*Manuscript
Click here to view linked References

NG peptides: a novel family of neurophysin-associated neuropeptides

Maurice R. Elphick

Queen Mary University of London,
School of Biological & Chemical Sciences,
Mile End Road,
London, E1 4NS, UK

Correspondence to: Maurice R. Elphick, Ph.D,
Professor of Animal Physiology & Neuroscience,
School of Biological & Chemical Sciences,
Queen Mary University of London,
Mile End Road,
London, E1 4NS, UK

Tel: 44 (0)207 882 5290
Fax: 44 (0)208 983 0973
E-mail: M.R.Elphick@qmul.ac.uk

Abstract

Neurophysins are prohormone-derived polypeptides that are required for biosynthesis of the neurohypophyseal hormones vasopressin and oxytocin. Accordingly, mutations in the neurophysin domain of the human vasopressin gene can cause diabetes insipidus. The association of neurophysins with vasopressin/oxytocin-type peptides dates back to the common ancestor of bilaterian animals and until recently it was thought to be unique. This textbook perspective on neurophysins changed with the discovery of a gene in the sea urchin *Strongylocentrotus purpuratus* (phylum Echinodermata) encoding a precursor protein comprising a neurophysin domain in association with NGFFFamide, a myoactive neuropeptide that is structurally unrelated to vasopressin/oxytocin-type neuropeptides (Elphick, M.R., Rowe, M.L., 2009. NGFFFamide and echinotocin: structurally unrelated myoactive neuropeptides derived from neurophysin-containing precursors in sea urchins. *J. Exp. Biol.* 212, 1067-1077). What is not known, however, is when and how the association of neurophysin with NGFFFamide-like neuropeptides originated. Here I report the discovery of genes encoding proteins comprising a neurophysin domain in association with putative NGFFFamide-like peptides in the hemichordate *Saccoglossus kowalevskii* (NGFWNamide and NGFYNamide) and in the cephalochordate *Branchiostoma floridae* (SFRNGVamide). Together with NGFFFamide, these peptides constitute a novel family of neuropeptides in invertebrate deuterostomes that are derived from neurophysin-containing precursors and that have the sequence motif NG – "NG peptides". Genes encoding NG peptides in association with neurophysin were not found in protostomes, urochordates or vertebrates. Interestingly, however, SFRNGVamide is identical to the N-terminal region of neuropeptide S, a peptide that

modulates arousal and anxiety in mammals, whilst NGFFFamide shares sequence similarity with SIFamide (AYRKPPFNGSIFamide), a neuropeptide that regulates sexual behaviour in *Drosophila*. Collectively, these data indicate that in an ancestor of extant deuterostomes a remarkable and unique event in the evolution of neuropeptide signalling systems occurred when a neurophysin-encoding exon(s) derived from a vasopressin/oxytocin-type neuropeptide gene became transcriptionally linked with another family of neuropeptides – NG peptides.

Keywords: *Saccoglossus kowalevskii*, *Branchiostoma floridae*, vasopressin, oxytocin, neuropeptide S, SIFamide

1. Introduction

Neurophysins are prohormone-derived polypeptides that are required for packaging, processing and protection of the neurohypophyseal hormones vasopressin and oxytocin. Accordingly, mutations in the neurophysin domain of the human vasopressin gene can cause diabetes insipidus due to impaired vasopressin biosynthesis (Ivell and Richter, 1984; Legros and Geenen, 1996; De Bree and Burbach, 1998; De Bree, 2000). Genes encoding precursor proteins comprising a vasopressin/oxytocin-like peptide and a C-terminal neurophysin domain have been identified in vertebrates, deuterostomian invertebrates and protostomian invertebrates. Thus, the association of neurophysins with vasopressin/oxytocin-type neuropeptides can be traced back to the common ancestor of bilaterian animals (Van Kesteren et al., 1992; Kawada et al., 2008; Stafflinger et al., 2008; Ukena et al., 2008).

Neurophysins were, until recently, thought to be uniquely associated with vasopressin/oxytocin-type neuropeptides. However, this textbook perspective on neurophysins changed with the discovery of a gene in the sea urchin *Strongylocentrotus purpuratus* (phylum Echinodermata) encoding a precursor protein comprising a neurophysin domain in association with a myoactive neuropeptide that is structurally unrelated to vasopressin/oxytocin-type neuropeptides – Asn-Gly-Phe-Phe-Phe-NH_2 or NGFFFamide (Elphick and Rowe, 2009). Thus, in *Strongylocentrotus purpuratus* there are two neuropeptide precursors that have a neurophysin domain: a precursor comprising neurophysin in association with a vasopressin/oxytocin-type neuropeptide ("echinotocin") and the NGFFFamide precursor (Elphick and Rowe, 2009).

The NGFFFamide precursor was discovered on account of the sequence similarity that NGFFFamide shares with NGIWYamide, a myoactive neuropeptide that is a potent inducer of oocyte maturation and spawning in the sea cucumber *Apostichopus japonicus* (Phylum Echinodermata) (Iwakoshi et al., 1995; Inoue et al., 1999; Kato et al., 2009). It is not known, however, how and when the association of neurophysin with NGFFFamide-like neuropeptides originated. Is it a peculiarity of sea urchins or echinoderms in general or do precursors comprising NGFFFamide-like peptides in association with neurophysin also occur in other phyla? To address this issue, here I have investigated the occurrence of genes encoding proteins related to the NGFFFamide precursor throughout the animal kingdom, utilising genome sequence data made publicly available recently.

