+}/Ca^{2+}-based action potential in a marine diatom. *PLoS ONE*, 4 (3): e4966. doi:10.1371/journal.pone.0004966.

Taylor, C.W. and Tovey, S.C. (2010) IP 3 receptors: toward understanding their activation. *Cold Spring Harbor Perspectives in Biology*, 2: a004010. doi:10.1101/cshperspect.a004010.

Thomas, P.D., Kejariwal, A., Guo, N., et al. (2006) Applications for protein sequence-function evolution data: mRNA/protein expression analysis and coding SNP scoring tools. *Nucleic Acids Research*, 34 (Web Server Issue): W645–W650. doi:10.1093/nar/gkl229.

Thomas, U., Kim, E., Kuhlendahl, S., et al. (1997) Synaptic clustering of the cell adhesion molecule Fasciclin II by discs- large and its role in the regulation of presynaptic structure. *Neuron*, 19 (4): 787–799. doi:10.1016/S0896-6273(00)80961-7.

Tochio, H., Mok, Y.K., Zhang, Q., et al. (2000) Formation of nNOS/PSD-95 PDZ dimer requires a preformed β-finger structure from the nNOS PDZ domain. *Journal of Molecular Biology*, 303 (3): 359–370. doi:10.1006/jmbi.2000.4148.

Topinka, J.R. and Bredt, D.S. (1998) N-terminal palmitoylation of PSD-95 regulates association with cell membranes and interaction with K+ channel Kv1.4. *Neuron*, 20: 125–134.

Torruella, G., De Mendoza, A. and Grau-Bové, X. (2015) Phylogenomics Reveals Convergent Evolution of Lifestyles in Close Relatives of Animals and Fungi. *Current Biology*, 25: 2404–2410. doi:10.1016/j.cub.2015.07.053.

Tournière, O., Dolan, D., Richards, G.S., et al. (2020) NvPOU4/Brain3 functions as a terminal selector gene in the nervous system of the cnidarian *Nematostella vectensis*. *bioRxiv*, p. 2020.01.08.898437. doi:10.1101/2020.01.08.898437.

Tsai, N.-P., Wilkerson, J.R., Guo, W., et al. (2012) Multiple autism-linked genes mediate synapse elimination via proteasomal degradation of a synaptic scaffold PSD-95. *Cell*, 151: 1581–1594. doi:10.1016/j.cell.2012.11.040.

Tu, J.C., Xiao, B., Naisbitt, S., et al. (1999) Coupling of mGluR/Homer and

PSD-95 complexes by the Shank family of postsynaptic density proteins. *Neuron*, 23 (3): 583–592. doi:10.1016/S0896-6273(00)80810-7.

Tu, J.C., Xiao, B., Yuan, J.P., et al. (1998) Homer binds a novel proline-rich motif and links group 1 metabotropic glutamate receptors with IP3 receptors. *Neuron*, 21 (4): 717–726. doi:10.1016/S0896-6273(00)80589-9.

Ueda, T., Koya, S. and Maruyama, Y.K. (1999) Dynamic patterns in the locomotion and feeding behaviors by the placozoan *Trichoplax adhaerence*. *BioSystems*, 54 (1–2): 65–70. doi:10.1016/S0303-2647(99)00066-0.

Valtschanoff, J.G. and Weinberg, R.J. (2001) Laminar organization of the NMDA receptor complex within the postsynaptic density. *The Journal of Neuroscience*, 21 (4): 1211–1217.

Vandanapu, R.R., Singh, A.K., Mikhaylova, M., et al. (2009) Structural differences between the SH3-HOOK-GuK domains of SAP90/PSD-95 and SAP97. *Protein Expression and Purification*, 68 (2): 201–207. doi:10.1016/j.pep.2009.07.007.

Varoqueaux, F. and Fasshauer, D. (2017) Getting Nervous: An Evolutionary Overhaul for Communication. *Annual Review of Genetics*, 51 (1): 455–476. doi:10.1146/annurev-genet-120116-024648.

