universität wien

DIPLOMARBEIT

Towards a molecular and functional characterisation of sensory-neurosecretory cells in *Platynereis dumerilii* and fish

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer.nat.)

Verfasser:	Stephan Kirchmaier
Matrikel-Nummer:	0301041
Studienrichtung:	Molekulare Biologie
Betreuerin:	Dr. Kristin Tessmar-Raible

Wien, am 14.12.2009

Acknowledgement

I want to thank everyone, who has contributed to that work. Especially, I want to thank Kristin Tessmar-Raible for her great support and many fruitful discussions during my thesis. Additionally, I want to thank her and Florian Raible for providing such an encouraging working environment. Furthermore, I want to thank all members of the Tessmar and Raible groups at the MFPL, especially Thomas Hoffmann and Ruth Fischer for helping me with my projects.

Additionally, I want to thank Jochen Wittbrodt and his group, especially Annette Schmidt, Petra Haas, Lazaro Centanin and Katherine Brown for helpful discussions, practical help and a nice time at the EMBL in Heidelberg which was funded by the University of Vienna with the "Förderungsstipendium". I am grateful to several facilities at the Vienna Biocenter, especially the MFPL animal care and the microscope facilities at the MFPL and the IMP.

Finally, I want to thank Isadora for supporting me in every aspect of my life and for being perfectly wonderful and inspiring.

Contents

1 Introduction

1.1 The Nonapeptide neurohormonal system

Neuronal hormones are used in all bilaterian animals to govern and regulate diverse bodily functions and processes. These hormones are produced in the brain and released into animal bodily fluids. Here, we focus on a highly conserved neurohormonal system, the Nonapeptide neurohormonal system.

The Nonapeptide system consists of the neuronal hormone, e.g. Vasopressin and Oxytocin in mammals, and the hormone receptors. As the name of the system suggests, the peptides themselves mostly consist of nine amino acids. It is encoded within a larger preprohormone, which is proteolytically cleaved to release the proper hormone. While the Nonapeptide family is conserved in lophotrochozoans and Deuterostomia, it is only present in basal ecdysozoans, like *Tribolium castaneum*. In higher ecdysozoans, like *Drosophila* species, it has been lost [108]. However, the receptors for the Nonapeptides are present in the genomes of all bilaterian species [54, 108]. This suggests that the receptors follow a different evolutionary course than their ligands.

In contrast to neurotransmitters used in synaptic signal transmission, the hormones are not synthesized in the axons but inside the cell soma. The hormones are transported in so-called large dense core vesicles (LCDVs) to the release sites. The release is not restricted to synapses but the Nonapeptide hormones are released from all sites along the cell membrane. Based on the long-lasting process of replenishing a secreted hormone pool, these are not suitable for fast responses to outer cues but rather represent a way to mediate longer-lasting processes [108].

In vertebrates, the Nonapeptide hormones are released into the blood stream from specific cells within the hypothalamus where they can act on the development and function of peripheral organs [109]. In addition, these hormones are also secreted into the CSF, the cerebro-spinal fluid. This so-called *liquor cerebrospinalis* represents the liquid inside and around the brain. In mammals, the Nonapeptides in the CSF are thought to act as neuromodulators to regulate diverse behaviours [25, 9].

1.2 Evolution of the different Nonapeptides

The peptide hormones of the Nonapeptide neuronal hormonal system have different names in different phyla [37, 87]. For example, the mammalian homologues are called Arginine-Vasopressin and Oxytocin while the homologues in teleost fishes are called Vasotocin and Isotocin. Therefore, we refer to the family here as to the Nonapeptide superfamily, since most homologues share a conserved hormone core of nine amino acids with only few exceptions. Recently, longer Nonapeptide homologues have been identified in two urochordate species [66, 112] (see also Figure 1.1). The mammalian peptides were first identified in 1952 [26] and since then homologous peptides in diverse bilaterian species were found. Due to a lack of proper genome information, the Nonapeptides themselves were mostly purified and sequenced on a peptide level [37]. This lead to a difficult, incoherent nomenclature since all homologues with only slightly different amino acid sequences were named differently. In figure 1.1, the nomenclature and sequence of the diverse members of the Nonapeptide superfamily is shown.

Nonapeptide hormone family members															
Name	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**Species**
Oxytocin	C	Y	I	Q	N	C	P	L	G						Mammalia
Mesotocin	C	Y	I	Q	N	C	P	I	G						Non-mammalian tetrapods, Marsupials
Isotocin	C	Y	I	S	N	C	P	I	G						Teleosts
Glumitocin	C	Y	I	S	N	C	P	G	G						Chondrichthyes
Valitocin	C	Y	I	Q	N	C	P	V	G						Chondrichthyes
Aspargtocin	C	Y	I	N	N	C	P	L	G						Chondrichthyes
Asvatocin	C	Y	I	N	N	C	P	V	G						Chondrichthyes
Phasvatocin	C	Y	F	N	N	C	P	V	G						Chondrichthyes
Phasitocin	C	Y	F	N	N	C	P	I	G						Chondrichthyes
Vasopressin	C	Y	F	Q	N	C	P	R	G						Mammalia
Lysipressin	C	Y	F	Q	N	C	P	K	G						Pigs, Marsupials. Warthogs
Phenypressin	C	F	F	Q	N	C	P	R	G						Marsupials
Vasotocin	C	Y	I	Q	N	C	P	R	G						Non-mammalian vertebrates
Ci-Vasopressin	C	F	F	R	D	C	S	N	M	D	W	Y	R		*Ciona intestinalis* (Urochordate)
Oxytocin-like peptide	X	Y	I	S	D	C	P	N	S	R	F	W	S	T	*Styela plicata* (Tunicate)
Annetocin	C	F	V	R	N	C	P	T	G						Annelids
Vasotocin	C	F	V	R	N	C	P	G	G						Annelids
Arg-Conopressin	C	I	I	R	N	C	P	R	G						Molluscs
Lys-Conopressin	C	F	I	R	N	C	P	K	G						Molluscs
Cephalotocin	C	Y	F	R	N	C	P	I	G						Molluscs
Octopressin	C	F	W	T	S	C	P	I	G						Molluscs

Figure 1.1: **The Nonapeptide superfamily** Diverse known orthologues of the Nonapeptide superfamily are shown. The highly conserved cystein residues are labelled red and positively charged amino acids are labelled yellow. Note that most invertebrates contain a positively charged amino acid, although mostly not at the position eight as in vertebrates (except for molluscs where the fourth position is also positively charged). The data for this table is taken from [37, 16, 66, 112].

Basal vertebrates, like lampreys and hagfish, as well as most invertebrates contain only one ortholog [63, 113, 109, 47], whereas all other vertebrates contain at least two homologues. It has been suggested that in vertebrates the eighth amino acid of vertebrate Nonapeptide homologues

defines if it is "vasopressin-like" or "oxytocin-like". The eighth amino acid of Vasopressin is positively charged, whereas the corresponding amino acid of Oxytocin is neutral (see Figure 1.1). This lead to the belief that all vertebrate Nonapeptides with a positively charged amino acid were related to Vasopressin and all hormones with a neutral amino acid were related to Oxytocin [1, 37]. One paralogue always has a neutral and the other one a positively charged amino acid at position eight. Thus, it is a long-standing belief that only one duplication occurred in the vertebrate lineage [74, 1, 36, 46, 47]. However, this means that the phylogeny of the Nonapeptide superfamily in vertebrates is inferred from a single amino acid.

We tested how the different Nonapeptides relate to each other by using whole preprohormone sequences for a phylogenetic analysis as well as by taking genomic arrangements, domain structure and regulatory similarities into account. In addition to that, we also examine other evolutionary aspects of the Nonapeptide system as well as the GnRH neurohormonal system by inferring evolutionary history from functional properties and vice versa.

1.3 Functions of the Nonapeptide system

Within the mammalian brain, Vasopressin and Oxytocin act as autocrin and paracrin neuromodulators that are thought to exert long-lasting behaviour changes (for reviews see [77, 78, 73] and also [37, 40, 7, 16]). For example, the expression pattern of the vasopressin receptor V1a within the brain of different mammals seems to be unique for each species. These differences are thought to be responsible for diverse species-specific behaviours, especially related to social bonding [40]. This was shown in a study where the expression pattern of the V1a receptor was artificially altered and changes in social bonding were detected [128, 25]. A connection between the development of reproductive organs and Vasopressin-like hormones was found in several studies [109, 39, 2, 33, 67]. Furthermore, the Nonapeptides are often implicated to affect several aspects of reproduction [7, 25, 78, 33] in diverse species. Additionally, secretion of the Nonapeptides is strongly responsive to changes in body fluid osmolarity. Thus, the Nonapeptides in many bilaterian species are involved in salt homeostasis, osmoregulation and therefore excretion [7]. For example, in mammalian species the Vasopressin receptor V2 in certain regions of the loop of Henle. There, Vasopressin helps in generating an osmotic gradient [6, 14]. It also plays important roles in salt homeostasis. For example, in rats it has been shown that Vasopressin can act antidiuretic as well as natriuretic. Similar effects are also observed in non-mammalian vertebrates. In the flounder and *Triakis scyllium* a direct connection between plasma osmolarity and Vasotocin content has been found [121, 61]. In summary, there is evidence that specific Nonapeptides play important roles in water conservation and excretion [7].

1.4 Linking the Nonapeptide system to light sensation

Animals living in fresh or sea water often reproduce via external fertilisation. This means that the gonadal products are expelled from the body into the open water. In order to ensure that the eggs are properly fertilised, the process of reproduction has to be synchronised. For that, the animals rely on external cues such as exogenous light cycles. These are either circadian, lunar or seasonal light cycles. Often circadian cycles are superimposed on either seasonal or lunar cycles.

The fish *Takifugu niphobles* has been shown to have a semi-lunar reproductive cycle and the amount of secreted Nonapeptides also cycles in correlation with reproductive capacity [85]. Furthermore, *Fundulus heteroclitus* has been shown to have a lunar reproduction cycle and the secretion of a Nonapeptide is thought to control this behaviour [103]. A seasonal reproductive cycle has also been shown for other fish species, for example *Gasterosteus aculeatus* [43], the catfish *Heteropneustes fossilis* [100] and Medaka [68].

In conclusion, a correlation of the amount of secreted Nonapeptide with light-dependent reproductive cycles has been found for several species. However, so far, it is unclear how the light-cycle dependent secretion of the Nonapeptide hormones in diverse species is regulated.

1.4.1 A conserved minimal module provides a possible explanation

In 2007, a study showed that a subset of vasotocinergic neurons, i.e. neurons that secrete the neuronal hormone Vasotocin, in the annelid *Platynereis dumerilii* and the vertebrate Zebrafish also express a presumptive light-sensitive ciliary Opsin molecule. In addition to that, in both species the cells are defined by the expression of a similar subset of transcription factors, namely Rx, nk2.1, otp and miR-7. This "molecular fingerprint" is conserved in annelids and vertebrates. Based on its complexity and the high degree of conservation on the level of specifying factors, the cell type can be regarded as ancient because it must have already been present in the common ancestor of *Platynereis dumerilii* and Zebrafish, namely Urbilateria [109].

The co-expression of the neuronal hormone Vasotocin with a presumptive light-sensitive Opsin molecule in a conserved cell type prompted the hypothesis that this cell type coordinates light-dependent Vasotocin secretion. Therefore, the response to light perception would be integrated within one single cell (see figure 1.2). In conclusion, these sensory-neurosecretory cells provide a possible explanation to the regulation of light cycle-dependent Vasotocin secretion identified in diverse species [109].

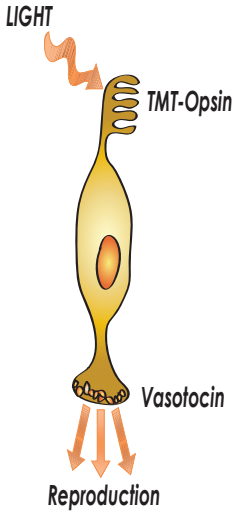


Figure 1.2: **A conserved sensory neurosecretory cell type** The current hypothesis is that a light input enters the cell and is taken up via a specific ciliary Opsin molecule (TMT Opsin in Zebrafish). The same cell then integrates this signal and responds to it via the secretion of the neuronal hormone Vasotocin. The hormone influences several aspects of reproduction. This cell type represents an ancient minimal module that must have already been present in Urbilateria [109] since it is conserved in diverse bilaterian species.

1.5 Opsins in vertebrates

1.5.1 Origin of Opsin molecules

Opsins belong to the superfamily of G-protein-coupled receptors (GPCRs). This is one of the largest protein families in mammals with more than 800 members identified alone in the human genome. An extensive phylogenetic analysis shows that within the GPCRs five families exist, namely the Adhesion-, Secretin-, Glutamate-, Frizzled/Tas2- and the Rhodopsin families [31]. The potential ligands are extremely variable and range from ions, odorants, lipids and peptides (e.g. Vasopressin and Oxytocin) to photons [31]. Also, the signal transduction cascade downstream of the receptor varies considerably. In my thesis, I only focus on members of the Rhodopsin family, namely the presumptive light-sensitive Opsin family (ALPHA group) as well as the receptors for the Nonapeptide hormone system (BETA group) [31].

1.5.2 Diverse Opsin families

The presumptive light-sensitive Opsin family is further subdivided into three different subtypes, namely the ciliary Opsins, the rhabdomeric Opsins and the photoisomerases (Figure 1.3). In addition, other Opsins have been identified that do not fit in these three subtypes, namely neuropsins [99]. I will not explicitly consider the photoisomerases nor neuropsins here.

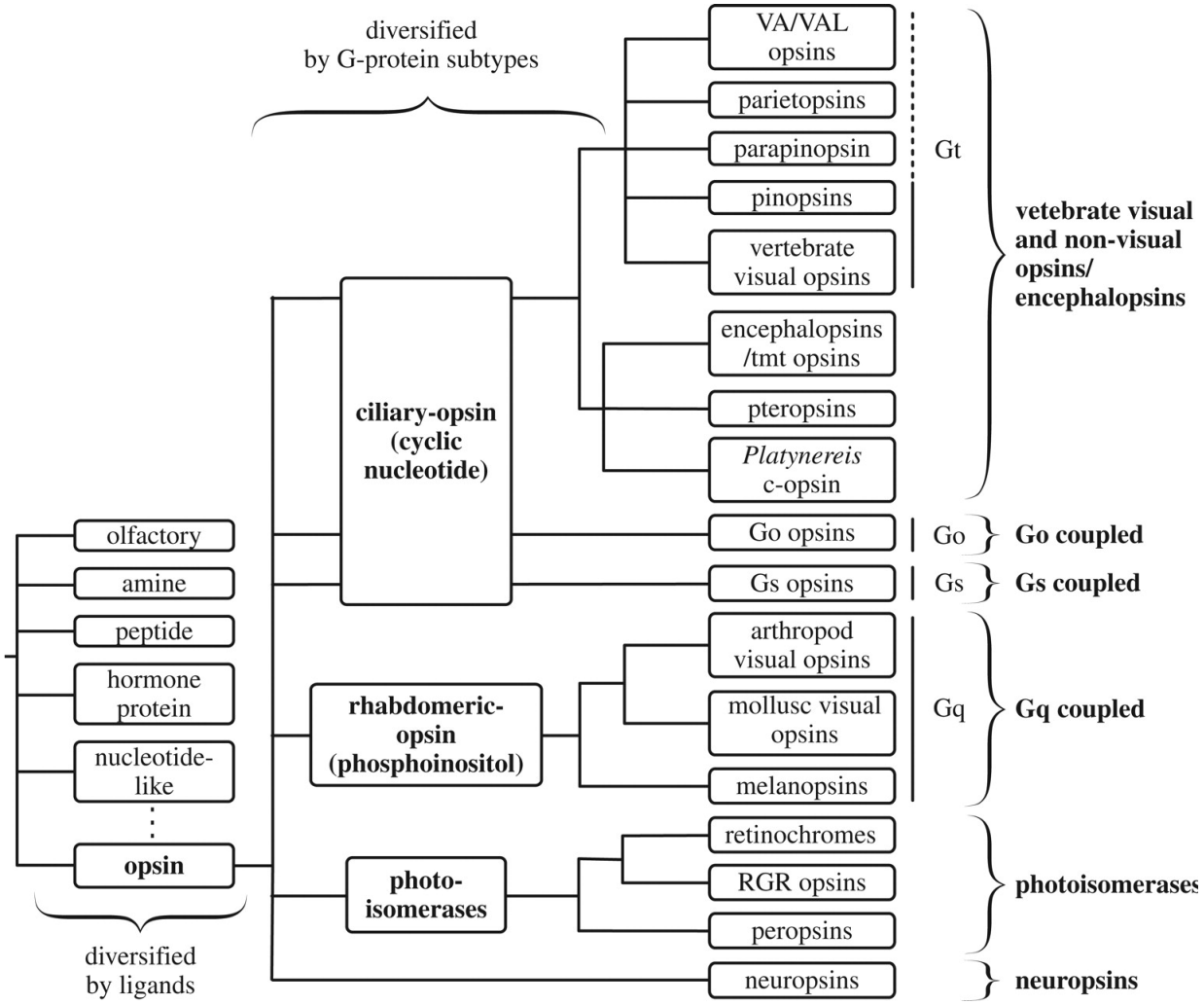


Figure 1.3: **Overview picture of Opsin phylogeny** Here, the phylogeny of Opsins is shown. At a first level, the G protein-coupled receptors can be distinguished by their ligands. Opsins usually bind a retinal chromophore that undergoes structural changes upon light absorption. Thereby, the structure of the receptor changes and activates or inactivates it. However, light absorbance has not been proven for all of them. The Opsins are further divided into three major families: ciliary- and rhabdomeric Opsins as well as photoisomerases. The ciliary and rhabdomeric Opsins are expressed in different cell types that are specified by different factors and they also employ different downstream signalling cascades [3]. In this figure, TMT Opsins, Encephalopsins and invertebrate ciliary Opsins cluster together in one group and are called "Encephalopsins". However, as I will argue below, these rather represent three different families with specific structural features. The figure is taken from [99] with modifications.

The names "ciliary" and "rhabdomeric" Opsins derive from the cell type morphology that express the respective Opsin molecule. Many photoreceptor cells enlarge their surface by folding the cell membrane. The rhabdomeric photoreceptor cells fold the apical cell surface while the ciliary photoreceptor cells fold the ciliary membrane [3]. In 2004, a ciliary photoreceptor molecule (C-Opsin) has been identified in the protostomian species *Platynereis dumerilii*. The ciliary photoreceptor cells also employ similar specifying gene regulatory networks as the cells in vertebrates. This is based on the finding that the ciliary photoreceptor cells in vertebrates and *Platynereis dumerilii* express for example a homologous Rx transcription factor. This changed the existing view that there are "vertebrate-type" Opsins (i.e. C-Opsins) and "invertebrate-type" Opsins (i.e. R-Opsins) [5, 109]. However, so far only few protostomian ciliary Opsins have been identified and the connection to a specific ciliary cell type morphology is not clear for all of them (for example see [115]).

It is believed, that Urbilateria already had specific rhabdomeric and ciliary Opsins as well as the corresponding rhabdomeric and ciliary photoreceptor cells. In invertebrates, the rhabdomeric cells formed simple eyes together with pigment cells to mediate directional photoresponses. The ciliary photoreceptor cells resided within the brain to mediate non-directional photoresponses, like circadian photoentrainment. However, in the vertebrate lineage, both photoreceptor cell types are present in the retina with the ciliary photoreceptors being the major visual photoreceptors. Thus, two independent lineages of light-sensitive cells reside within the vertebrate retina [5] and the vertebrate eye therefore represents a composite structure. In the course of this thesis, I will focus on specific ciliary Opsins, namely TMT Opsins and Encephalopsins that are present in the inner vertebrate brain.

1.5.3 TMT Opsins and Encephalopsins

Encephalopsin has been identified in 1999 as an Opsin expressed in a wide variety of regions within the mouse brain and in the testes [10]. In another study, mouse Encephalopsin was also detected in the developing and adult eyes [102]. Later on, the expression of *encephalopsin* in humans was assessed. A northern blot confirmed the expression in a wide variety of tissues, like the retina, lung, liver and kidney [48]. Thus, there are species-specific differences between the expression pattern of *encephalopsin* in humans and mice [10]. Additionally, differences in the structural organisation between the human and mouse *encephalopsin* orthologue have been found. While the murine *encephalopsin* gene seems to encode for only one isoform, it was shown that at least six different splice isoforms of human *Encephalopsin* exist. However, only two of these seem to encode for proteins since the other four splice variants contain a premature stop codons [65]. Therefore, it is likely that these transcripts are degraded by nonsense-mediated mRNA decay [101, 80]. Furthermore, the *encephalopsin* gene is intertwined with the adjacent KMO gene (Kynurenine 3-monooxygenase, NP_003670 (*H. sapiens*), NP_598570 (*M. musculus*)) in humans

but not in mice. Thus, there are differences in the *encephalopsin* gene structure and expression between humans and mice which so far have not been implicated to any function [65].

In 2003, a protein similar to Encephalopsin has been discovered in Zebrafish and Fugu, namely TMT Opsin (Teleost Multiple Tissue Opsin) [86]. The authors cloned one TMT Opsin orthologue in each species and tested the expression using semi-quantitative RT-PCR as well as a cross-species RNAse Protection Assay. They could identify expression in the heart, brain, liver, kidney, eye and one zebrafish cell line (PAC2 fibroblast cell line). However, they did not test the expression in developing embryos. Furthermore, the authors did not find any particular function for the identified Opsin molecules. An involvement in the circadian clock based on the broad expression pattern was suggested [86]. However, the broad expression pattern the authors identified is the only support for that. The authors did not use *in situ* hybridisation techniques to identify the cells and tissues that express the TMT Opsin. In 2007, it was found that the identified Zebrafish TMT Opsin homologue is not expressed in a wide variety of tissues in the developing embryo but has very distinct expression domains within the brain, especially in the hypothalamus [109] where it is in part co-expressed with the neuronal hormone Vasotocin.

1.6 Model organisms

1.6.1 *Platynereis dumerilii*

Platynereis dumerilii is an annelid species that lives along the coasts in the Mediterranean, but also a wide geographic distribution in warm seas has been reported. It is a marine model organisms which can be cultured in the lab in plastic boxes in a mixture of natural and artificial sea water. The animal has been cultured in the lab since the early 1950ies [30].

Platynereis dumerilii has a generation time of approximately three month. After a longer growth phase in self-spun tubes, the animals start to mature and undergo metamorphosis. The animal changes from a pelagic lifestyle to a benthic one [30]. This means that it begins to swim in the search of a mate. Having found a mate, the worms engage in a so-called nuptial dance. The animals swim very fast in narrow circles around each other until the eggs and sperm are released into the sea water. The eggs are fertilised externally. One couple produces up to about 3000 eggs which develop synchronously over a longer time period [30]. Thus, the animals invest a lot of resources into the generation of the gametes, which they expel all at once. Mature females mostly consist of eggs and after these are released the worm shrinks dramatically and dies shortly after their first mating. Mature males can mate more often, yet they also die at most one day after their first mating.

Why choosing *Platynereis dumerilii* as a model organism?

There are several reasons why *Platynereis dumerilii* is an interesting model organism. First, it is a lophotrochozoan species. Lophotrochozoans are a major protostomian group. They represent a sister group to ecdysozoans. Protostomia belong to the Bilateria along with the Deuterostomia (see Figure 1.4). All Bilateria with spiralian cleavage belong to the Lophotrochozoa. However, the absence of spiralian cleavage does not exclude species from the lophotrochozoans. Within the lophotrochozoans, two subgroups emerge, namely Trochozoa and Platyzoa. Trochozoa are characterised by the existence of a trochophore larva. For Platyzoa, the relationship is less clear and still under debate. Thus, I will not further adress these. Ecdysozoans comprise other protostomes, for example arthropods and nematodes. The other bilaterian branch, the deuterostomes, consist of the Chordata (Vertebrata, Cephalochordata and Urochordata), Echinodermata, Hemichordata and Xenoturbellida [38]. In today's molecular research, the major model organsims are either deuterostomes, like mice, or ecdysozoans, like *Drosophila melanogaster* and *Caenorhabditis elegans*. Although lophotrochozoans represent the biggest group [38] , they are widely ignored in today's molecular biological research except for research done on few molluscan species. Thus, *Platynereis dumerilii* represents a model organism important for comparative research since it belongs to a broad and widely ignored bilaterian group.

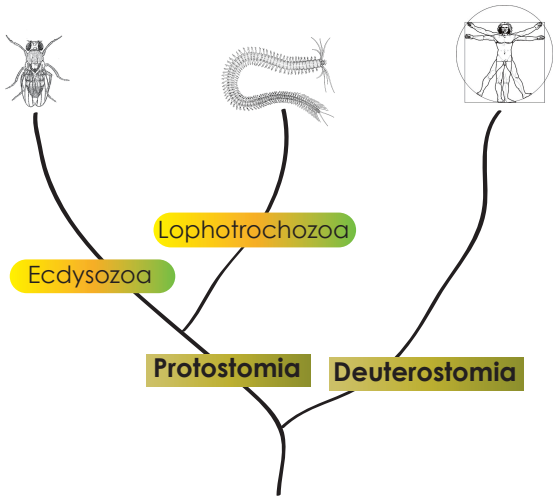


Figure 1.4: **Bilaterian phylogeny** Within bilaterians, two major branches exist, one of the Deuterostomes (e.g. *Homo sapiens*)) and one of the Protostomes. The protostomes are further subdivided into Lophotrochozoans (e.g. *Platynereis dumerilii*) and Ecdysozoans (e.g. *Drosophila melanogaster*).

It has been shown that striking similarities between the annelid *Platynereis dumerilii* and vertebrates exist on the genomic level. A global phylogeny of protein sequences that are present in all bilaterian species was used to show that on the exon-level, *Platynereis dumerilii* rather re-

sembles vertebrates than Ecdysozoans do. In addition, the gene structure of *Platynereis dumerilii* is complex as is the gene structure in humans. Thus, the last common ancestor of humans and *Platynereis dumerilii* likely shared the complex gene structures. However, other vertebrates as well as Ecdysozoans have simple gene structures. This suggests that in the course of evolution the gene structures were simplified in these animal [93]. Thus, also in this respect, *Platynereis dumerilii* is an important species for comparative research.

Another important factor is that *Platynereis dumerilii* is a marine species. Although there are huge numbers of marine species, they are most often not applicable for molecular biological research. This is in part due to the fact that most marine animals cannot be cultured in the lab, let alone in defined medium. *Platynereis dumerilii* can not only be kept in the laboratory, but it also does not need much space per individual. Also, the generation time of approximately three months makes a lab culture possible. Furthermore, many other marine organisms have a much longer generation time. One couple produces vast amount of offspring that develop synchronously and can therefore also be used for many developmental studies. In addition, several molecular biological techniques have been established, like whole-mount *in situ* hybridisation.

Since the fertilisation is external, the maturation of males and females has to be synchronised. As many other marine organisms, *Platynereis dumerilii* synchronises the reproduction using external light cycles. They employ a lunar cycle for their maturation. This means that during full moon, the animals usually do not mature. The number of maturing animals peaks around new moon. The exact time of the peak or the number of peaks is dependent on the genomic background, however the number of maturing animals declines around full moon in all lines. This rhythmic behaviour was shown to be governed by an inner lunar clock [51, 52]. This means that the animal is entrained by the full moon and it shows this behaviour also in free-running conditions (i.e. the animals encounter no moon anymore after entrainment with several moon phases). In the lab culture, the full moon can be simulated using a white lamp that is on seven consecutive nights of every fourth week [30].

1.6.2 Medaka and Zebrafish

The teleost fishes Medaka and Zebrafish represent two commonly used vertebrate model organisms in molecular biology. The fishes can be easily and, compared to mice, cheaply held in the laboratory. The genome sequences of both species are freely available online [57], although the Zebrafish genome sequencing project still awaits completion. Both species are amenable for diverse embryological and molecular procedures like cell transplantation, micro-injection and *in situ* hybridisation.

After a teleost-specific genome duplication, the lineages within the teleostei evolved independently and so did the duplicated genes of the different species. In general, after the a duplication event, a gene could either be affected by gene loss, sub- or neofunctionalisation [88, 34]. Since

the species evolved independently, the duplicated genes evolved independently. Functions that are carried out by one gene in other vertebrates might be split up between these duplicates in a species-specific way in teleosts. Thus, the analysis of homologous genes from the two teleost fish species Medaka and Zebrafish might provide interesting insights into the ancestral function of this gene [124, 34]. Also, the last common ancestor of Medaka and Zebrafish lived around 115-200 million years ago (Myr). The phylogenetic position of these species makes them well suited for comparative analyses of gene regulation and gene expression [34].

Medaka originates from Japan, Korea and eastern China. These regions have seasons and consistently Medaka shows seasonality in certain physiological processes, for example concerning reproductive capacity [68] but also locomotor activity [126]. On the other hand, Zebrafish is a subtropical fresh-water fish that never experiences seasons and no seasonality has been reported in Zebrafish so far. Based on the idea that deep brain photoreceptors are important to mediate light-dependent non-visual tasks like photoperiodic time measurement [91, 106], it would be very interesting to know if this is also the case for the TMT Opsins in Medaka and, if yes, what function they have in Zebrafish. Thus, trying to answer the question why this cell type is conserved in Zebrafish and Medaka possibly provides means to assay the presence and absence of seasonality in both species.

1.7 Goal of my thesis

The primary goal of my thesis is to further investigate the conserved ancient cell type that is characterised by the co-expression of a presumptive light-sensitive Opsin molecule and the neuronal hormone Vasotocin. First, I want to further characterise the Nonapeptide hormonal system in *Platynereis dumerilii*. I compare the *vasotocin* expression between new and full moon phases and between mature and immature worms. Additionally, I identify two receptors for the Vasotocin hormone in the annelid's genome. A second part concerns the characterisation of the transcriptional regulation of the *vasotocin* gene in vertebrates. We compare the expression domains in Medaka and Zebrafish as well as establish and analyse transgenic cross-species reporter lines. Finally, I work on an functional *in vivo* characterisation of the TMT Opsin positive cells. This subproject led to the discovery of many novel TMT Opsin members, which I further characterised by using a bioinformatic as well as a molecular approach.

2 Materials and Methods

2.1 Animal culture

2.1.1 *Platynereis dumerilii*

The animals were cultured in plastic boxes with a mixture of natural and artificial sea water at 18°C. The light regime is 16 hours light followed by a 8 hours dark period. Three days before and after the real full moon light was turned on during the dark period in order to entrain the animals in their lunar rhythm. Two rooms are available, one with a full moon-phase corresponding to the real full moon and one with a full moon-phase corresponding to the real new moon. During my experiments I did not distinguish between the two rooms. The animals I used for the quantitative RT-PCR experiments were from the B321xxx(x)-inbred line. The B stands for Berlin, where the strain was initially kept and each number represents one further generation of inbreeding. For the cloning experiments, only wildtype animals were used. This means that they do not have a controlled inbred genomic background.

2.1.2 Fish

The fishes were cultured in freshwater tanks at 26°C. For all cloning experiments and the whole-mount *in situ* hybridisation experiments, the inbred Cab strain (Medaka) and the WIK, AB, TL or Tüb strains were used. Embryos were obtained by scraping them off the female belly (Medaka) or by standard breeding in breeding tanks (Zebrafish). These breeding tanks have an inlay with a "rough" surface needed to induce matings. The eggs fall through the inlay and are easily collected. The embryos were kept in 1xERM (Medaka) or 1xE3 (Zebrafish) fish medium. After the embryos were capable of swimming (approx. after 7 days), they were set out in larger tanks and fed with dust food. After few weeks, the diet was changed to dust food and artemia and later on to flake food and artemia.

