Gene sequence and related structure of neuropeptides in invertebrates

M Mandrioli, E Ottaviani
Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy

Abstract

Numerous neuropeptides were studied in different invertebrate species, but the presence of most of them was evaluated only by an immunocytochemical approach using antibodies against the vertebrate homologues. As a consequence, several authors referred to the presence of neuropeptide-like molecules. In view of the availability of numerous wholly sequenced invertebrate genomes, here we reviewed the data on the gene sequence and the related structure of four neuropeptides (corticotrophin-releasing hormone, tachykinins, bombesin and insulin) in invertebrates, analyzing their functions in comparison to what reported in vertebrates.

Key Words: neuropeptides; invertebrates; gene sequence; structure; functions

Introduction

In vertebrates the term neuropeptide refers to all peptides that are produced by neurons and by the pituitary gland. The same terminology has been used also in invertebrates. As known, in the latter the anterior pituitary gland is lacking and the peptides are produced mainly by neurons and immunocytes (Ottaviani and Franceschi, 1997; Tascedda and Ottaviani, 2016).

Numerous studies have shown that the neuroendocrine and the immune systems are extremely interconnected. In invertebrates and vertebrates, both systems produce several soluble mediators, including hormones, neurotransmitters, cytokines and peptides. Immune stimuli induce the immunocytes to synthesize neuropeptides, which in turn may influence the activity of the neuroendocrine system. So that, a peptide, such as adrenocorticotropic hormone, could be considered an hormone or a neurotransmitter depending on the target which is involved. These findings suggest that nature followed the same general strategy for the construction of the immune and the neuroendocrine systems (Ottaviani and Franceschi, 1997; Tascedda and Ottaviani, 2016).

In invertebrates, the presence of numerous neuropeptides was studied (Boer et al., 1979; Marchand et al., 1989; Ottaviani and Cossarizza, 1990; Sonetti et al., 1990; Ottaviani et al., 1997; Tascedda and Ottaviani, 2016), but for some of them it was impossible to define the gene sequence. The only evidence was determined by an immunocytochemical approach using antibodies against the vertebrate homologues. Additionally, functional tests using the corresponding vertebrate molecules revealed an overlapping with the vertebrate responses. An example of such behaviour is represented by the pro-opiomelanocortin (POMC) molecules. In this context, the evidence of gene expression was only detected in the parasitic flat worm, Schistosoma mansoni (Duvaux-Miret et al., 1990), in the leech, Theromyzon tessulatum (Salzet et al., 1997) and in bivalve mollusc, Mytilus edulis (Stefano et al., 1999). Regardless of negative molecular data, results of extreme interest were obtained from functional studies supporting not only the presence of peptides or similar molecules, but also their involvement in both immune and neuroendocrine responses, as in vertebrates (Weigent and Blalock, 1987).

In this perspective, we reviewed data on the gene sequence and the related structure of four neuropeptides (corticotrophin-releasing hormone, tachykinins, bombesin and insulin) in invertebrates, comparing their functions to those of the vertebrate homologues.

Corticotrophin-releasing hormone (CRH) gene and related peptide

CRH is an ancient regulatory molecule found in fish, amphibians, and birds, as well as in mammals and it shows a remarkable degree of conservation (Seasholtz et al., 2002). In both mammalian and non-mammalian vertebrates, central CRH plays multiple roles in regulating and coordinating the
responses to external and internal challenges to viability, both through its effects on the pituitary gland and as a neurotransmitter/neuromodulator (Seasholtz et al., 2002).

The structure of the pre-procorticotrophin-releasing hormone (pre-proCRH) generally consists of various domains including the N-terminal proCRH (amino acids 27 - 122), proCRH (125 - 151) and corticotrophin-releasing hormone (1 - 41) (Brar et al., 1997; Perone et al., 1998). CRH (also referred as corticoliberin) is then cleaved from pre-proCRH (125 - 194) by the action of specific endopeptidases (Perone et al., 1998). The DNA sequence of the CRH gene has been studied in several vertebrates, such as pig, fish, mouse and human cells (Robinson et al., 1989), whereas to date, a single CRH gene was isolated in invertebrates and in particular in the cabbage moth Mamestra brassicae (Malagoli et al., 2002).