Varoqueaux, F., Williams, E.A., Grandemange, S., et al. (2018) High cell diversity and complex peptidergic signaling underlie placozoan behavior. *Current Biology*, 28 (21): 3495-3501.e2. doi:10.1016/j.cub.2018.08.067.

te Velthuis, A.J.W., Admiraal, J.F. and Bagowski, C.P. (2007) Molecular evolution of the MAGUK family in metazoan genomes. *BMC Evolutionary Biology*, 7: 129. doi:10.1186/1471-2148-7-129.

Verpelli, C., Schmeisser, M.J., Sala, C., et al. (2012) “Scaffold proteins at the postsynaptic density.” In Kreutz, M.R. and Sala, C. (eds.) *Synaptic plasticity*. Wien: Springer. pp. 29–62. doi:10.1007/978-3-7091-0932-8.

Villanueva, C. and Giulivi, C. (2010) Subcellular and cellular locations of nitric oxide synthase isoforms as determinants of health and disease. *Free Radical Biology and Medicine*. 49 (3) pp. 307–316. doi:10.1016/j.freeradbiomed.2010.04.004.

Voglis, G. and Tavernarakis, N. (2006) The role of synaptic ion channels in synaptic plasticity. *EMBO reports*, 7: 1104–1110. doi:10.1038/sj.embor.7400830.

Walter, T.S., Meier, C., Assenberg, R., et al. (2006) Lysine methylation as a routine rescue strategy for protein crystallization. *Structure*, 14 (11): 1617–1622. doi:10.1016/j.str.2006.09.005.

Wang, S.J.H., Tsai, A., Wang, M., et al. (2014) Phospho-regulated *Drosophila* adducin is a determinant of synaptic plasticity in a complex with Dlg and PIP2 at the larval neuromuscular junction. *Biology Open*, 3 (12): 1196–1206. doi:10.1242/bio.20148342.

Watanabe, H., Fujisawa, T. and Holstein, T.W. (2009) Cnidarians and the evolutionary origin of the nervous system. *Development Growth and Differentiation*, 51 (3): 167–183. doi:10.1111/j.1440-169X.2009.01103.x.

Weerth, S.H., Holtzclaw, L.A. and Russell, J.T. (2007) Signaling proteins in raft-like microdomains are essential for Ca^{2+} wave propagation in glial cells. *Cell Calcium*, 41: 155–167. doi:10.1016/j.ceca.2006.06.006.

Weigert, F. (1920) Uber polarisiertes Fluoreszenzlicht. *Verh. d. D. Phys. Ges.*, 1: 100–102.

Wernimont, A. and Edwards, A. (2009) *In situ* proteolysis to generate crystals for structure determination: an update Song, H. (ed.). *PLoS ONE*, 4 (4): e5094. doi:10.1371/journal.pone.0005094.

Westfall, J.A. (1996) Ultrastructure of synapses in the first-evolved nervous systems. *Journal of Neurocytology*, 25: 735–746. doi:10.1007/bf02284838.

Westfall, J.A. and Grimmelikhuijzen, C. (1993) Antho-RFamide immunoreactivity in neuronal synaptic and nonsynaptic vesicles of sea anemones. *The Biological Bulletin*, 185 (1): 109–114. doi:10.2307/1542134.

Wetzel, L.A., Levin, T.C., Hulett, R.E., et al. (2018) Predicted glycosyltransferases promote development and prevent spurious cell clumping in the choanoflagellate *S. rosetta*. *eLife*, 7. doi:10.7554/eLife.41482.

Wickham, H. (2017) *stringr: simple, consistent wrappers for common string operations*. p. https://CRAN. R-project. org/package= stringr.

Won, S., Levy, J.M., Nicoll, R.A., et al. (2017) MAGUKs: multifaceted synaptic organizers. *Current Opinion in Neurobiology*, 43: 94–101. doi:10.1016/j.conb.2017.01.006.

Woods, D.F. and Bryant, P.J. (1991) The discs-large tumor suppressor gene of Drosophila encodes a guanylate kinase homolog localized at septate junctions. *Cell*, 66 (3): 451–464. doi:10.1016/0092-8674(81)90009-X.

Woods, D.F. and Bryant, P.J. (1993) Apical junctions and cell signalling in epithelia. *Journal of Cell Science*, 17: 171–181.

Woods, D.F., Hough, C., Peel, D., et al. (1996) Dig protein is required for junction structure, cell polarity, and proliferation control in *Drosophila* epithelia. *The Journal of Cell Biology*, 134 (6): 1469–1482.