2.2 Vasotocin system in *Platynereis dumerilii*

2.2.1 Cloning the Vasotocin receptors

Several protein sequences of known Vasotocin receptors were downloaded from the NCBI database. The *Lymnaea stagnalis* orthologues (AAA91998.1 and AAC46987) were used to blast the genomes of *Lottia gigantea* (a mollusc as *Lymnaea stagnalis*), the crustacean *Daphnia pulex* and the annelids *Helobdella robusta* and *Capitella capitata*. The best hits were assessed by reblasting them against the non-redundant NCBI protein database. With the obtained protein sequences a multiple sequence alignment was prepared using ClustalX2 [75]. Highly conserved regions were manually identified. These conserved regions were subsequently used to design degenerate primer pairs. Degenerate primers are mixtures of primers with specific differences in the nucleotide composition that in combination correspond to the backtranslated conserved stretches of several homologous proteins. The identified conserved amino acid sequence was manually backtranslated using the common genetic code. Three different nested forward primers and four different reverse primers were designed. These four primers were used to potentially discriminate between two distinct Vasotocin receptor homologues, since two homologues have also been discovered in *Lymnaea stagnalis* [114]. The melting temperature (later on referred to as X) of the degenerate primer mixture were calculated using the program Oligo6 made by Molecular Biology Insights, Inc..

The degenerate primers were used for several nested PCR reactions to clone the Vasotocin receptors of *Platynereis dumerilii*. RNA for the cDNA conversion was used from two different stages, namely 50hpf and 75hpf. The RNA was extracted by J. Zantke using the RNAeasy Mini standard RNA extraction protocol for animal tissues. The cDNA was generated from 1μg total RNA using the Roche Transcriptor High Fidelity Kit as described in the manual. After the whole procedure, 40μl ddH_2O were added to the cDNA preparation. This cDNA was used to amplify the *Platynereis* Vasotocin receptor fragments by degenerated PCR using the Quiagen HotStarTaq as in the manual with certain changes. The final mix had a total volume of 50μl. From a 100μM primer stock, 1.5μl were used for one reaction from each primer. 3μl of the diluted cDNA were used per reaction. The program was as follows: 98°C for 15 minutes to activate the TAQ polymerase. Then 5 cycles of 95°C for 1 minute (Denaturing phase), $X - 5$°C for 2 minutes (Annealing phase) and 72°C for 4 minutes (Elongation phase). These were followed by 35 cycles using the same scheme as before except using the calculated melting temperature of X as the annealing temperature. After a final elongation phase at 72°C for 10 minutes, the reactions were stored at 10°C. One μl of the resulting PCR reactions were used as template for a nested PCR reaction. Otherwise the reaction mixture and the cycling scheme was the same as above. The reactions were loaded on a 2% agarose-metaphor gel with a 1 : 1 ratio. Metaphor is a special type of agarose used to properly distinguish small bands. Single bands with a size roughly corresponding to the fragment sizes identified in the various invertebrate receptors were gel-

eluted from a SYBRsafe-containing gel (Invitrogen) and subcloned into the pGemT-Easy vector system as described in the manual. The vectors were tested by EcoRI digestion (Fermentas) as well as by HinfI digestion (Fermentas). The EcoRI digestion releases the whole insert to check the correct size of the insert. HinfI has a recognition sequence of GANTC where N stands for any nucleotide. Therefore, the enzyme cuts rather often in random DNA stretches and thus can be used for DNA fingerprinting. Inserts with the same overall size as shown with the EcoRI digest but a different sequence can be identified. All vectors that showed a different HinfI restriction pattern were sent to sequencing. The presence of the correct Vasotocin receptor sequence was confirmed by blasting the obtained sequences against the non-redundant NCBI protein database using blastX. Two different receptor fragments have been cloned.

The partial sequences of the Vasotocin receptors were used to generate threefold nested primers to amplify the 5′ and 3′ end of the corresponding genes via RACE PCR [32]. For that, the RNA extracted at 75hpf was used for the cDNA conversion using the Clontech SMART Race Kit. Two different cDNA preparations were made, one to amplify the 5′ and the other to amplify the 3′ end as described in the protocol. For both reactions, 1μg total RNA was used as starting material. The Quiagen HotStarTaq was used for the RACE PCR and 10μl of Q-Solution were added to each reaction. Additionally, 5μl of the universal primer from the Clontech SMART RACE Kit was used together with 3μl gene-specific primer (5μM working solution). The same PCR cycling scheme as for the degenerate PCR was used. Irrespective of the high melting temperature calculated for the used gene-specific primers the annealing temperature was set to 55°C for the first 5 cycles and to 59°C for the other 35 cycles. For the nested PCR reactions, one μl of the primary PCR reaction was used and the annealing temperature of the 35 cycles were set to 60°C. Otherwise, the reactions were carried out as described above. 10μl of the nested RACE PCR reactions were loaded on one 2% agarose-metaphor gel (mixed 1 : 1) for each gene. The gel was blotted overnight on a nylon membrane using a dry blot technique. In short, the DNA was denatured by soaking the gel in 0.5M NaOH,1.5M NaCl for 20 minutes. The gel was placed on saran wrap with the slots down. The nylon membrane (Amersham) was placed on the gel followed by two thick blotting Whatman papers and a stack of paper towels. A bottle was used to weight the whole stack. The next day, the blot was unbuilt and the wells marked using a pencil. The membrane was dried and a radioactive probe was generated using the RadPrime labeling kit (Invitrogen) with the cloned Vasotocin receptor fragments obtained via degenerated PCRas described in the manual. This probe was used to visualize fragments in the gel that contain an overlap with the cloned receptor fragments. The probe was purified using Illustra ProbeQuant G50 sephadex columns as in the manufacturer's descriptions. Subsequently, the probe was hybridised using a standard southern blot protocol [98]. The largest identified bands were cut out of a preparative SYBRSafe-containing gel, subcloned into the pGemT-Easy vector and transformed into Top10 chemical competent cells (Invitrogen) as in the manufacturer's description. The grown plates were used for a colony lift. A colony lift membrane (Whatman Optitran $BA-585$, 0.45μm d : 82mm) was placed shortly on

the plate containing the bacteria, then lifted and the bacteria were denatured on 3MM Whatman paper soaked in 0.5M NaOH, 1.5M NaCl for 5 minutes. The membranes were neutralised using 1.5M NaCl, 1M Tris-HCl and finally equilibrated on 2xSSC. The membranes were dried and the bacteria containing the correct fragment were identified by hybridising the same radioactive probes as before to the attached DNA. In this way, the bacteria containing the fragment of interest could be isolated immediately and sent to sequencing to confirm the correct clones.

2.2.2 Sequence analysis

Since the Vasotocin receptors belong to the seven transmembrane GPCRs, the seven transmembrane topology was tested using the program TMHMM. A multiple sequence alignment was prepared using Muscle [28, 29] with several known vertebrate and invertebrate Nonapeptide receptors (Figure 3.1). Mammalian GnRH receptors were used as an outgroup since it has been shown that these are closely related sister molecules with a common ancestral receptor [31, 56]. The phylogenetic analysis was calculated using the neighbour-Joining algorithm implemented in the program ClustalX2 [75] excluding position with gaps and correcting for multiple substitutions. 1000 bootstrap replicas were calculated and the tree was visualised using FigTree.

2.2.3 Expression analysis

Immature animals were collected at the same circadian timepoint (13:00-14:00) at new moon and full moon phases (from the In- and Outphase rooms, respectively). The whole experimental setup was prepared twice during different full moon phases. This was done because it was found that *vasotocin* mRNA levels fluctuate over a day and do not oscillate (J. Zantke, personal communication). Therefore, it is likely that the expression levels are different between different days and circadian timepoints. Analysing the expression in two independent experiments might also provide the possibility to detect a certain trend.

Mature males were used for the analysis as well as immature worms irrespective of gender. The number of pooled heads of mature males differed between the different experimental setups. The first time, 15 mature males were pooled prior RNA extraction. The second time, only 5 were pooled. Thus, the first timepoint for mature males averages over 15 animals and the second over 5 animals. This happened because at the second time, only five animals were available in the culture. Also, this affects the statistical significance of the outcome since no biological replicates were treated as two biological replicates. For the statistical analysis, the values obtained for the mature males in both experimental setups were pooled. For all immature animals, three biological replicates of five randomly chosen heads were prepared. For each biological replicate five randomly chosen heads of immature animals were pooled and three technical replicates were prepared for each biological replicate.

The RNA was extracted using the RNAeasy Mini Kit according the standard animal tissue RNA extraction protocol. The RNA was eluted twice with the same 30μl of RNAse-free water. The RNA concentration was measured using Nanodrop and 0.5μg were used to generate cDNA using the Quantitect Reverse Transcription Kit (Quiagen) according to the manufacturer's instructions. After the cDNA conversion, 40μl PCR-grade water was added. The qRT-PCR primers for *vasotocin* have already been generated and tested previously (unpublished). As an internal control, I used the gene *cdc5*, for which primers were also already available. The qRT-PCR reaction was prepared using the Power SYBRGreen PCR MasterMix from Applied BioSystems. The mastermix was mixed with the respective primer pair (final concentration of 0.5pM/μl) and usually 5μl cDNA. The reaction had a final volume of 20μl. A standard qRT-PCR reaction was started using the Applied BioSystems StepONEPlus Real Time PCR System.

The resulting data were subsequently analysed using the REST MCS Beta 2006 software with a pair wise fixed reallocation randomisation test [92] and the resulting graph was prepared using Microsoft Excel 2007. A randomisation test is a non-parametric method, which does not assume any kind of distributions about the data. On the other hand, the T-test is a parametric method and assumes a normal data distribution. Such a non-parametric approach is essentially important for data sets generated by quantitative RT-PCR since these often do not follow a normal distribution [92]. The means of both experimental groups (e.g. full moon vs. new moon) were compared and the probability that we observed the differences in the samples, if the population means would be the same, was calculated. A randomisation test initially assumes that the data was randomly allocated to the two groups. The number of randomisations could be chosen and was always set to 2000. For each of these permutations the test statistic was calculated and thereby, a distribution of the test statistic was generated. One compares how extreme the found difference between the two real data groups was in relation to the generated distribution. The p-value then indicates the probability of obtaining test statistics that exceed, i.e. are more extreme, the obtained value for the real, non-randomised input data. A p-value below 0.05 is thought to indicate significantly different results (see [92] and G.W. Horgan and J. Rouault, http://www.bioss.ac.uk/smart/unix/mrandt/slides/frames.htm).

2.3 Regulation of *vasotocin* expression in vertebrates

2.3.1 Analysis of *nonapeptide* expression in Medaka

The genes encoding for the Nonapeptides Vasotocin and Isotocin were predicted using FGeneSH+ from Softberry using the identified Fugu homologues as templates and the genomic region, that I identified via TBlastN at EnsEMBL [57]. I cloned the genes using a nested PCR, subcloned it into the pCRII-Topo cloning vector (Invitrogen) as described in the manual. Midipreps of the correct vector were prepared and from that DIG- and Fluo-labelled cRNA probes were made as described in the appendix. Then different whole-mount *in situ* hybridisations were prepared to test the expression of the Nonapeptides in the developing animal. Pictures were either taken using the Zeiss Axioplan2 microscope (for NBT/BCIP staining) or the Leica LSM 510 Meta (fluorescent staining).

2.3.2 Reporter lines

Reporter lines are useful tools to label diverse cell types specifically. The labelled cell type can be used for several types of studies. For example, it is possible to isolate the labelled cells using laser microdissection. The isolated cells can subsequently be used for a number of analyses, like cell-specific transcriptome studies. However, reporter lines can also be used to study enhancer elements to assess in which cells or tissues the enhancer element drives the expression. Here, we use cross-species reporter lines. These utilise enhancer/promotor elements from one species and we test their functionality in another species. Thereby, we assess if the enhancer/promotor elements can be interpreted by the cross-species proteoms and thus we test the conservation of the enhancer/promotor elements across species.

2.3.3 Analysis of the *vasotocin* reporter line in Medaka

A Medaka *vasotocin* reporter line, that was prepared by Kristin Tessmar-Raible (unpublished) was analysed. The construct originates from Fugu and a similar construct was already used to generate transgenic mice [36] (See Figure 2.1). Four different Medaka insertion lines were available in the fourth generation. In mice, endogenous *vasopressin* is expressed in the SCN (suprachiasmatic nucleus), the PVN (paraventricular nuclei) and the SON (supraoptic nuclei) within the hypothalamus. The authors of this study found that the Fugu *vasotocin* gene is expressed in the PVN as well as SON, but fails to be expressed in the SCN. Additionally, the authors found that the potential regulatory regions reside 3' of the Fugu *vasotocin* gene. The used construct is shown in figure 2.1. We analysed the expression of the reporter gene by *in vivo* microscopy using the Leica LMD6500. The animals were embedded in sucrose with 0.2% Tricaine to anaesthesize the fishes. Pictures were taken as well as enhanced using the

accompanying microscope software. Additionally, I analysed the lines using different whole-mount *in situ* hybridization techniques (see appendix). Pictures were taken using the Leica LMD6500 for Medaka as well as the Zeiss Axioplan2 for Zebrafish.

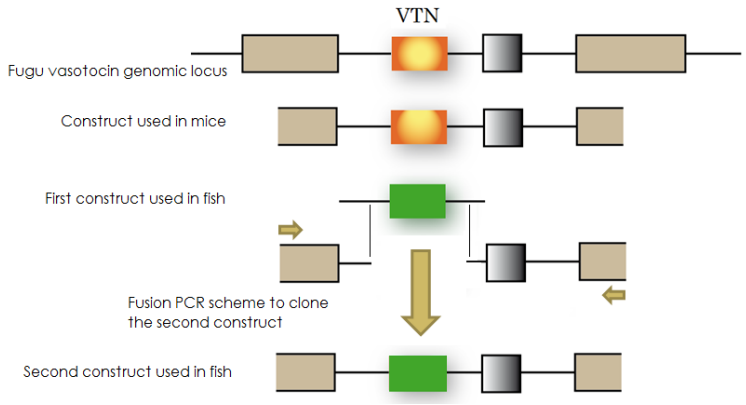


Figure 2.1: **Schematic representation of the reporter constructs** At the top, a schematic representation of the Fugu *vasotocin* genomic locus is shown. Three different genes are located in this locus next to the vasotocin gene. The construct used in mice contained the complete Fugu *vasotocin* gene as well as the adjacent ORF (2) completely and fragments of the genes 1 and 3 [36]. The reporter construct used in this study is much shorter and does not contain any ORF except a fluorescent marker protein (GFP[2] or mCherry) integrated within the Fugu *vasotocin* locus. In order to properly compare the fish with the mouse reporter line, the construct used for the mouse reporter line was recloned via fusion PCR. We fused the missing 3′ and 5′ fragments to the already existing *vasotocin* locus with the integrated reporter gene.

2.3.4 Generation of a new *vasotocin* reporter line in *Danio rerio*

The same construct used to generate the Vasotocin reporter lines in Medaka was cloned into a Zebrafish transgenesis injection vector. This vector is the destination vector of the Tol2Kit [70]. 5′ and 3′ of the insert, two specific sites are present within the vector, namely Tol2 sites. This vector is co-injected with mRNA encoding for the Tol2 transposase. This transposase helps integrating the genetic material between the Tol2 sites into the Zebrafish genome via a so far unknown mechanism [94, 70]. The mRNA of the Tol2 transposase was made using the mMessage mMachine Kit from Ambion as in the manufacturer's instructions. As a template for the *in vitro* transcription, a vector from the Tol2Kit (pCS2FA-transposase), linearised with NotI (NEB) and purified with the PCR purification Kit from Quiagen was used.

The whole construct was amplified with primers that contained specific restriction sites at the 5′ end. After the PCR reaction, the product and the injection vector was digested using the same enzymes. After dephosphorylation of the vector using the Antarctic Phosphatase (NEB) as described in the manual, the vector and the cleaved PCR product were ligated using a 1:3

molar ratio and a standard T4 ligase (NEB). The correct vector was partially sequenced and tested using a restriction digest. Then, an endotoxin-free maxi prep was made using the Quiagen Endotoxin-Free Maxiprep Kit according to the manual. $25ng/\mu$l of the vector were injected into Zebrafish embryos together with $20ng/\mu$l Tol2 mRNA (see Appendix).

The injected embryos were tested for ectopic expression of the reporter gene anywhere in the body one day after the injection. The presence of the fluorescent reporter protein was assessed using the Zeiss SteREO Lumar with appropriate filters and HiLite. At this time, the reporter gene is often misexpressed and all embryos that showed any expression where raised. Usually, only in the F1 generation the reporter gene is properly expressed. The fishes were mated and several embryos of one batch were used to extract genomic DNA using the Nucleospin Tissue kit from Macherey-Nagel. The genomic DNA was eluted with 100μl elution buffer and 200ng of this genomic DNA were used for a genotyping PCR in order to verify if the construct was transmitted to the next generation. For each genotyping reaction, three independent PCR reactions were made. One that amplified a fragment at the 5' end, one in the middle and one at the 3' end of the construct. The FirePol Taq form Solis BioDyne was used according to standard PCR conditions in a total volume of 50μl. The cycling scheme was as follows: $95°C$ for 4 minutes, then 35 cycles with $95°C$ for one minute, $62°C$ for one minute and $72°C$ for one minute. After a final elongation phase at $72°C$ for 10 minutes the reactions were stored at $10°C$ until gel analysis on a 1.2% agarose gel. Each amplicon for all three independent reactions should have a size of about 500 nucleotides.

2.3.5 Other *vasotocin* reporter lines

The constructs used for the Medaka and Zebrafish reporter lines were shorter than the one used for the transgenic mouse *vasotocin* reporter line ([36], Figure 2.1). In particular, it lacked partially the genomic region 3' of the *vasotocin* gene that was shown to be necessary for the expression of Fugu *vasotocin* in the hypothalamus of the transgenic mice. In order to compare the mouse and teleost reporter lines properly, a construct as the one used in mice was prepared. The only difference was the incorporation of the fluorescent reporter gene into the *vasotocin* gene locus as in the aforementioned shorter constructs used in fish. This construct was prepared using different fusion PCRs (see Figure 2.1). In short, the 5'end and the 3'end of the longer construct were used to design primers to amplify two regions, beginning from the respective ends of the longer construct and ending within roughly 100bp of the shorter construct. These primers contained an ISceI-site at the 5'ends. The two amplicons were then purified, mixed with the injection vector of the shorter construct and the primers at the ends of the longer construct were used to amplify the complete final construct with the GFP[2] inserted (see Figure 2.1). The PCR was digested using ISceI (NEB) and the band with the correct size (approx. 11kbp) was gel-eluted and cloned into the pDestISceI-Tol2 Vector. This vector represents a Gateway destination vector that contains

ISceI sites as well as Tol2 sites. This means that in Zebrafish, the ISceI sites will additionally be integrated into the genome. Since ISceI originates from yeast [110], no major complications are expected. However, when injecting this vector into Medaka, the existing Tol2 sites might harm because the Tol2 transposae originates from Medaka and might lead to the incorporation of the vector backbone into the genome where it might jump, thereby potentially mutagenizing the animals. This problem was overcome by amplifying the construct from the vectors with primers that contain an attached ISceI site at the 5' end using a high-fidelity polymerase for the PCR reaction. The amplified product was gel-eluted using the Quiagen Gel Extraction Kit as in the manual and quantified using Nanodrop. 25ng/μl were used from the PCR product to inject Medaka embryos. The Zebrafish embryos were injected with the vector as described above.

2.3.6 Phylogenetic analysis

cDNA and protein sequences from many Nonapeptide genes across bilateria were retrieved from NCBI and the EnsEMBL genome browser. The cDNA as well as the protein sequences were used for the phylogenetic analysis. The protein sequences were aligned using ClustalX2 [75] and the cDNA sequences were aligned using Muscle [28, 29]. The phylogeny was assessed using the neighbour-joining algorithm (ClustalX2, [75]), the maximum-likelihood algorithm (PhyML3, [45]) and Baysian phylogenetic inference (MrBayes, [58]). For the neighbour-joining analysis, multiple substitutions were corrected and positions with gaps were ignored. 1000 Bootstrap replicas were calculated. PhyML3 was used with standard settings, estimating all values possible to estimate. The LG model was used for the analysis of the protein sequences and the GTR model for the analysis of the nucleotide sequences. For the analysis with MrBayes the GTR model was used for the nucleotide sequences with gamma-distributed rate variation across sites and a proportion of invariable sites. For the proteins, a fixed rate model was estimated by setting the *aamodelpr* parameter to "mixed". Thereby, regularly new models are assessed and they contribute to the final result according to their posterior probability values. The tree was calculated until the standard deviation of split frequencies fell below 0.01, which happened usually after 100.000 generations. 25% were discarded as burnin. The resulting trees were visualised using FigTree and manually tested for similarities. Important nodes that are identical in all three trees were numbered using the obtained bootstrap values from the neighbour-joining and the maximum likelihood analysis as well as the posterior probability values from the baysian phylogenetic inference. The same protein alignment was used for a domain comparison analysis between the diverse Nonapeptide orthologs.

2.4 TMT Opsin family in vertebrates

2.4.1 Sequence analysis

The published Fugu TMT Opsin protein sequence [86] was obtained via the NCBI database (Acc.: NP_001027778.1) and blasted against the genomes of diverse vertebrate species using TBlastN (i.e. Blast a protein sequence against a translated DNA database) within the EnsEMBL genome browser [57]. The species were: **Mammals:** Human, Mouse, Rat and Chimpanzee. **Non-mammalian tetrapods:** Anole Lizard, Zebra Finch, Chicken and Frog. **Teleosts:** Stickleback, Tetraodon, Medaka, Fugu, Zebrafish. The results were sorted according to the lowest E-Values so that the best hits were shown on top. Since each exon usually represents a single entry, all entries in close proximity to each other (i.e. in similar region on chromosome, group or scaffold) were treated as one. The genomic DNA, often also including neighbouring genes, was then used for a homology-based gene prediction using the program FGeneSH+ from Softberry with standard settings. The previously identified Fugu TMT Opsin protein [86] was used to enhance the prediction. The genes in the DNA stretch are predicted using hidden markov models and then the prediction fitting best to the provided homologous protein is reassessed taking the provided homologous protein sequence into account. The resulting proteins were then blasted against the Non-redundant Protein database on the NCBI webserver to verify the results. If a TMT Opsin-like protein was identified, the best hits of the blast search also represented TMT Opsin like sequences. As soon as two different proteins were discovered (most often another vertebrate visual or non-visual ciliary Opsin, like Vertebrate Ancient Opsin, Pineal Opsin or Rhodopsin) for one species, the search within the initial TBlastN results from the EnsEMBL server was stopped.

The obtained protein sequences were used for a phylogenetic analysis. The proteins were aligned using Muscle [28, 29] and a phylogenetic analysis, as described for the Nonapeptide protein sequences, was prepared. Several members of the RGR photoisomerases were used as an outgroup. In addition, several visual and non-visual ciliary Opsin sequences were included to find the relation of the TMT Opsins with these Opsin families. Another tree was prepared including invertebrate ciliary and rhabdomeric Opsins as well as Melanopsins. This tree was prepared with a neighbour-joining algorithm. 1000 bootstrap replicas were calculated using the program ClustalX2 [75]. The alignment was also used to analyse amino acid residues identified in other Opsin molecules as being important for Opsin function. Additionally, for each Opsin group the seven transmembrane topology was assessed for one completely predicted Opsin homologue using the program TMHMM. In order to compare longer stretches of amino acids between different Opsin families, the tool WebLogo [20] was used with standard settings. All other Opsin sequences were obtained from the NCBI database or via personal communication (Enrique Arboleda). Furthermore, several incompletely predicted TMT Opsin sequences were omitted from this analysis (see Figures 3.16, 3.17 and 3.18). However, all were included for another phylogenetic analysis using the neighbour-joining algorithm to see to which cluster the

truncated sequences would fit (data not shown). Like this, the clustering of the incompletely predicted sequences was assessed.

In order to generate reporter lines, regulatory regions were predicted using the phylogenetic footprinting assay. For that, only the teleost homologues of one group were used. This was prepared for the TMT Opsin groups A2 and B. All genomic regions were downloaded from the EnsEMBL genome browser. The genes were predicted again using the Softberry FGeneSH+ program, which then also gives the orientation of the gene as well as the exact position of the exons within to the genomic fragment. This was used to build an annotation file according to the requirements of the Mulan software [89, 76]. The genomic regions as well as the annotation were uploaded to the Mulan webserver and repeats were masked according to Fugu genomic DNA for all. The suggested phylogenetic tree was accepted and the output was tested using a minimal ECR length of 70 and minimal ECR identity of 50%. Non-coding conserved regions were blasted against the protein database using blastX on the NCBI server to exclude non-annotated conserved genes from the analysis.

2.4.2 Cloning of the fish TMT Opsins

All identified Opsin homologues from Medaka and Zebrafish were cloned in order to verify that the different TMT Opsins of Medaka and Zebrafish do not represent pseudogenes. Subsequently the obtained fragments were used to generate an antisense DIG-labelled cRNA probe for whole-mount *in situ* hybridisation to assess the expression domains of the genes.

The TMT Opsin cDNA sequences were used to predict primer pairs with the program Primer3 using the standard settings [97]. The primer locations were then compared to the multiple sequence alignment to assess if these primers were located in properly predicted regions. The primers were used to clone the TMT Opsin homologues using RT-PCR. The used cDNA was prepared from 1μg total RNA of several seven day old Medaka embryos and from 1μg adult Zebrafish brain total RNA using the Roche Transcriptor High Fidelity cDNA conversion kit. The resulting cDNA was diluted using 40μl water. The Medaka RNA was isolated using the standard RNA extraction protocol from the Quiagen RNAeasy Mini Kit. The Zebrafish adult brain RNA was already prepared by Kristin Tessmar-Raible. The PCR was done using the Fusion Polymerase by Finnzymes with 3μl of the cDNA and 2μl of a 5μM primer solution in a final volume of 50μl. The annealing temperature was set to 60°C for all reactions and the buffer HF was used with standard conditions. Some Zebrafish TMT Opsins could not be cloned using the Phusion Polymerase but were successfully cloned using FirePol from Solis Biodyne in a standard reaction. 8μl of the PCR reactions were loaded on a 1.2% agarose-TAE gel to test if a fragment with the correct size was amplified. Then, the remaining PCR reactions were loaded on a 1.2% preparative gel containing SYBRSafe (Invitrogen). The bands were eluted using the Quiagen Gel Extraction kit as in the manual. The DNA was eluted using the same 30μl Elution buffer

twice. Then 3μl were used for subcloning into the pJet vector using the Fermentas CloneJet Blunt-End cloning Kit as in the manufacturer's instructions. The vector was transformed into Top10 chemical competent cells (Invitrogen) according to the manual.

2.4.3 Expression analysis

The expression was analysed using the standard whole-mount *in situ* protocol (see Appendix). For Zebrafish, only three day old embryos were used for the analysis. For medaka, three and five day old embryos were used. Animals were mounted in 87% glycerol and pictures were taken using the Zeiss Axioplan 2 microscope (Zebrafish) and the Leica LMD650 (Medaka). The images were further enhanced using MacBiophotonics ImageJ (for the minimum projection of Z-stacks) or Adobe Photoshop for brightness and contrast enhancement.

3 Results and Discussion

3.1 The Nonapeptide system in *Platynereis dumerilii*

3.1.1 Cloning of two Nonapeptide receptors in *Platynereis dumerilii*

A Nonapeptide homologue, called Vasotocin, has already been identified in *Platynereis dumerilii*[109]. However, so far no receptor to that neuronal hormone has been cloned. Since the expression domains of the receptors indicate the target tissue of the neuronal hormone, I cloned fragments of the Vasotocin receptors using degenerate PCR. We found several fragments potentially encoding Vasotocin receptors in the published genome of the annelids *Capitella capitata* and *Helobdella robusta*. Furthermore, two distinct Nonapeptide receptors have been identified in *Lymnaea stagnalis* [114]. This lead us believe that also in *Platynereis dumerilii* more than one receptor for the Nonapeptide might exist. Thus, we designed the degenerate primers in a way that potentially two distinct Vasotocin receptors could be identified.

Indeed, we managed to clone two fragments that encode for two different putative Vasotocin receptors. Their homology to other already identified Nonapeptide receptors has been assessed by blasting them against the non-redundant protein database at the NCBI server. In order to get the full-length coding sequence, a 3'- and 5'-RACE PCR was prepared using primers that annealed within the two receptor fragments. Thereby, more sequence information was obtained, however the full-length coding sequence still remains undetermined. The obtained sequences were then analysed thoroughly.

First, a phylogenetic analysis was prepared using a neighbour-joining algorithm. Vertebrate GnRH receptors were used as an outgroup since these represent the closest sister group to known Nonapeptide receptors [31]. There are two main branches. One contains all the vertebrate Nonapeptide receptors while the other one contains the invertebrate homologues from the lophotrochozoans *Octopus vulgaris*, *Lymnaea stagnalis* and the novel identified homologues from *Platynereis dumerilii*. In *Octopus vulgaris* the Nonapeptide receptors underwent a species-specific duplication as did the Nonapeptides (see below).

The Nonapeptide receptors belong to the Rhodopsin family of G-protein coupled receptors. They also have seven transmembrane structures. These can be predicted using a software called TMHMM. Obviously, not all seven transmembrane domains were properly predicted although as in the case of one *Lymnaea stagnalis* orthologue all were predicted correctly (Figure 3.2).