Human, rodent, carnivore and equid CRH are identical in amino-acid sequences, and the identity between mammalian and fish CRH ranges from 75 % in tilapia and salmon to 95 % in the suckerfish (Seasholtz et al., 2002; Power and Schulkin, 2006).

The comparison between vertebrate proCRH amino acid sequences and the M. brassicae proCRH (Figs 1, 2) shows a similarity of 36 % between the insect vs vertebrate peptides. The most conserved region ranges from amino acids 41 - 90 of the human sequence, corresponding to a portion of the N-terminal proCRH. Furthermore, comparison of the vertebrate CRH sequence with the insect homologue shows a conserved position of the cleavage sites usually denoted by pairs of dibasic

---

**Fig. 1** Alignment showing the presence of few and short highly conserved regions between the invertebrate and vertebrate proCRH. HS: Homo sapiens. SS: Sus scrofa; XL: Xenopus laevis; MB: M. brassicae; DR: Danio rerio.

---

**Fig. 2** Reconstruction of the relationship amongst the unique invertebrate proCRH compared to some the most studied vertebrate corticoliberins. HS: H. sapiens; SS: S. scrofa; XL: X. laevis; MB: M. brassicae; DR: D. rerio.
Fig. 3 Alignment showing the absence of highly conserved regions between the invertebrate and vertebrate tachykinins. HS: H. sapiens; MM: Mus musculus; CI: Ciona intestinalis; DM: D. melanogaster; BD: Bactrocera dorsalis; BI: Bombus impatiens; PC: Polistes Canadensis; TC: T. castaneum; CE: C. elegans.

amino acids. It should be emphasised that a BLAST analysis performed against the highly conserved region of the insect putative peptide revealed the highest degree of homology with proCRH sequences of vertebrate species, such as human, rat, sheep, frog and fish. Furthermore, the comparison of the insect CRH (1 - 41) putative fragment with the human and Tilapia mossambica CRH shows a similarity of 34.2 and 30.8 %, respectively. This moderate sequence identity is not surprising as the evolutionary distance between insects and vertebrates is estimated at between 700 and 993 million years (Gu, 1998).

According to several authors, CRH derived from a molecule that was found in the common ancestor of vertebrates (Lovejoy and Balment, 1999; Power and Schulkin, 2006). The descendants of that original molecule include CRH, the urocortins (found in mammals), sauvagine (found in frogs), and urotensin I (found in fish). The urocortins are related to sauvagine and urotensin-I (Seasholtz et al., 2002), which implies that the separation of CRH from the urocortins likely occurred before the separation of mammals from other vertebrates. However, the presence of a proCRH-like gene in the cabbage moth and the occurrence of an urotensin-I like peptide in the nematode Caenorhabditis elegans suggest that the origin of this peptide hormone family could be more ancient than previously supposed. Furthermore, the presence of remarkably conserved CRH-binding proteins in insect genomes (Huising and Flik, 2005; Westphal...
and Seasholtz 2006; Mandrioli et al., 2007; Liu et al., 2011), and the occurrence of the insect diuretic hormone I and its receptors sharing similarities with the vertebrate CRH hormone system (Huising and Flik, 2005) add substantial weight to the supposition that the CRH system probably evolved in a common ancestor of insects and vertebrates (Malagoli et al., 2002; Huising and Flik, 2005).

**Fig. 4** Phylogenetic tree showing the relation amongst the various invertebrate tachykinins compared to the vertebrate homologues. HS: H. sapiens; MM: M. musculus; CI: C. intestinalis; DM: D. melanogaster; BD: B. dorsalis; BI: B. impatiens; PC: P. canadensis; TC: T. castaneum; CE: C. elegans.

**Tachykinin gene and related peptides**

Substance P (SP), the first member of the tachykinin family of peptides (also known as neurokinins), has been called a "pioneering neuropeptide", since knowledge gained from studies of tachykinins has informed our understanding of many neuropeptides (Steinhoff et al., 2014). Indeed, the discovery of SP as an activity in extracts of horse brain and intestine with effects on intestinal contractility and blood pressure marked the identification of the first of many "brain-gut neuropeptides", which are present in enteric neurons and entero-endocrine cells as well as in neurons of the brain (Von Euler and Gaddum, 1931).