2. Material And Methods

To investigate the phylogenetic distribution of genes encoding orthologs of the NGFFFamide precursor, its 266 amino-acid residue sequence (GenBank accession number XP_001177223) was submitted as a query for Basic Local Alignment Search Tool (BLAST) analysis of genome sequence databases for the following species: *Nematostella vectensis* (Phylum Cnidaria; http://genomeportal.jgi-psf.org/Nemve1), *Helobdella robusta* (Phylum Annelida; http://genomeportal.jgi-psf.org/Helro1), *Caenorhabditis elegans* (Phylum Nematoda; http://www.sanger.ac.uk/Projects/C_elegans), *Apis mellifera*, *Bombyx mori*, *Daphnia pulex*, *Drosophila melanogaster*, *Tribolium castaneum* (Phylum Arthropoda; http://flybase.org/blast, http://wfleabase.org/genomics), *Saccoglossus kowalevskii* (Phylum Hemichordata; http://www.hgsc.bcm.tmc.edu/project-species-o-Acorn%20worm.hgsc), *Branchiostoma floridae* (Phylum Chordata, sub-phylum cephalochordata; http://genomeportal.jgi-psf.org/Brafl1) *Ciona intestinalis* (Phylum Chordata, sub-phylum urochordata; http://genomeportal.jgi-psf.org/Cioin2), *Petromyzon marinus* (Phylum Chordata, sub-phylum vertebrata; http://genome.wustl.edu/tools/blast), *Callorhinchus milii* (Phylum Chordata, sub-phylum vertebrata; http://esharkgenome.imcb.a-star.edu.sg/), *Danio rerio* (Phylum Chordata, sub-phylum vertebrata http://www.sanger.ac.uk/Projects/D_rerio), *Xenopus tropicalis* (Phylum Chordata, sub-phylum vertebrata; http://genomeportal.jgi-psf.org/Xentr4), *Gallus gallus* (Phylum Chordata, sub-phylum vertebrata; http://www.chick.manchester.ac.uk/), *Mus musculus* (Phylum Chordata, sub-phylum vertebrata; http://www.sanger.ac.uk/cgi-bin/blast/submitblast/m_musculus) and *Homo sapiens* (Phylum Chordata, sub-phylum vertebrata; http://www.sanger.ac.uk/cgi-

bin/blast/submitblast/hgp). Proteins sharing the highest sequence similarity with NGFFFamide precursor in each species analysed were then submitted as BLAST queries against the GenBank nr database to establish whether the NGFFFamide precursor was a mutual best hit indicative of orthology. Proteins identified by this approach as orthologs of the NGFFFamide precursor were then subject to more detailed analysis using clustalw2 for multiple sequence alignment (http://www.ebi.ac.uk/Tools/clustalw2).

3. Results

3.1. Discovery of an ortholog of the NGFFFamide precursor in the hemichordate Saccoglossus kowalevskii

To investigate if NGFFFamide-type neuropeptide precursors are unique to echinoderms or also occur in other deuterostomian invertebrates, genome sequence data for the hemichordate species *Saccoglossus kowalevskii* was first analyzed because the phylum Hemichordata is a sister phylum to the echinoderms in a clade of the animal kingdom known as the Ambulacraria (Bromham and Degnan, 1999). BLAST analysis of *Saccoglossus kowalevskii* genomic contigs identified two contigs containing sequences that encode putative polypeptides sharing substantial sequence similarity with the NGFFFamide precursor, contigs 94209 and 42727.

Analysis of the sequence of contig 42727 revealed that it contains a gene encoding a vasopressin/oxytocin-type precursor comprising a vasopressin/oxytocin-like peptide with a predicted amino acid sequence CFISDCARGamide. Discovery of this peptide is of interest because it is the first vasopressin/oxytocin-like peptide to be identified in a hemichordate and I suggest that it is named “hemitocin”.

Analysis of the sequence of contig 94209 revealed that it contains a DNA sequence encoding a polypeptide that shares 43% amino acid identity with the neurophysin domain of the NGFFFamide precursor. However, this contig did not contain sequences encoding NGFFFamide-like peptides. To specifically search for *Saccoglossus* genomic sequences encoding NGFFFamide-like peptides, a BLAST search was performed using the query sequence KRNGFFFGKRNGFFFGKR, which corresponds to the NGFFFamide-containing region of the NGFFFamide precursor in

Strongylocentrotus purpuratus, but this search yielded no hits. A BLAST search of the *Saccoglossus* contig database was then performed using the sequence KRNGIWYGKRNGIWYGKR, which is a hypothetical partial precursor sequence for the sea cucumber neuropeptide NGIWYamide. This yielded a single hit (contig 31671), which contained a 54 bp sequence encoding two copies of a putative NGIWYamide-like peptide NGFWNamide (i.e. KRNGFWNGKRNGFWNGKR). The 19360 bp sequence of contig 31671 was then submitted as a query to the gene prediction program genscan (http://genes.mit.edu/GENSCAN.html), which predicted a partial gene comprising two exons that encode a 222-residue protein sequence. This protein sequence contains five copies of the sequence NGFWNG and a single copy of the sequence NGFYNG, each flanked by putative dibasic cleavage sites (KR). Thus, contig 31671 contains a partial gene encoding the putative NGIWYamide/NGFFFamide-like peptides NGFWNamide and NGFYNamide. Furthermore, analysis of the 222-residue protein sequence using SignalP (http://www.cbs.dtu.dk/services/SignalP/) indicated the presence of a N-terminal signal peptide sequence, demonstrating that this protein has one of the characteristic features of neuropeptide precursors.

Collectively these data indicate that in *Saccoglossus* there is a gene encoding a protein that is a homolog of the *Strongylocentrotus* NGFFFamide precursor, comprising a signal peptide and NGFWNamide/NGFYNamide neuropeptides in its N-terminal domain (contig 31671) and a neurophysin-like polypeptide forming its C-terminal domain (contig 94209). However, the fact that the N-terminal and C-terminal domains of this putative protein are encoded by exons on two different contigs suggests that there may be a large intron separating these exons in the *Saccoglossus* genome. Importantly, this hypothesis is consistent with the structure of the gene

encoding the NGFFFamide precursor in *Strongylocentrotus purpuratus*, where there is a 16,420 base intron separating the exon encoding two copies of NGFFFamide and the exon encoding the neurophysin domain of the NGFFFamide precursor (Elphick and Rowe, 2009). Further investigation of this issue will be possible when a final assembly of the *Saccoglossus* genome sequence has been produced. Nevertheless, the data available at present are consistent with a gene model that is strikingly similar in structure to the NGFFFamide gene in *Strongylocentrotus*. Thus, as illustrated in Fig. 1A, the putative *Saccoglossus* NGFWNamide/NGFYNamide gene encodes a 303 residue protein and comprises three protein-coding exons. The first of these exons encodes residues 1-32, which includes the N-terminal signal peptide; the second of these exons encodes residues 33-222, which includes five copies of NGFWNamide and one copy of NGFYNamide; and the third of these exons encodes residues 223-303, which forms the neurophysin domain of the protein. This gene model is supported by the presence of 5' donor (gt) and 3' acceptor (ag) sequences flanking the introns that separate the putative exons in contigs 31671 and 94209.