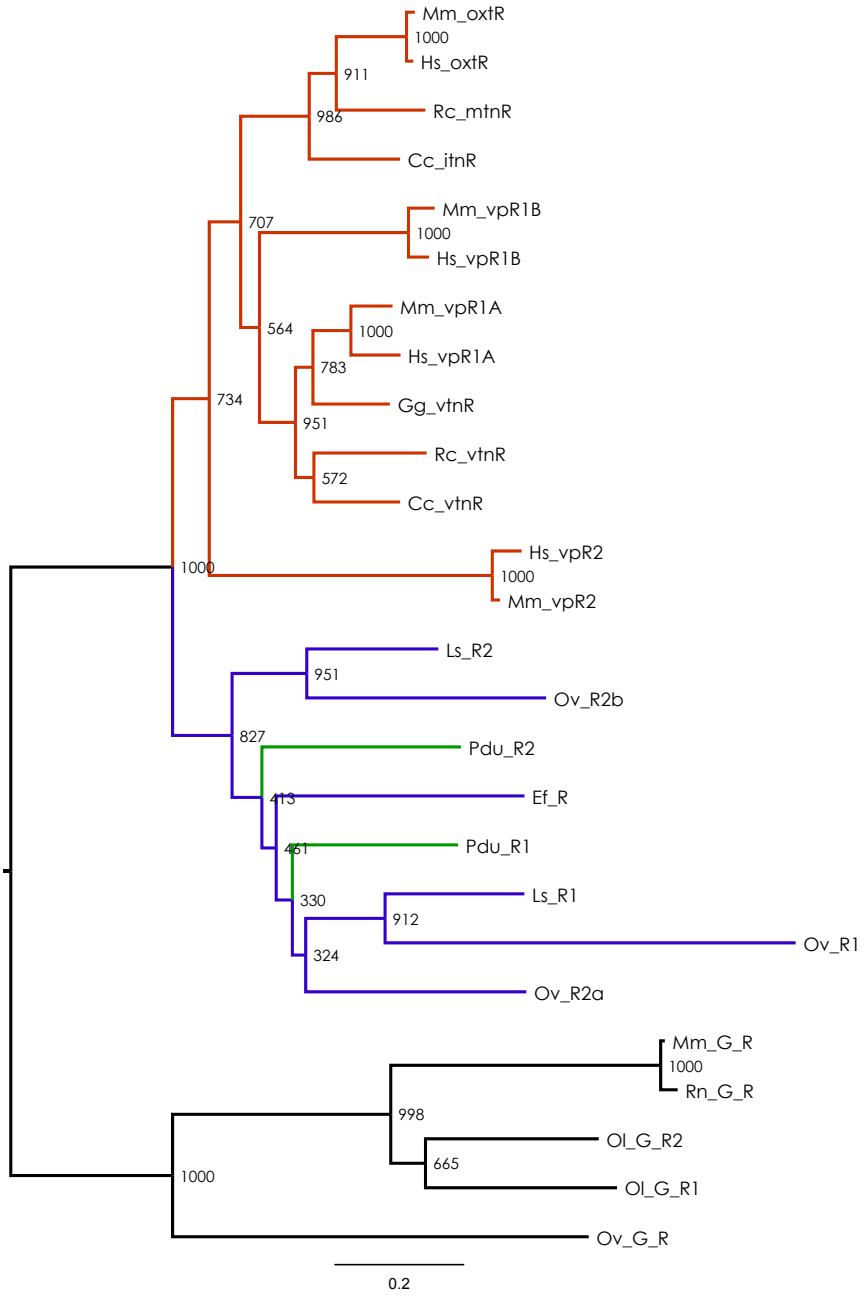


Figure 3.1: **Phylogenetic analysis of the Nonapeptide receptors** A phylogenetic analysis of diverse Nonapeptide receptors identified from diverse species was prepared using ClustalX2 [75]. GnRH receptors (black) were used as an outgroup since these represent a closely related sister group to the Nonapeptide receptors [31]. Two main branches appear, one contains all vertebrate Nonapeptide receptors (red) and in the other one all invertebrate homologues cluster (blue, green for the *Platynereis dumerilii* homologues identified in this study). Abbreviations are described in the appendix.

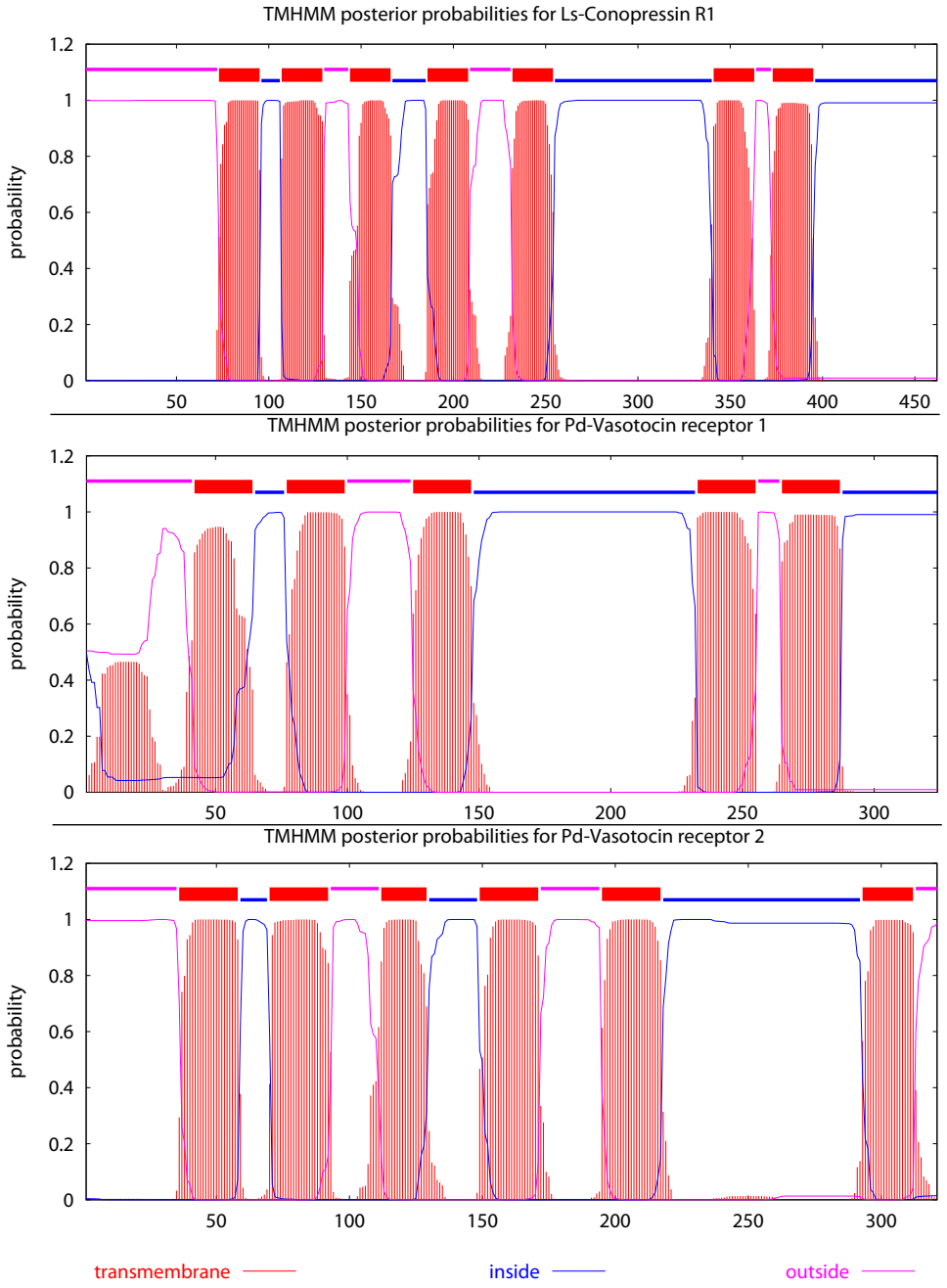


Figure 3.2: **Transmembrane domain analysis of the identified Vasotocin receptors** Here, the output of the transmembrane domain analysis using the online tool TMHMM for the two cloned Vasotocin receptors is shown. The analysis at the top was prepared using one Nonapeptide receptor of *Lymnaea stagnalis* as a positive control. The seven transmembrane domains that should be highly conserved are present (indicated by the red peaks), however I could only identify six transmembrane domains for both *Platynereis dumerilii* Vasotocin receptors. This is most likely due to incomplete sequence information obtained with the RACE PCR reactions. However, the position of the transmembrane domains seem to be conserved between the different homologues.

Based on the alignment we deduce that the RACE PCR did not give us the complete coding sequence with the 5′ end missing in the vtnR1 homologue and the 3'end missing in the vtnR2 homologue. This could have several reasons. The most likely explanation is that the 3′ RACE PCR primer annealed unspecifically to a stretch that has some complementary in the 3′ region of the Vasotocin receptor 2. The other receptor lacks a proper 5′ end which indicates that likely the reverse transcription during the cDNA conversion starting from the 3'end stopped prematurely. This could happen because of rigid secondary structures formed by the mRNA. Furthermore, it could be that the *Platynereis dumerilii* homologues really lack these domains, which is rather unlikely based on the high conservation of related genes. Another possibility would be that we discovered additional transciption initiation and termination sites, respectively. However, similar changes in the transcriptional regulation have not been reported for other homologues.

3.1.2 Is *vasotocin* expression under lunar control?

The gene for the neuronal hormone Vasotocin in *Platynereis dumerilii* has already been cloned previously [109]. A subset of the Vasotocin secreting cells in *Platynereis dumerilii* also express a ciliary Opsin. Since this cell type is highly conserved it was suggested that it regulates the secretion of Vasotocin according to exogenous light cycles that correlate to reproductive cycles. For *Platynereis dumerilii*, a lunar reproductive cycle has been identified [51]. Thus, I tested the expression of the *vasotocin* gene using quantitative RT-PCR. The mRNA levels were measured in the heads of immature animals during new moon and full moon as well as in mature males during new moon.

The expression of the *vasotocin* gene has been assessed in order to determine if the mRNA levels cycle according to the lunar rhythm and also if Vasotocin can be implicated to play a role in maturation. The experiments were prepared twice during different moon phases and the worms were beheaded at the same circadian timepoint. The experiment was repeated because it was already found that *vasotocin* expression is variable during a 48h cycle and worms at the same timepoints at different days can have highly variable *vasotocin* mRNA levels (J. Zantke, personal communication). By repeating the experiment independently at the same circadian timepoint and at the same day of different new and full moon phases, at least a trend can be estimated. For each sample, three different biological replicas were prepared using immature animals. On the other hand several heads of mature males were pooled but no biological replicates were made. Thus this data has to be treated as preliminary data.

In both experimental setups, it was found that the *vasotocin* levels of immature worms does not change significantly corresponding to the moon phase. However, mature males always had significantly upregulated levels of *vasotocin* expression (Figure 3.3). This suggests a role of the neuronal hormone Vasotocin in maturation but excludes it from being a direct downstream target of the lunar clock. It has to be stressed that this experiment has to be repeated to gain proper

statistical significance concerning the upregulation of *vasotocin* levels in mature animals.

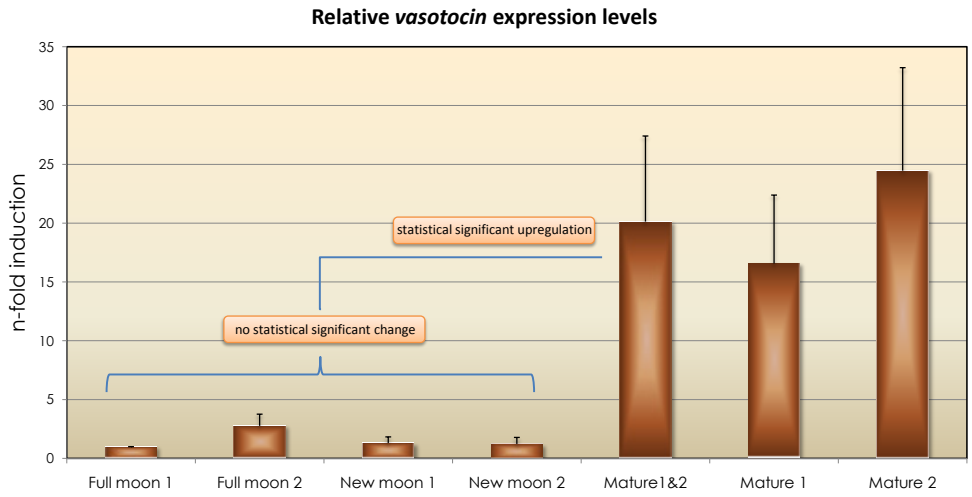


Figure 3.3: **Expression levels of *vasotocin* in *Platynereis dumerilii* heads** Here, the qRT-PCR results of two different experimental setups are shown (labelled with 1 and 2, respectively). The expression was analysed using the REST software. Error bars represent the standard error calculated by the REST software taking all biological and technical replicates into account [92]. All expressions were related to the level of Full moon 1, that is the expression level of *vasotocin* under full moon conditions in heads of immature animals during the first experiment. The expression of immature animals is highly similar in all four sampling points. In mature males, Vasotocin is significantly upregulated. Although the standard deviation is very high a clear increase in *vasotocin* mRNA levels was detected in both independent experiments (labelled with 1 and 2). For the statistical analysis, both experiments with the mature males were treated alone and as biological replicates. In all of these cases, they are significantly different from the *vasotocin* levels in immature males at a 5% significance level.

A problem could be that this assay does not take sex-specific differences into account. Nonapeptides are known to be differentially regulated between males and females. Immature males and females are impossible to distinguish based on outer morphological cues. The only way to distinguish them is to cut them and analyse a coelomic extract under the microscope. In males, single gamete precursors can be identified while they are clustered in females. However, this is only possible as soon as the worm reaches a certain size. Before that, the gametes cannot be seen or distinguished. Although mature males and females can be easily distinguished I focused on mature males. The females are very likely to start spawning when they are transferred out of the raising cup into fresh natural sea water. Thus, potentially more females have to be collected and usually the needed amount of mature females was not available at the specific timepoint. Therefore, I decided to carry out these experiments with mature males although sex-specific differences might exist that might provide important insights into the mechanism of maturation in *Platynereis dumerilii*. Furthermore, I did not test the expression in mature males during full moon conditions. Thus, since usually almost no worm matures during a full moon phase, it

cannot be excluded that the *vasotocin* gene is regulated in a moon-dependent manner in mature animals.

3.1.3 Outlook

We wanted to assess the expression domains of both receptors in developing embryos as well as adult tissues. Thereby, we could have identified cells that are responsive to secreted Vasotocin. For whole mount *in situ* hybridisation experiments, three different probes were used per gene, the fragment obtained via degenerate PCR and the two RACE PCR fragments. However, although often tried at diverse stages and circadian timepoints, a specific expression pattern could never be identified. This might be due to a very low expression level in the cell which can be detected by highly sensitive PCR-based methods but fails to give a signal when using whole-mount *in situ* hybridisation.

Furthermore, the quantitative RT-PCR experiments only assess mRNA abundance. However, the secretion of neuropeptides can also regulated on other levels. For example, the hormone-containing vesicles are often already primed at the membrane surface and await a signal for the fusion of the vesicle with the target membrane. Thereby, the peptide release response can be accelerated. Since these hormones are stored in these vesicles, the mRNA does not necessarily correlate with secreted hormone levels. However, it would be possible to assess the proper hormone either by purifying and later on quantifying it or by ELISA, which would require a specific antibody against the *Platynereis dumerilli* Nonapeptide. Nevertheless, the significant upregulation of *vasotocin* mRNA levels suggests an important role of Vasotocin in the maturation of *Platynereis dumerilii*. Furthermore, a direct moon-dependent regulation on the level of transcription can be excluded based on the presented qRT-PCR results.

3.2 The regulation of *vasotocin* expression in vertebrates

3.2.1 Independent duplication of the Nonapeptide hormone genes in vertebrates

The Nonapeptide hormones exist in diverse non-vertebrate species such as echinoderms and annelids, but has been lost in nematodes and higher insects [108]. I will focus especially on the evolutionary history of the Nonapeptide in the vertebrate lineage. Basal vertebrates, like lampreys and hagfish, as well as most invertebrates possess only one orthologue [63, 113, 109, 47, 46], whereas all higher vertebrates possess at least two homologues. The correct phylogenetic grouping of homologues is important, as often conclusions about functional conservation or divergence are drawn from it. Many authors assume that this increase in gene number reflects an early duplication of an ancestral gene in the vertebrate lineage that gave rise to independent Arginine-Vasopressin and Oxytocin lineages [47, 46, 117, 36, 18, 37, 74]. There are few indications for that. First, when comparing the different vertebrate Nonapeptide sequences, one homologue always has a neutral amino acid at the position eight ("Oxytocin-like" hormones), whereas the corresponding amino acid in the other homologue is positively charged ("Vasopressin"-like hormones) [37]. Additionally, similarities in the genomic regions endcoding the Nonapeptides between diverse vertebrate species have been found [46, 47]. In transgenic mice and rats, the Fugu *vasotocin* and *isotocin* genomic regions show expression of the Fugu Nonapeptides in vasopressinergic and oxytocinergic neurons, respectively [117, 36]. However, we re-assess the phylogeny of the vertebrate Nonapeptides by using whole preprohormone sequences for a phylogenetic analysis as well as by taking genomic arrangements, domain structure and the evolution of regulatory elements into account. We analysed the data present in the literature and examined the phylogeny of the Nonapeptide superfamily using whole preprohormone cDNA and protein sequences (Figure 3.4), the domain structure of the Nonapeptide homologues (Figure 3.5 and the arrangement of the different genomic loci encoding the Nonapeptide genes. In contrast to current belief, we conclude that rather several independent duplications occurred during vertebrate evolution and we also discuss important functional implications.

First, we obtained protein and cDNA sequences for identified Nonapeptide orthologues of different species through the NCBI and EnsEMBL databases [57]. Upon alignment of these sequences, trees were generated using neighbour-joining, maximum-likelihood and Bayesian algorithms implemented in the programs ClustalX2, PhyML3.0 and MrBayes3.1.2 [75, 45, 58] (Figure 3.4). All trees consistently show an independent duplication of mammalian Arginine-Vasopressin and Oxytocin on the one hand, and teleost Isotocins and Vasotocins on the other hand. Node support is strong, independent of the method and the usage of cDNA or protein sequences (Figure 3.4). Our conclusion confirms previous analyses [60, 59], but is in contrast to other more recent interpretations [47].

We next asked if the independent duplications of the ancestral vertebrate Nonapeptide locus are reflected in the domain structure of the different homologues. The preprohormone consists

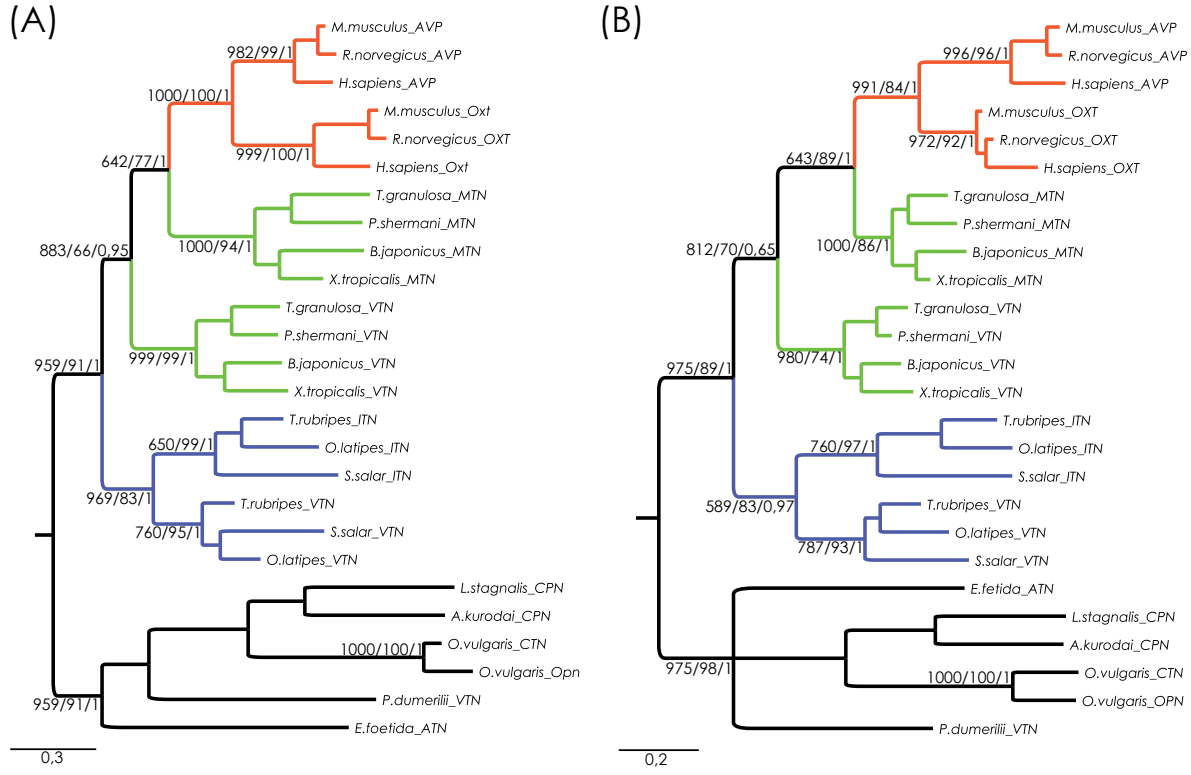


Figure 3.4: **Revisited phylogeny of the vertebrate Nonapeptide orthologs** (a) A phylogenetic analysis has been prepared using the DNA sequences of several vertebrate Nonapeptide homologues. Invertebrate Nonapeptide preprohormone sequences were used as an outgroup. The tree was generated upon a multiple sequence alignment with Muscle using three different methods: neighbour-Joining (ClustalX2), Maximum-likelihood (PhyML3.0) and Baysian phylogenetic inference (MrBayes3.1.2) [45, 58, 75]. The numbers indicate the bootstrap values (1000 for NJ- and 100 for ML-analysis) and the posterior probability values obtained with the baysian analysis after 100000 generation with 25% burn-in. The tree has the topology of the baysian analysis. It clearly shows highly supported independent duplications within the teleost and mammalian lineage. The names represent the Genus and the capital letter indicates the species (e.g. MusM: Mus musculus). Abbreviations: AVP: Arg-Vasopressin, OXT: Oxytocin, VTN: Vasotocin, MTN: Mesotocin, ITN: Isotocin, ATN: Annetocin, CPN: Conopressin, CTN: Cephalotocin, OPN: Octopressin. (b) The phylogenetic analysis was repeated using corresponding protein sequences (compare to (a)) except that the multiple sequence alignment was prepared using ClustalX2 [75]. In these trees we see independent duplications within the teleost and mammalian lineage as well. The position of the members from the non-mammalian tetrapods is less clear. In addition, we tested Vasotocin and Mesotocin sequences of the chicken *Gallus gallus* but these sequences seem to be diverged, because they jumped between branches and led to lower support values.

of four different parts: the signal peptide, the Nonapeptide (the actual hormone), Neurophysin and Copeptin. The region that includes the Nonapeptide sequence and the Neurophysin domain is highly conserved in all homologues, whereas the Copeptin domain with a leucin-rich core segment is more variable (Figure 3.5). Consistent with their close relatedness, teleost Isotocins and Vasotocins both contain the Copeptin domain, whereas it is absent in vertebrate Mesotocins and mammalian Oxytocins. A basic cleavage site and an N-linked glycosylation site preceding the Copeptin domain have been identified in human Vasopressin [21]. These sites are also present in Vasotocins of hagfish, coelacanths and non-mammalian tetrapods except for birds [46, 47], although their functionality is unclear [82]. Teleost *vasotocin* and *isotocin* genes lack these C-terminal sequence features [46, 47] (Figure 3.5). In summary, the domain structure of the vertebrate Nonapeptide family members alone does not allow a clear conclusion about the orthology relationships of the different members. On the one hand, it supports an independent teleost duplication. On the other hand, Mesotocin and mammalian Oxytocin resemble each other by the lack of the Copeptin domain, which is, conversely, present in Vasotocin and Vasopressin. However, as we argue below, the loss of the Copeptin domain can be interpreted as a case of functional convergence.

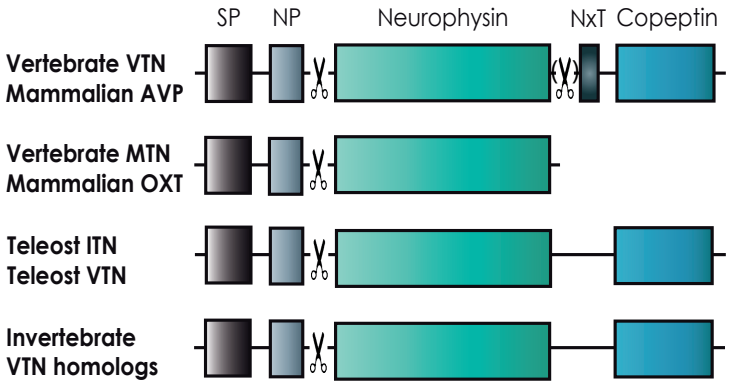


Figure 3.5: **Domain comparison between the different Nonapeptide homologues** A comparison of the domain structure of the Nonapeptide superfamily is shown. Vasotocin in non-mammalian amniotes and mammalian Vasopressin contain the widely conserved copeptin domain while it is not present in the Mesotocin and Oxytocin homologues in the same species. However, it is retained in teleost Isotocins. In addition, it was shown that the Copeptin is cleaved in mammalian Vasopressin. Although the cleavage site is conserved on a sequence level, it was shown in frogs that it is not functional [82]. The VTN orthologue of *Gallus gallus* contains the N-linked glycosylation site (NxT) but lacks a homologous cleavage site. Thus, the differences in the domain structure and the preprohormone processing also indicate a divergent evolutionary history of the vertebrate Nonapeptide superfamily. Abbreviations: SP: Signal peptide, NP: Nonapeptide, NxT: N-linked glycosilation site, Scissors indicate a proteolytic cleavage site.

Finally, we assessed the organisation of the genomic loci in mammals, non-mammalian tetrapods and teleosts. Mammalian Vasopressin and Oxytocin genes are oriented in a head-to-head fashion, whereas their orthologues in non-mammalian tetrapods are found in tandem orientation. The teleost Nonapeptide gene loci are more variable. In some teleosts, the *vasotocin* and *isotocin* genes are not directly linked to each other, but five genes are present in between. In zebrafish, the genes are even located on two different chromosomes [46, 47]. This divergent organization of the genomic loci is in accordance with the proposed independent duplications in vertebrates. One set of data, however, seems to support orthology between *vasotocin/vasopressin* and *isotocin/oxytocin*. The regulatory regions of the teleost Takifugu rubripes *vasotocin* and *isotocin* genes are capable of driving the expression of a gene in transgenic mice and rats in vasopressinergic and oxytocinergic cells, respectively. This finding has been interpreted to indicate a closer relatedness of *vasotocin/vasopressin* and *isotocin/oxytocin* genes [117, 36]. Based on the outlined history of gene duplications, we propose an alternative scenario: the single ancestral chordate precursor locus could have already contained distinct cis-regulatory elements, as they are found in the current teleost and mammalian loci. Upon duplication of this ancestral gene locus, including its regulatory regions, similar, but convergent subfunctionalisation might have occurred in different vertebrate groups. Such a scenario would also explain the ability of mammals to distinctly interpret the cis-regulatory regions of the Fugu Nonapeptide genes [116, 36]. In conclusion, the evolution of the Nonapeptide superfamily is more divergent than it has been assumed previously. Our re-analysis suggests that the ancestral Nonapeptide preprohormone gene duplicated independently at least two times during vertebrate evolution, giving rise to *vasopressin* and *oxytocin* in mammals and to *vasotocin* and *isotocin* in teleosts. These independent duplications have to be accounted for when discussing the functional evolution of the Nonapeptide superfamily.

Particularly, the in-depth analysis of the prohormone domain structure has interesting functional implications. So far, physiological functions have only been described for the name-giving Nonapeptides. The C-terminal rest of the prohormones is believed to serve as a carrier/folding aid for the respective Nonapeptide [21]. However, as mentioned above, the Copeptin domain follows its own evolutionary path, indicative of a separate physiological function. It is present in both teleost homologues [46, 47]. In non-mammalian tetrapods, however, it can be either present (Vasotocin) or absent (Mesotocin) from the prohormone [82], indicating that it is not strictly required for Nonapeptide function. Similarly, mammalian Oxytocin lacks the entire Copeptin amino acid stretch, whereas it is present in the Vasopressin precursor. Together, these data suggest that the different components of the prohormone precursor could carry different functions. In line with this, mammalian vasopressinergic cells secrete Copeptin separately from Vasopressin and Neurophysin [21, 84].

3.2.2 Regulation of the Nonapeptide gene expression in vertebrates

We analysed so-called reporter lines to assess conservation of regulatory elements across different species. In 2003, a transgenic reporter mouse was established that contained the Fugu *vasotocin* genomic locus [36]. In mouse, the Nonapeptide ortholog is usually expressed in the SCN, the PVN and SON of the hypothalamus. The authors of this study showed that the genomic region of the *vasotocin* locus can induce expression of the Fugu *vasotocin* gene in a mouse background. Additionally, the authors could show that the regulatory regions that drive the expression of *vasotocin* are located at the 3'end of the gene and might even span the adjacent open reading frame. A scheme of the used construct is shown in Figure 2.1 [36].

Kristin Tessmar-Raible generated a reporter line in the species Medaka (unpublished). For that, a similar albeit shorter fragment of the Fugu *vasotocin* genomic region was used (see Figure 2.1). Four independent insertion lines were already available in the fourth generation. In order to analyse the reporter line, I first cloned the Medaka Nonapeptide orthologues using standard RT-PCR. The obtained fragment was subsequently used to generate antisense probes for whole-mount *in situ* hybridisation. Additionally, the ontogeny of the Vasotocin expressing cells was determined.

Differences in the ontogeny of *vasotocin* expression domains between Medaka and Zebrafish

The ontogeny of vasotocinergic cells has already been established for Zebrafish [27]. In this system, the cells first appear in 24hpf embryos in the ventral hypothalamus. One day later, cells that express *vasotocin* in 48hpf embryos were identified in the dorsal preoptic area. In addition to that, cell movements in the ventral hypothalamus domain have been reported. The other Nonapeptide gene, *isotocin*, is only expressed in the dorsal preoptic area, in cells that are in close vicinity to the *vasotocin* expressing cells [27]. As can be seen in Figure 3.6, the first Vasotocin expressing cells form at stage 33 to 34. These seem to reside within the dorsal preoptic area. During the same day, the expression domain broadens, more cells activate the *vasotocin* expression within the dorsal preoptic area. Shortly after that, cells in the ventral hypothalamus start to express the *vasotocin* gene. This shows that there are temporal differences between the *vasotocin* expression in Medaka and Zebrafish since the two prominent expression domains in the Zebrafish brain appear in a reversed order in Medaka. However, whole-mount *in situ* hybridisation experiments represent snapshots of the development. Thus, I cannot exclude that the domains are moving and changing position. One way to do that would be via two-color whole mount *in situ* hybridisation using probes against vasotocin and isotocin. Since *isotocin* is only expressed in the dorsal preoptic area, this domain can be identified and correlated to the developing *vasotocin* gene expression. Nevertheless, if the developmental timing of the occurence of both *vasotocin*-expressing domains is correct, then the ontogeny of vasotocinergic

cells in Medaka is different from the one in Zebrafish, where it has been shown that the ventral hypothalamic region appears first and only approximately one day later, the dorsal preoptic area starts to express *vasotocin* [27]. This suggests that there are differences in the temporal regulation of the development of *vasotocin*-secreting domains. However, the site where *vasotocin* seems to be eventually expressed in similar regions in both species.

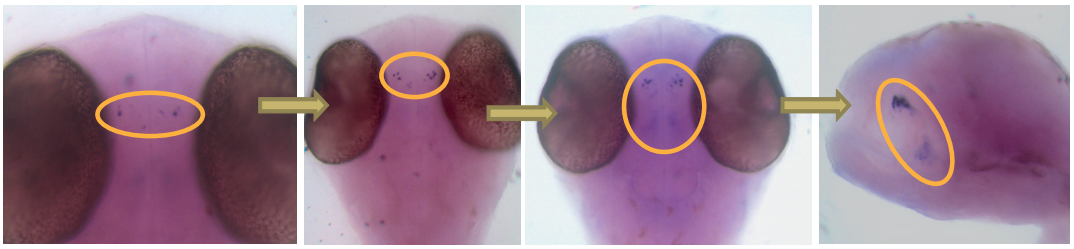


Figure 3.6: **Ontogeny of vasopressinergic cells in Medaka** A whole mount *in situ* hybridisation was prepared using approximately five days old embryos (stage 34 - 36). The embryos develop slightly asynchronously within one petri dish. The *vasotocin* mRNA was visualised using DIG-labelled antisense cRNA probes and a NBT-BCIP staining. In several fishes no staining was visible (not shown) but additionally, there were fishes with expression domains at the same site but with different sizes. The domain that appears alone is most likely located at the dorsal preoptic area. There is also a growth of this domain that seemed to correlate with the developmental stage of the embryos. Furthermore, a ventral hypothalamic region could only be identified together with the dorsal region. If the ventral hypothalamic region occurred, the dorsal expression site was rather large in comparison to others. The expression domains are encircled.