Worley, P.F., Zeng, W., Huang, G., et al. (2007) Homer proteins in Ca^{2+} signaling by excitable and non-excitable cells. *Cell Calcium*, 42 (4–5): 363–371. doi:10.1016/j.ceca.2007.05.007.

Woznica, A., Gerdt, J.P., Hulett, R.E., et al. (2017) Mating in the closest living relatives of animals is induced by a bacterial chondroitinase. *Cell*, 170: 1175–1183. doi:10.1016/j.cell.2017.08.005.

Wu, H., Reissner, C., Kuhlendahl, S., et al. (2000) Intramolecular interactions regulate SAP97 binding to GKAP. *The EMBO journal*, 19 (21): 5740–5751.

doi:10.1093/emboj/19.21.5740.

Xavier, R., Rabizadeh, S., Ishiguro, K., et al. (2004) Discs large (Dlg1) complexes in lymphocyte activation. *The Journal of cell biology*, 166 (2): 173–8. doi:10.1083/jcb.200309044.

Xiao, B., Tu, J.C., Petralia, R.S., et al. (1998) Homer regulates the association of group 1 metabotropic glutamate receptors with multivalent complexes of Homer-related, synaptic proteins. *Neuron*, 21 (4): 707–716. doi:10.1016/S0896-6273(00)80588-7.

Xing, X. and Wu, C.F. (2018) Unraveling synaptic GCaMP signals: dDifferential excitability and clearance mechanisms underlying distinct Ca^{2+} dynamics in tonic and phasic excitatory, and aminergic modulatory motor terminals in *Drosophila*. *eNeuro*, 5 (1): e0362-17. doi:10.1523/ENEURO.0362-17.2018.

Xu, W. (2011) PSD-95-like membrane associated guanylate kinases (PSD-MAGUKs) and synaptic plasticity. *Current Opinion in Neurobiology*, 21: 306–312. doi:10.1016/j.conb.2011.03.001.

Yamada, K.H., Hanada, T. and Chishti, A.H. (2007) The effector domain of human Dlg tumor suppressor acts as a switch that relieves autoinhibition of Kinesin-3 motor GAKIN/KIF13B. *Biochemistry*, 46: 10039–10045. doi:10.1021/bi701169w.

Yang, X., Xie, X., Chen, L., et al. (2010) Structural basis for tandem L27 domain-mediated polymerization. *The FASEB Journal*, 24 (12): 4806–4815. doi:10.1096/fj.10-163857.

Yang, Z. (1997) PAML : a program package for phylogenetic analysis by maximum likelihood. *CABIOS APPLICATIONS NOTE*, 13 (5): 555–556.

Yang, Z. (2007) PAML 4 : phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24 (8): 1586–1591. doi:10.1093/molbev/msm088.

Yu, F., Kuo, C.T. and Jan, Y.N. (2006) Drosophila neuroblast asymmetric cell division: recent advances and implications for stem cell biology. *Neuron*, 51: 13–20.

Yuan, J.P., Kiselyov, K., Shin, D.M., et al. (2003) Homer binds TRPC family channels and is required for gating of TRPC1 by IP3 receptors. *Cell*, 114 (6): 777–789. doi:10.1016/S0092-8674(03)00716-5.

Yum, S., Takahashi, T., Koizumi, O., et al. (1998) A novel neuropeptide, Hym-176, induces contraction of the ectodermal muscle in *Hydra*. *Biochemical and Biophysical Research Communications*, 248 (3): 584–590. doi:10.1006/bbrc.1998.8831.

Zakelj-Mavri, M., Kastelic-Suhadolc, T., Plemenita, A., et al. (1995) Steroid hormone signalling system and fungi. *Comparative Biochemistry and Physiology*, 12 (2): 637–642.

Zakhvatkin, A.A. (1949) *The comparative embryology of the low invertebrates.*

Sources and Method of the origin of metazoan development. Moscow: Soviet Science.

Zakon, H.H. (2012) Adaptive evolution of voltage-gated sodium channels: The first 800 million years. *Proceeding of the National Academy of Sciences*, 109: 10619–10625. doi:10.1073/pnas.1201884109.