An interesting feature of the putative *Saccoglossus* NGFWNamide/NGFYNamide precursor is that it comprises six putative neuropeptide molecules (5 x NGFWNamide and 1 x NGFYNamide), whereas the *Strongylocentrotus* NGFFFamide precursor comprises only two copies of NGFFFamide. Furthermore, comparison of the sequences of the spacer peptides that follow copies of NGFWNamide or NGFYNamide in the *Saccoglossus* NGFWNamide/NGFYNamide precursor reveal sequence similarities, which are indicative of intragenic duplication (Fig. 1B).

Importantly, submission of the putative 303 residue *Saccoglossus* NGFWNamide/NGFYNamide precursor sequence as a BLAST query against the

NCBI non-redundant protein database revealed that the *Strongylocentrotus* NGFFFamide precursor was the best hit, indicating that these two proteins are orthologs. Therefore, the discovery of the NGFWNamide/NGFYNamide gene in *Saccoglossus* indicates that neurophysin-associated NGFFFamide-type neuropeptides date back at least as far as the common ancestor of echinoderms and hemichordates.

3.2. Discovery of an ortholog of the NGFFFamide precursor in the cephalochordate Branchiostoma floridae

Having obtained evidence that NGFFFamide-type neuropeptide precursors date back to the common ancestor of the Ambulacraria, the occurrence of related proteins in the chordate clade of the deuterostomes was investigated. The NGFFFamide precursor sequence was submitted as a BLAST query against a database of predicted proteins in the cephalochordate *Branchiostoma floridae* ("amphioxus") (Putnam et al., 2008). The best hit was a 256-residue protein (Brafl1-84803) and the second best hit was a 167-residue protein (Brafl1-84802).

Analysis of Brafl1-84802 revealed that it is a vasopressin/oxytocin-type precursor comprising a vasopressin/oxytocin-like peptide with a predicted amino acid sequence CYIINCPRGamide, which has been reported previously (Gwee et al., 2009).

Submission of the Brafl1-84803 amino acid sequence as a BLAST query against the GenBank non-redundant (nr) protein database revealed that the NGFFFamide precursor has the highest level of sequence similarity with the Brafl1-84803 protein, suggesting that these two proteins are orthologs. Analysis of the sequence of the Brafl1-84803 protein revealed that it has an N-terminal signal

peptide, as would be expected for a putative neuropeptide precursor. Furthermore, the C-terminal region of the Brafl1-84803 protein comprises a neurophysin-like polypeptide, which largely accounts for the sequence similarity shared by the NGFFFamide precursor and the Brafl1-84803 protein. Analysis of the sequence of the Brafl1-84803 protein between the N-terminal signal peptide and the C-terminal neurophysin domain revealed the presence of two copies of the sequence KRSFRNGVGKR. This sequence is noteworthy because of the presence of the flanking putative dibasic cleavage sites (KR) and a putative substrate (G) for C-terminal amidation. Thus, it appears that the Brafl1-84803 protein is the precursor of a putative neuropeptide – SFRNGVamide (Fig. 2). Interestingly, this peptide contains the dipeptide sequence NG, a feature that is a characteristic of NGFFFamide, the related sea cucumber neuropeptide NGIWYamide and the putative peptides identified here in *Saccoglossus* (NGFWNamide and NGFYNamide). Collectively, these characteristics of the Brafl1-84803 protein indicate it is a neurophysin-containing neuropeptide precursor that shares a common ancestry with the *Strongylocentrotus* NGFFFamide precursor and the *Saccoglossus* NGFWNamide/NGFYNamide precursor.

To further assess relationships between the putative SFRNGVamide precursor in *Branchiostoma*, the putative *Saccoglossus* NGFWNamide/NGFYNamide precursor and the *Strongylocentrotus* NGFFFamide precursor, the sequences of these proteins were aligned using the online multiple sequence alignment program clustalw2 (Fig. 3). The alignment shows the positions of introns in the genes encoding these proteins (see pairs of amino acid residues underlined), revealing similarity in the structure of the genes and providing further evidence that these proteins are orthologs. The protein-coding region of the NGFFFamide gene is interrupted by two introns. The

first intron separates the codons for residues 36 and 37 and an intron is located at a similar position in the NGFWNamide/NGFYNamide and SFRNGVamide precursors, separating the codons for residues 32 and 33 in both precursors. The second intron separates the codons for residue 184 and the first cysteine residue (185) of the C-terminal neurophysin domain in the NGFFFamide precursor and an intron is located at the same position in the NGFWNamide/NGFYNamide and SFRNGVamide precursors, preceding the codon for the cysteine at residue positions 223 and 180, respectively. Thus, there are two introns located in similar positions in all three genes and both of the introns are in the same phase (phase 0) in all three genes. This provides important evidence that the genes encoding the NGFFFamide precursor, the NGFWNamide/NGFYNamide precursor and the SFRNGVamide precursor are derived from a common ancestral gene.