Next, I tested the expression of the homologous Nonapeptide gene *isotocin* in comparison to *vasotocin* using 2-color fluorescent whole-mount *in situ* hybridisation. In zebrafish, *isotocin* is only expressed in the dorsal preoptic area in cells that are in close vicinity to vasotocinergic neurons. An *isotocin* expression domain in the ventral hypothalamus has not been identified [27]. In medaka, *isotocin* is also only expressed within the dorsal preoptic area. As in Zebrafish, the medaka isotocinergic neurons are in close vicinity to the the vasotocinergic neurons and there is no co-expression of both Nonapeptides in any cell (Figure 3.7).

Analysis of the Medaka *vasotocin* reporter lines

Furthermore, I analysed the expression of the fluorescent reporter gene in the reporter lines. I used in vivo fluorescence microscopy as well as whole-mount *in situ* hybridisation techniques. Two distinct expression domains are visible at the region where also the neuronal hormone is expressed. Additional ectopic expression can be seen around the anterior commissure and the lenses. However, no GFP expression was detected in the ventral hypothalamus.

In order to verify the expression domains seen *in vivo*, I prepared whole-mount *in situ* hybridisation using DIG labelled antisense probes against GFP. In Figure 3.9, a representative staining is

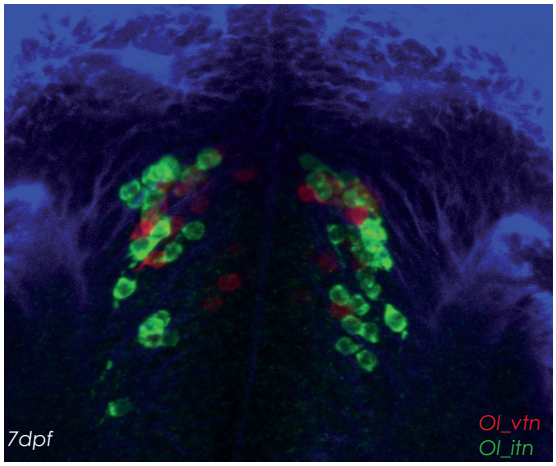


Figure 3.7: **Co-expression of Medaka *vasotocin* and *isotocin*** Here, a 2-color fluorescent whole mount *in situ* hybridisation is shown. Vasotocinergic neurons are labelled red while isotocinergic neurons are labelled green. The background label is anti-acetylated tubulin, which did not penetrate the embryo well for unknown reasons. A Z-stack was recorded using the Zeiss LSM 510 Meta with a 20x objective. A maximum intensity projection was prepared using the MacBiophotonics ImageJ package. Because a maximum intensity projection is used, cells seem to partially overlap (yellow spots). However, this is due to the overlay of several images and analysing the whole stack in 3D makes it clear that there is no co-expression.

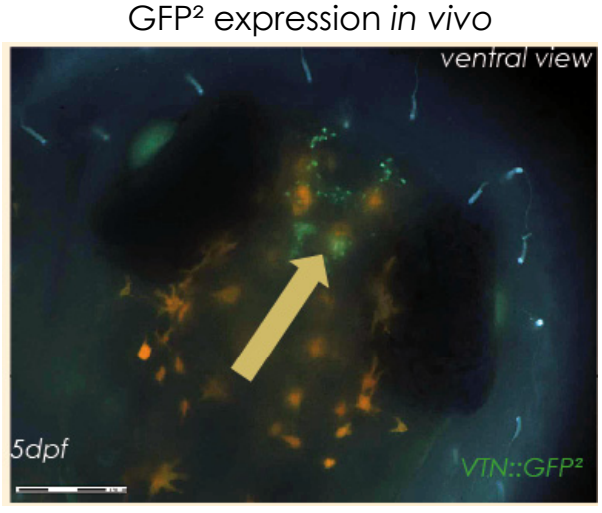


Figure 3.8: **In vivo analysis of GFP^2 expression** The arrow indicates the two expression domains of GFP^2 that are within the dorsal preoptic area between the eyes. The picture represents a projection of a Z-Stack. Thus, note that the ventral hypothalamic expression domain is not detectable. Furthermore, ectopic expression within cells along the anterior commissure is present.

depicted. The three different pictures show different focus planes in the same animal. In all three different insertion lines, ectopic expression can be detected in the anterior commissure. Additionally, the expression domain in the dorsal preoptic area is also stained. However, the expression domain seems to be larger than the endogenous *vasotocin* expression domain I identified.

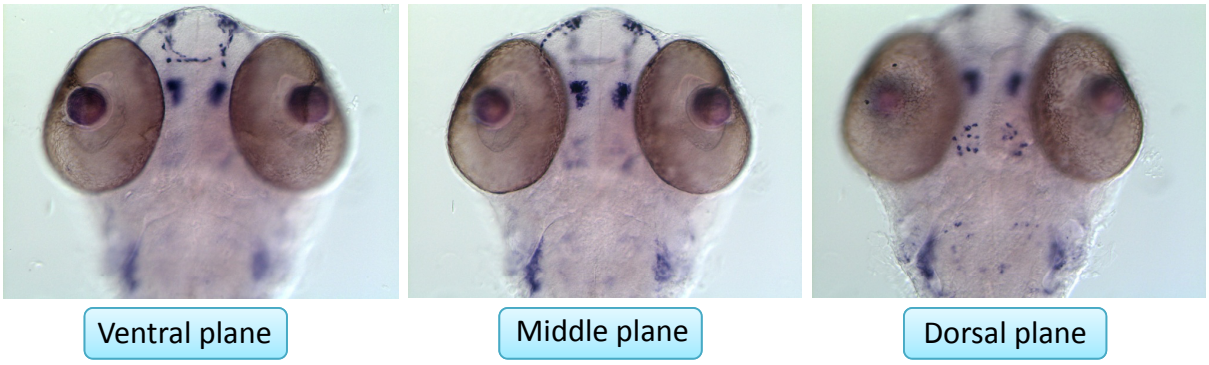


Figure 3.9: **Whole mount *in situ* hybridisation of GFP[2]** A whole mount *in situ* hybridisation shows where the reporter gene is expressed. The left picture indicates the ventral expression domains, the middle picture shows the focal plane in the middle of the embryo and the right picture shows the dorsal focal plane. As in Figure 3.8, the dorsal preoptic area is stained while the ventral hypothalamic domain is not stained at all. In addition many other ectopic expression domains are present. The ectopic expression domain within the anterior commissure is present in all insertion lines while the other domains are variable and therefore most likely due to position effects.

In none of the insertion lines, any expression of the reporter gene was visible in the ventral hypothalamus. However, all had a very broad expression domain at the site of the dorsal preoptic area. In order to test, if the reporter gene is expressed in vasotocinergic neurons, a 2-color fluorescent whole mount *in situ* hybridisation against *vasotocin* or *isotocin* and the reporter gene was used to clearly distinguish the cells under the confocal microscope. Since the expression is only visible within the dorsal preoptic area, the same batch of fishes was also used to stain *isotocin* along with the reporter gene. Representative results are shown in figure 3.10.

In addition to the 2-color fluorescence whole mount *in situ* hybridisation, stainings for the *vasotocin* mRNA was prepared using NBT-BCIP (data not shown). In few, no staining was detected at all and in some the expression domains are as in the wildtype. However, in few also ectopic expression of *vasotocin* was found in the anterior commissure. This could mean that the construct somehow induces the expression of *vasotocin*. This would mean that potentially the Vasotocin levels in these fishes are higher, providing that the cells contain the machinery to process and secrete the neuronal hormone. An integration site-dependent mechanism could also be possible. However, I have seen that in all four insertion lines and thus such a position effect is rather unlikely. Another possibility would be that with my probes against Medaka *vasotocin* I could also label the Fugu *vasotocin* gene. Although the gene is not expressed, it is

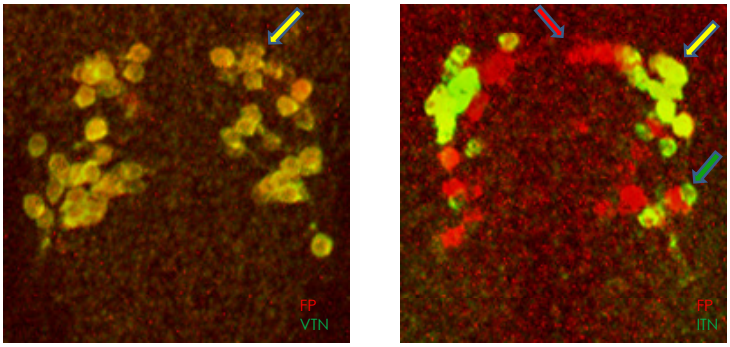


Figure 3.10: **Co-expression of the Nonapeptides with the Reporter genes** Here, the co-expression of the reporter gene is compared to the expression of the Nonapeptides Vasotocin and Isotocin. The pictures were taken using a confocal microscope and a maximum intensity projection was prepared using MacBiophotonics ImageJ. The *vasotocin* gene seems to be always coexpressed with the reporter gene whereas the *isotocin* gene has some overlap (yellow) but there are also cells that express *isotocin* or the reporter gene alone (red and green arrows). Furthermore, the results for the *vasotocin* gene are inconsistent with the fact that the *isotocin* and *vasotocin* genes are not co-expressed. Furthermore, other whole-mount in situ hybridisation experiments gave inconsistent results (discussed in the text).

likely still present on the reporter gene mRNA. However, this would also mean that the probes cross-react differently between different fishes. A possible explanation for the 2-color fluorescent whole mount *in situ* hybridisation results would be that, the antibody-conjugated peroxidase was not inactivated properly after the first staining. This could result in remaining peroxidase activity in cells that should not be labelled with the second staining. However, increasing the inactivation time did not result in changes. Furthermore, the 2-color fluorescence whole mount *in situ* hybridisation of the genes *vasotocin* and *isotocin* were prepared in the same run and did not show overlapping cells. Thus, the way to test the reporter lines still has to be optimised.

Generation and analysis of Zebrafisch reporter lines

The same construct as was used to generate the Medaka *vasotocin* reporter lines was used to generate transgenic zebrafish lines. For that, the construct was recloned into a Tol2 Zebrafish injection vector belonging to the Tol2Kit [70]. It was then injected into one- to two-cell stage embryos along with mRNA encoding for the Tol2 transposase. The injected fishes were out-crossed or incrossed and their offspring (F1) was analysed using the SteREO Lumar microscope. However, in none of the embryos from several couples, the expression of the reporter gene could be detected. This is surprising since the construct worked in Medaka and a similar construct even works in mice.

In the next step, I used several embryos from one batch for genomic DNA extraction. This

DNA was then used for genotyping PCR (Figure 3.11). Three different reactions per batch were prepared. These reactions should amplify fragments of the construct at three different sites. In this way, it was tested if the complete transgene was transmitted. Using this approach, I could identify at least seven couples that transmitted the transgene to the next generation. The readout was very clear since three independent PCR reactions were used and always a positive and negative control was prepared. Although I found the gene to be transmitted to at least seven different F1 generations, I could not identify the expression of the reporter gene in any of them. This also makes possible integration site-dependent inhibition of gene expression rather unlikely because at least seven different insertion lines exist. Nevertheless, genotyping via a PCR reaction also provides several difficulties. Since I extracted the genomic DNA from a group of embryos, it is still possible that only partial fragments that very between individuals were transmitted to the next generation. This would require that the gametes of the parents are heterogenous concerning the construct insertion site. Another possibility would be that the construct was disrupted and integrated partially at different sites in the genome. Thus, I can still detect the fragments but the close spatial relationship is missing and thus the reporter gene cannot be expressed.

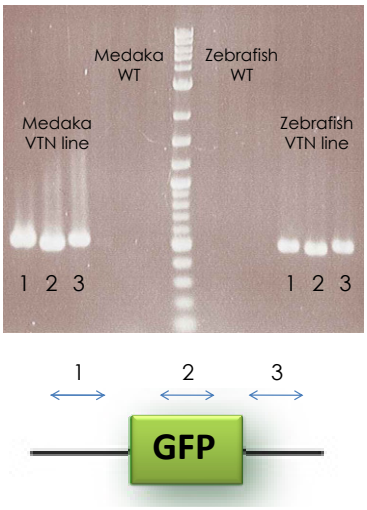


Figure 3.11: **Representative genotyping PCR** At the bottom, the locations of the three amplicons is depicted according to the construct used to generate transgenic animals. At the top, a genotyping PCR reaction was analysed on a 1.2% agarose gel. The Medaka reporter lines were used as a positive control. The genomic DNA of wildtype Medaka and Zebrafish was used for the negative control. On the right side, the genotyping PCR results for the offspring of one couple is shown.

This is of course a very surprising result. While the construct that originates from Fugu can be used to drive the expression of a reporter gene at least partially in mice [36] and Medaka, it fails to express the reporter gene in a Zebrafish background. This could also mean that Zebrafish cannot interprete the Fugu promotor. However, this can be tested by using a promotor that is known to

work in Zebrafish instead of the Fugu promotor. Another explanation is that Zebrafish changed the cis-regulatory factors that drive the expression of *vasotocin*. This is also supported by other facts. First, the development of *vasotocin* expressing domains is reversed between Medaka and Zebrafish as shown above. Although the expression domains seem to be conserved, this could be an indication for different gene regulatory networks that drive the expression. Another hint comes from the genomic structure of the Nonapeptide genes locus. In several teleost species, the *vasotocin* and *isotocin* genes are linked with several genes in between. These genes show conserved synteny relationships. In other vertebrate species, these two genes are even more closely linked. This linkage is also thought to be important for the regulation of Nonapeptide gene expression [46, 47, 117]. However, Zebrafish is the only species in which these genes are not linked but reside on different chromosomes [47]. If this linkage really is important for the regulation of *vasotocin* expression, then Zebrafish must have evolved different mechanisms to drive the expression in the proper neurons.

One way to test this would be to identify regulatory elements using the computational approach phylogenetic footprinting. However, this cannot be done because the Zebrafish *vasotocin* gene locus is not completely sequenced yet. Another possibility would be to inhibit known upstream regulatory factors. For example, *rx3*, *otp* and *sim1* have been identified to be upstream of *vasotocin* in Zebrafish [27, 109]. Their expression in Medaka could be downregulated by injecting specific morpholinos, thereby inhibiting translation. Additionally, natural occuring mutants could be used. In these animals, the *vasotocin* expression could be identified using standard whole-mount *in situ* hybridisation. In this way, it can be tested if these factors are also important for *vasotocin* expression in Medaka. If not, then this would represent a strong indication for differences in the upstream regulatory machinery and would explain why the Fugu construct works in Medaka and mice but not in Zebrafish.

In order to correctly compare the different reporter lines, I recloned the construct that was used in mice with the integrated reporter gene using fusion PCR as shown in figure 2.1. The construct was cloned into the pDestISceITol2 injection vector. This vector was used to inject zebrafish embryos. Because the Tol2 transposase is present in Medaka, the fragment with the ISceI sites at both ends had to be amplified via PCR and injected together with ISceI protein as a linear fragment. Unfortunately, in the course of the diploma thesis, the injected fishes did not reproduce yet and thus I could not analyse them.

3.2.3 Evolutionary links between the GnRH and the Nonapeptide system

A common evolutionary origin links the Nonapeptide and GnRH neurohormonal systems

Several functional similarities exist between the Nonapeptide system and the GnRH-GAP decapeptide system (Figure 3.14). We compared both systems to better understand their functional and evolutionary relationship. A first level of comparison between the systems concerns their

role in reproduction. Most prominently, members of the Nonapeptide family have been well established for their control of intraspecies communication/behavioural changes [111, 25]. These behaviours are often connected to bonding, parental care or mating (see below, [84, 111, 83, 130, 13, 39, 40, 50, 79]). In addition, a connection between the development/maturation of reproductive organs and Nonapeptide hormones has been suggested by several studies for vertebrates and invertebrates [109, 39, 2, 33, 67]. The GnRH-system also governs several aspects crucial for gonadal development and reproduction directly and indirectly [62, 120, 42, 81]. Expression and functional studies also suggest a role for GnRH in the reproduction in invertebrates [64, 19, 104, 95]. In mammals, Vasopressin and Oxytocin control a variety of behaviours, including different male and female reproductive behaviours (see for example [37, 40, 7, 16]). A similar behavioural function has also been described for frogs [25]. Thus, GnRH and the Nonapeptide system have so far predominantly been connected to the regulation of reproduction. In addition, secretion of Vasopressin and its orthologues is directly responsive to changes in body fluid osmolarity [7]. Thus, many Nonapeptides are involved in salt homeostasis, osmoregulation and therefore excretion [7]. Given their functional similarities, we wondered if these derive from a shared evolutionary ancestry. Although there is no apparent sequence similarity between the GnRH and Nonapeptide preprohormones, both encode two proteolytically released peptides. Their major functional parts are of similar lengths (Nonapeptide vs. Decapeptide) and contain additional C-terminal co-released peptides.

The evolutionary relationship of hormonal systems is often difficult to assess based on the hormones alone, as these are relatively short and tend to evolve faster. In contrast, receptors are often more reliable to track the history of hormonal systems, as they tend to be larger and easier to detect by similarity searches. Indeed, both Nonapeptides and/or GnRH-GAP have been lost from multiple invertebrate groups, whereas their receptors were maintained [112, 108, 54, 23, 15]. A closer analysis of the GnRH- and Nonapeptide receptors reveals that they belong to the same Rhodopsin-family of G-protein coupled receptors (GPCRs) and are more closely related to each other than to any other GPCR [54, 31]. Therefore, both receptors most likely derive from a common precursor molecule, which suggests a common origin of the GnRH and Nonapeptide neuronal hormone systems early in evolution [108].

Ontogeny and evolution mirrors the functional diversification of the Nonapeptide superfamily

The seemingly unrelated functions of the Nonapeptide neurohormonal system in reproduction and excretion are mirrored by the evolution and ontogeny of the reproductive and excretory organs in diverse bilaterian species. These organs develop from the same anlage (Figure 3.12) The vertebrate urogenital system consists of the kidney, testes or ovaries, as well as the duct system that connects the excretory and the reproductive systems to the outside. All parts of this

system originate from the intermediate mesoderm [35]. The ontogenetic connection between osmoregulatory and reproductive organs is also conserved in many protostomes, most prominently in lophotrochozoans. The postlarva of the "annelid stem species" [8] most likely had tubular segmental organs that functioned as excretory as well as genital organs. Also, several modern annelid species use their metanephridia (i.e. excretory systems) to expel their gonadal products. It has been proposed that these organs share a common evolutionary heritage [8]. These relations leave us with the puzzling question, why two seemingly divergent functions like reproduction and excretion as well as their organs originally evolved together.

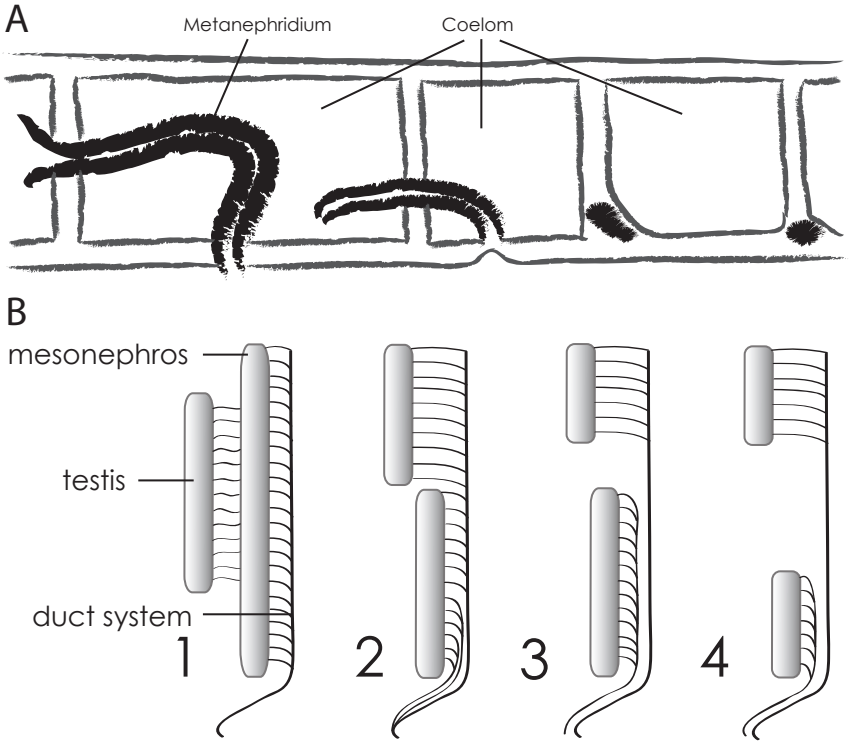


Figure 3.12: **Comparative view of urogenital development in annelids and vertebrates** (A) The development of metanephridia is depicted. These are excretory organs in annelids and they are often used to expel gonadal products as well. As can be seen in the different segments, the metanephridia develop from one single spot and grow into a tubule-like structure that connects the coelom with the outer world. [8] (B) The phylogenetic development of the urogenital system in amniotes is shown here. The schemes represent the urogenital system in several sturgeon species (A), shark and urodele species (B and C) and in amniotes (D). In the ancient urogenital system, the testis was still connected directly to the mesonephros. As can be seen in B-D, the testis and mesonephros seperate during evolution. Therefore, the excretory and reproductive organs share a common evolutionary origin. This is still apparent by the fact that both systems develop from the same tissue anlage. Pictures reproduced after [96].

Why did reproductive and excretory organs evolve together?

The first hint that sheds light onto this relationship is the possible pheromone-like role of the Nonapeptides conserved in basal chordates as well as in higher vertebrates. The CSF contains a considerable amount of neuronal hormones. Whereas in vertebrates, the CSF is always disconnected from the environment, in the basal chordate amphioxus, both larvae and adults show a continuity between their CSF and the outer sea water ([119, 55, 72], Fig. 3). During amphioxus development, the edges of the neural plate bend up and fuse to form the neural tube. The anterior part of the neural tube does not close entirely but rather leaves a porus, the larval anterior neuropore [125]. In the adult, a remnant of the anterior neuropore still exists, namely Kölliker's pit (Fig. 3). The structure is filled with cilia [123] that are moving [55] [54]. The CSF/sea water continuum allows vital dyes to be taken up via the anterior neuropore and to stain neurons within the body [55]. This finding suggests that diffusion of small molecules, like peptide hormones, through the anterior neuropore is possible.

Remarkably, pheromone-related functions of the Nonapeptide system exist in vertebrates. Vasopressin accumulates in the CSF after intranasal application [12] and sex-dependent differences in the response to administered Vasopressin and Oxytocin were found. Human males rate faces as unfriendlier after Vasopressin application, whereas females rate faces as friendlier [111]. Additionally, intranasal administration of Oxytocin in human males modulates trust [69, 44, 9] as well as the perception of displayed emotions [24]. The aforementioned possibility that neuropeptides can be taken up and released by marine organisms and the fact that accumulation of Vasopressin and Oxytocin in human CSF after intranasal application occurs [12] and affects behaviour, indicates that uptake and release of Nonapeptides could be an evolutionary ancient feature.

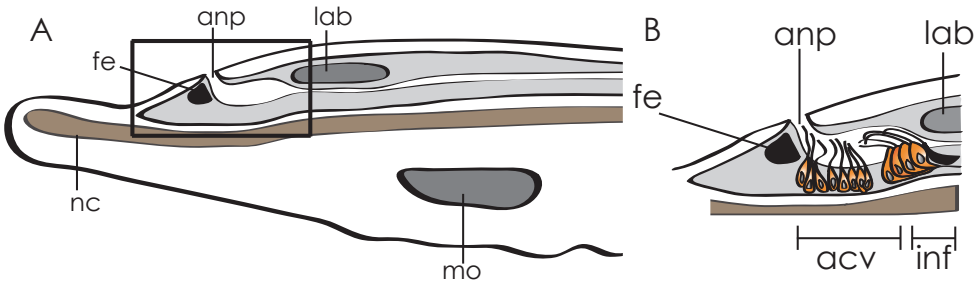


Figure 3.13: **Basal chordate anatomy is compatible with an ancestral pheromonal function for neurohormones.** a) Schematized head of an amphioxus larva (dorsal to the top, anterior to the left). b) Close-up of the anterior neuropore (anp) region (boxed area in a). The anp allows diffusion of small molecules between the CSF in the anterior cerebral vesicle and the surrounding seawater. The figure is based on [71, 72], redrawn with permission. Abbreviations: anp: anterior neuropore, fe: frontal eye, nc: notochord, lab: lamellar body, mo: mouth, acv: anterior cerebral vesicle, inf: infundibulum

A second aspect that provides a link between the seemingly divergent functions of ancient neurohormonal factors considers the ontogenetic relationship of GnRH neurons and olfaction. In mammals, a subset of the gnrh-gap expressing neurons appears first in the nasal placode and subsequently migrates through the cribriform plate into the basal forebrain. Finally, the cells will be positioned in the hypothalamus and innervate gonadotropin-secreting cells in the brain [17]. Interestingly, these GnRH neurons do not only originate together with olfactory neurons from the same placode, but will later receive direct input from olfactory neurons [127, 11]. Thus, it has been proposed that the olfactory-hypothalamus-adenohypophysis axis originates from an ancestral cell type that combined both olfactory and GnRH-secreting properties [4].

An ancestral chemosensory-neurosecretory system evolved in prebilaterian species

So far, we discussed that the GnRH and the Nonapeptide systems are functionally similar and originated from one common ancestral hormonal system. Both play prominent roles in reproduction, including their function as pheromones and neuromodulators to influence several intraspecies behaviours. Additionally, the Nonapeptide system is crucial for osmosensation and subsequently, osmoregulation/excretion. In light of the additional link between the GnRH and the olfactory system, we suggest that the ancestral GnRH/Nonapeptide hormonal system was able to sense olfactory/pheromonal and chemical cues (Figure 3.14). We hypothesize that this ancient chemosensory-neurosecretory system was used to sense outer environmental conditions, and subsequently adjusted the organism to these conditions. Furthermore, this system would have been able to respond to proper environmental mating conditions with the secretion of neuronal peptide pheromones. These pheromones were sensed by other individuals and influenced their reproductive behaviours. The same peptides also functioned as hormones and influenced reproduction in both sender and receiver, thus providing a possible feedback-loop that ultimately led to mating in favourable conditions. The predicted pheromone-like function of this ancestral system seems to be still present in some of today's metazoans. For example, secreted GnRH induces mass spawning in the mollusc *Mopalia sp.* [41]. As outlined above, pheromone-like transport of Nonapeptides is also possible. Such a broadly chemosensory system regulating different body functions presumably had little possibilities to regulate gonadal growth or reproductive behaviour in a more specific or detailed way. Molecular duplication of the ancestral system then gave way to a more fine-tuned subfunctionalization.

When did this ancestral chemosensory-neurosecretory system evolve?

The molecules of both systems are present in the two major branches of the bilaterian animal tree. Therefore, we conclude that the common ancestor of all bilaterian species (Urbilateria) must have had both systems. In order to test if both were already in place at the dawn of metazoa, we searched for the existence of related peptide hormones and their receptors in the available

genomes of three non-bilaterian metazoa, namely *Nematostella vectensis* (Cnidaria), *Trichoplax adherens* (Placozoa) and *Monosiga brevicollis* (protist). However, despite reported immunoreactivty for GnRH in cnidarian species [62], we did not find any indication for the presence of GnRH- and Nonapeptide- related neurohormones or receptors in the published genomes. As these molecules must have evolved from a common precursor, at least this precursor molecule should be detectable. Our failure to do so could be explained by incomplete genome sequences/assembly or that these species lost the respective orthologues in the course of evolution.

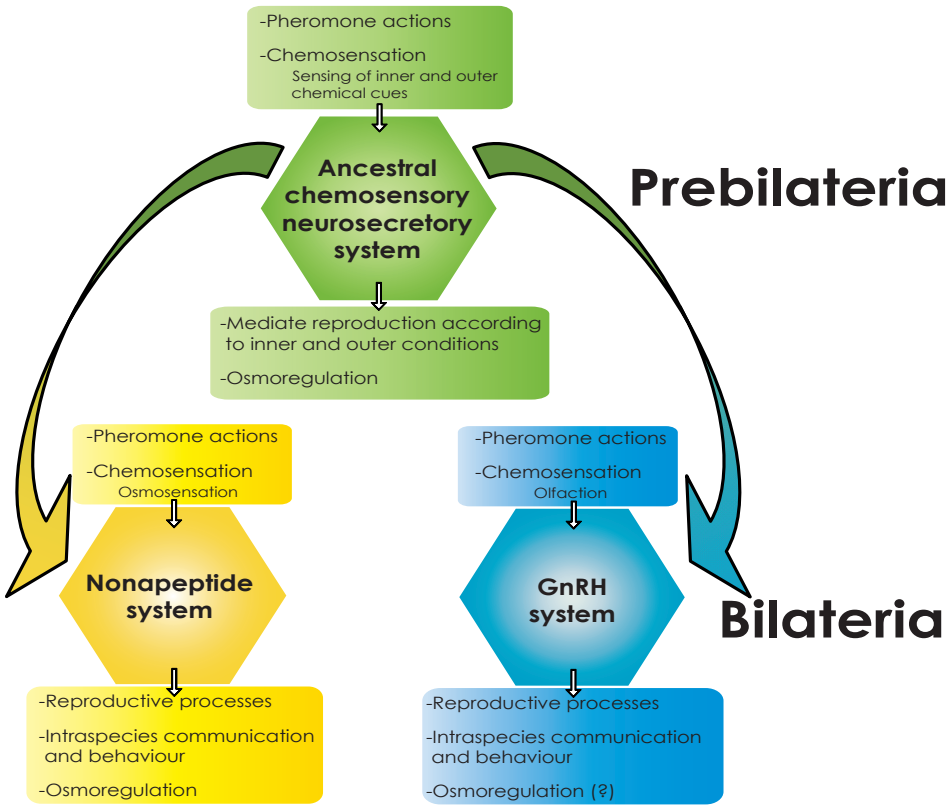


Figure 3.14: **A scenario for the evolutionary diversification of the Nonapeptide and the GnRH systems from a common precursor** In this scenario, an ancestral prebilaterian chemosensory and neurosecretory system (top) acted to sense environmental conditions and respond to them by secretion of peptide pheromones/hormones to regulate homeostasis and various aspects of reproduction. In the course of bilaterian evolution, these functions were subdivided between the Nonapeptide and GnRH neurohormonal systems (bottom). Shared functions between the systems therefore reflect ancestry rather than convergence.

3.2.4 Outlook

The proposed independent duplications of the Nonapeptide genes are very important, since often assumptions concerning the function of diverse genes are drawn from evolutionary relationships. This is especially important, since experiments for the functional characterisation of the Nonapeptides in diverse species are influenced by the assumed evolutionary history.