SP belongs to a large family of structurally related peptides, the tachykinins, that derive from alternative processing of three Tac genes (Steinhoff et al., 2014). The tachykinins interact with three neurokinin receptors (NKRs) encoded by three Tacr genes. Knowledge of the structure, function, signalling, and trafficking of these receptors has guided studies of other G protein-coupled receptors (GPCRs) and, in this sense, the NKRs may be considered "pioneering receptors."

The tachykinins are expressed throughout the nervous and immune systems, regulate an extraordinarily diverse range of physiological processes, and have been implicated in important pathological conditions (Steinhoff et al., 2014).

In vertebrates tachykinin and the tachykinin-related peptides (TRPs) form a group of ancestral neuropeptides that are found in a wide range of animals, from octopus to human (Severini et al., 2002; Satake et al., 2003; Zhou et al., 2012).

In insects, immunoreactivity against vertebrate tachykinins has been demonstrated multiple times, both in neuronal and intestinal tissues (Verhaert and De Loof, 1985; Nässel et al., 1990). Successive molecular studies on insect neuropeptides and their GPCRs have described TRPs and two GPCRs as the receptors for the TRPs in D. melanogaster and other insect species (Schoof et al., 1990a, b; Nässel, 2002; Nässel and Winther, 2010; Van Loy et al., 2010; Steinhoff et al., 2014) (Figs 3, 4). The insect multiple paracopies of the TRP gene contain the C-terminal FxGxRamide motif, whereas vertebrate tachykinins typically contain the FxGLMamide motif, whereas vertebrate tachykinins typically contain the FxGLMamide motif (Schoof et al., 1990a, b; Nässel, 2002; Nässel and Winther, 2010; Van Loy et al., 2010; Steinhoff et al., 2014). Two closely related TRP receptors (TRPRs) in D. melanogaster were described previously: *Drosophila* tachykinin receptor and neurokinin K receptor. These receptors were identified using a hybridization-based homology search followed by functional assays (Li et al., 1991; Monnier et al., 1992). In a subsequent study, however, NKD activity was not recapitulated with the typical TRPs, whereas DTKR was activated by the TRPs of *D. melanogaster* (Poels et al., 2007, 2009).

An arthropod-specific peptidergic system, the neuropeptide designated here as natalisin and its receptor, was identified and investigated in *D. melanogaster*, *Tribolium castaneum* and *Bombyx mori* (Jiang et al., 2013). In all three species, natalisin expression was observed in 3 - 4 pairs of the brain neurons: the anterior dorso-lateral interneurons, inferior contralateral interneurons, and small pars intercerebralis neurons (Jiang et al., 2013). In *B. mori*, natalisin was also expressed in two additional pairs of contralateral interneurons in the subesophageal ganglion. Natalisin-RNAi and the activation or silencing of the neural activities in the natalisin-specific cells in *D. melanogaster* induced significant defects in the mating behaviours of both males and females. Knockdown of natalisin expression in *T. castaneum* resulted in significant reduction in the fecundity (Jiang et al., 2013). The
Fig. 5 Alignment of the bombesin-like molecules showing the high sequence identity among the insect peptides in respect to the poorly conserved sequence isolated in mollusc. HI: Haematobia irritans; SC: Stomoxys calatrans; DM: D. melanogaster; DR: Deroceras reticulatum; BV: Bombina variegata.

similarity of the natalisin C-terminal motifs to those of vertebrate tachykinins and of tachykinin-related peptides in arthropods led to the identification of the natalisin receptor. A G protein-coupled receptor, previously known as tachykinin receptor 86C (also known as the neurokinin K receptor of D. melanogaster), now has been recognized as a bona fide natalisin receptor. Taken together, the taxonomic distribution pattern of the natalisin gene and the phylogeny of the receptor suggest that natalisin is an ancestral sibling of tachykinin that evolved only in the arthropod lineage (Jiang et al., 2013).

During the past decade, it has become clear that the tachykinin family of peptides has been well preserved in a broad range of animal species belonging to different phylogenetic clades. However, there is a huge discrepancy between the efforts that have been performed to identify and isolate invertebrate tachykinin-related peptides and the detailed characterization of their corresponding receptors, a group of structurally related G protein-coupled receptors. Indeed, only five invertebrate receptors for TKRPs have been properly analyzed to date. Some of these receptors seem to display only moderate or no TK-like ligand specificity, while others can only be activated by specific peptide isoforms. Indeed, diverse in vitro cell-based signal transduction experiments have been employed to study signal transduction induced by distinct TKRPs. These studies also made clear that separate peptide isoforms are sometimes capable of inducing and/or stabilizing different receptor conformations upon binding that result in distinct receptor signalling properties.