3.3. Orthologs of the NGFFFamide precursor are not present in urochordates and vertebrates but SFRNGVamide shares sequence identity with neuropeptide S

NGFFFamide-type neuropeptide precursors were not discovered in the urochordate *Ciona intestinalis* or in several vertebrate species analyzed and the best hits in BLAST searches were precursors of vasopressin/oxytocin-type neuropeptide precursors. Interestingly, however, the putative *Branchiostoma* neuropeptide SFRNGVamide shares 100% sequence identity with the N-terminal region of a neuropeptide in humans and other tetrapod vertebrates that is known as neuropeptide S (e.g. the sequence of neuropeptide S in humans is SFRNGVGTGMKKTSFQRAKS) (Xu et al., 2004) (Fig. 4).

3.4. Orthologs of the NGFFFamide precursor are not present in cnidarians or in protostomes but NGFFFamide shares sequence similarity with arthropod SIFamide neuropeptides

BLAST searches of genome sequence databases for cnidarian and protostomian invertebrate species did not reveal NGFFFamide precursor-like genes and the best hits were genes encoding proteins comprising a C-terminal neurophysin domain in association with a vasopressin/oxytocin-like peptide. However, NGFFFamide does share sequence similarity with members of the SIFamide neuropeptide family in arthropods, which have the C-terminal sequence NGSIFamide (Verleyen et al., 2009) (Fig. 4).

4. Discussion

4.1. NG peptides: a novel family of neurophysin-associated neuropeptides in invertebrate deuterostomes

The evolutionary origin of a gene in the sea urchin *Strongylocentrotus purpuratus* encoding a neuropeptide precursor comprising two copies of the myoactive peptide NGFFFamide and a neurophysin domain was investigated. Genes encoding related proteins comprising NGFFFamide-like peptides in association with neurophysin were discovered in the hemichordate *Saccoglossus kowalevskii* and the cephalochordate *Branchiostoma floridae* but not in cnidarians, protostomes, urochordates or vertebrates.

The *Saccoglossus* gene encodes five copies of the putative peptide NGFWNamide and one copy of the putative peptide NGFYNamide. These peptides are clearly structurally similar to NGFFFamide and the sea cucumber neuropeptide NGIWYamide, which is consistent with the close phylogenetic relationship between echinoderms and hemichordates (Bromham and Degnan, 1999). Furthermore, it can be inferred that the common ancestor of echinoderms and hemichordates would have had a gene encoding a neuropeptide precursor with a C-terminal neurophysin domain and a N-terminal domain containing one or more copies of a C-terminally amidated pentapeptide with the N-terminal residues Asn-Gly (NG). Accordingly, one can speculate that NGIWYamide, the NGFFFamide-like peptide discovered in the sea cucumber *Apostichopus japonicus* (Iwakoshi et al., 1995), may be derived from a precursor protein that has a C-terminal neurophysin domain.

The precursor protein identified here in the cephalochordate *Branchiostoma* is also an ortholog the NGFFFamide precursor, comprising a C-terminal neurophysin domain and a N-terminal domain that contains two copies of the putative peptide SFRNGVamide. Like NGFFFamide in *Strongylocentrotus* and NGFWNamide/NGFYNamide in *Saccoglossus*, this *Branchiostoma* peptide contains the dipeptide sequence Asn-Gly. Thus, it can be inferred that the deuterostomian common ancestor of cephalochordates and the Ambulacraria would have had a gene encoding a precursor protein comprising a C-terminal neurophysin domain and a N-terminal domain containing one or more copies of an amidated peptide containing the dipeptide sequence Asn-Gly. Furthermore, as the presence of the Asn-Gly (NG) sequence appears to be a characteristic feature of this novel family of neurophysin-associated neuropeptides in deuterostomes, I propose that this family of peptides are collectively named "NG peptides". Interestingly, in the *Strongylocentrotus* and *Saccoglossus* NG peptides the NG sequence is located at the N-terminus whereas in *Branchiostoma* it forms an internal sequence. It is not possible to infer from these data what the structural characteristics of a NG peptide would have been in a common ancestor of the Ambulacraria and cephalochordates, but this may be possible if genes encoding NG peptides are identified in a wider range of species.

The alignment in Fig. 3 shows that the neurophysin domains of the three NG peptide precursors each have fourteen cysteine residues, which is also a characteristic feature of neurophysins associated with vasopressin/oxytocin-type peptides (De Bree and Burbach, 1998). Furthermore, the positions of these cysteine residues are conserved between the *Strongylocentrotus* and the *Saccoglossus* NG peptide precursors. Interestingly, however, the seventh and eighth cysteines of the neurophysin domain of the *Branchiostoma* NG peptide precursor are shifted by three

residue positions with respect to the *Strongylocentrotus* and the *Saccoglossus* NG peptide precursors. This is interesting because a glutamate residue located between the seventh and eighth cysteine residues in neurophysins associated with vasopressin/oxytocin-type peptides is known to form part of a peptide-binding pocket (De Bree and Burbach, 1998). Therefore, structural differences in this region of the *Strongylocentrotus/Saccoglossus* NG peptide-associated neurophysins and the *Branchiostoma* NG peptide-associated neurophysin may reflect differences in the structural characteristics of their putative neuropeptide partners - namely, that the putative *Strongylocentrotus* and *Saccoglossus* NG peptides are amidated pentapeptides with an N-terminal NG motif, whereas the putative *Branchiostoma* NG peptide is an amidated hexapeptide with an internal NG motif.

4.2. Neuropeptide S: a vertebrate homolog of NG peptides?

The occurrence of genes encoding neurophysin-containing NG peptide precursors in Ambulacraria and in a cephalochordate indicates that this gene family dates back at least as far as the common ancestor of the deuterostomes. How then can the absence of genes encoding this type of neuropeptide precursor in urochordates and vertebrates be explained? The most parsimonious explanation is that the gene encoding a NG peptide-type neuropeptide precursor with a neurophysin domain was lost in a common ancestor of urochordates and vertebrates. Interestingly, however, the putative *Branchiostoma* peptide SFRNGVamide shares 100% sequence identity with neuropeptide S, a neuropeptide that is involved in regulation of arousal and anxiety in mammals and which belongs to a family of neuropeptides identified in tetrapod vertebrates (Reinscheid, 2007). Furthermore, it is the N-terminal region of

neuropeptide S sharing sequence identity with SFRNGVamide (see Fig. 4) that is critical for its biological activity (Roth et al., 2006).