Furthermore, I analysed the expression of *vasotocin* and its regulation in two different teleost fish species. Striking differences between the *vasotocin* expression were identified. The newly established reporter lines with the complete construct that was used in mice [36] will show if the important regulatory regions are located outside of the construct that was used to generate the analysed fish reporter lines. In addition to that, the *vasotocin* gene locus of the teleost species should be analysed using a phylogenetic footprinting approach to identify conserved regulatory regions. Unfortunately, this locus has not been fully annotated in the published Zebrafish genome and thus differences in conserved non-coding and potentially regulatory elements could not be identified. However, a comparative analysis of this locus should provide further indications for the way the *nonapeptide* genes are regulated in vertebrates.

Finally, we compared the functions and the phylogenetic relationship of the Nonapeptide and the GnRH systems. Thereby, we deduced that both derive from a common prebilaterian hormonal system that governs several aspects of reproduction and that this system was broadly chemosensory in nature. Analysing the pheromone-like aspects of the GnRH and Nonapeptide system in chordate species might provide further insights into the functions of the proposed ancient hormonal system. In addition, it has been found that some GnRH- and Nonapeptide releasing cells express a presumptive light-sensitive Opsin molecule [109, 49]. Thus, the ancient hormonal system might also have responded to outer light clues. However, this is often dependent on exogenous light cycles and thus, light might have an effect over a longer time-period while the chemosensory nature of this system likely initiates immediate responses.

3.3 Identification of a novel deep brain photoreceptor family

In the previous sections, I discussed my results concerning the Nonapeptide hormonal system in diverse species and as stated previously, the hormones are secreted according to exogenous light cycles. Our hypothesis is that deep brain photoreceptor cells are responsible for that. These cells co-express the Nonapeptide and a member of a specific Opsin family [109]. In this section, I want to discuss the results I obtained during the characterisation of this specific Opsin family.

In 1999, two studies independently reported that a deep brain photoreceptor-like molecule exists in mammalian brains [10, 48]. The gene and its expression domains were identified in adult human and mouse tissues. In 2003, a group reported that they discovered a potential homologue to Encephalopsin, namely TMT Opsin [86]. The authors suggested that based on the overall similarities the TMT Opsin and Encephalopsin genes are closely related but based on the percentage of sequence identity these Opsins rather represent two different families. However, Encephalopsins and TMT Opsins were regarded as true homologues in another study [122]. One member of the TMT Opsin family has been identified in Zebrafish and one member in Fugu [86]. The authors found that it is expressed in a variety of tissues and organs in adult fishes. However, in 2007 it was shown that the gene is not expressed in a variety of tissues and organs in developing fishes but is restricted to neuronal regions, most prominently in the hypothalamus. Since these Opsins are deep brain photoreceptors they might also play important roles in photoperiodic time measurement. Additionally, a conserved co-expression with the neuronal hormone Vasotocin in some cells was found. This lead to the hypothesis that this Opsin might be implicated in the regulation of seasonal Vasotocin secretion [109].

3.3.1 Initial identification of the novel TMT Opsin homologues

I blasted several vertebrate genomes (see Figure 3.18) and found that only one ortholog exists in diverse mammalian species, namely Encephalopsin. In all other vertebrate species, I could identify one Encephalopsin homologue. Additionally, in the genomes of marsupials, non-mammalian tetrapods and teleosts I found members of the TMT Opsin family. I predicted the identified homologues and prepared a multiple sequence alignment to test if some Opsins were likely to be falsly predicted. Indeed, some seemed to be truncated, which can be explained by incomplete assembly of the respective genome sequences or by incorrect gene prediction.

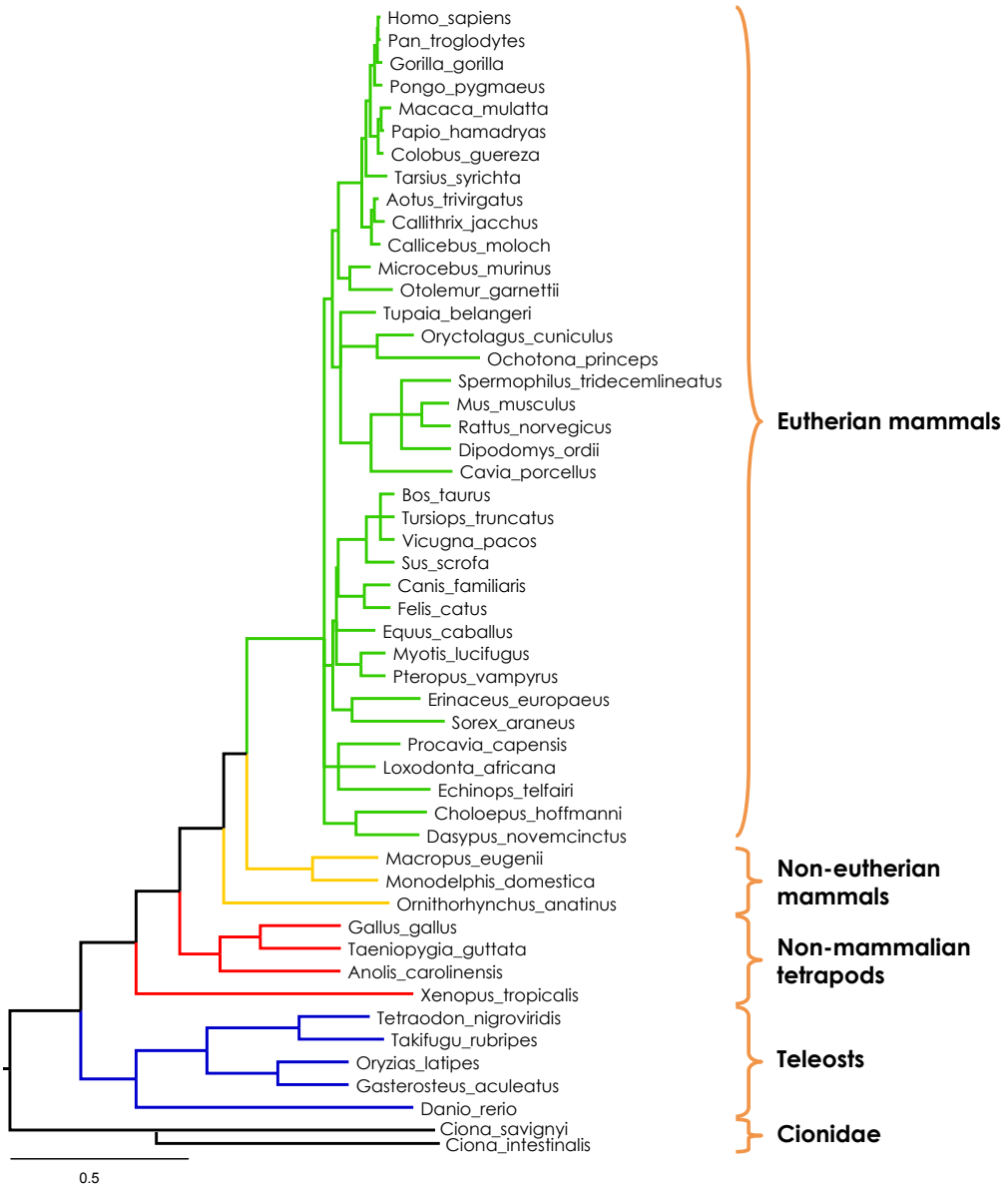


Figure 3.15: **Vertebrate phylogeny** A phylogenetic analysis of the vertebrates is shown. The tree was taken from the EnsEMBL genome browser [57] and modified using FigTree. The genome sequences for the species used to build the tree are available. Eutheria (shown in green) seem to lack any evidence for the existence of TMT Opsins while in all other species I could identify either several complete paralogs or at least fragments of TMT Opsin genes (as for two marsupial species). I tested the eutherian species *Tursiops truncatus*, *Bos taurus*, *Loxodonta africana*, *Mus musculus*, *Rattus norvegicus*, *Pan troglodytes* and *Homo sapiens*. This means I cannot exclude the existence of TMT Opsins in other eutherian species, but based on the phylogenetic position of all tested species, they likely represent all eutherian groups. Furthermore, for almost all of these species, one Encephalopsin homologue has been automatically annotated, except for *Sus scrofa* which most likely is due to incomplete genome assembly. Interestingly, in all of these species one Encephalopsin homologue has been identified despite several genome duplications.

In order to see how these homologues relate to each other and to other already identified visual and non-visual rhabdomeric and ciliary Opsins a phylogenetic analysis was prepared (Figure 3.16 and 3.17). For the calculation of the first tree, the neighbour-joining algorithm was used. Members of the ciliary and rhabdomeric Opsins were included to see where the TMT Opsin and Encephalopsin family clusters. The second tree was calculated using three different algorithms and the output always had a similar topology. Furthermore, all important nodes are highly supported by all three methods. The tree in figure 3.17 shows the topology resulting from the baysian phylogenetic inference.

The TMT Opsins and Encephalopsins cluster together with the ciliary Opsins when compared to rhabdomeric Opsins (Figure 3.16). Furthermore, the bootstrap values around the invertebrate ciliary Opsins show a very low support for the existing branches although the support is high when omitting this group. Also, the branch clusters rather to the other ciliary Opsins when calculating the tree using a maximum likelihood algorithm (data not shown). This suggests that the invertebrate ciliary Opsins represent an outgroup to the vertebrate ciliary Opsins and diverged before the divergence of the ciliary Opsins. The branch that includes all TMT Opsins and Encephalopsins is always separated from other identified visual and non-visual ciliary Opsins (Figure 3.17). Furthermore, four distinct groups form. One contains the mammalian Encephalopsin homologues as well as one homologue for each species. The other three groups represent three different TMT Opsin families, namely TMT Opsin family A, B and C. In these groups, teleost-specific duplications were identified. However, it has to be stressed that the identification of these Opsin genes highly depend on genome coverage. Therefore, the identified members might not represent all existing TMT Opsin or Encephalopsin orthologues. In table 3.18, a summary of the phylogenetic clustering in the diverse species is shown which also highlights teleost specific duplications.

I identified clear TMT Opsin homologues from non-mammalian tetrapod species and from five different teleost fish species (Figure 3.18). Interestingly, I also found fragments of TMT Opsin-like proteins in *Macropus eugenii* and *Ornithorhynchus anatinus* when blasting the already annotated TMT Opsin like-protein of *Monodelphis domestica* (XP_001372147.1) against both genomes. I only could find fragments of the TMT Opsin homologues in both species, which is most likely due to a low genome coverage. These are marsupials, an infraclass to placental mammals. Placental mammals (Eutheria) seem to have lost the TMT Opsin family, since I could not find any indications for its presence in available genomes although many have a very high coverage, like the human and mouse genomes.

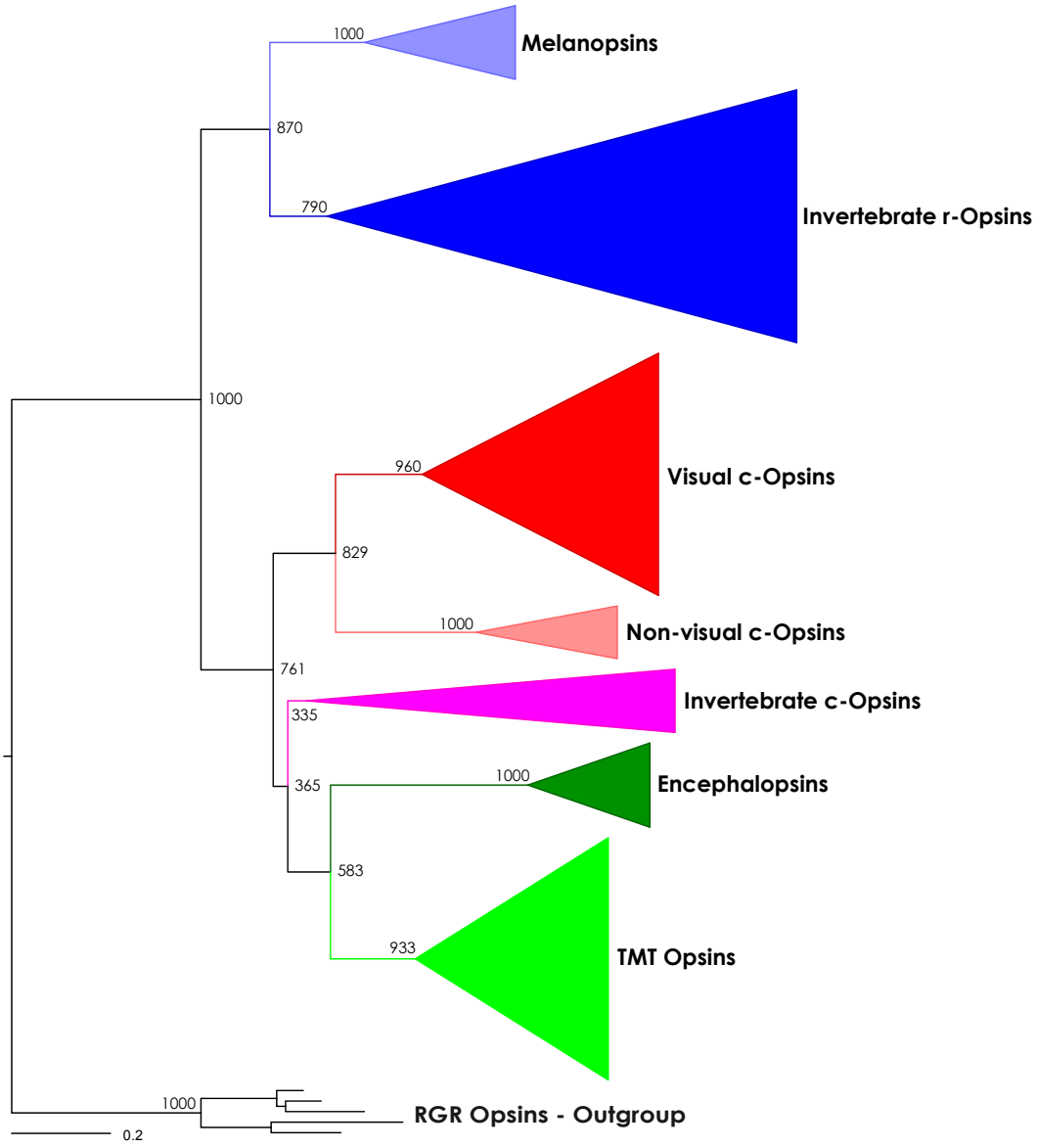


Figure 3.16: **Phylogeny of the TMT Opsins and Encephalopsins in relation to other Opsin families** A phylogenetic analysis using different rhabdomeric Opsins, Melanopsins and ciliary Opsins in addition to TMT Opsins and Encephalopsins was prepared using ClustalX2 [75]. The TMT Opsins and Encephalopsins cluster together with ciliary Opsins and outside of rhabdomeric Opsins. RGR Opsins were used as an outgroup.

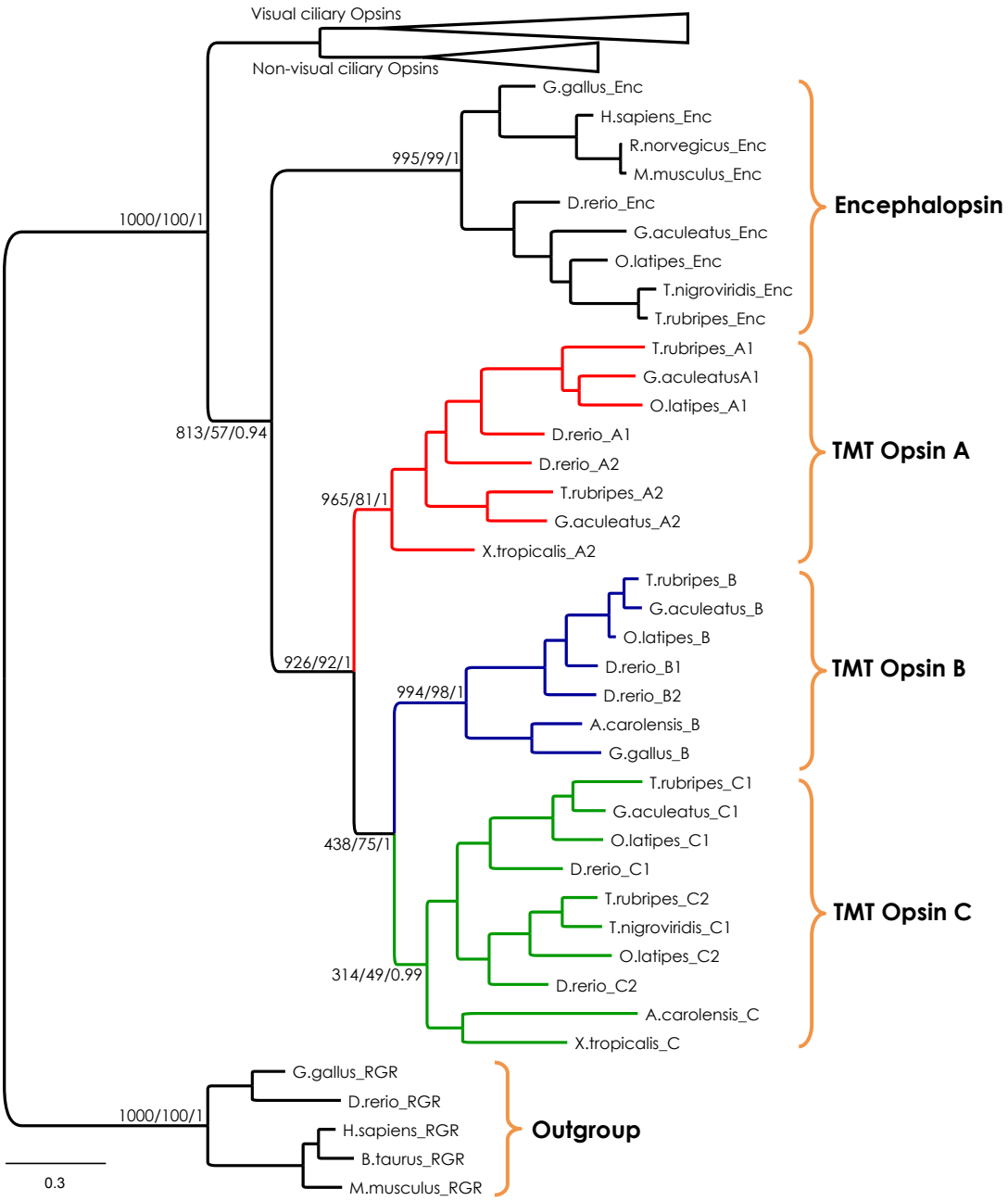


Figure 3.17: **Phylogeny of the TMT Opsins and Encephalopsins in relation to visual and non-visual vertebrate ciliary Opsins** This phylogenetic analysis was prepared to compare the relation between the different TMT Opsin and Encephalopsin homologues to each other. The tree was calculated using three different methods: neighbour-joining, Maximum likelihood and Baysian phylogenetic inference. The support values for all important nodes are indicated and the topology of the tree is according to the baysian phylogenetic inference. It has to be stressed that not all homologues were included since not for all I could predict a complete seven transmembrane domain. The tree clearly shows that Encephalopsin homologues cluster outside of the TMT Opsins. Three different TMT Opsin families were identified that contain homologues from different groups. Again, RGR Opsins were used as an outgroup.

		Not used in tree	Incomplete sequence (7TM)	Encephalopsin Blue clade	TMT-Opsin A Red clade	TMT-Opsin B Yellow clade	TMT-Opsin B Green clade
Anolis carolinensis							
1	Scaffold 66	x	x		x		
2	Scaffold 193					x	
3	Scaffold 586	x	x	x			
4	Scaffold 772						x
Gallus gallus							
1	Chrom. 1					x	
2	Chrom. 3			x			
Chimpanzee							
1	Chrom. 1			x			
Oryzias latipes							
1	Chrom. 10						x
2	Chrom. 14						x
3	Chrom. 15			x			
4	Chrom. 20				x		
5	Chrom. 21					x	
6	Scaffold 906	x	x		x		
Rattus norvegicus							
1	Chrom. 13			x			
Gasterosteus aculeatus							
1	Group III				x		
2	Group VI			x			
3	Group VII						x
5	Group XVI					x	
6	Group XXI				x		
Takifugu rubripes							
1	Scaffold 42					x	
2	Scaffold 52			x			
3	Scaffold 71						x
4	Scaffold 132						x
5	Scaffold204				x		
6	Scaffold 263				x		
Tetraodon nigroviridis							
1	Chrom. 1						x
2	Chrom. 2	x	x				x
3	UnRandom			x			
Xenopus tropicalis							
1	Scaffold 31						x
2	Scaffold 55				x		
3	Scaffold 162	x	x	x			
4	Scaffold 307	x	x				x
Taeniopygia guttata							
1	Chrom. 1						x
2	Chrom. 9	x	x				
3	UnRandom	x	x				
Danio rerio							
1	Chrom. 2				x		
2	Chrom. 6					x	
3	Chrom. 9					x	
4	Chrom. 10						x
5	Chrom. 13			x			
6	Chrom. 14						x
7	Chrom. 24				x		
Homo sapiens							
1	Chrom. 1			x			
Mus Musculus							
1	Chrom. 1			x			

Figure 3.18: **Summary of the identified TMT Opsin and Encephalopsin family** This table shows the animals that were used for the initial blast search for novel TMT Opsin and Encephalopsin family members. It also indicates on which chromosome, scaffold or group the genes are encoded and which have been predicted properly and thus also have been used for the phylogenetic analysis. Furthermore, the affiliation of the specific Opsins to the Encephalopsin or the different TMT Opsin families is indicated and species-specific duplications are highlighted.

Furthermore, Encephalopsin seems to be highly conserved in vertebrate species. This is deduced from the automatic annotation of the genomes accessible via the EnsEMBL genome browser [57]. In almost every vertebrate genome that can be accessed online, one Encephalopsin homologue has been annotated (except for *Sus scrofa* which could also be explained by incomplete genome coverage). This conservation of gene number despite whole-genome duplications that occurred during vertebrate evolution suggests that the gene number might also be under selective pressure. However, in order to test this, the function of the Encephalopsin homologues have to be unravelled. Despite a not well supported role in asthma susceptibility [122], no function has ever been suggested for Encephalopsins.

3.3.2 Analysis of the novel Opsin sequences

In order to verify that the identified genes represent real Opsin genes I analysed specific amino acid residues that are known to be crucial for Opsin function and therefore characterise Opsin proteins. This was done by analysing a multiple sequence alignment with all TMT Opsin homologues and other important Opsin families. All compared amino acids are highlighted in the scheme of the bovine Rhodopsin structure that was published in 2000 [90] (Figure 3.19). All numbered sequence positions mentioned later on refer to the corresponding position in bovine Rhodopsin.

In Figure 3.20 the main conserved sequence characteristics are compared. The overall seven transmembrane structure is conserved as for all GPCRs. All of the novel identified Opsins contain a highly conserved Lysine residue that is needed to covalently attach the light-absorbent retinal group to the Opsin protein (see figure 3.21). Furthermore, the transducin binding domain as well as the cystein residues that form a cystein bridge are highly conserved. The transducin binding domain is usually ERY for the TMT Opsins and Encephalopsins. However, all homologues that cluster within the TMT Opsin B group have the sequence DRY. This might not only have functional implications, but also indicates that the phylogenetic clustering obtained with the TMT Opsin B family is correct, since all homologues that cluster in this group have a DRY at this position while all others an ERY. Furthermore, I tested the Exon-Exon boundaries of the homologues I identified in Zebrafish and Medaka. For all homologues, the three exon-exon boundaries are at the same position within the gene. Interestingly, these intron positions are also conserved in other visual and non-visual ciliary Opsins [86]. An additional second intron is present in vertebrate ancient Opsins, pinopsins as well as the visual Opsins. However, the position of this second exon is not conserved. Other Opsin families have a completely different intron-exon structure. Therefore, the genomic structure of the TMT Opsins and Encephalopsins also indicate that these genes are closely related to ciliary Opsins.

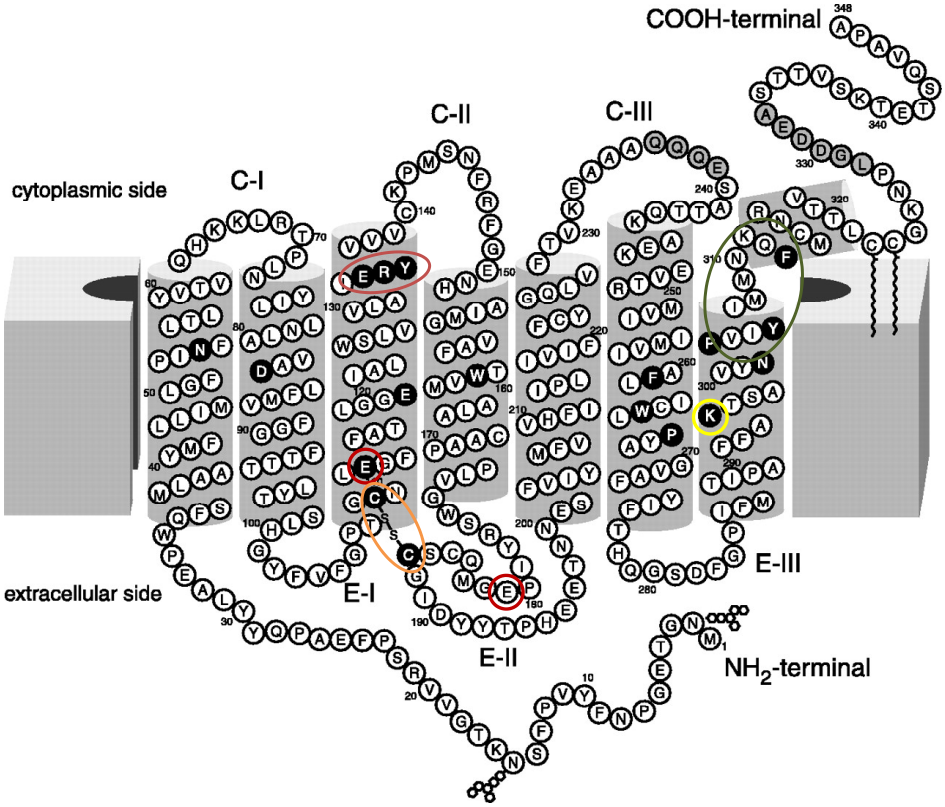


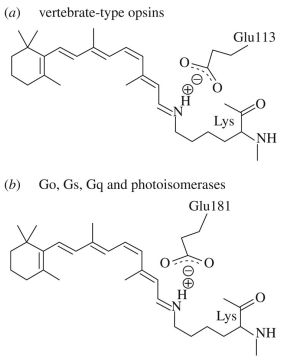
Figure 3.19: **Schematic representation of the structure of bovine Rhodopsin** In 2000, the structure of the bovine Rhodopsin homologue has been elucidated [90]. This schematic representation of the structure is taken from [90] and important amino acid residues were highlighted. In dark red, the highly conserved transducin binding domain is highlighted. In red, the two residues that act as a counterion in diverse Opsin molecules is labelled. The orange ellipse marks the highly conserved cystein residues that form a cystein bridge and in yellow, the retinal-binding lysine residue is marked. Finally, the sequence within the green ellipse represents a highly conserved C-terminal stretch that is thought to be crucial for G protein binding and also serves as a signature sequence that distinguishes ciliary Opsins from other Opsin families. [4]

Sequence feature	Encephalopsin	TMT-Opsin A	TMT-Opsin B	TMT-Opsin C
7TM domain	conserved	conserved	conserved	conserved
Disulfide bond	conserved	conserved	conserved	conserved
Schiff base	K	K	K	K
Transducin binding domain	ERY	ERY	DRY	ERY

Figure 3.20: **Comparison of potentially functional amino acid residues in the novel identified Opsin families** This table shows the comparison of highly important structural features and amino acid residues within the Encephalopsin and TMT Opsin family. All members highly resemble the light-sensitive bovine Rhodopsin and can therefore potentially act as light-absorbent molecules as well.

At position 113 in the bovine Rhodopsin the amino acid Glutamate (E) serves as a so-called counterion. This counterion is needed to balance a positive charge that is present when 11-cis retinal binds covalently to the highly conserved lysine residue at position 296. Usually, the retinal group absorbs UV light. However, in the presence of a counterion the wavelength of the absorbed light shifts to a visible range [105, 107]. In vertebrate ciliary Opsins, the counterion is at the position 113. The corresponding amino acid in invertebrate Opsins is neutral and therefore cannot act as a counterion. It was discovered, that the amino acid at the position 181, which is a negatively charged Glutamate in all rhabdomeric Opsins in invertebrates acts as a counterion [105, 107]. It has been proposed that the ancestor of the rhabdomeric and ciliary also used the counterion at the position 181 and vertebrate ciliary Opsins evolved in a way that the counterion is at the position 113. This is supported by the fact that cephalopodian retinochrome, a Rhodopsin-like protein, uses the glutamate at position 181 as a counterion [105, 107, 99]. I compared the amino acids at the corresponding positions in the novel identified Opsin proteins. The results and a structural scheme of the counterion function in vertebrate and invertebrate Opsins are shown in Figure 3.21.

Encephalopsins have a negatively charged amino acid at the position 113. The amino acid is not glutamate but rather an aspartate. This change in amino acid likely does not have an effect since both amino acids are quite similar and aspartate can also function as a counterion [99]. Furthermore, this charged amino acid residue lies within a predicted transmembrane region. It is likely that the negative charge has to be counterbalanced, which suggests that this residue likely acts as a counterion. As can be seen in the table, TMT Opsins have the same amino acids at the position 113 and 181 as the invertebrate rhabdomeric as well as ciliary Opsins. This suggests that TMT Opsins have different structural features than other so far identified vertebrate ciliary Opsins. Based on the occurence of the same counterion setup in vertebrate TMT Opsins, it is likely that the counterion setup identifed in protostomian rhabdomeric and ciliary Opsins represents



Opsin group	Counterion 113	Counterion 181
Melanopsin	Y	E
Insect R-Opsin	Y / F	E
Lophotrochozoan R-Opsin	Y	E
Peropsin	Y	D
RGR-Opsin	H	E
Insect c-Opsin	Y	E
Visual Opsins	E	E / H
Pineal Opsins	E	S
VA Opsin	E	S
Lophotrochozoan C-Opsin	Y	E
TMT-Opsin	Y	E
Encephalopsin	D	D (Mammals), E
Sea urchin/Amphioxus*	Y	E

Figure 3.21: **Potential counterions in diverse Opsin families** A counterion is a negatively charged amino acid residue that interacts and balances a positive charge that emerges after the covalent binding of the retinal group to a highly conserved lysine residue (shown on the left, picture taken fron [99]). The counterion in vertebrate ciliary Opsins (visual and nonvisual) usually is a negatively charged amino acid at the position 113. However, in other families, like invertebrate Opsins, Melanopsins, Peropsins, RGR Opsins as well as TMT Opsins, the amino acid at the corresponding position 113 is not negatively charged but neutral. However, for some of them, it has been shown that they employ a highly conserved glutamate at position 181 as a counterion [105, 107]. The table on the left illustrates the amino acid residues at the respective positions 113 and 181 for diverse Opsin families. The asterisk indicates that most but not all Amphioxus Opsin homologues share the indicated counterion setup.

the ancestral setup. Other vertebrate ciliary Opsins evolved in a way that the counterion is shifted to the position 113. In summary, divergent Opsin families like the vertebrate TMT Opsin and the rhabdomeric Opsins share this specific regulatory feature while it has been changed only in other vertebrate ciliary Opsins. Thus, it is most likely that the ancestral Opsin employed a counterion at position 181 and that many vertebrate ciliary Opsin families shifted the counterion to the position 113 [99].