Zeng, M., Chen, X., Guan, D., et al. (2018) Reconstituted postsynaptic density as a molecular platform for understanding synapse formation and plasticity. *Cell*, 174: 1172–1187. doi:10.1016/j.cell.2018.06.047.

Zettler, L.A.A., Nerad, T.A., O'Kelly, C.J., et al. (2001) The nucleariid amoebae: more protists at the animal-fungal boundary. *The Journal of Eukaryotic Microbiology*, 48 (3): 293–297. doi:10.1111/j.1550-7408.2001.tb00317.x.

Zhou, W., Chung, I., Liu, Z., et al. (2004) A voltage-gated calcium-selective channel encoded by a sodium channel-like gene. *Neuron*, 42: 101–112.

Zhu, J., Shang, Y. and Zhang, M. (2016) Mechanistic basis of MAGUK-organized complexes in synaptic development and signalling. *Nature Reviews Neuroscience*, 17 (4): 209–223. doi:10.1038/nrn.2016.18.

Zimber, A., Nguyen, Q.D. and Gespach, C. (2004) Nuclear bodies and compartments: Functional roles and cellular signalling in health and disease. *Cellular Signalling*, 16 (10): 1085–1104. doi:10.1016/j.cellsig.2004.03.020.

Zucker, R.S. (1999) Calcium- and activity-dependent synaptic plasticity. *Current Opinion in Neurobiology*. doi:10.1016/S0959-4388(99)80045-2.

frontiers in Synaptic Neuroscience

PMC full text: Front Synaptic Neurosci. 2016; 8: 23.

Published online 2016 Aug 19. doi: 10.3389/fnsyn.2016.00023

<< Prev Figure 1 Next >>

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 04, 2020

This Agreement between University of Exeter -- Tarja Hoffmeyer ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4881770086030
License date	Aug 04, 2020
Licensed Content Publisher	Elsevier
Licensed Content Publication	Molecular Cell
Licensed Content Title	Structure of the SH3-Guanylate Kinase Module from PSD-95 Suggests a Mechanism for Regulated Assembly of MAGUK Scaffolding Proteins
Licensed Content Author	Aaron W McGee,Srikanth R Dakoji,Olav Olsen,David S Bredt,Wendell A Lim,Kenneth E Prehoda
Licensed Content Date	Dec 1, 2001
Licensed Content Volume	8
Licensed Content Issue	6
Licensed Content Pages	11
Start Page	1291
End Page	1301
Type of Use	reuse in a thesis/dissertation

Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	PhD thesis: The evolutionary origin of postsynaptic signalling machineries – Insights from the single-celled sister group to the animals
Institution name	University of Exeter
Expected presentation date	Aug 2020
Portions	Figure 1a (used from PDB deposit, colours changed) Picture of monomer in Figure 7
Requestor Location	University of Exeter Stocker Road Exeter, EX4 4QD United Kingdom Attn: University of Exeter
Publisher Tax ID	GB 494 6272 12
Total	0.00 EUR
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions

ELSEVIER LICENSE
TERMS AND CONDITIONS

Sep 13, 2020

This Agreement between University of Exeter -- Tarja Hoffmeyer ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4885170674240
License date	Aug 10, 2020
Licensed Content Publisher	Elsevier
Licensed Content Publication	Cell
Licensed Content Title	The Postsynaptic Density Proteins Homer and Shank Form a Polymeric Network Structure
Licensed Content Author	Mariko Kato Hayashi,Chunyan Tang,Chiara Verpelli,Radhakrishnan Narayanan,Marissa H. Stearns,Rui-Ming Xu,Huilin Li,Carlo Sala,Yasunori Hayashi
Licensed Content Date	Apr 3, 2009
Licensed Content Volume	137
Licensed Content Issue	1
Licensed Content Pages	13
Start Page	159
End Page	171
Type of Use	reuse in a thesis/dissertation

Portion	figures/tables/illustrations
Number of figures/tables/illustrations	2
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	PhD thesis: The evolutionary origin of postsynaptic signalling machineries – Insights from the single-celled sister group to the animals
Institution name	University of Exeter
Expected presentation date	Aug 2020
Portions	Figure 1A Shank domain architecture Figure 3B Homer domain architecture
Requestor Location	University of Exeter Huettenstrasse 26 Cologne, 50823 Germany Attn: University of Exeter
Publisher Tax ID	GB 494 6272 12
Total	0.00 EUR
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions

Copyright Clearance Center Marketplace™

The Company of Biologists Ltd. - License Terms and Conditions

This is a License Agreement between University of Exeter - Tarja Hoffmeyer ("You") and The Company of Biologists Ltd. ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by The Company of Biologists Ltd., and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	10-Aug-2020	**Type of Use**	Republish in a thesis/dissertation
Order license ID	1054365-1		
ISSN	1477-9145	**Publisher**	COMPANY OF BIOLOGISTS LTD.
		Portion	Image/photo/illustration

LICENSED CONTENT

Publication Title	Journal of experimental biology	**Country**	United Kingdom of Great Britain and Northern Ireland
Author/Editor	Company of Biologists.	**Rightsholder**	The Company of Biologists Ltd.
Date	01/01/1930	**Publication Type**	e-Journal
Language	English	**URL**	http://jeb.biologists.org/

REQUEST DETAILS

Portion Type	Image/photo/illustration	**Distribution**	Worldwide
Number of images / photos / illustrations	2	**Translation**	Original language of publication
Format (select all that apply)	Print, Electronic	**Copies for the disabled?**	No
		Minor editing privileges?	No
Who will republish the content?	Academic institution	**Incidental promotional use?**	No
Duration of Use	Life of current edition	**Currency**	EUR
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

NEW WORK DETAILS

Title	PhD thesis: The evolutionary origin of postsynaptic signalling machineries – Insights from the single-celled sister group to the animals	**Institution name**	University of Exeter
		Expected presentation date	2020-08-31
Instructor name	Tarja T. Hoffmeyer		

ADDITIONAL DETAILS

Order reference number	N/A	**The requesting person / organization to appear on the license**	University of Exeter - Tarja Hoffmeyer

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure 2B, Figure 4D	**Title of the article/chapter the portion is from**	The origin and evolution of synaptic proteins – choanoflagellates lead the way
Editor of portion(s)	The company of Biologists.	**Author of portion(s)**	Pawel Burkhardt
Volume of serial or monograph	N/A	**Issue, if republishing an article from a serial**	N/A
Page or page range of portion	508 + 510	**Publication date of portion**	2015-02-18

PUBLISHER TERMS AND CONDITIONS

The acknowledgement should state "Reproduced / adapted with permission" and give the source journal name. The acknowledgement should either provide full citation details or refer to the relevant citation in the article reference list. The full citation details should include authors, journal, year, volume, issue and page citation. Where appearing online or in other electronic media, a link should be provided to the original article (e.g. via DOI): Development: dev.biologists.org Disease Models & Mechanisms: dmm.biologists.org Journal of Cell Science: jcs.biologists.org The Journal of Experimental Biology: jeb.biologists.org

CCC Republication Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

   3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

   3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically set forth in the Order Confirmation or in a separate written agreement signed by CCC, invoices are due and payable on "net 30" terms. While User may exercise the rights licensed immediately upon issuance of the Order Confirmation, the license is automatically revoked and is null and void, as if it had never been issued, if complete payment for the license is not received on a timely basis either from User directly or through a payment agent, such as a credit card company.

   3.3. Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) is "one-time" (including the editions and product family specified in the license), (ii) is non-exclusive and non-transferable and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Aug 10, 2020

This Agreement between University of Exeter -- Tarja Hoffmeyer ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4885190345432
License date	Aug 10, 2020
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Invertebrate Biology
Licensed Content Title	Evolutionary emergence of synaptic nervous systems: what can we learn from the non-synaptic, nerveless Porifera?
Licensed Content Author	Michael Nickel
Licensed Content Date	Mar 11, 2010
Licensed Content Volume	129
Licensed Content Issue	1
Licensed Content Pages	16
Type of use	Dissertation/Thesis

Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No
Title	PhD thesis: The evolutionary origin of postsynaptic signalling machineries – Insights from the single-celled sister group to the animals
Institution name	University of Exeter
Expected presentation date	Aug 2020
Portions	Figure 4 (excluding part 3)
Requestor Location	University of Exeter Stocker Road Exeter EX4 4PY Exeter, EX4 4PY United Kingdom Attn: University of Exeter
Publisher Tax ID	EU826007151
Total	0.00 EUR

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work

Copyright Clearance Center Marketplace™

University of Chicago Press - Journals - License Terms and Conditions

This is a License Agreement between University of Exeter - Tarja Hoffmeyer ("You") and University of Chicago Press - Journals ("Publisher") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by University of Chicago Press - Journals, and the CCC terms and conditions.