How can this striking sequence identity shared between neuropeptide S and the *Branchiostoma* NG peptide be explained? One possible explanation is convergent protein evolution, which is a statistically plausible explanation because other proteins containing the sequence SFRNGV are present in the GenBank database. However, the *Branchiostoma* NG peptide precursor and the vertebrate neuropeptide S precursors are the only proteins in the GenBank database that contain the sequence SFRNGV preceded by the dibasic cleavage site sequence Lys-Arg (KR).

An alternative explanation is that the genes encoding SFRNGVamide in *Branchiostoma* and neuropeptide S in tetrapods evolved from a common ancestral gene encoding a precursor comprising a NG peptide and a C-terminal neurophysin domain but the neurophysin-encoding exon was lost in the lineage that gave rise to neuropeptide S in tetrapods. Genes encoding neuropeptide S have been sequenced, revealing that it is derived from a precursor protein comprising a single copy of the neuropeptide located in the C-terminal region of the precursor. However, because genes encoding neuropeptide S-type precursors are not found in urochordates or in basal vertebrates such as agnathans, elasmobranchs and teleosts (Reinscheid, 2007), the hypothesis that neuropeptide S precursors and the *Branchiostoma* NG peptide precursor evolved from a common ancestral gene in chordates requires invoking loss of genes encoding neuropeptide S precursors in multiple lineages. Therefore, based on the data available at present, it is not possible to draw firm conclusions on the nature of the relationship between the *Branchiostoma* NG peptide and neuropeptide S in tetrapods. However, further insights on this issue may emerge as more genome sequence data is obtained from chordate species.

4.3. The evolutionary origin of neurophysin-associated NG peptides

An important issue that invites explanation is how genes encoding neurophysin-associated NG peptides originated. Genes encoding precursors of vasopressin/oxytocin neuropeptides with a C-terminal neurophysin domain have been identified throughout the bilateria in both protostomian and deuterostomian invertebrate species (Van Kesteren et al., 1992; Stafflinger et al., 2008; Ukena et al., 2008; Elphick and Rowe, 2009), so the association of neurophysin with vasopressin/oxytocin type peptides is evolutionarily more ancient than the association of neurophysins with NG peptides. Therefore, the neurophysin-encoding exon of the NG peptide precursors in deuterostomian invertebrates may have originated following partial or complete duplication of a gene encoding a vasopressin/oxytocin-type precursor. Alternatively, the neurophysin-encoding exon of NG peptide precursors may have originated as a consequence of retrotranscription of a partial or full-length mRNA encoding a vasopressin/oxytocin-type neuropeptide precursor and incorporation of the resulting cDNA into the genome.

If NG peptide precursor genes evolved from a complete vasopressin/oxytocin precursor gene then one would expect to observe sequence similarity in N-terminal neuropeptide-containing regions of NG peptide precursors and vasopressin/oxytocin precursors. However, there is little if any such sequence similarity. Furthermore, one might also expect to see similarity in the exon/intron organisation of genes encoding NG peptides and vasopressin/oxytocin peptides. However, this is not the case; the genes are similar in having an intron that precedes the codon for the first residue of the neurophysin domain but the positions of other introns are different. In NG peptide

precursor genes there is an intron separating the exon encoding the signal peptide and NG peptides, whereas in vasopressin/oxytocin-type neuropeptide precursor genes the signal peptide and vasopressin/oxytocin peptide are encoded by a single exon (Ivell and Richter, 1984; Elphick and Rowe, 2009). Conversely, in vertebrate and invertebrate vasopressin/oxytocin-type neuropeptide precursor genes the neurophysin-encoding domain is typically interrupted by an intron, (Ivell and Richter, 1984; Stafflinger et al., 2008; Elphick and Rowe, 2009), whereas the neurophysin-encoding domain of NG peptide precursor genes is intronless. Collectively, these character differences argue against the hypothesis that the NG peptide precursor evolved from a complete vasopressin/oxytocin-type precursor gene. A more likely scenario for the origin of a NG peptide precursor gene would be fusion of neurophysin-encoding exon(s) with a gene encoding a neuropeptide precursor comprising a signal peptide encoding exon and a NG peptide encoding exon. Furthermore, the absence of an intron in the neurophysin-encoding exon of NG peptide genes suggests perhaps that this exon may have originated by retrotranscription of mRNA encoding a vasopressin/oxytocin-type neuropeptide precursor, followed by incorporation of a partial cDNA encoding a neurophysin domain into the genome at a position 3' to a gene encoding a NG peptide.

If this model is correct then, whilst the association of NG peptides with neurophysin may date back to the common ancestor of the deuterostomes, NG peptides may be evolutionarily even more ancient. It is interesting, therefore, that there is a family of neuropeptides in arthropods known as SIFamides (Verleyen et al., 2009) that have a conserved C-terminal sequence NGSIFamide that is structurally similar to NGFFFamide. As with the similarity between the putative *Branchiostoma* peptide SFRNGVamide and neuropeptide S, this structural similarity between

SIFamides and NGFFFamide may be a consequence of convergent evolution. However, if this structural similarity does reflect common ancestry, then the NG peptides identified here in deuterostomian invertebrates may represent an evolutionary link between a neuropeptide (SIFamide) involved in regulation of sexual arousal in *Drosophila* (Terhzaz et al., 2007) and an anxiolytic neuropeptide (neuropeptide S) that promotes arousal in mammals (Xu et al., 2004).