A C-terminal stretch close to the seventh transmembrane domain important for the binding of G-proteins exists. This stretch has also been implicated in representing a signature sequence that distinguishes ciliary from rhabdomeric Opsins [5]. I compared the corresponding sequence stretches between the different Opsin families using a sequence logo. As can be seen in figure 3.22, this C-terminal sequence stretch is highly conserved between the vertebrate visual and non-visual Opsins as well as TMT Opsins. This suggests that despite the invertebrate-type counterion setup shown above the TMT Opsins contain that ciliary-type C-terminal stretch. However, the corresponding C-terminal stretch of Encephalopsins differs considerably from the stretch in other vertebrate ciliary Opsins. It also does not resemble any other Opsin group. This means that the C-terminal stretch does not represent a clear signature sequence, since Encephalopsins are ciliary Opsins based on the phylogenetic clustering and the gene structure. Furthermore, it suggests that differences in the binding of a G-protein exist between Encephalopsins and other vertebrate ciliary Opsins. Taken together, based on the phylogenetic clustering, the differences

in the counterion setup and the differences between the C-terminal stretches of TMT Opsins and Encephalopsins, I conclude that these rather represent two different Opsin families. Thus, TMT Opsins are not homologous to Encephalopsins but rather represent two distinct Opsin families (as suggested by Moutsaki et al., [86]). Furthermore, since no TMT Opsin homologue has been found in the genome of eutherians but in all other vertebrate groups, I deduce that the TMT Opsin family likely was lost in eutherians.

3.3.3 Opsin expression in two teleost species, Medaka and Zebrafish

In the next step, I wanted to assess the gene expression to determine if the novel members of the TMT Opsin family are expressed in diverse tissues and organs as suggested earlier [86]. If yes, this could mean that the authors of this study rather assessed the expression of several TMT Opsin homologues at once instead of only the one they identified. Since the authors used a cross-species RNA protection assay to assess the expression pattern, similar Opsin proteins might also have been detected. Furthermore, for one TMT Opsin homologue it has been shown that it is discretely expressed in the hypothalamus of Zebrafish. It would be interesting to know if the identified novel members of this family are always expressed sites from where neuronal hormones are secreted.

I cloned fragments of all TMT Opsin and Encephalopsin homologues I found in Medaka and Zebrafish. This also means that all identified members of this family are in fact expressed since the mRNA must have been present. Therefore they do not represent pseudogenes that emerged from a recent duplication and currently degenerate via genetic drift. I prepared antisense DIG-labelled cRNA probes to test where these Opsins are expressed via whole-mount in situ hybridisation. However, not for all cloned Opsins a clear expression pattern could be identified. This suggests that either the genes are expressed at a very low level or unknown technical difficulties hindered the whole-mount *in situ* hybridisation procedure for some unknown reason. Interestingly, I almost exclusively found expression in distinct and confined areas within the central nervous system for all of the novel members from the TMT Opsin family. Only the zebrafish TMT Opsin B2 homologue seems to be expressed within the swim bladder in two stripes.

I never found any expression in other tissues and organs. It is also apparent that paralogues that cluster within one group also show different expression patterns. This suggests that extensive sub- and neofunctionalisation occurred after the duplication event. Thus, the closely related homologues likely exert different functions or similar functions in different cells of the brain. This of course contradicts earlier results [86]. As mentioned above, it would be possible that the authors rather detected several homologues instead of only the one they were looking at. However, since no expression in any other organ for any TMT Opsin homologue was detected, it could also be that the TMT Opsin expressing cells form later in development, sometime after seven days post fertilization for Medaka and five days post fertilization for Zebrafish. In fact, it

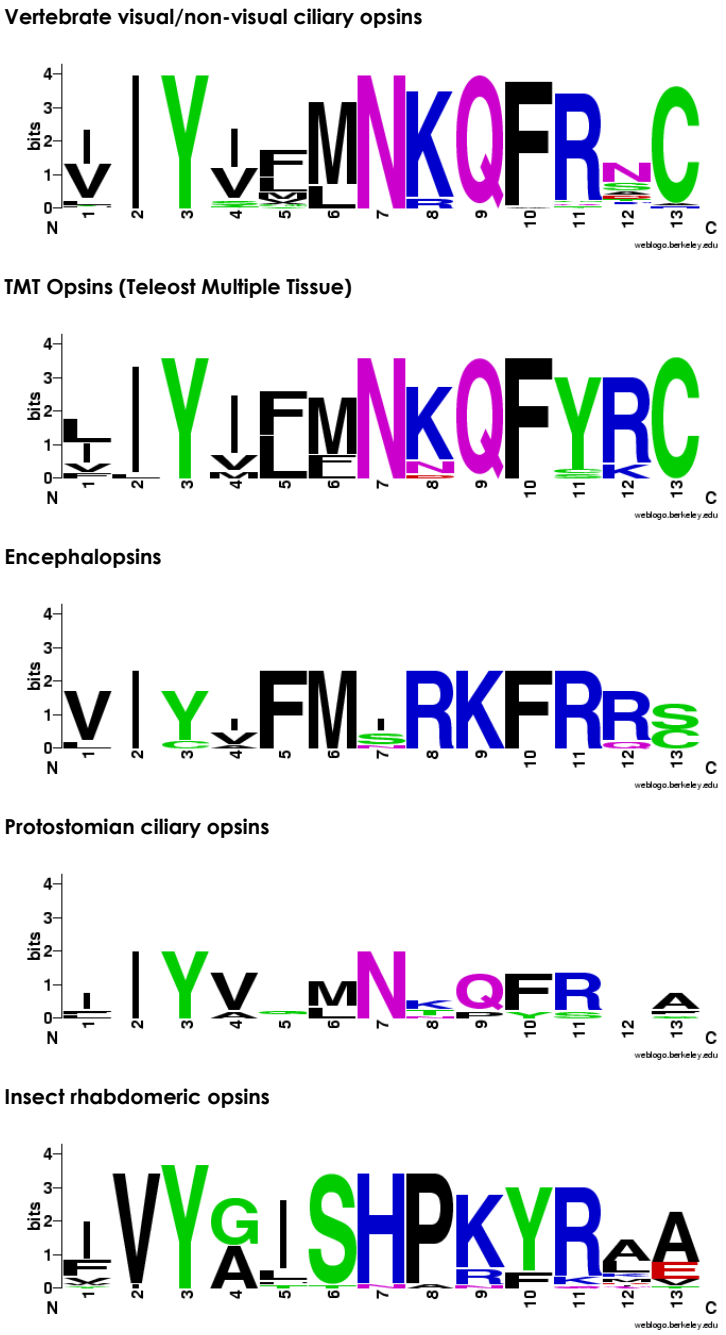


Figure 3.22: **Sequence logos of a C-terminal stretch important for G protein binding** Here, a comparison between the conserved sequence stretch at the C-terminos from several Opsin homologues is shown. The sequences were aligned and the stretch within the single groups was analysed using the tool WebLogo. The TMT Opsin family strongly resembles the other vertebrate visual and non-visual ciliary Opsins. However, the Encephalopsin family does not resemble any other family. Thus, differences in the downstream signalling are expected since this region is important for G protein binding. Furthermore, this does not represent a signature sequence [5] to determine if an Opsin molecule belongs to the ciliary Opsins, since the Encephalopsins represent clear ciliary Opsins based on the phylogenetic clustering and the gene structure.

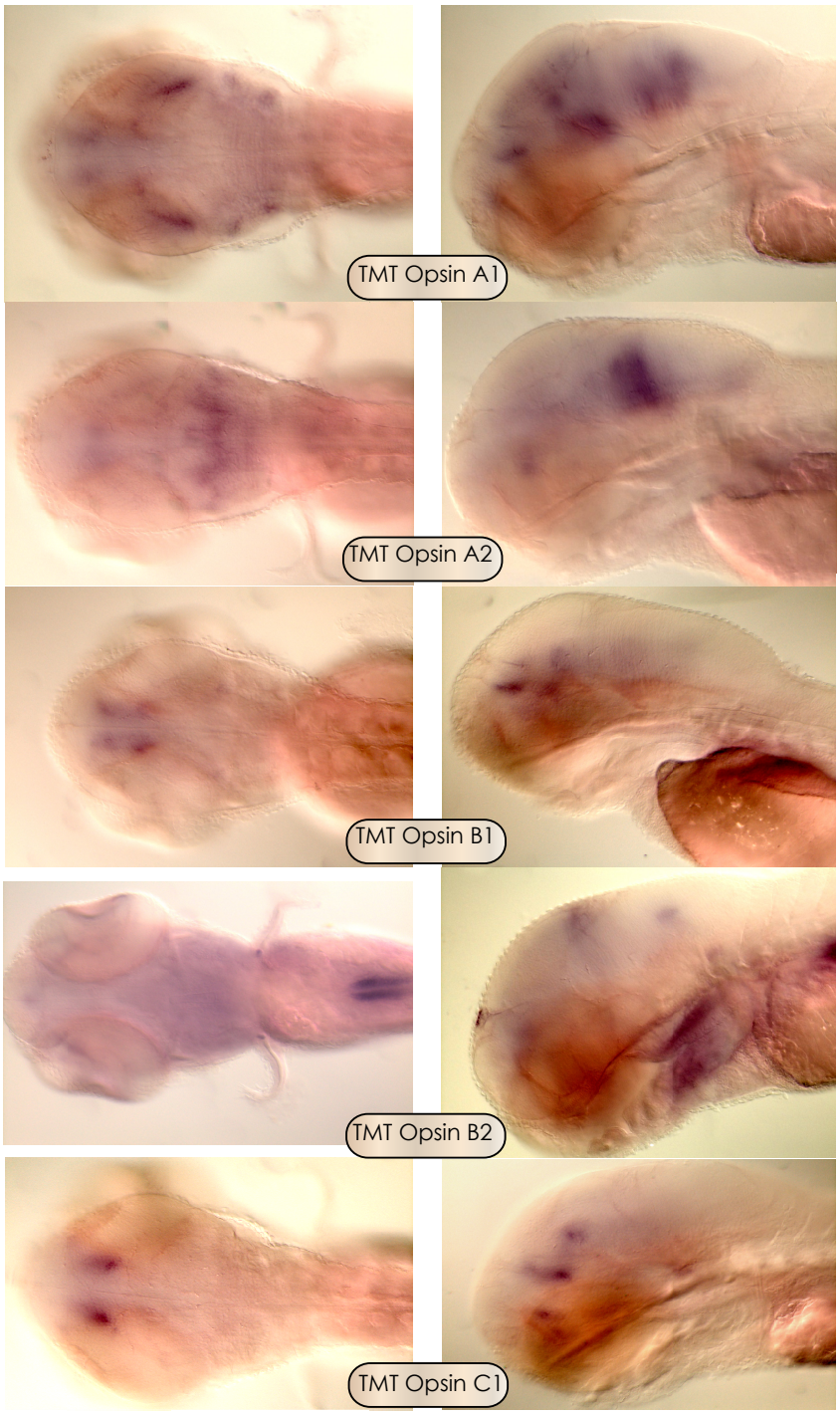


Figure 3.23: **Expression pattern of Zebrafish TMT Opsins** The expression of the indicated TMT Opsin homologues was analysed in 3 days old embryos. Deep brain photoreceptor cells could be identified in all regions of the brain. Interestingly, the TMT Opsin homologues are all expressed in distinct confined regions. Interestingly, even closely related paralogues show very divergent expression domains. Unfortunately, not for all identified Opsin genes an expression pattern was identified. The left column shows the lateral view with anterior to the left and dorsal to the top. The right column shows a dorsal view with anterior to the left.

seems that the number of expression domains increase with age in Zebrafish and Medaka. This is shown for the expression of Medaka homologues in Figure 3.24 and we found a similar increase in expression domains also in developing Zebrafish (data not shown). In Medaka, most domains that are present in five days old embryos are also present in seven days old embryos, however additional domains appear.

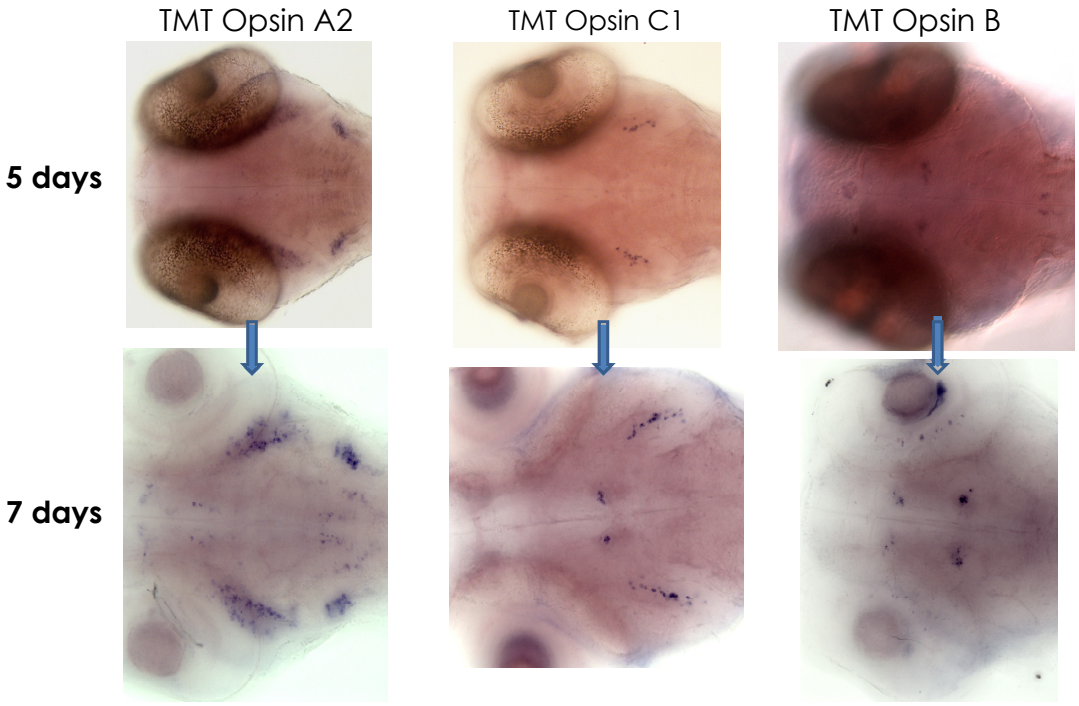


Figure 3.24: **Expression pattern of Medaka TMT Opsins** The upper row shows the expression in five days old embryos and the lower row shows the expression on seven days old embryos. The pictures in the lower row represent minimal intensity projections of a Z-Stack to show all domains at once. The pictures of the upper row show one focal plane. However, all cells that were present are visible. The top right picture was taken by Thomas Hoffmann. It illustrates the highly conserved pineal expression domain (compare to Figure 3.23. Note, that after seven days of development several new expression domains are visible. The fishes are shown with a dorsal view with anterior to the left.

Interestingly, only few expression domains of homologous TMT Opsins seem to be conserved between different species. One example for that is the TMT Opsin B family. In this family, homologues from medaka, chicken and zebrafish exist among homologues of other species. The medaka homologue is specifically expressed in few cells within the amacrine layer of the eye. The cells seem to be enriched behind the lense, thus making light excitation despite the eye pigment possible. Furthermore, the expression of the homologous TMT Opsin in chicken can also be found in the amacrine layer of the eye (personal communication). In the zebrafish genome two homologues that cluster into the TMT Opsin B group were identified. One of these

homologues shows expression in the pineal gland. A similar pineal expression was also detected with the TMT Opsin B homologue in Medaka. However, the pineal expression in medaka seems to be regulated by unknown cues because the expression could not always be detected although other expression domains were always found. In the course of this thesis, it was not possible to elucidate the regulatory mechanisms for that. A pineal expression of a presumptive light-sensitive Opsin molecule is very interesting since the pineal gland is known to be light-receptive and also highly neurosecretory [118]. It is well known that cells from the pineal gland secrete melatonin in order to translate an outer circadian cycle into an internal melatonin-level rhythm [119, 129]. A study from 2007 shows that the zebrafish pineal gland expresses two ciliary Opsins, namely exo-Rhodopsin and red-cone Opsin. Furthermore, it was shown that the amount of light that is absorbed for each wavelength cannot be explained with the absorption of both Opsins. Thus, the authors suggested that other visual pigments are co-expressed in the pineal gland of Zebrafish [129]. The function of the specific Opsins in the pineal gland were not addressed. Since the Zebrafish TMT Opsin B2 homologue and the Medaka TMT Opsin B homolog are expressed in the pineal gland, it is possible that these represent the other identified Opsins that correspond to the pineal absorption spectrum. However, this still awaits verification.

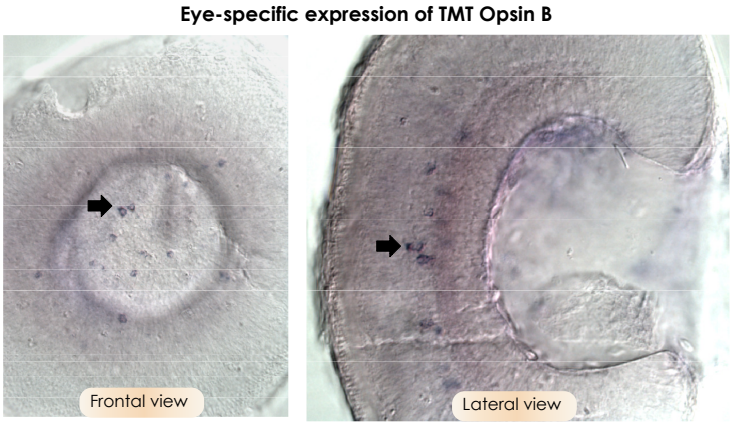


Figure 3.25: **The Medaka TMT Opsin B homologue is expressed in the eye** The expression of the TMT Opsin B homologue in the eyes is shown. The lense was removed to better visualize the staining. At the left, it can be seen that the TMT Opsin B-positive cells are restricted to the amacrine cell layer. The expression at this site is also conserved in the chicken (personal communication) with the homologous TMT Opsin. The picture on the left indicates that the cells are expressed behind the lense and thus should receive proper light input despite the eye pigment.

3.3.4 Conserved alternative splicing events

For subsequent functional analysis we cloned the full-length coding sequence of the novel Opsins I found in Medaka and Zebrafish using a combination of RACE PCR and nested RT-PCR. The full-length sequences of most Opsin homologues from Medaka were cloned by Ruth Fischer (All except A2 and C2, which so far could not be cloned).

Interestingly, the nested PCR reactions from Zebrafish spanning the whole coding sequence represented one single band on an agarose gel, whereas fragments with different sizes were cloned from Medaka cDNA. When analysing the sequences, we found only one isoform for all Zebrafish TMT Opsins. However, in Medaka alternative splicing has been detected in several of the Opsin genes.

Interestingly, for the human and mouse Encephalopsin homologues a similar relationship was found. Whereas in mice no alternative splicing was detected, six different splice isoforms for the human Encephalopsin have been detected [65]. The human Encephalopsin gene has at least six different exons. The sequence used in this study is the complete sequence that is encoded by the four exons 1, 2, 5 and 6. In alternatively spliced isoforms, two additional exons are present after the second exon. Four of these six isoforms contain a premature stop codon and therefore should be degraded via NMD [101, 80]. For the Medaka Encephalopsin homologue we discovered two isoforms. One of them corresponds to the predicted sequence, whereas the other one lacks the complete exon 2. This type of alternative splicing is called exon skipping. This leads to a frameshift and subsequently to a premature stop codon before the last exon-exon junction. This is usually an indication for the degradation of the mRNA via non-sense mediated mRNA decay (NMD) [101, 80].

For the Medaka TMT Opsin A1 and B1 homologues three different isoforms were discovered. In addition to the complete predicted protein we found two different splice isoforms for both genes. In both genes, a cryptic splice site in exon one is used to generate mRNA that contains a premature stop codon. Therefore, members of the TMT Opsin families A and B share the exact same cryptic splice site despite the phylogentic clustering that is backed up by specific differences in the transducin binding domain (as discussed above). This suggests that this alternative splicing isoform dates back to the common precursor of both TMT Opsin families and was retained during evolution. However, no alternative splicing event was identified in the TMT Opsin C family so far. This strongly suggests a functional role for these different isoforms. In addition to the alternative splicing described above, both homologues have additional cryptic splice sites in the first and second exon, respectively. Both isoforms again contain premature stop codons and therefore cannot produce functional proteins.

The function of any of these isoforms is not clear. Since the alternatively spliced transcripts with premature stop codons can be detected very efficiently via RT-PCR, they seem to be relatively abundant when compared to the complete transcript. Furthermore, the conservation of the

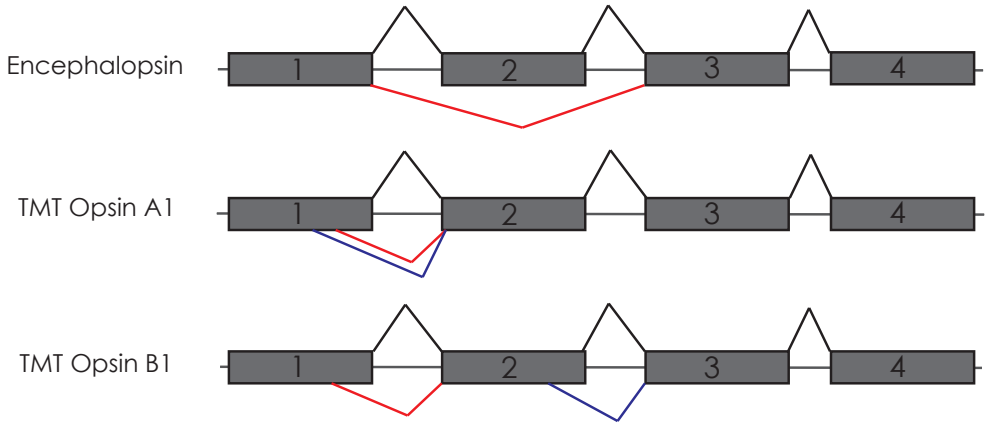


Figure 3.26: **Alternative splicing identified in Medaka TMT Opsin and Encephalopsin homologues** The gene structure identified for the two TMT Opsins A1 and B1 as well as for the Encephalopsin homologue is shown. While we could identify the correctly spliced isoform (black lines) for all of them we found additional alternative splicing events (red and blue lines). In the case of Encephalopsin, the exon 2 is skipped (Exon skipping) while the TMT Opsins have additional cryptic splice sites within the exons 1 or 2, respectively. No alternative splicing has been detected for Zebrafish TMT Opsin or Encephalopsin homologues.

alternative splicing sites in different TMT Opsin families suggest that an important functional component might be present. It could be that nonsense-mediated mRNA decay is used to regulate the Encephalopsin and TMT Opsin mRNA pools in the cell. However, a role of NMD in the regulation of gene expression is under debate [101, 80]. Another important factor to consider is that the alternatively spliced transcripts might not be present in all expression domains. As shown above, I identified that the Opsins are expressed in a several distinct domains within the developing brain. However, some domains might not express the complete protein based on the occurence of the alternatively spliced isoform. Additionally, we could not identify these splice isoforms in Zebrafish. This might be due to the fact that adult brain cDNA was used to clone the TMT Opsin and Encephalopsin homologues in Zebrafish and whole 22 days old embryos were used for Medaka. Thus, we cannot exclude that these splice isoforms are present during Zebrafish development, but we did not detect them in adult brain tissue.

3.3.5 Generation of reporter lines

An interesting evolutionary question opens up when analysing the different expression domains of the homologous members of the TMT Opsin family in teleosts. Although the genes derive from a common precursor molecule, their expression domains show that even paralogous genes are expressed in different domains. This means that the members of this family were subjected to extensive sub- and neofunctionalisation that followed gene duplication. Therefore, the upstream

regulatory sequences that guide a cell type-specific expression of the TMT Opsin homologues must have changed. We analysed this by trying to identify potential regulatory regions of some TMT Opsin groups via phylogenetic footprinting. In short, this is the comparison of homologous genomic DNA loci encoding for the gene of interest from diverse species. All non-coding yet still conserved sequence stretches likely represent regulatory regions. I analysed members of the TMT Opsin A and B family. Genomic regions of the five available teleost fish genomes were analysed using the program Mulan [76]. In figure 3.27, the results are depicted.

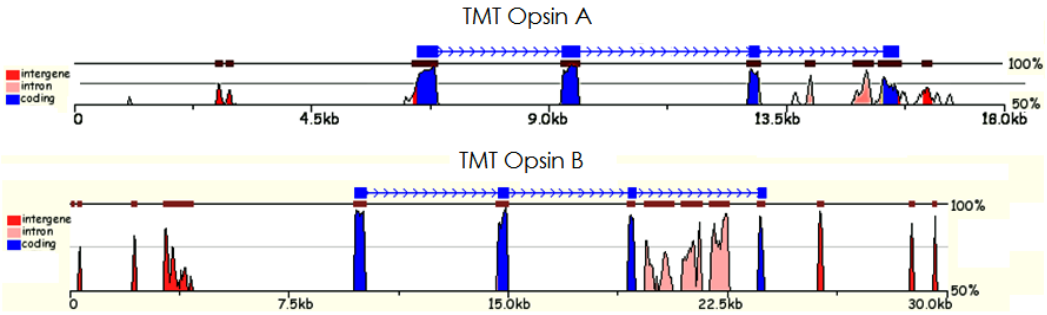


Figure 3.27: **Phylogenetic footprinting for two TMT Opsin families** Here, the results for two phylogentic footprinting analyses are shown. The Fugu genomic region was used as a template and it was compared to the corresponding genomic regions of four different teleost species. The genomic regions that were used were chosen according to the phylogenetic clustering identified with the TMT Opsin homologues. Blue peaks indicate the highly conserved exons. Pink and red peaks indicate non-coding conserved regions. However, the peaks in the third intron and 3' of the TMT Opsin homologue B rather represent another gene that was not annotated. In the analysis of the TMT Opsin A family, these peaks are also present, yet a lower degree of conservation is identified. Also, the gene is not known to be expressed from this locus based on EnsEMBL annotations and I therefore conclude that these represent remnants of the gene that changed via genetic drift. Together with the phylogenetic clustering of TMT Opsins, I conclude that the ancestral TMT Opsin gene was intertwined with the other gene before the paralogues emerged after several duplication events.

I could identify several peaks upstream of the TMT Opsin homologues of the A1 and B family. In the same plots, conserved elements are also identified in the third exon and 3' of the gene. When blasting them against the NCBI non-redundant protein database, I found that the conserved sequences represent the protein *α-2,6-sialyltransferase II*. Interestingly, the *α-2,6-sialyltransferase II* gene is intertwined with the TMT Opsin B orthologue. However, similar peaks are also identified between the TMT Opsin A1 family. Not only are the peaks lower, indicating less conservation across species but also no other *α-2,6-sialyltransferase II* paralogue has been annotated at this locus [57]. However, this does not exclude the conserved remnants from being regulatory regions important to guide TMT Opsin expression.

This suggests that the common ancestral molecule that gave rise to the TMT Opsin A, B and C

family already was already linked to *α-2,6-sialyltransferase II*. Several copies of that gene were not needed and the genes were lost via genetic drift. This must have been a relatively recent event since the remnants of this gene are still detectable in the TMT Opsin family A. I did not prepare phylogenetic footprinting for the TMT Opsin C family, but based on the phylogenetic analysis it seems that the B and C families are more related to each other than to the A family. Therefore, the B and C families likely shared a common ancestral molecule with the TMT Opsin A family.

In order to characterise the identified potential regulatory regions I cloned the complete third intron from the Medaka TMT Opsin A1 homologue as well as 6.5 kilobases upstream of the TMT Opsin homologues in front of a fluorescent reporter gene. I included all upstream sequences until the endogenous start codon. In this way I cloned all identified non-coding conserved regions as well as the highly conserved third intron into a reporter vector. This vector was subsequently used to generate transgenic Zebrafish and Medaka in order to label specifically the cell types that express the TMT Opsin A1 homolog. The resulting reporter lines should then be used to functionally adress the TMT Opsin expressing cell type. For example, it is possible to isolate the reporter gene expressing cells using laser microdissection or fluorescence activated cell sorting (FACS). Additionally, one important step would be the characterisation of the cell type morphology. As discussed above, ciliary photoreceptor cells contain a cilium with a large number of membrane stacks to increase their own surface and therefore the light sensitivity. This is because more Opsins can act within a larger membrane surface. If the specific TMT Opsin-expressing cells are labelled with GFP, then an immonugold-labelled antibody against the GFP transgene can be used to identify the cells using electron microscopy. In this way, the morphology of the cells can be analysed and might also provide hints for a potential light-receptive capacity of the TMT Opsin proteins.

3.3.6 Outlook

Deep-brain photoreceptors are often implicated to guide photoperiodic time measurement. Thereby the animals can discriminate between differences in daylength and therefore seasons. In order to identify if these Opsins are important for this process, fishes could be raised under long- and short-day conditions (usually 16 hours light, 8 hours dark and vice versa). The expression levels in specific expression domains could be assessed via *in situ* hybridisation in fishes kept under the different light regimes.

The identification of several new members of the TMT Opsin family and their distinct and confined expression domains within the brain provide us with the problem that a functional characterisation will be very difficult. At first, the hypothesis that the TMT Opsin A2 homologue that is co-expressed with the neuronal hormone Vasotocin could regulate Vasotocin secretion is still valid. However, it is difficult to hypothesise about similar functions of the other homologues, since they are expressed in various regions within the brain, some of which are not known to be

neurosecretory.

There are three different Rx transcription factors encoded in the teleost genome. In Medaka, the Rx3 homologue is expressed in the eye and also in the hypothalamus and neurohypophysis [22]. The other Rx homologues are only expressed within the eye during development ([53], MEPD). It has been shown that Rx genes are important for the specification of ciliary Opsin photoreceptor cells in the eye and a hypothesis exists that Rx genes are crucial for the specification of ciliary photoreceptor cells (reviewed in [3]). However, the dispersed confined expression pattern of the TMT Opsin genes within the brain suggest that this might not always be true. Based on the expression pattern and the only partial overlap with Rx genes suggest that the TMT Opsin-expressing cells might be regulated differently than other ciliary photoreceptor cells. Another possibility would be that the cells acquire their cell identity via Rx genes before but later on move to their respective target sites. An interesting experiment would be to test with whole mount *in situ* hybridisation the expression of TMT Opsins in Rx3 deficient animals. Either natural occurring mutants can be used for that or the respective Rx genes can be knocked down using a morpholino-based approach.

As suggested above, the alternative splicing that we detected in Medaka might have a potential regulatory role. It would be interesting to test if the splice isoforms that contain a premature stop codon are localised to different regions within the brain. Thereby, it might be possible to characterise domains where the correct isoform is expressed and thus also the correct protein is produced. This could be done by whole-mount in situ hybridisation using small antisense LNA probes. These LNA probes should be used against the alternative exon-exon junctions. LNAs are characterised by a very high melting temperature that specifically drops when a mismatch is present. Thus, the size of the LNA can be very small and it can still be effectively used to selectively label the cells that express either isoform. Furthermore, the alternative splicing might also regulate the amount of functional mRNA within a cell. Therefore, it could also be tested if the levels of both transcripts change with outer cues. For example, a circadian or photoperiodic regulation of TMT Opsin alternative splicing is imaginable. This could be tested by using quantitative RT-PCR specifically detecting either Opsin isoform. A method for that would be to detect the specific isoforms with Taqman probes that span the respective exon-exon junctions. However, when doing so, the different expression domains within the brain have to be taken into account.