All payments must be made in full to CCC.

Order Date	10-Aug-2020	Type of Use	Republish in a thesis/dissertation
Order license ID	1054374-1		
ISSN	0033-5770	Publisher	UNIVERSITY OF CHICAGO PRESS
		Portion	Chart/graph/table/figure

LICENSED CONTENT

Publication Title	The quarterly review of biology	Publication Type	Journal
Article Title	Neuroid conduction and the evolution of conducting tissues.	Start Page	319
		End Page	332
		Issue	4
Author/Editor	AMERICAN INSTITUTE OF BIOLOGICAL SCIENCES., AMERICAN SOCIETY OF NATURALISTS., STATE UNIVERSITY OF NEW YORK AT STONY BROOK.	Volume	45
		URL	http://www.press.uchicago.edu/ucp/journals/journal/qrb.html
Date	01/01/1926		
Language	English		
Country	United States of America		
Rightsholder	University of Chicago Press - Journals		

REQUEST DETAILS

Portion Type	Chart/graph/table/figure	Distribution	Worldwide
Number of charts / graphs / tables / figures requested	1	Translation	Original language of publication
		Copies for the disabled?	No
Format (select all that apply)	Print, Electronic	Minor editing privileges?	Yes
Who will republish the content?	Academic institution	Incidental promotional use?	No
Duration of Use	Life of current edition	Currency	GBP
Lifetime Unit Quantity	Up to 499		
Rights Requested	Main product		

NEW WORK DETAILS

Title	PhD thesis: The evolutionary origin of postsynaptic signalling machineries - Insights from the single-celled sister group to the animals	**Institution name**	University of Exeter
		Expected presentation date	2020-08-31
Instructor name	Pawel Burkhardt		

ADDITIONAL DETAILS

Order reference number	N/A	**The requesting person / organization to appear on the license**	University of Exeter - Tarja Hoffmeyer

REUSE CONTENT DETAILS

Title, description or numeric reference of the portion(s)	Figure 1	**Title of the article/chapter the portion is from**	Neuroid conduction and the evolution of conducting tissues.
Editor of portion(s)	The University of Chicago Press Journals: The quarterly review of biology	**Author of portion(s)**	Mackie, G O
		Issue, if republishing an article from a serial	4
Volume of serial or monograph	45	**Publication date of portion**	1970-12-01
Page or page range of portion	319-332		

CCC Republication Terms and Conditions

1. Description of Service; Defined Terms. This Republication License enables the User to obtain licenses for republication of one or more copyrighted works as described in detail on the relevant Order Confirmation (the "Work(s)"). Copyright Clearance Center, Inc. ("CCC") grants licenses through the Service on behalf of the rightsholder identified on the Order Confirmation (the "Rightsholder"). "Republication", as used herein, generally means the inclusion of a Work, in whole or in part, in a new work or works, also as described on the Order Confirmation. "User", as used herein, means the person or entity making such republication.

2. The terms set forth in the relevant Order Confirmation, and any terms set by the Rightsholder with respect to a particular Work, govern the terms of use of Works in connection with the Service. By using the Service, the person transacting for a republication license on behalf of the User represents and warrants that he/she/it (a) has been duly authorized by the User to accept, and hereby does accept, all such terms and conditions on behalf of User, and (b) shall inform User of all such terms and conditions. In the event such person is a "freelancer" or other third party independent of User and CCC, such party shall be deemed jointly a "User" for purposes of these terms and conditions. In any event, User shall be deemed to have accepted and agreed to all such terms and conditions if User republishes the Work in any fashion.

3. Scope of License; Limitations and Obligations.

    3.1. All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The license created by the exchange of an Order Confirmation (and/or any invoice) and payment by User of the full amount set forth on that document includes only those rights expressly set forth in the Order Confirmation and in these terms and conditions, and conveys no other rights in the Work(s) to User. All rights not expressly granted are hereby reserved.