5. Conclusions

The data presented in this paper indicate that in an ancestor of the deuterostomes a gene encoding a precursor protein comprising a NG peptide acquired a 3' exon encoding a neurophysin-type protein. This presumably occurred as a consequence of duplication and translocation of the neurophysin-encoding exon(s) of a gene encoding a vasopressin/oxytocin-type neuropeptide precursor. Consequently, neurophysin became associated with, possibly for the first and only time in the evolutionary history of the animal kingdom, neuropeptides that are not members of the vasopressin/oxytocin family. The structural and functional significance of this association with respect to biosynthesis and release of NG peptides in echinoderms, hemichordates and cephalochordates now remains to be explored. Importantly, if neurophysins are required for biosynthesis of NG peptides, then investigation of this neuropeptide system in invertebrate deuterostomes may provide new insights on mechanisms of neurophysin-dependent neuropeptide biosynthesis and hence why mutations in the neurophysin-encoding exons of the vasopressin gene can result in the defects in vasopressin biosynthesis that cause diabetes insipidus.

Acknowledgements

I am grateful to the Baylor College of Medicine Human Genome Sequencing Center for providing online access to a draft assembly of the genome sequence of the acorn worm *Saccoglossus kowalevskii* (http://www.hgsc.bcm.tmc.edu/project-species-o-Acorn%20worm.hgsc).

References

Bromham, L.D. and Degnan, B.M.: Hemichordates and deuterostome evolution: robust molecular phylogenetic support for a hemichordate + echinoderm clade. Evol Dev 1 (1999) 166-71.

De Bree, F.M.: Trafficking of the vasopressin and oxytocin prohormone through the regulated secretory pathway. J Neuroendocrinol 12 (2000) 589-94.

De Bree, F.M. and Burbach, J.P.: Structure-function relationships of the vasopressin prohormone domains. Cell Mol Neurobiol 18 (1998) 173-91.

Elphick, M.R. and Rowe, M.L.: NGFFFamide and echinotocin: structurally unrelated myoactive neuropeptides derived from neurophysin-containing precursors in sea urchins. J Exp Biol 212 (2009) 1067-77.

Gwee, P.C., Tay, B.H., Brenner, S. and Venkatesh, B.: 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 (2009) 47.

Inoue, M., Birenheide, R., Koizumi, O., Kobayakawa, Y., Muneoka, Y. and Motokawa, T.: Localization of the neuropeptide NGIWYamide in the holothurian nervous system and its effects on muscular contraction. Proc Biol Sci 266 (1999).

Ivell, R. and Richter, D.: Structure and comparison of the oxytocin and vasopressin genes from rat. Proc Natl Acad Sci U S A 81 (1984) 2006-10.

Iwakoshi, E., Ohtani, M., Takahashi, T., Muneoka, Y., Ikeda, T., Fujita, T., Minakata, H. and Nomoto, K.: Comparative aspects of structure and action of bioactive peptides isolated from the sea cucumber *Stichopus japonicus*. In: Ohno, M. (Ed.), Peptide Chemistry 1994. Protein Research Foundation, Osaka, 1995, pp. 261-264.

Kato, S., Tsurumaru, S., Taga, M., Yamane, T., Shibata, Y., Ohno, K., Fujiwara, A., Yamano, K. and Yoshikuni, M.: Neuronal peptides induce oocyte maturation and gamete spawning of sea cucumber, Apostichopus japonicus. Dev Biol 326 (2009) 169-76.

Kawada, T., Sekiguchi, T., Itoh, Y., Ogasawara, M. and Satake, H.: Characterization of a novel vasopressin/oxytocin superfamily peptide and its receptor from an ascidian, *Ciona intestinalis*. Peptides 29 (2008) 1672-8.

Legros, J.J. and Geenen, V.: Neurophysins in central diabetes insipidus. Horm Res 45 (1996) 182-6.

Putnam, N.H., Butts, T., Ferrier, D.E., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.K., Benito-Gutierrez, E.L., Dubchak, I., Garcia-Fernandez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin, I.T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W., Satoh, N. and Rokhsar, D.S.: The amphioxus genome and the evolution of the chordate karyotype. Nature 453 (2008) 1064-71.

Reinscheid, R.K.: Phylogenetic appearance of neuropeptide S precursor proteins in tetrapods. Peptides 28 (2007) 830-7.

Roth, A.L., Marzola, E., Rizzi, A., Arduin, M., Trapella, C., Corti, C., Vergura, R., Martinelli, P., Salvadori, S., Regoli, D., Corsi, M., Cavanni, P., Calo, G. and Guerrini, R.: Structure-activity studies on neuropeptide S - Identification of

the amino acid residues crucial for receptor activation. Journal of Biological Chemistry 281 (2006) 20809-20816.
Stafflinger, E., Hansen, K.K., Hauser, F., Schneider, M., Cazzamali, G., Williamson, M. and Grimmelikhuijzen, C.J.: Cloning and identification of an oxytocin/vasopressin-like receptor and its ligand from insects. Proc Natl Acad Sci U S A 105 (2008) 3262-7.
Terhzaz, S., Rosay, P., Goodwin, S.F. and Veenstra, J.A.: The neuropeptide SIFamide modulates sexual behavior in Drosophila. Biochem Biophys Res Commun 352 (2007) 305-10.
Ukena, K., Iwakoshi-Ukena, E. and Hikosaka, A.: Unique form and osmoregulatory function of a neurohypophysial hormone in a urochordate. Endocrinology 149 (2008) 5254-61.
Van Kesteren, R.E., Smit, A.B., De With, N.D., Van Minnen, J., Dirks, R.W., Van Der Schors, R.C. and Joosse, J.: A vasopressin-related peptide in the mollusc *Lymnaea stagnalis*: peptide structure, prohormone organization, evolutionary and functional aspects of *Lymnaea* conopressin. Prog Brain Res 92 (1992) 47-57.
Verleyen, P., Huybrechts, J. and Schoofs, L.: SIFamide illustrates the rapid evolution in Arthropod neuropeptide research. Gen Comp Endocrinol 162 (2009) 27-35.
Xu, Y.L., Reinscheid, R.K., Huitron-Resendiz, S., Clark, S.D., Wang, Z., Lin, S.H., Brucher, F.A., Zeng, J., Ly, N.K., Henriksen, S.J., de Lecea, L. and Civelli, O.: Neuropeptide S: a neuropeptide promoting arousal and anxiolytic-like effects. Neuron 43 (2004) 487-97.