4 Conclusion

In conclusion, I analysed various aspects of a highly conserved system of sensory-neurosecretory cells in three different species, namely *Platynereis dumerilii*, Medaka and Zebrafish.

In *Platynereis dumerilii*, I cloned the so far unknown homologues of the Vasotocin receptors. In addition, I analysed the expression of *vasotocin* in respect to the lunar rhythm. It seems that the expression of *vasotocin* does not correlate with a lunar cycle. It likely plays a role in maturation since the hormone is upregulated in maturing males. This suggests that the *vasotocin* gene is not directly connected to the lunar clock but rather to maturation, which on the other hand is governed by a lunar clock.

In vertebrates the sensory and neurosecretory aspects of this system was analysed independently from each other. In the first part, we analysed the evolutionary history of the Nonapeptide gene family in vertebrates where we found that likely independent duplications gave rise to the Nonaptide genes in today's vertebrates. Furthermore, we deduced a common evolutionary origin of the GnRH and the Nonapeptide system based on functional and phylogenetic comparisons. We conclude that both systems evolved from an ancient pre-bilaterian broadly chemosensory neurosecretory system that regulated different aspects of reproduction. In addition to the theoretical part, we analysed cross-species reporter lines in Medaka and Zebrafish that use Fugu enhancer and promotor elements. We conclude that the expression of the *vasotocin* gene in Zebrafish is differentially regulated since the reporter gene fails to be expressed in Zebrafish but works at least partly in Medaka and also in mice [36].

In the final project, I identified several members of the TMT Opsin and Encephalopsin deep brain photoreceptor families in vertebrates. After intensive sequence analysis I conclude that both are indeed seperate Opsin families that have distinct structural features that clearly distinguishes them from other so far identified ciliary Opsins. However, both families are highly conserved in vertebrates except the TMT Opsin family was lost in eutherian species. I analysed the expression of several of these novel Opsin homologues in Medaka and Zebrafish and found that the genes are expressed in confined domains in all brain regions. The expression was found in regions that are known to be neurosecretory, like the hypothalamus [109] or the pineal gland. However, the genes are also expressed in other domains that so far are not known to be neurosecretory. Based on the assumption that the identified genes encode for light-responsive proteins, many more deep brain photoreceptor regions within the teleost brain than previously assumed were discovered.

Bibliography

[1] R. Acher and J. Chauvet. The neurohypophysial endocrine regulatory cascade: precursors, mediators, receptors, and effectors. *Front Neuroendocrinol*, 16(3):237–289, Jul 1995.

[2] E. Y. Adashi and A. J. Hsueh. Direct inhibition of testicular androgen biosynthesis revealing antigonadal activity of neurohypophysial hormones. *Nature*, 293(5834):650–652, Oct 1981.

[3] Detlev Arendt. Evolution of eyes and photoreceptor cell types. *Int J Dev Biol*, 47(7-8): 563–571, 2003.

[4] Detlev Arendt. The evolution of cell types in animals: emerging principles from molecular studies. *Nat Rev Genet*, 9(11):868–882, Nov 2008.

[5] Detlev Arendt, Kristin Tessmar-Raible, Heidi Snyman, Adriaan W Dorresteijn, and Joachim Wittbrodt. Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. *Science*, 306(5697):869–871, Oct 2004.

[6] R. J. Balment, M. J. Brimble, M. L. Forsling, and C. T. Musabayane. Natriuretic response of the rat to plasma concentrations of arginine vasopressin within the physiological range. *J Physiol*, 352:517–526, Jul 1984.

[7] R. J. Balment, W. Lu, E. Weybourne, and J. M. Warne. Arginine vasotocin a key hormone in fish physiology and behaviour: a review with insights from mammalian models. *Gen Comp Endocrinol*, 147(1):9–16, May 2006.

[8] Thomas Bartolomaeus. Structure, function and development of segmental organs in annelida. *Hydrobiologia*, 402:21–37, 1999.

[9] Thomas Baumgartner, Markus Heinrichs, Aline Vonlanthen, Urs Fischbacher, and Ernst Fehr. Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron*, 58(4):639–650, May 2008.

[10] S. Blackshaw and S. H. Snyder. Encephalopsin: a novel mammalian extraretinal opsin discretely localized in the brain. *J Neurosci*, 19(10):3681–3690, May 1999.

[11] Ulrich Boehm, Zhihua Zou, and Linda B Buck. Feedback loops link odor and pheromone signaling with reproduction. *Cell*, 123(4):683–695, Nov 2005.

[12] Jan Born, Tanja Lange, Werner Kern, Gerard P McGregor, Ulrich Bickel, and Horst L Fehm. Sniffing neuropeptides: a transnasal approach to the human brain. *Nat Neurosci*, 5(6): 514–516, Jun 2002.

[13] S. K. Boyd. Brain vasotocin pathways and the control of sexual behaviors in the bullfrog. *Brain Res Bull*, 44(4):345–350, 1997.

[14] W. J. Burgess, R. J. Balment, and J. S. Beck. Effects of luminal vasopressin on intracellular calcium in microperfused rat medullary thick ascending limb. *Ren Physiol Biochem*, 17(1): 1–9, 1994.

[15] R. D. Burke, L. M. Angerer, M. R. Elphick, G. W. Humphrey, S. Yaguchi, T. Kiyama, S. Liang, X. Mu, C. Agca, W. H. Klein, B. P. Brandhorst, M. Rowe, K. Wilson, A. M. Churcher, J. S. Taylor, N. Chen, G. Murray, D. Wang, D. Mellott, R. Olinski, F. Hallböök, and M. C. Thorndyke. A genomic view of the sea urchin nervous system. *Dev Biol*, 300(1):434–460, Dec 2006.

[16] H. K. Caldwell and W. S. Young III. *Oxytocin and Vasopressin: Genetics and Behavioral Implications*. Springer US, 2006.

[17] Anna Cariboni, Roberto Maggi, and John G Parnavelas. From nose to fertility: the long migratory journey of gonadotropin-releasing hormone neurons. *Trends Neurosci*, 30(12): 638–644, Dec 2007.

[18] Hyun Ju Cho, Sujata Acharjee, Mi Jin Moon, Da Young Oh, Hubert Vaudry, Hyuk Bang Kwon, and Jae Young Seong. Molecular evolution of neuropeptide receptors with regard to maintaining high affinity to their authentic ligands. *Gen Comp Endocrinol*, 153(1-3):98–107, 2007.

[19] Carlo Di Cristo, Marina Paolucci, Josè Iglesias, Javier Sanchez, and Anna Di Cosmo. Presence of two neuropeptides in the fusiform ganglion and reproductive ducts of octopus vulgaris: Fmrfamide and gonadotropin-releasing hormone (gnrh). *J Exp Zool*, 292(3):267–276, Feb 2002.

[20] Gavin E Crooks, Gary Hon, John-Marc Chandonia, and Steven E Brenner. Weblogo: a sequence logo generator. *Genome Res*, 14(6):1188–1190, Jun 2004.

[21] F. M. de Bree and J. P. Burbach. Structure-function relationships of the vasopressin prohormone domains. *Cell Mol Neurobiol*, 18(2):173–191, Apr 1998.

[22] K. Deschet, F. Bourrat, F. Ristoratore, D. Chourrout, and J. S. Joly. Expression of the medaka (oryzias latipes) ol-rx3 paired-like gene in two diencephalic derivatives, the eye and the hypothalamus. *Mech Dev*, 83(1-2):179–182, May 1999.

[23] Carole Deyts, Didier Casane, Philippe Vernier, Franck Bourrat, and Jean-Stéphane Joly. Morphological and gene expression similarities suggest that the ascidian neural gland may be osmoregulatory and homologous to vertebrate peri-ventricular organs. *Eur J Neurosci*, 24(8):2299–2308, Oct 2006.

[24] Gregor Domes, Markus Heinrichs, Andre Michel, Christoph Berger, and Sabine C Herpertz. Oxytocin improves "mind-reading" in humans. *Biol Psychiatry*, 61(6):731–733, Mar 2007.

[25] Zoe R Donaldson and Larry J Young. Oxytocin, vasopressin, and the neurogenetics of sociality. *Science*, 322(5903):900–904, Nov 2008.

[26] V. du Vigneaud. *A trail of research*. Cornell University Press, 1952.

[27] Jennifer L Eaton, Bo Holmqvist, and Eric Glasgow. Ontogeny of vasotocin-expressing cells in zebrafish: selective requirement for the transcriptional regulators orthopedia and single-minded 1 in the preoptic area. *Dev Dyn*, 237(4):995–1005, Apr 2008.

[28] Robert C Edgar. Muscle: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5:113, Aug 2004.

[29] Robert C Edgar. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32(5):1792–1797, 2004.

[30] Albrecht Fischer and Adriaan Dorresteijn. The polychaete platynereis dumerilii (annelida): a laboratory animal with spiralian cleavage, lifelong segment proliferation and a mixed benthic/pelagic life cycle. *Bioessays*, 26(3):314–325, Mar 2004.

[31] Robert Fredriksson, Malin C Lagerström, Lars-Gustav Lundin, and Helgi B Schiöth. The g-protein-coupled receptors in the human genome form five main families. phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol*, 63(6):1256–1272, Jun 2003.

[32] M. A. Frohman. Rapid amplification of complementary dna ends for generation of full-length complementary dnas: thermal race. *Methods Enzymol*, 218:340–356, 1993.

[33] Y. Fujino, T. Nagahama, T. Oumi, K. Ukena, F. Morishita, Y. Furukawa, O. Matsushima, M. Ando, H. Takahama, H. Satake, H. Minakata, and K. Nomoto. Possible functions of oxytocin/vasopressin-superfamily peptides in annelids with special reference to reproduction and osmoregulation. *J Exp Zool*, 284(4):401–406, Sep 1999.

[34] Makoto Furutani-Seiki and Joachim Wittbrodt. Medaka and zebrafish, an evolutionary twin study. *Mech Dev*, 121(7-8):629–637, Jul 2004.

[35] Scott F. Gilbert. *Developmental Biology*. Sinauer Associates, 2000.

[36] P. Gilligan, S. Brenner, and B. Venkatesh. Neurone-specific expression and regulation of the pufferfish isotocin and vasotocin genes in transgenic mice. *J Neuroendocrinol*, 15(11): 1027–1036, Nov 2003.

[37] G. Gimpl and F. Fahrenholz. The oxytocin receptor system: structure, function, and regulation. *Physiol Rev*, 81(2):629–683, Apr 2001.

[38] Gonzalo Giribet. Assembling the lophotrochozoan (=spiralian) tree of life. *Philos Trans R Soc Lond B Biol Sci*, 363(1496):1513–1522, Apr 2008.

[39] J. L. Goodson and A. H. Bass. Forebrain peptides modulate sexually polymorphic vocal circuitry. *Nature*, 403(6771):769–772, Feb 2000.

[40] J. L. Goodson and A. H. Bass. Social behavior functions and related anatomical characteristics of vasotocin/vasopressin systems in vertebrates. *Brain Res Brain Res Rev*, 35(3):246–265, Jul 2001.

[41] Aubrey Gorbman, Arthur Whiteley, and Scott Kavanaugh. Pheromonal stimulation of spawning release of gametes by gonadotropin releasing hormone in the chiton, mopalia sp. *Gen Comp Endocrinol*, 131(1):62–65, Mar 2003.

[42] A. Gore. *The Master Molecule of Reproduction*. Kluwer Academic Publishers, 2002.

[43] M. Gozdowska, A. Kleszczy?ska, E. Soko?owska, and E. Kulczykowska. Arginine vasotocin (avt) and isotocin (it) in fish brain: diurnal and seasonal variations. *Comp Biochem Physiol B Biochem Mol Biol*, 143(3):330–334, Mar 2006.

[44] Adam J Guastella, Philip B Mitchell, and Mark R Dadds. Oxytocin increases gaze to the eye region of human faces. *Biol Psychiatry*, 63(1):3–5, Jan 2008.

[45] Stephane Guindon, Franck Lethiec, Patrice Duroux, and Olivier Gascuel. Phyml online–a web server for fast maximum likelihood-based phylogenetic inference. *Nucl. Acids Res.*, 33: 557–559, 2005.

[46] Pai-Chung Gwee, Chris T Amemiya, Sydney Brenner, and Byrappa Venkatesh. Sequence and organization of coelacanth neurohypophysial hormone genes: evolutionary history of the vertebrate neurohypophysial hormone gene locus. *BMC Evol Biol*, 8:93, 2008.

[47] Pai-Chung Gwee, Boon-Hui Tay, Sydney Brenner, and Byrappa Venkatesh. Characterization of the neurohypophysial hormone gene loci in elephant shark and the japanese lamprey: origin of the vertebrate neurohypophysial hormone genes. *BMC Evol Biol*, 9(1): 47, Feb 2009.

[48] S. Halford, M. S. Freedman, J. Bellingham, S. L. Inglis, S. Poopalasundaram, B. G. Soni, R. G. Foster, and D. M. Hunt. Characterization of a novel human opsin gene with wide tissue expression and identification of embedded and flanking genes on chromosome 1q43. *Genomics*, 72(2):203–208, Mar 2001.

[49] Stephanie Halford, Susana S Pires, Michael Turton, Lei Zheng, Irene González-Menéndez, Wayne L Davies, Stuart N Peirson, José M García-Fernández, Mark W Hankins, and Russell G Foster. Va opsin-based photoreceptors in the hypothalamus of birds. *Curr Biol*, 19(16):1396–1402, Aug 2009.

[50] Cheryl F Harding and Sandra A Rowe. Vasotocin treatment inhibits courtship in male zebra finches; concomitant androgen treatment inhibits this effect. *Horm Behav*, 44(5):413–418, Dec 2003.

[51] C. Hauenschild. Lunar periodicity. *Cold Spring Harb Symp Quant Biol*, 25:491–497, 1960.

[52] C. Hauenschild. The endocrine control of reproduction in annelids. *Arch Anat Microsc Morphol Exp*, 54:429–452, 1965.

[53] Thorsten Henrich, Mirana Ramialison, Beate Wittbrodt, Beatrice Assouline, Franck Bourrat, Anja Berger, Heinz Himmelbauer, Takashi Sasaki, Nobuyoshi Shimizu, Monte Westerfield, Hisato Kondoh, and Joachim Wittbrodt. Mepd: a resource for medaka gene expression patterns. *Bioinformatics*, 21(14):3195–3197, Jul 2005.

[54] R. S. Hewes and P. H. Taghert. Neuropeptides and neuropeptide receptors in the drosophila melanogaster genome. *Genome Res*, 11(6):1126–1142, Jun 2001.

[55] Nicholas Holland and Jr-Kai Yu. Epidermal receptor development and sensory pathways in vitally stained amphioxus (*Branchiostoma floridae*). *Acta Zoologica*, 83:309–319, 2002.

[56] C. H. Hoyle. Neuropeptide families and their receptors: evolutionary perspectives. *Brain Res*, 848(1-2):1–25, Nov 1999.

[57] T. J. P. Hubbard, B. L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, T. Down, S. C. Dyer, S. Fitzgerald, J. Fernandez-Banet, S. Graf, S. Haider, M. Hammond, J. Herrero, R. Holland, K. Howe, K. Howe, N. Johnson, A. Kahari, D. Keefe, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, C. Melsopp, K. Megy, P. Meidl, B. Ouverdin, A. Parker, A. Prlic, S. Rice, D. Rios, M. Schuster, I. Sealy, J. Severin, G. Slater, D. Smedley, G. Spudich, S. Trevanion, A. Vilella, J. Vogel, S. White, M. Wood, T. Cox, V. Curwen, R. Durbin, X. M. Fernandez-Suarez, P. Flicek, A. Kasprzyk, G. Proctor, S. Searle, J. Smith, A. Ureta-Vidal, and E. Birney. Ensembl 2007. *Nucleic Acids Research*, 1:gkl996v2, 2006.

[58] J. P. Huelsenbeck and F. Ronquist. Mrbayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8):754–755, Aug 2001.

[59] S. Hyodo, S. Ishii, and J. M. Joss. Australian lungfish neurohypophysial hormone genes encode vasotocin and [phe2]mesotocin precursors homologous to tetrapod-type precursors. *Proc Natl Acad Sci U S A*, 94(24):13339–13344, Nov 1997.

[60] Susumu Hyodo, Yukio Kato, Masao Ono, and Akihisa Urano. Cloning and sequence analyses of cdnas encoding vasotocin and isotocin precursors of chum salmon,oncorhynchus keta: evolutionary relationships of neurohypophysial hormone precursors. *Journal of Comparative Physiology*, 160:601–608, 1991.

[61] Susumu Hyodo, Takehiro Tsukada, and Yoshio Takei. Neurohypophysial hormones of dogfish, triakis scyllium: structures and salinity-dependent secretion. *Gen Comp Endocrinol*, 138(2):97–104, Sep 2004.

[62] O. Kah, C. Lethimonier, G. Somoza, L. G. Guilgur, C. Vaillant, and J. J. Lareyre. Gnrh and gnrh receptors in metazoa: a historical, comparative, and evolutive perspective. *Gen Comp Endocrinol*, 153(1-3):346–364, 2007.

[63] Atsuhiro Kanda, Kyoko Takuwa-Kuroda, Eiko Iwakoshi-Ukena, and Hiroyuki Minakata. Single exon structures of the oxytocin/vasopressin superfamily peptides of octopus. *Biochem Biophys Res Commun*, 309(4):743–748, Oct 2003.

[64] Atsuhiro Kanda, Toshio Takahashi, Honoo Satake, and Hiroyuki Minakata. Molecular and functional characterization of a novel gonadotropin-releasing-hormone receptor isolated from the common octopus (octopus vulgaris). *Biochem J*, 395(1):125–135, Apr 2006.

[65] Grit Kasper, Stefan Taudien, Eike Staub, Detlev Mennerich, Melissa Rieder, Bernd Hinzmann, Edgar Dahl, Uta Schwidetzky, André Rosenthal, and Andreas Rump. Different structural organization of the encephalopsin gene in man and mouse. *Gene*, 295(1):27–32, Jul 2002.

[66] Tsuyoshi Kawada, Toshio Sekiguchi, Yoshiyuki Itoh, Michio Ogasawara, and Honoo Satake. Characterization of a novel vasopressin/oxytocin superfamily peptide and its receptor from an ascidian, ciona intestinalis. *Peptides*, 29(10):1672–1678, Oct 2008.

[67] R. E. Van Kesteren, A. B. Smit, R. P. De Lange, K. S. Kits, F. A. Van Golen, R. C. Van Der Schors, N. D. De With, J. F. Burke, and W. P. Geraerts. Structural and functional evolution of the vasopressin/oxytocin superfamily: vasopressin-related conopressin is the only member present in lymnaea, and is involved in the control of sexual behavior. *J Neurosci*, 15(9):5989–5998, Sep 1995.

[68] C. S. Koger, S. J. Teh, and D. E. Hinton. Variations of light and temperature regimes and resulting effects on reproductive parameters in medaka (oryzias latipes). *Biol Reprod*, 61(5): 1287–1293, Nov 1999.

[69] Michael Kosfeld, Markus Heinrichs, Paul J Zak, Urs Fischbacher, and Ernst Fehr. Oxytocin increases trust in humans. *Nature*, 435(7042):673–676, Jun 2005.

[70] Kristen M Kwan, Esther Fujimoto, Clemens Grabher, Benjamin D Mangum, Melissa E Hardy, Douglas S Campbell, John M Parant, H. Joseph Yost, John P Kanki, and Chi-Bin Chien. The tol2kit: a multisite gateway-based construction kit for tol2 transposon transgenesis constructs. *Dev Dyn*, 236(11):3088–3099, Nov 2007.

[71] T. C. Lacalli and S. J. Kelly. Ventral neurons in the anterior nerve cord of amphioxus larvae. i. an inventory of cell types and synaptic patterns. *J Morphol*, 257(2):190–211, Aug 2003.

[72] Thurston C Lacalli. Basic features of the ancestral chordate brain: a protochordate perspective. *Brain Res Bull*, 75(2-4):319–323, Mar 2008.

[73] Rainer Landgraf and Inga D Neumann. Vasopressin and oxytocin release within the brain: a dynamic concept of multiple and variable modes of neuropeptide communication. *Front Neuroendocrinol*, 25(3-4):150–176, 2004.

[74] Dan Larhammar, Görel Sundström, Susanne Dreborg, Daniel Ocampo Daza, and Tomas A Larsson. Major genomic events and their consequences for vertebrate evolution and endocrinology. *Ann N Y Acad Sci*, 1163:201–208, Apr 2009.

[75] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. Clustal w and clustal x version 2.0. *Bioinformatics*, 23(21):2947–2948, Nov 2007.

[76] Gabriela G Loots and Ivan Ovcharenko. Mulan: multiple-sequence alignment to predict functional elements in genomic sequences. *Methods Mol Biol*, 395:237–254, 2007.

[77] M. Ludwig. Dendritic release of vasopressin and oxytocin. *J Neuroendocrinol*, 10(12):881–895, Dec 1998.

[78] Mike Ludwig and Gareth Leng. Dendritic peptide release and peptide-dependent behaviours. *Nat Rev Neurosci*, 7(2):126–136, Feb 2006.

[79] C. A. Marler, S. K. Boyd, and W. Wilczynski. Forebrain arginine vasotocin correlates of alternative mating strategies in cricket frogs. *Horm Behav*, 36(1):53–61, Aug 1999.

[80] Nicholas J McGlincy and Christopher W J Smith. Alternative splicing resulting in nonsense-mediated mrna decay: what is the meaning of nonsense? *Trends Biochem Sci*, 33(8):385–393, Aug 2008.

[81] Chryssa Metallinou, Byron Asimakopoulos, Andreas Schröer, and Nikos Nikolettos. Gonadotropin-releasing hormone in the ovary. *Reprod Sci*, 14(8):737–749, Dec 2007.

[82] G. Michel, J. Chauvet, M. T. Chauvet, and R. Acher. One-step processing of the amphibian vasotocin precursor: structure of a frog (rana esculenta) "big" neurophysin. *Biochem Biophys Res Commun*, 149(2):538–544, Dec 1987.

[83] F. L. Moore and L. J. Miller. Arginine vasotocin induces sexual behavior of newts by acting on cells in the brain. *Peptides*, 4(1):97–102, 1983.

[84] Nils G Morgenthaler, Joachim Struck, Stefan Jochberger, and Martin W Dünser. Copeptin: clinical use of a new biomarker. *Trends Endocrinol Metab*, 19(2):43–49, Mar 2008.

[85] Eiji Motohashi, Tomoko Hamabata, and Hironori Ando. Structure of neurohypophysial hormone genes and changes in the levels of expression during spawning season in grass puffer (takifugu niphobles). *Gen Comp Endocrinol*, 155(2):456–463, Jan 2008.

[86] Paraskevi Moutsaki, David Whitmore, James Bellingham, Katsuhiko Sakamoto, Zoë K David-Gray, and Russell G Foster. Teleost multiple tissue (tmt) opsin: a candidate photopigment regulating the peripheral clocks of zebrafish? *Brain Res Mol Brain Res*, 112(1-2): 135–145, Apr 2003.

[87] D.O. Norris. *Vertebrate endocrinology*. Elsevier Academic Press, 2007.

[88] S. Ohno. *Evolution by gene duplication*. Springer Verlag, 1970.

[89] Ivan Ovcharenko, Gabriela G Loots, Belinda M Giardine, Minmei Hou, Jian Ma, Ross C Hardison, Lisa Stubbs, and Webb Miller. Mulan: multiple-sequence local alignment and visualization for studying function and evolution. *Genome Res*, 15(1):184–194, Jan 2005.

[90] K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto, and M. Miyano. Crystal structure of rhodopsin: A g protein-coupled receptor. *Science*, 289(5480):739–745, Aug 2000.

[91] Stuart N Peirson, Stephanie Halford, and Russell G Foster. The evolution of irradiance detection: melanopsin and the non-visual opsins. *Philos Trans R Soc Lond B Biol Sci*, 364 (1531):2849–2865, Oct 2009.

[92] Michael W Pfaffl, Graham W Horgan, and Leo Dempfle. Relative expression software tool (rest) for group-wise comparison and statistical analysis of relative expression results in real-time pcr. *Nucleic Acids Res*, 30(9):e36, May 2002.

[93] Florian Raible, Kristin Tessmar-Raible, Kazutoyo Osoegawa, Patrick Wincker, Claire Jubin, Guillaume Balavoine, David Ferrier, Vladimir Benes, Pieter de Jong, Jean Weissenbach, Peer Bork, and Detlev Arendt. Vertebrate-type intron-rich genes in the marine annelid platynereis dumerilii. *Science*, 310(5752):1325–1326, Nov 2005.

[94] Martina Rembold, Kajori Lahiri, Nicholas S Foulkes, and Joachim Wittbrodt. Transgenesis in fish: efficient selection of transgenic fish by co-injection with a fluorescent reporter construct. *Nat Protoc*, 1(3):1133–1139, 2006.

[95] Franck Rodet, Christophe Lelong, Marie-Pierre Dubos, Katherine Costil, and Pascal Favrel. Molecular cloning of a molluscan gonadotropin-releasing hormone receptor orthologue specifically expressed in the gonad. *Biochim Biophys Acta*, 1730(3):187–195, Sep 2005.

[96] Alfred Romer and Thomas Parsons. *Vergleichende Anatomie der Wirbeltiere - The vertebrate body*. Verlag Paul Parey, 1991.

[97] S. Rozen and H. Skaletsky. Primer3 on the www for general users and for biologist programmers. *Methods Mol Biol*, 132:365–386, 2000.

[98] J Sambrook and W Russell. *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory Press Cold Spring Harbor, 1989.

[99] Yoshinori Shichida and Take Matsuyama. Evolution of opsins and phototransduction. *Philos Trans R Soc Lond B Biol Sci*, 364(1531):2881–2895, Oct 2009.

[100] V. Singh and K. P. Joy. Effects of hcg and ovarian steroid hormones on vasotocin levels in the female catfish heteropneustes fossilis. *Gen Comp Endocrinol*, 162(2):172–178, Jun 2009.

[101] Lukas Stalder and Oliver Mühlemann. The meaning of nonsense. *Trends Cell Biol*, 18(7): 315–321, Jul 2008.

[102] Emma E Tarttelin, James Bellingham, Lindsay C Bibb, Russell G Foster, Mark W Hankins, Kevin Gregory-Evans, Cheryl Y Gregory-Evans, Dominic J Wells, and Robert J Lucas. Expression of opsin genes early in ocular development of humans and mice. *Exp Eye Res*, 76(3):393–396, Mar 2003.

[103] Malcom H. Taylor. Environmental and endocrine influences on reproduction of fundulus heteroclitus. *Amer. Zool.*, 26(1):159–171, 1986.

[104] K. Terakado. Induction of gamete release by gonadotropin-releasing hormone in a protochordate, ciona intestinalis. *Gen Comp Endocrinol*, 124(3):277–284, Dec 2001.

[105] A. Terakita, T. Yamashita, and Y. Shichida. Highly conserved glutamic acid in the extracellular iv-v loop in rhodopsins acts as the counterion in retinochrome, a member of the rhodopsin family. *Proc Natl Acad Sci U S A*, 97(26):14263–14267, Dec 2000.

[106] Akihisa Terakita. The opsins. *Genome Biol*, 6(3):213, 2005.

[107] Akihisa Terakita, Mitsumasa Koyanagi, Hisao Tsukamoto, Takahiro Yamashita, Takashi Miyata, and Yoshinori Shichida. Counterion displacement in the molecular evolution of the rhodopsin family. *Nat Struct Mol Biol*, 11(3):284–289, Mar 2004.

[108] Kristin Tessmar-Raible. The evolution of neurosecretory centers in bilaterian forebrains: insights from protostomes. *Semin Cell Dev Biol*, 18(4):492–501, Aug 2007.

[109] Kristin Tessmar-Raible, Florian Raible, Foteini Christodoulou, Keren Guy, Martina Rembold, Harald Hausen, and Detlev Arendt. Conserved sensory-neurosecretory cell types in annelid and fish forebrain: insights into hypothalamus evolution. *Cell*, 129(7):1389–1400, Jun 2007.

[110] Violette Thermes, Clemens Grabher, Filomena Ristoratore, Franck Bourrat, André Choulika, Jochen Wittbrodt, and Jean-Stéphane Joly. I-scei meganuclease mediates highly efficient transgenesis in fish. *Mech Dev*, 118(1-2):91–98, Oct 2002.

[111] R. R. Thompson, K. George, J. C. Walton, S. P. Orr, and J. Benson. Sex-specific influences of vasopressin on human social communication. *Proc Natl Acad Sci U S A*, 103(20):7889–7894, May 2006.

[112] Kazuyoshi Ukena, Eiko Iwakoshi-Ukena, and Akira Hikosaka. Unique form and osmoregulatory function of a neurohypophysial hormone in a urochordate. *Endocrinology*, 149(10): 5254–5261, Oct 2008.

[113] R. E. van Kesteren, A. B. Smit, R. W. Dirks, N. D. de With, W. P. Geraerts, and J. Joosse. Evolution of the vasopressin/oxytocin superfamily: characterization of a cdna encoding a vasopressin-related precursor, preproconopressin, from the mollusc lymnaea stagnalis. *Proc Natl Acad Sci U S A*, 89(10):4593–4597, May 1992.

[114] R. E. van Kesteren, C. P. Tensen, A. B. Smit, J. van Minnen, L. F. Kolakowski, W. Meyerhof, D. Richter, H. van Heerikhuizen, E. Vreugdenhil, and W. P. Geraerts. Co-evolution of ligand-receptor pairs in the vasopressin/oxytocin superfamily of bioactive peptides. *J Biol Chem*, 271(7):3619–3626, Feb 1996.

[115] Rodrigo A Velarde, Colin D Sauer, Kimberly K O Walden, Susan E Fahrbach, and Hugh M Robertson. Pteropsin: a vertebrate-like non-visual opsin expressed in the honey bee brain. *Insect Biochem Mol Biol*, 35(12):1367–1377, Dec 2005.

[116] B. Venkatesh and S. Brenner. Structure and organization of the isotocin and vasotocin genes from teleosts. *Adv Exp Med Biol*, 395:629–638, 1995.

[117] B. Venkatesh, S. L. Si-Hoe, D. Murphy, and S. Brenner. Transgenic rats reveal functional conservation of regulatory controls between the fugu isotocin and rat oxytocin genes. *Proc Natl Acad Sci U S A*, 94(23):12462–12466, Nov 1997.

[118] B. Vigh, M. J. Manzano, A. Zádori, C. L. Frank, A. Lukáts, P. Röhlich, A. Szél, and C. Dávid. Nonvisual photoreceptors of the deep brain, pineal organs and retina. *Histol Histopathol*, 17(2):555–590, Apr 2002.

[119] Béla Vigh Ingeborg Vigh-Teichmann. *Neurosecretion: Molecules, Cells, Systems*, chapter The cerebrospinal fluid-contacting neurosecretory cell - A protoneuron, pages 458–460. Plenum Publishing Corporation, 1982.

[120] Kathryn Walters, Iga N Wegorzewska, Yue-Pui Chin, Manan G Parikh, and T. J. Wu. Luteinizing hormone-releasing hormone i (lhrh-i) and its metabolite in peripheral tissues. *Exp Biol Med (Maywood)*, 233(2):123–130, Feb 2008.

[121] J. M. Warne and R. J. Balment. Effect of acute manipulation of blood volume and osmolality on plasma [avt] in seawater flounder. *Am J Physiol*, 269(5 Pt 2):R1107–R1112, Nov 1995.