    3.2. General Payment Terms: You may pay by credit card or through an account with us payable at the end of the month. If you and we agree that you may establish a standing account with CCC, then the following terms apply: Remit Payment to: Copyright Clearance Center, 29118 Network Place, Chicago, IL 60673-1291. Payments Due: Invoices are payable upon their delivery to you (or upon our notice to you that they are available to you for downloading). After 30 days, outstanding amounts will be subject to a service charge of 1-1/2% per month or, if less, the maximum rate allowed by applicable law. Unless otherwise specifically

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Aug 10, 2020

This Agreement between University of Exeter -- Tarja Hoffmeyer ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4885200443134
License date	Aug 10, 2020
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Genetics
Licensed Content Title	The evolution of cell types in animals: emerging principles from molecular studies
Licensed Content Author	Detlev Arendt
Licensed Content Date	Dec 31, 1969
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no

Circulation/distribution	30 - 99
Author of this Springer Nature content	no
Title	PhD thesis: The evolutionary origin of postsynaptic signalling machineries – Insights from the single-celled sister group to the animals
Institution name	University of Exeter
Expected presentation date	Aug 2020
Portions	Figure 5a
Requestor Location	University of Exeter Stocker Road Exeter, EX4 4PY United Kingdom Attn: University of Exeter
Total	0.00 GBP

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 11, 2020

This Agreement between University of Exeter -- Tarja Hoffmeyer ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4885910679033
License date	Aug 11, 2020
Licensed Content Publisher	Elsevier
Licensed Content Publication	Neuron
Licensed Content Title	Structure of the Homer EVH1 Domain-Peptide Complex Reveals a New Twist in Polyproline Recognition
Licensed Content Author	Jutta Beneken,Jian Cheng Tu,Bo Xiao,Mutsuo Nuriya,Joseph P. Yuan,Paul F. Worley,Daniel J. Leahy
Licensed Content Date	Apr 1, 2000
Licensed Content Volume	26
Licensed Content Issue	1
Licensed Content Pages	12
Start Page	143
End Page	154
Type of Use	reuse in a thesis/dissertation

Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Title	PhD thesis: The evolutionary origin of postsynaptic signalling machineries – Insights from the single-celled sister group to the animals
Institution name	University of Exeter
Expected presentation date	Aug 2020
Portions	Homer EVH1 structure from PDB entry
Requestor Location	University of Exeter Stocker Road Exeter, EX4 4PY United Kingdom Attn: University of Exeter
Publisher Tax ID	GB 494 6272 12
Total	0.00 GBP

Terms and Conditions

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Aug 11, 2020

This Agreement between University of Exeter -- Tarja Hoffmeyer ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4885920170286
License date	Aug 11, 2020
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	THE FASEB JOURNAL
Licensed Content Title	Structural basis for tandem L27 domain-mediated polymerization
Licensed Content Author	Xue Yang, Xingqiao Xie, Liu Chen, et al
Licensed Content Date	Aug 11, 2010
Licensed Content Volume	24
Licensed Content Issue	12
Licensed Content Pages	10
Type of use	Dissertation/Thesis

Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Will you be translating?	No
Title	PhD thesis: The evolutionary origin of postsynaptic signalling machineries – Insights from the single-celled sister group to the animals
Institution name	University of Exeter
Expected presentation date	Aug 2020
Portions	Images of protein structure (from PDB)
Requestor Location	University of Exeter Huettenstrasse 26 Cologne, 50823 Germany Attn: University of Exeter
Publisher Tax ID	EU826007151
Total	0.00 GBP

Terms and Conditions

TERMS AND CONDITIONS

This copyrighted material is owned by or exclusively licensed to John Wiley & Sons, Inc. or one of its group companies (each a"Wiley Company") or handled on behalf of a society with which a Wiley Company has exclusive publishing rights in relation to a particular work

Structures of the L27 Domain of Disc Large Homologue 1 Protein Illustrate a Self-Assembly Module

Author: Agnidipta Ghosh, Udupi A. Ramagopal, Jeffrey B. Bonanno, et al

Publication: Biochemistry

Publisher: American Chemical Society

Date: Feb 1, 2018

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

If credit is given to another source for the material you requested, permission must be obtained from that source.

BACK CLOSE WINDOW