Bombesin-related peptides and their genes in invertebrates

The bombesin-like neuropeptides, originally isolated from the frog skin (Erspamer et al., 1972; Erspamer, 1988), include bombesin, gastrin releasing peptide (GRP) and neuromedin B (NMB) and exert a wide variety of physiological actions in the CNS and the periphery through a class of related receptors (Sun et al., 2000; Gonzalez et al., 2008).

The first clear molecular evidence for bombesin/GRP signalling in invertebrates has been reported in D. melanogaster (Randall et al., 2001; Sano et al., 2015), where a bombesin-like peptide
CCHa2 might have a bilaterian evolution and they are absent from non-clearly assessed that they diversified during the bilaterian genomes. Currently available data could (Seung-Joon similarity corresponds to a conserved function possible to confirm that the observed sequence terrestrial molluscs, but a functional characterization (Gastropoda: Pulmonata), one of the most common a CCHa2-like molecule has also been observed in the gray garden slug homeostasis under volatile nutritional conditions. important role in the maintenance of energy metabolic regulator analogous to the mammalian appears that CCHa2 functions as a short-acting metabolic status (Sano that CCHa2 mediates relatively rapid changes in metabolic status. The expression of CCHa2 responds to yeast and glucose within 6 h, indicating that CCHa2 mediates relatively rapid changes in metabolic status (Sano et al., 2015). Thus, it appears that CCHa2 functions as a short-acting metabolic regulator analogous to the mammalian gut- or stomach-derived hormones described above, and that D. melanogaster CCHa2 might have an important role in the maintenance of energy homeostasis under volatile nutritional conditions. CCHa2 homologues have been successively identified in different insect species (Figs 5 - 6) and a CCHa2-like molecule has also been observed in the gray garden slug Deroceras reticulatum (Gastropoda: Pulmonata), one of the most common terrestrial molluscs, but a functional characterization of this molecule is still absent so that it is not possible to confirm that the observed sequence similarity corresponds to a conserved function (Seung-Joon et al., 2017).

Molecular analyses on bombesin receptors clearly assessed that they diversified during the bilateral evolution and they are absent from non-bilaterian genomes. Currently available data could therefore favour the understanding of the functional diversification of this ancient family of neuropeutal peptides and their receptors during the animal evolution.

Insulin/IGF signaling in Drosophila and other invertebrates

Insulin is probably one of the most extensively investigated peptide hormones, due to its critical role in the carbohydrate metabolism and thus importance in diabetes and obesity for humans (Claeys et al., 2002; Garofalo, 2002; Gronke et al., 2010; Antonova et al., 2012). Since its discovery (Banting and Best, 1922), insulin and insulin-like peptides have been identified in a large number of animals from invertebrates, such as nematodes, molluscs and insects, to chordates (Claeys et al., 2002; Garofalo, 2002; Gronke et al., 2010; Antonova et al., 2012). A single type of insulin is present generally in mammals, together with two insulin-like growth factors (IGFs) and one relaxin. These peptides display a variety of functions in different tissues both during development and in the mature organism.

In insects, varying numbers of well conserved insulin-like peptides (ILPs) have been identified in different species, ranging from one in the locusts, Locusta migratoria and Schistocerca gregaria, to 38 in the silkworm B. mori (Lagueux et al., 1990; Yoshida et al., 1998; Badisco et al., 2008; Mizoguchi and Okamoto, 2013; Veenstra, 2014). The classification of insect ILPs as insulin-like is based on similarities in the amino acid sequence of the mature peptides to those of mammalian insulins, especially the number and positions of cysteine residues (Brogiolo et al., 2001) (Figs 7 - 8). At present however the role of these peptides has been investigated in depth only in Drosophila where eight ILPs (DILP1-8), but only two receptors (dlnR and Lgr3), are known (Näs­sel and Vanden Broeck, 2016). DILP2, 3 and 5 are produced by a set of neurosecretory cells (IPCs) in the brain and their biosynthesis and release are controlled by a number of mechanisms differing between larvae and adults. Adult IPCs display cell-autonomous sensing of circulating glucose, coupled to evolutionarily conserved mechanisms for DILP release. The glucose-mediated DILP secretion is modulated by neurotransmitters and neuropeptides, as well as by factors released from the intestine and adipocytes. Larval IPCs, however, are indirectly regulated by glucose-sensing endocrine cells