[122] Julia H White, Mathias Chiano, Mark Wigglesworth, Robert Geske, John Riley, Nicola White, Simon Hall, Guohua Zhu, Frank Maurio, Tony Savage, Wayne Anderson, Joanna Cordy, Melissa Ducceschi, G. A. I. N. investigators, Jorgen Vestbo, and Sreekumar G Pillai. Identification of a novel asthma susceptibility gene on chromosome 1qter and its functional evaluation. *Hum Mol Genet*, 17(13):1890–1903, Jul 2008.

[123] Helmut Wicht and Thurston Lacalli. The nervous system of amphioxus: structure, development, and evolutionary significance. *Can. J. Zool.*, 83:122–150, 2005.

[124] Joachim Wittbrodt, Akihiro Shima, and Manfred Schartl. Medaka–a model organism from the far east. *Nat Rev Genet*, 3(1):53–64, Jan 2002.

[125] K. Yasui, S. Tabata, T. Ueki, M. Uemura, and S. C. Zhang. Early development of the peripheral nervous system in a lancelet species. *J Comp Neurol*, 393(4):415–425, Apr 1998.

[126] T. Yokota and T. Oishi. Seasonal change in the locomotor activity rhythm of the medaka, oryzias latipes. *Int J Biometeorol*, 36(1):39–44, Mar 1992.

[127] Hayan Yoon, L. W. Enquist, and Catherine Dulac. Olfactory inputs to hypothalamic neurons controlling reproduction and fertility. *Cell*, 123(4):669–682, Nov 2005.

[128] L. J. Young, R. Nilsen, K. G. Waymire, G. R. MacGregor, and T. R. Insel. Increased affiliative response to vasopressin in mice expressing the v1a receptor from a monogamous vole. *Nature*, 400(6746):766–768, Aug 1999.

[129] Limor Ziv, Adi Tovin, Daniel Strasser, and Yoav Gothilf. Spectral sensitivity of melatonin suppression in the zebrafish pineal gland. *Exp Eye Res*, 84(1):92–99, Jan 2007.

[130] R. T. Zoeller and F. L. Moore. Brain arginine vasotocin concentrations related to sexual behaviors and hydromineral balance in an amphibian. *Horm Behav*, 22(1):66–75, Mar 1988.

5 Appendix

5.1 Whole mount *in situ* hybridisation in Medaka and Zebrafish

Whole mount *in situ* hybridisation is a method used to visualize RNA within the whole body of an animal. This can be done in Medaka as well as in Zebrafish embryos. First, the embryos at the proper stage of interest have to be fixed in 4% PFA in 2x PTW (2xPBS, 0.1% Tween20) for 4-6 hours at roomtemperature on a shaker and optionally also up until 6 days at $4°C$. After that, the embryos were washed in 1xPTW and bleached using 0.5% KOH, 3% H_2O_2 for roughly 30 minutes in a basket until the pigmentation is completely dissolved. Then, the embryos were put to 100% MeOH and put on a shaker for 1 hour. During this hour, the MeOH was replaced once and after the hour, the MeOH was replaced and the fishes stored at $-20°C$. At this stage, the embryos can be stored for several months.

The probes were prepared from two possible types of templates. Either the gene was cloned within a vector that contained T7 or SP6 transcription start sites in the proper orientation or a PCR was prepared to isolate the fragment with specific primers. The reverse primer is attached to a T7 or SP6 site. If a vector was used, then the vector had to be cut with a restriction enzyme that linearizes the vector at the opposite site of the antisense probe template than the transcription start site. The vector and PCR products should be checked on a gel if linearisation or amplification worked efficiently. Then, after purification with either the Quiagen PCR Purification Kit or the Quiagen Gel Extraction kit as in the manual, the concentration is assessed and $1\mu g$ of the vector is used for the *in vitro* transcription reaction. For the PCR, 200ng were used for the reaction. Then, the reaction was started by mixing $2\mu l$ of the DIG-UTP dNTP mix (Roche) with $1\mu l$ RNAsin (RNAse inhibitor, Promega), $2\mu l$ of the RNA polymerase (SP6 or T7, NEB) and $2\mu l$ Transcription buffer (NEB) and water to $20\mu l$. The reaction is incubated at $37°C$ for $3-4$ hours and then $1\mu l$ RNAse-free DNAse (NEB) was added. After 20 minutes, the reaction was stopped by cleaning it up using the RNAsy Mini Kit (RNA Cleanup protocol). It wis eluted with $50\mu l$ RNAse-free water. Then, $2\mu l$ are tested on a gel in formamide containing loading buffer. Prior to loading, the sample is heated to $94°C$ for 5 minutes. Only one thick band should be visible. If so, $150\mu l$ Hyb-Mix are added to the probe and then stored at $-20°C$.

All following steps have to be made at room temperature and in 2ml SafeLock tubes unless otherwise indicated. The embryos are rehydrated in a series of MeOH, 1xPTW mixtures (75% MeOH, 50% MeOH, 25% MeOH; 5 minutes each). Then the embryos are washed twice with

1xPTW for five minutes. The embryos are treated with ProteinaseK to facilitate probe and antibody penetration. The time of ProteinaseK digest is dependent on species, developmental stage and ProteinaseK batch. For Zebrafish, it takes usually 11 minutes and for Medaka usually $20 - 2$ minutes for the used stages. However ProteinaseK digestion times have to be found empirically for each stage. After the ProteinaseK digest, the digest was stopped by rinsing the embryos shortly in 2mg/ml glycine/1xPTW. Then, the embryos were postfixed for 20-30 minutes in 4% PFA/1xPTW. After postfixation the embryos were washed with 1xPTW 6 times (Medaka) or 4 times (Zebrafish). The embryos were prehybridised by incubating them in Hyb-Mix (50% formamide, 5xSSC, 0.1% Tween20, 0.5mg/ml torula RNA, 50μg/ml Heparin) at 65°C for 2-4 hours. After that, 20μl of the probes were denatured in 175μl Hyb-Mix at 80°C. Then, the Hyb-Mix with the fish embryos was replaced by the 200μl probe mixture. The probe was hybridised to the mRNAs at 65°C overnight.

The next day, the embryos were washed to remove non-bound probe. The washes were all performed at 65°C. For zebrafish, they were as follows: 5 minutes in Hyb- (50% formamide, 5xSSC, 0.1% Tween20), 3x10 minutes in 25% Hyb- in 2xSSCT, 5 minutes in 2xSSCT and 2x30 minutes in 0.2xSSCT. These steps are followed by one wash step for five minutes in 0.2xSSCT,MABT (1:1) at roomtemperature and then five minutes MABT. For Medaka, the following was used: 2x30 minutes in 50% formamide/2xSSCT, 15 minutes in 2xSSCT and 2x30 minutes in 0.2xSSCT.

After the washes, the embryos have to be blocked to exclude unspecific binding of the antibody. This is done for $1 - 2$ hour(s) using either MABT + 2% DIG block (Roche) (Maleic acid buffer plus Tween) (Zebrafish) or 5% sheep serum in 1xPTW (Medaka). Then, the embryos are incubated with anti-DIG antibody ($1 : 2000$ (Medaka) or $1 : 4000$ (Zebrafish) dilution in the respective blocking solution) either overnight at 4°C or for 2 hours at room temperature. After the incubation, the antibody is washed off either six times with 1xPTW or 5 times with MABT for 10 minutes each. If the antibody was applied at roomtemperature, then one of the washes can be put to 4°C overnight. The embryos are transferred to a 6- or 24-well plate and then stained. Medaka embryos are first equilibrated in staining buffer (2M TrisHCl, 5M NaCl, 1M $MgCl_2$, 10% Tween20). The embryos are then stained in SB+NBT (337.5μg/ml)/BCIP(175μg/ml) for several hours until staining appeared. Zebrafish embryos are put to BM Purple, a commercially available NBT/BCIP substrate (Roche) until staining was visible. The reaction is stopped by washing the embryos twice in 1xPTW (one wash overnight at 4°C). The next day, the embryos are refixed in 4% PFA for 30 minutes, washed with 1xPTW three times and then put to 87% glycerol until they were mounted on microscope slides for image acquisition.

Fluorescent 2-color WM-ISH – "CHECK PROTOCOL"

DAY 1

- Fix as usual, MeOH treated animals at -20°C, all steps in 2ml tubes
- Rehydrate in 75% MeOH/PTW for 5 minutes at RT
- Rehydrate in 50% MeOH/PTW for 5 minutes at RT
- Rehydrate in25% MeOH/PTW for 5 minutes at RT
- Wash in PTW for 5 minutes
- Wash in PTW for 5 minutes
- Digest with Proteinase K at 22°C
    - ([10µg/ml PTW] >>> 5µl stock [20mg/ml] in 10ml PTW)
    - Stage 30 >>> 16 minutes
    - Stage 32-34 >>> 20 minutes
    - Stage 37 >>> 25 minutes
- Rinse shortly with 2mg/ml glycine/PTW (1:100 dilution of Stock in PTW)
- Fix in 4%PFA/PTW for 20 minutes
- Wash in PTW for 5 minutes
- Wash in PTW for 5 minutes
- Wash in PTW for 5 minutes
- Wash in PTW for 5 minutes
- Wash in PTW for 5 minutes
- Prehybridize in Hyb-Mix at 65°C for 2h
- Mix 20µl of each probe (XXX-DIG/XXX-FLUO) with 160µl Hyb-Mix/sample, heat to 80°C
- Remove Hyb-Mix and add probe with Hyb-Mix
- Incubate at 65°C o/n

DAY 2

- Remove the probe mix
- Wash embryos with 50% formamide/2xSSCT for 30 minutes at 65°C
- Wash embryos with 50% formamide/2xSSCT for 30 minutes at 65°C
- Wash embryos with 2xSSCT for 15 minutes at 65°C
- Wash embryos with 0,2xSSCT for 30 minutes at 65°C
- Wash embryos with TNT for 5 minutes at RT
    - (100mM TrisHCl, pH 7.5; 350mM NaCl;0.1%Tween20)
- Wash embryos with TNT for 5 minutes at RT
- Block embryos in TNB (TNT plus 2% DIG Block) for 2h at RT
- Add 200µl Anti-Fluo-Fab-POD (Roche) (1:50 diluted in TNB) o/n at 4°C,
    - Dissolve lyophilized antibody with 1ml water to obtain 150U/ml

DAY 3

- Rinse with TNT
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Rinse in 100µl TSA Amplification Diluent
- Dilute Fluorescein Fluorophore Tyramide 1:50 in TSA Amplification Diluent (i.e. staining solution)
- Add staining solution (covered embryos, ~100µl)
- ALL STEPS IN THE DARK!!!
- Stain for 30 minutes to 1.5 hours, wash few fishes with TNT twice and check fluorescence using the SteREO Lumar
- After sufficient staining, rinse embryos using TNT
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Incubate in 1%H2O2/TNT without shaking in the dark for 20 minutes
- Wash with TNT for 5 minutes
- Wash with TNT for 5 minutes
- Wash with TNT for 5 minutes
- Wash with TNT for 5 minutes
- Wash with TNT for 5 minutes
- Block embryos with 1% Roche Blocking reagent/TNT at RT on a shaker
- Add Anti-Dig-POD antibody (Roche) at a 1:100 dilution AND Anti ac. tub. AB (1.2:100) in TNB overnight at 4°C

DAY 4

- ALL STESPS IN THE DARK!!!
- Rinse with TNT
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Rinse in 100µl TSA Amplification Diluent
- Dilute Cy3 Fluorophore Tyramide 1:50 in TSA Amplification Diluent
- Add staining solution (covered embryos, ~100µl)
- Stain for 30 minutes, wash few fishes with TNT twice and check fluorescence using the SteREO Lumar
- After sufficient staining, rinse embryos using TNT
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Incubate embryos using anti mouse Cy5 (0.5:100 in TNT) for 2h at RT in the dark
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Wash with TNT for 10 minutes
- Remove all TNT and add DABCO Glycerol, let stay at 4°C o/n

5.2 Injection of Zebrafish and Medaka

The injection solution was different between Medaka and Zebrafish. In Medaka, the vectors or PCR products were injected ($20 - 25ng/\mu$l) with the meganuclease ISceI (NEB, stored at $-80°C$), 0.5x ISceI buffer (NEB) in a total volume of 20μl water. For Zebrafish, the vector DNA was injected ($20 - 25ng/\mu$l) along with mRNA from the transposase Tol2. This transposase is encoded on a vector and part of the Tol2Kit [70]. The vector was cut using NotI (Fermentas), the cutting was verified on a gel and the vector was purified using the Quiagen PCR Purification Kit. Then, an *in vitro* transcription has been started using the linearized vector as the template. The reaction was prepared using the Ambion mMessage mMachine SP6 Kit according to the protocol. This kit adds a cap-structure to the transcribed mRNA and it can therefore be injected in cells where it is translated. Thus, while the meganuclease ISceI is injected as a protein, the Tol2 transposase is injected as mRNA and is only translated in the injected embryo. At the beginning, several injection markers were added (FITC or TRITC, Molecular Probes). However, this was stopped since the needle tended to clog more often and the expression of the transgene was not well detectable because of bleedthrough. Thus, later on, the injection markers were not added anymore to the mix. The mix was centrifuged at full speed for 5 minutes and then stored on ice. 3μl were used to fill the needles.

In medaka, the fertilized eggs were scraped off of the belly from several females. They were rolled on a sheet of Whatman paper in order to dry and separate them by removing the hair-like structures. Medaka embryos can be kept at $4°C$ or on ice, which slows down their development. This is not possible with Zebrafish eggs that are much more vulnerable to such cold temperatures. Zebrafish couples were set up the day before the injection in so-called mating cages. In these cages, the fishes are seperated by a plastic disc and mating is induced in the morning after the disc is removed. After 30 minutes, the eggs can be collected easily because they fall through an inlay and remain seperated from the parents.

The eggs were fixed in a petri dish that contains a solid agarose layer (1.5% agarose in fish medium) with furrows where the embryos can be placed in using forceps. The embryos were positioned so that the eggs face the upper right corner from where the needle will penetrate the embryo. The petri dish with the fishes was put under a Zeiss Binocular. The capillaries (Science Products, GB100F-10 ($0.58x1x100$mm) for Medaka and GB100TF10 ($0.78x1x100$mm). The numbers indicate inner diameter x outer diameter x length) were pulled to needles using the Sutter Instrument Flaming/Brown Micropipette Puller P97 with the following settings: (Medaka) Pressure: 500, Heat: 560, Pull: 28, Velocity: 30, Time: 150. (Zebrafish) Pressure: 500, Heat: 510, Pull: 100, Velocity: 170, Time: 120. The needle was attached to a needle holder connected to the InjectMan NI2 to move the needle properly and the FemtoJet -express- to regulate the air supply. The injection solution was injected in one to two-cell stage embryos. The amount was variable but it usually was one third of the cell volume. The embryos were then kept in the same

petridish after the injection at 28°C. The next day, expression of the reporter gene was assessed using the Zeiss SteREO Lumar. Any animal with expression in any part of the body were raised to adulthood.

Used primer sequences

	stock #	Oligo name	Derived from	Sequence	length	Purpose
1	1	pducdc5_L1	Platynereis	CCTATTGACATGGACGAAGATG	29	Q-RT-PCR reference gene
2	2	pducdc5_R1		TTCCCTGTGTGTTCGCAAG	20	
3	264	Pdu_vtnL2		GGAGTGTGTTGTACAGACGAGAA	23	Q-RT-PCR for vasotocin
4	265	Pdu_vtnR2		GGCTTGTCCATGTCCAGTTT	20	
5	117	degVTNRU1	deg. PCR - multiple species	CAYYTXWSXATHGCXGAYYT	20	cloning of Platynereis VT receptors X=Inosine
6	118	degVTNRL2		ACRAARAAXGGXSWCCARCA	20	
7	119	degVTNRU2		GCNTTYTTYAAYRTXYTXCCXCA	23	
8	120	degVTNRU3		GAYMGNTAYYTXGCXATHTGYCAYCC	26	
9	121	degVTNRL1VTR2		RTADATCCAXGGRTTXGTRCA	21	
10	122	degVTNRL2VTR2		RTCCCAXGCXGCCCACATYTG	21	
11	123	degVTNRL1VTR1		RTADATCCAXGGRTTRCARCA	21	
12	147	vtnR1_RF1	Platynereis	GCAATGGCCTGGTTATTGTCTCTGC	25	nested 3' -F- and 5' -R- RACE PCR Vasotocin receptors
13	148	vtnR1_RR1		GTAACATATCCTTCCGTAAGCAATGAC	27	
14	149	vtnR1_RF2		GCATTCAATGATTTAGACGACCGA	24	
15	150	vtnR1_RR2		GTACAGAGAACAGCAGAGACAATAACC	27	
16	151	vtnR2_RF1		GTTATCACGAAGTGGAGTCAGGAGTG	26	
17	152	vtnR2_RF2		GACAGCCAATCAGATCGCTTCATCG	25	
18	153	vtnR2_RR1		GATGTAGATAGAAACAGTGAAACAAG	26	
19	154	vtnR2_RR2		GAGGCGTACTCAAAGTTAAGATAATCC	27	
20	168	TI_f	Medaka	GTATCGATGTCGAGGCAGAAACTTCCTT	28	PCR to amplifiy genomic region of tmt-Opsin, third intron
21	171	TI_r		GGATCCAGAGAGGACAGGTGTTACT	25	
22	155	PE_f_dirTOPO		CACCGCCTGCAGGGTTCAAGAAATG	25	PCR to amplifiy genomic region of tmt-Opsin, PE element-TOPO cloning into middle entry vector (Gateway system)
23	170	PE_r		GCCGCCAACTGAGGACCAGTAC	22	
24	433	XhoI-ISceI-pDestTol2Af	pDestTol2A	CAGCTCGAGTTAGGGATAACAGGGTAATCAGCTATGACCATGATTACG	52	Addition of an ISceI site to the vector (use one vector in Medaka and Zebrafish)
25	434	KpnI-ISceI-pDestTol2Ar		TACGGTACCATAGGGATAACAGGGTAATAACGACGGCCAGTGAATTATC	53	
26	755	Fugu_vtn_R1BP_attB2R	Fugu	GGGGACCACTTTGTACAAGAAAGCTGGGTGCTGGAAATCAGCACCAGTTC	50	Cloning of the vtn Reporter construct used in mice, fusion pcr to get GFP incorporation
27	413	Fugu_vtn_R2		ACAGGCTCAGAGGATGTGAGG	21	
28	415	Fugu_vtn_R4		TCTCTCTGCTTTCGCCATCTC	21	
29	416	Fugu_vtn_R5_attB1		GGGGACAAGTTTGTACAAAAAAGCAGGCTAGAGTCCAGGAGAGCTCCAAA	50	
30	164	mCF	mCherry	TTCGCCTGGGACATCCTGTC	20	transgene control PCR - genotyping
31	165	mCR		TGTACAGCCGTCCATGCCG	19	
32	166	mGF	GFP²	CACCTACGGCAAGCTGACCCTGAA	24	
33	167	mGR		CAGGACCATGTGATCGCGCT	20	
34	417	Fugu_vtnUP_f2	Fugu	CTGTGTGGTTTGGGTTGGTTT	21	
35	418	Fugu_vtnUP_r2		ACCTTGCCTGGTTCCATACCT	21	
36	419	Fugu_vtnDOWN_f2		ACTACGTGCTGCATGCTGAGA	21	
37	420	Fugu_vtnDOWN_r2		ACATCAACGCCTCACATCCTC	21	
38	156	F1vtnOUT	Medaka	AGCGATGCATCCCTCCCTCAG	21	Cloning of Medaka Nonapeptide homologues, nested PCR
39	157	F2vtnIN		CTGGGATTCCTCGCTCTGTC	20	
40	158	R1vtnOUT		TTCAAGCAGAGAGCAAAGAGC	21	
41	159	R2vtnIN		GCAATGTCTCATTTGGCTGA	20	
42	160	F1itnOUT		ACTGGAGCAACTGTATCCGTG	21	
43	161	F2itnIN		TTCATCATGTCCGTGTGCTCT	21	
44	162	R1itnOUT		TCAAAGTTGATGATGGACCAA	21	
45	163	R2itnIN		GATTTGTTGCCACGGGACT	19	
46	189	R15RACE_R1	Platynereis	CAGCAGTCATAGACCCCCGAACC	30	RACE PCR primers used to clone the vasotocin receptors R1 and R2
47	190	R15RACE_R2		CGTAGGACGAAGCGTACATCGCAAC	33	
48	191	R15RACE_R3		CAGATGAGTTGCGGCAACACATTGAAG	35	
49	192	R13RACE_P3		TCTTGCAGTAATCGCGTGCTACCTGA	26	
50	193	R13RACE_P2		CGGAAGGATATGTTACGTCGTGTGGA	26	
51	194	R13RACE_P1		CCCGAGTGGACACTGCAGCTTTATG	25	
52	195	R25RACE_R1		CGACCGTACAGCACGGCGAGAATA	24	
53	196	R25RACE_R2		CCGGGTAGAAGTACGCCCAGCAGT	24	
54	197	R25RACE_R3		ATGCCGATCATCACCTTGACCCTCT	25	
55	198	R23RACE_P1		GTACGGTCGCATATGCTGGGAGGT	24	
56	199	R23RACE_P2		GACAGCCAATCAGATCGCTTCATCG	25	
57	200	R23RACE_P3		GATACCCTGTGTCCGCGAGCTCAT	24	

Used abbreviations

ATN	Annetocin
AVP	Arginine-Vasopressin
C Opsin	Ciliary Opsin
Cc	*Catostomus commersoni*
cDNA	complementary Deoxyribonucleic acid
CPN	Conopressin
cRNA	complementary Ribonucleic acid
CSF	Cerebrospinal fluid/Central spinal fluid
CTN	Cephalotocin
DNA	Deoxyribonucleic acid
ECR	Evolutionary Conserved Region
Ef	*Eisenia fetida*
ELISA	Enzyme-linked immunosorbent assay
ERM	Embryo Rearing Medium
FACS	Fluorescence-activated cell sorting
G_R	GnRH receptor
GFP	Green fluorescent protein
Gg	*Gallus Gallus*
GnRH	Gonadotropin Releasing Hormone
GPCRs	G protein-coupled receptors
Hs	*Homo sapiens*
ITN	Isotocin
itnR	Isotocin receptor
LCDVs	Large Dense Core Vesicles
LNA	locked nucleic acid
Ls	*Lymnaea stagnalis*
Mm	*Mus musculus*
MTN	Mesotocin
mtnR	Mesotocin receptor
Myr	Million years
NMD	Nonsense-mediated RNA decay
NP	Nonapeptide
NxT	N-linked glycosilation site
Ol	*Oryzias latipes*
OPN	Octopressin
Otp	Orthopedia
Ov	*Octopus vulgaris*
OXT	Oxytocin
oxtR	Oxytocin receptor
PCR	Polymerase Chain Reaction
Pdu	*Platynereis dumerilii*
PVN	Paraventricular Nucleus
qRT-PCR	quantitative Reverse Transcriptase-PCR
R Opsin	Rhabdomeric Opsin
RACE PCR	Rapid Amplification of cDNA Ends
Rc	*Rana catesbeiana*
RGR Opsins	RPE retinal G protein-coupled receptor
RNA	Ribonucleic acid
RT-PCR	Reverse Transcriptase PCR
Rx	Retinal homeobox
SCN	Suprachiasmatic Nucleus
Sim1	Single-minded homolog 1
SON	Supraoptic nucleus
SP	Signalpeptide
TMT Opsin	Teleost Multiple Tissue Opsin
vpR	Vasopressin receptor
VTN	Vasotocin

Amino acids according to IUPAC rules

5.3 Used software

TMHMM *- Prediction of transmembrane domains*
http://www.cbs.dtu.dk/services/TMHMM/
Softberry *- Homology-based gene prediction (and many other tools)*
http://linux1.softberry.com/berry.phtml
Mulan *- Phylogenetic footprinting*
http://mulan.dcode.org/
Vista *- Phylogenetic footprinting*
http://genome.lbl.gov/vista/index.shtml
REST *- qRT-PCR data analysis*
http://www.gene-quantification.de/
Primer3 *- Primer design software*
http://frodo.wi.mit.edu/primer3/
Oligocalc *- Primer properties calculations*
http://www.basic.northwestern.edu/biotools/oligocalc.html
IDT DNA Tools *- Misc. tools, Primer properties calculations*
http://eu.idtdna.com/Scitools/Scitools.aspx
EBI Tools *- Misc. tools, Muscle*
http://www.ebi.ac.uk/Tools/
Expasy Tools *- Misc. tools*
http://www.expasy.ch/tools/
CLC Main Workbench 5 *- Overall sequence management*
http://www.clcbio.com/
NCBI webserver
http://www.ncbi.nlm.nih.gov/
EnsEMBL genome browser
http://www.ensembl.org/index.html
JGI genome browser
http://genome.jgi-psf.org/
PhyML3.0 *- Phylogenetic analysis*
http://www.atgc-montpellier.fr/phyml/
MrBayes3 *- Phylogenetic analysis*
http://mrbayes.csit.fsu.edu/
ClustalX2 *- Sequence alignment and phylogenetic analysis*
no homepage available

6 Abstract

In today's neuroscience, the brain is mostly seen as an integrative organ that processes information perceived by peripheral sensory organs. However, besides its processing function, the vertebrate brain contains many cells that have been suggested to be directly sensory. We focus on one specific type of such sensory-neurosecretory cells. These cells express the presumptive light-sensitive Opsin molecule and secrete the neuronal Nonapeptide hormone Vasotocin. These cells are highly conserved and present in vertebrates, as was shown for Zebrafish, as well as in invertebrate species, such as the annelid *Platynereis dumerilii*. Thus, they presumably already existed in Urbilateria and form the ancient core of the vertebrate brain (Tessmar-Raible, 2007).

In many species, the secretion pattern of the neuronal hormone Vasotocin correlates with reproductive cycles that are governed by exogenous light cycles. It has been suggested that this cell type acts as a conserved minimal module that controls the cycling Vasotocin hormone levels (Tessmar-Raible, 2007).

The goal of my diploma thesis is to further investigate different aspects of this system to finally come closer towards a functional characterisation of this highly conserved, sensory-neurosecretory cell type. For that I work on three different projects. First, I further characterise the Vasotocin hormonal system in *Platynereis dumerilii*. Second, I investigate how the expression of *vasotocin* is regulated in diverse vertebrate species and finally, I characterise novel presumptive light-sensitive members of a deep brain photoreceptor family in vertebrates.

7 Zusammenfassung

Das Gehirn wird oft als integratives Organ beschrieben, welches Informationen von peripheren sensorischen Organen bekommt, prozessiert und einen bestimmten Output erzeugt. Jedoch kann das Gehirn nicht nur Informationen prozessieren, es ist auch in der Lage selbst Informationen aufzunehmen. Es wurden im Gehirn von einigen Spezies Zellen beschrieben, die direkt sensorisch sind. Diese Zellen sekretieren oft auch neuronale Hormone und bilden den anzestralen Kern des Wirbeltiergehirns. Wir arbeiten an einem bestimmten Typ dieser sensorischen neurosekretorischen Zellen. Diese exprimieren ein wahrscheinlich licht-sensitives Opsin Protein und sekretieren ein neuronales Hormon der Nonapeptid Superfamilie, nämlich Vasotocin. Diese Zellen sind höchst konserviert und wurden in Fischen sowie in Würmern entdeckt. Daher müssen diese Zellen im gemeinsamen Vorläufer der beiden Spezies bereits existiert haben (Urbilateria) (Tessmar-Raible, 2007).

Unsere Hypothese ist, dass dieser Zelltyp ein anzestrales, hochkonserviertes Minimalmodul repräsentiert, das die Sekretion von Vasotocin reguliert. Das beruht auf der Beobachtung dass in einigen Spezies die Sekretion von Vasotocin mit den spezifischen Reproduktionszyklen korreliert. Diese Reproduktionszyklen werden meist über exogene Lichtzyklen reguliert. Das Ziel meiner Diplomarbeit war es verschiedene Aspekte dieses Systems genauer zu untersuchen, um schlussendlich zur funktionellen Charakterisierung von diesem sensorisch-neurosekretorischem Zelltyp beizutragen.

Ich arbeitete an drei verschiedenen Projekten. Das erste Projekt umfasst die Charakterisierung einiger Aspekte des Nonapeptid Hormonsystems in *Platynereis dumerilii*. Im Rahmen des zweiten Projektes habe ich untersucht wie die Expression von *vasotocin* in Vertebraten reguliert wird. Abschließend habe ich neue wahrscheinlich licht-sensitive Mitglieder einer Photorezeptorfamilie in Vertebraten identifiziert und charakterisiert.

CURRICULUM VITAE

STEPHAN EMANUEL KIRCHMAIER

PERSONAL DETAILS

Date of birth:	11/12/1982
Place of birth:	Innsbruck/Austia

EDUCATION

Elementary school in Villach	1989 – 1993
Grammar school in Villach	1993 – 2002
Study of molecular biology at the University of Vienna	2003 – 2009

INTERNATIONAL EXPERIENCE

Erasmus exchange year at the University of Groningen/The Netherlands	2006 – 2007
Research training at the EMBL/Heidelberg, Group of Dr. Jochen Wittbrodt	October 2008

WORKING EXPERIENCE

Biomay — Summer 2005
Vienna Competence Center
Purification of several recombinant proteins expressed in *E. coli*

Group of T. Czerny — Summer 2006
University of Veterinary Medicine Vienna
Molecular cloning and plasmid injection into *Oryzias latipes* (Medaka) embryos

Cell biochemistry group – P.J.M. van Haastert — Fall 2006
University of Groningen / Faculty of life sciences
Quantification of *Dictyostelium discoideum* chemotaxis behaviour

Evolutionary genetics group – L.W. Beukeboom — Winter 2006
University of Groningen / Faculty of life sciences
Sequence analysis of sex determining genes in different *Nasonia* species

Evolutionary genetics group – L.W. Beukeboom — Jan. – Jul. 2007
University of Groningen / Faculty of life sciences
Master project: Identification and characterisation of a *period* ortholog in the parasitic wasp *Nasonia vitripennis*

Group of J. Penninger — March, May 2008
IMBA – Institute for Molecular Biotechnology in Vienna
Characterisation of Fam3C knockout and overexpression mice and starting a fly screen for the identification of hematopoietic regulators

Group of C. Seiser June 2008
MFPL – Max F. Perutz Laboratories in Vienna
Characterisation of epidermis-specific HDAC1/DNMT1 knockout mice

Group of A. & M. Matzke July 2008
GMI – Gregor-Mendel-Institute
Analysis of *Arabidopsis thaliana* mutants with defects in the RdDM-pathway

Group of K. Tessmar-Raible since September 2008
MFPL – Max F. Perutz Laboratories in Vienna
Diploma thesis: Sensory – neurosecretory cells in *Platynereis dumerilii* and fish

CONFERENCES

EMBO Meeting 2009
Amsterdam, The Netherlands
Abstract and poster presentation "Characterization of photosensory-neurosecretory cells in Medaka and Zebrafish"

SPECIAL HONOURS

Unilever Research Prize 2007
granted for the master project "Identification and characterisation of a period ortholog in the parasitic wasp Nasonia vitripennis"

Achievement Scholarship
granted by the University of Vienna for the academic year 2006/2007

Travel grant – "Förderungsstipendium"
granted by the University of Vienna for the research training at the EMBL in Heidelberg, October 2008