Figure Legends

Fig. 1. The *Saccoglossus kowalevskii* NGFWNamide/NGFYNamide precursor. A. The predicted coding sequence of a gene (lowercase) that encodes the NGFWNamide/NGFYNamide precursor protein (uppercase) is shown. The positions of introns in the gene are shown by highlighting the pairs of bases (bold and underline) in the sequence that are interrupted by an intron. The predicted signal peptide is shown in italics and underlined. The five copies of the NGFWNG sequence and the single NGFYNG sequence are underlined, flanked by putative dibasic cleavage sites (KR or KK), which are shown in bold. The fourteen cysteine residues of the C-terminal neurophysin domain of the precursor are shown in bold and underlined. The asterisk shows the position of the stop codon. B. Alignment of the sequences of the spacer peptides that follow the KRNGFWNGKR or KKNGFWNGKR or KRNGFYNGKR sequences (underlined) in the NGFWNamide/NGFYNamide precursor reveals sequence similarity (identical residues are shown in bold) indicative of intragenic duplication. The amino acid residue positions in the precursor are shown for the N- and C-terminal residues of each peptide segment.

Fig. 2. The *Branchiostoma* SFRNGVamide precursor. The predicted coding sequence of a gene (lowercase) that encodes the SFRNGVamide precursor protein (uppercase) is shown. The positions of introns in the gene are shown by highlighting the pairs of bases (bold and underline) in the sequence that are interrupted by an intron. The predicted signal peptide is shown in italics and underlined. The two copies of the SFRNGVG sequence are underlined, flanked by putative dibasic cleavage sites (KR), which are shown in bold. The fourteen cysteine residues of the C-terminal

neurophysin domain of the precursor are shown in bold and underlined. The asterisk shows the position of the stop codon.

Fig. 3. Multiple sequence alignment of the *Strongylocentrotus purpuratus* (*Sp*) NGFFFamide precursor, the *Saccoglossus kowalevskii* (*Sk*) NGFWNamide/NGFYNamide precursor and the *Branchiostoma floridae* (Bf) SFRNGVamide precursor. Signal peptide sequences are shown in blue, neuropeptides are shown in red flanked by putative dibasic cleavage sites (green) and the neurophysin domain is shown in purple. Underlined pairs of amino acids show the similarity in the positions of introns in the genes encoding these proteins. The symbol * shows the positions of residues that are identical in all of the sequences whilst the symbols : and . show the positions of strongly and weakly conserved residues, respectively. This clustalw2 alignment shows that there are much higher levels of sequence conservation in the C-terminal neurophysin domains than in the N-terminal neuropeptide-containing domains. This is partly due to lineage-specific intragenic duplication of neuropeptide-encoding segments in the N-terminal domains, which has caused sequence divergence. Therefore, sequence identity is restricted to just a few residues in the N-terminal domain and some of these amino acid alignments may be arbitrary.

Fig. 4. Comparison of the sequences of NG peptides in invertebrate deuterostomes with neuropeptide S in vertebrates and SIFamides in arthropods.

NG peptides from invertebrate deuterostomes (shown in grey shaded box) share sequence similarity with the C-terminal region of arthropod SIFamide neuropeptides (shown above the grey shaded box) or with N-terminal region of neuropeptide S in vertebrates (shown below the grey shaded box). A shared feature of all the peptides is the Asn-Gly (NG) dipeptide sequence, which is shown in bold. Full names and common names of the species listed are: *Drosophila melanogaster* (fruit fly), *Procambarus clarkii* (red swamp crayfish), *Strongylocentrotus purpuratus* (purple sea urchin), *Apostichopus japonicus* (Japanese sea cucumber), *Saccoglossus kowalevskii* (acorn worm), *Branchiostoma floridae* (Florida lancelet), *Xenopus (Silurana) tropicalis* (western clawed frog), *Anolis carolinensis* (green anole lizard), *Gallus gallus* (chicken) and *Homo sapiens* (human).

Figure 1

Figure 1

A

```
1   atgcttagaaagatcaatgcagtggtctttttcctcgtggctatatgtaccctgtccaga
     M  L  R  K  I  N  A  V  V  F  F  L  V  A  I  C  T  L  S  R   20
61  gcaacatttggtgaagatggcatgaccgagaagcaggtacttaagttgcacaaatattgg
     A  T  F  G  E  D  G  M  T  E  K  Q  V  L  K  L  H  K  Y  W   40
121 ccaaaagaagatatctcagaattaggaagttcatcaacctcgggggatagtggcgagaac
     P  K  E  D  I  S  E  L  G  S  S  S  T  S  G  D  S  G  E  N   60
181 gaagcagtgaaaatgggattctgggctgttgatggaaaacgaaatgggttttggaatgga
     E  A  V  K  M  G  F  W  A  V  D  G  K  R  N  G  F  W  N  G   80
241 aaacgcaatggtttctggaatggcaagaggaattttgatgatctgagaatatttggaaca
     K  R  N  G  F  W  N  G  K  R  N  F  D  D  L  R  I  F  G  T  100
301 aataaaaaacattcaagcatcaacaatcaaaaaggaatggtttttggaacggaaagcgg
     N  K  K  H  S  S  I  N  N  Q  K  R  N  G  F  W  N  G  K  R  120
361 aattttgatgattttgaaataccggagaagaaggaacaaccacactggagagatgaaaaa
     N  F  D  D  F  E  I  P  E  K  K  E  Q  P  H  W  R  D  E  K  140
421 aagaatgggttctggaacgggaaaaggaattttgacaactttaaaacagatgacctgcaa
     K  N  G  F  W  N  G  K  R  N  F  D  N  F  K  T  D  D  L  Q  160
481 taccctagtatagaagacaaaaggaatggcttttggaatgggaaaaggaactttgaaatg
     Y  P  S  I  E  D  K  R  N  G  F  W  N  G  K  R  N  F  E  M  180
541 tctatgaataagaaggcgtccgcaagtactgaaagtgaaaaaagaaacgggttctataat
     S  M  N  K  K  A  S  A  S  T  E  S  E  K  R  N  G  F  Y  N  200
601 ggcaagagaagtgttgataacacaatgaacatatcttcatatcactataaagaatcagca
     G  K  R  S  V  D  N  T  M  N  I  S  S  Y  H  Y  K  E  S  A  220
661 aaaaagtgtaccacatgtggaccaggaggcaagggacaatgtgttatgtacggagtatgc
     K  K  C  T  T  C  G  P  G  G  K  G  Q  C  V  M  Y  G  V  C  240
721 tgtagtcttgaaattggttgttcaatgcttaccaaagaaaccgaagagtgtacgactagt
     C  S  L  E  I  G  C  S  M  L  T  K  E  T  E  E  C  T  T  S  260
781 cctctcgttggtgaatgcggtagaagcgacgttcaatgtggtaacggaggaagatgcgtt
     P  L  V  G  E  C  G  R  S  D  V  Q  C  G  N  G  G  R  C  V  280
841 gctaatggcgtatgttgcaccaaagagacccagtcatgtaaaattgaccaggagtgtaac
     A  N  G  V  C  C  T  K  E  T  Q  S  C  K  I  D  Q  E  C  N  300
901 gttaggtggtag
     V  R  W  *                                                  303
```