![Fig. 6 Phylogenetic tree evidencing the stronger similarity of the insect bombesin-like molecules to the B. variegata bombesin in respect to the mollusc annotated homologue. HI: H. irritans; SC: S. calatans; DM: D. melanogaster; DR: D. reticulatum; BV: B. variegata.](image-url)
producing adipokinetic hormone, or by circulating factors from the intestine and fat body. Furthermore, IIS is situated within a complex physiological regulatory network that also encompasses the lipophilic hormones, 20-hydroxyecdysone and juvenile hormone. After release from IPCs, the ILP action can be modulated by circulating proteins that act either as protective carriers (binding proteins), or competitive inhibitors. Some of these proteins appear to have additional functions that are independent of ILPs. Taken together, the signalling with multiple ILPs is under complex control, ensuring tightly regulated IIS in the organism.

The insulin signalling pathway may also play relevant role in the development of trait allometry in insects because levels of both insulin and growth factor signals are sensitive to larval nutrition, and because these signal levels affect overall rates of cell proliferation in imaginal discs during the period of disc growth (Emlen et al., 2006). In particular the insulin pathway seems to be associated with the development of beetle horns and it has been suggested that a differential response to insulin could explain why imaginal discs exhibit nutrition-dependent (plastic) variation in growth and final trait sizes, whereas other discs are less sensitive to insulin signals (Emlen et al., 2006). Stated another way, traits sensitive to insulin signals should display steep and positive allometries in natural populations (trait size tightly correlated with among-individual variation in body size), and traits insensitive to insulin signals should have shallow/flat allometry slopes (trait size not correlated with variation in body size).
Concluding remarks

Neuropeptides represent the largest single class of signal compounds and are involved in regulation of development, growth, reproduction, metabolism and behaviour of invertebrates (Altstein and Nässel, 2010). Over the last forty years there has been a tremendous increase in our knowledge of neuropeptide signalling in different invertebrate species mainly though the immunocytochemical approach using antibodies against the vertebrate homologues of the studied neuropeptide (Ottaviani and Franceschi, 1997; Tascedda and Ottaviani, 2016).

In the last decade, the fast improvement in the techniques for genome sequencing, peptidomic analysis, receptor characterization and targeted gene interference (combined with physiological and behavioural analyses) prompted a revision of the previously published results in order to update our knowledge about the presence of genes coding for neuropeptides in invertebrates.

As reported in our present review, the signalling pathways based on CRH, tachykinins, bombesin and insulin are highly conserved not only in vertebrates, but also in molluscs and insects. Interestingly, currently available molecular data could suggest that some of these neuropeptides could be more ancient than previously suggested.

On the contrary, we have not found any trace of homologous genes in invertebrates for some neuropeptides, such as gastrin, enkephalin and vasoactive intestinal polypeptide, whose presence has been suggested in several papers by immunocytochemical procedures or by functional tests. Regarding the apparently controversial correlation between molecular evidence and the reported immunoreactivity for these peptides up till now the suggested explanations are scarce. Our analyses evidenced for some of these molecules, such as VIP and enkephalin, the presence of invertebrate peptides that share some similarity to the vertebrate homologues, but lack the functional domains typically observed in these neuropeptides. These results can explain the positive immunoreactivity observed using antibodies developed to bind the vertebrate VIP and enkephalin and could also justify the use of the term VIP- and encephalin-like molecules recurrently used in literature.

References
Gu X. Early metazoan divergence was about 830 million years ago. J. Mol. Evol. 47: 369-37, 1998.
Huising MO, Flik G. The remarkable conservation of corticotropin-releasing hormone (CRH)-binding protein in the honeybee (Apis mellifera) dates the CRH system to a common ancestor of


Schoofs L, Holman GM, Hayes TK, Kochansky JP, Nachman RJ, De Loof A. Locusta tachykinin III and IV: Two additional insect neuropeptides with homology to peptides of the vertebrate
Veenstra JA. The contribution of the genomes of a termite and a locust to our understanding of insect neuropeptides and neurohormones. Front. Physiol. 5: 454, 2014.