B

```
81   KRNGFWNGKRNFDDLRIFGTNKKHSSINNQ  110
111  KRNGFWNGKRNFDDFEIP-EKKEQPHWRDE  139
140  KKNGFWNGKRNFDNFKTD--DLQYPSIED-  166
167  KRNGFWNGKRNFEM-SMN--KKASASTESE  193
194  KRNGFYNGKRSVDN-TMN--ISSYHYKESA  220
```

Figure 2

Figure 2

```
1   atgatgcagactccaatcttcctgtgttctgttgtcctcgtcggagctgtttgcgggcaa
     M  M  Q  T  P  I  F  L  C  S  V  V  L  V  G  A  V  C  G  Q   20
61  ctgtcggagacgaacaactttccacaaaacggcaataggcggctgtcccctgagcgctcg
     L  S  E  T  N  N  F  P  Q  N  G  N  R  R  L  S  P  E  R  S   40
121 gcgaccgtgctacgacagtttctccacctggaaggggcggtcggctcgccagtttctccg
     A  T  V  L  R  Q  F  L  H  L  E  G  A  V  G  S  P  V  S  P   60
181 tccgacgggcgggccctggaaaccggagacaagagaagcttccggaacggtgtggggaaa
     S  D  G  R  A  L  E  T  G  D  K  R  S  F  R  N  G  V  G  K   80
241 agaagggacagcgaggaggaacgacttccacagaaccgcggagcaaccgagttgaaggct
     R  R  D  S  E  E  E  R  L  P  Q  N  R  G  A  T  E  L  K  A  100
301 gaggccaccattttctcgcaaaacggagaccctcacgacgagggtgcaaaggcagcctca
     E  A  T  I  F  S  Q  N  G  D  P  H  D  E  G  A  K  A  A  S  120
361 gagaagcggtcgttccgcaacggagttggaaaacgaacccactttcgcattgtcgcagac
     E  K  R  S  F  R  N  G  V  G  K  R  T  H  F  R  I  V  A  D  140
421 gcgtcactgggtggactggatgaaccggaagcacttcgtcaaacgggaggcgacgcaaac
     A  S  L  G  G  L  D  E  P  E  A  L  R  Q  T  G  G  D  A  N  160
481 agtccgtcactgtcacgtgatctatgggcggaagttcagggaaaagacgatgaacagtgt
     S  P  S  L  S  R  D  L  W  A  E  V  Q  G  K  D  D  E  Q  C  180
541 cccgcctgtggatcagacggcagcggtgtgtgcgtgctgaaaggagtgtgctgccgcctt
     P  A  C  G  S  D  G  S  G  V  C  V  L  K  G  V  C  C  R  L  200
601 gactccggctgtgtgctacggaaggacgtctgctccagtcttcccgaccgcgccctctgc
     D  S  G  C  V  L  R  K  D  V  C  S  S  L  P  D  R  A  L  C  220
661 gccagccttcagtacagcgccacctgtcggacggacgggaagtgcgtcgcgccgggagtc
     A  S  L  Q  Y  S  A  T  C  R  T  D  G  K  C  V  A  P  G  V  240
721 tgttgccgtgccgctgaccactcttgcttcctggaccccgagtgcgactag
     C  C  R  A  A  D  H  S  C  F  L  D  P  E  C  D  *           256
```

Figure 3

Figure 3

Figure 4

Figure 4

Peptide	Sequence	Source
SIFamide	AAYRKPPF**NG**SIFa	*Drosophila* (Arthropoda)
SIFamide	AGYRKPPF**NG**SIFa	*Procambarus* (Arthropoda)
NGFFFamide	**NG**FFFa	*Strongylocentrotus* (Echinodermata)
NGIWYamide	**NG**IWYa	*Apostichopus* (Echinodermata)
NGFWNamide	**NG**FWNa	*Saccoglossus* (Hemichordata)
NGFYNamide	**NG**FYNa	*Saccoglossus* (Hemichordata)
SFRNGVamide	SFR**NG**Va	*Branchiostoma* (Cephalochordata)
Neuropeptide S	SFR**NG**VGSGIKKNSFRRAKL	*Xenopus* (Vertebrata)
Neuropeptide S	SFR**NG**VGSGMKKTSFRRARL	*Anolis* (Vertebrata
Neuropeptide S	SFR**NG**VGSGIKKTSFRRAKS	*Gallus* (Vertebrata)
Neuropeptide S	SFR**NG**VGTGMKKTSFQRAKS	*Homo* (Vertebrata)