General and Comparative Endocrinology 171 (2011) 237–244

Contents lists available at ScienceDirect

General and Comparative Endocrinology

journal homepage: www.elsevier.com/locate/ygcen

# Identification of members of the gonadotropin-releasing hormone (GnRH), corticotropin-releasing factor (CRF) families in the genome of the holocephalan, *Callorhinchus milii* (elephant shark)

Tanya G. Nock, Dhan Chand, David A. Lovejoy *

*Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, Canada M5G 3G5*

A R T I C L E  I N F O

*Article history:*
Received 21 July 2010
Revised 25 January 2011
Accepted 2 February 2011
Available online 15 February 2011

*Keywords:*
Peptide structure evolution
GnRH
CRF
HPA axis
Reproduction
Stress
Chondrichthyes
Gnathostomes
Vertebrates

A B S T R A C T

The gonadotropin-releasing hormone (GnRH) and corticotropin-releasing family (CRF) are two neuropeptides families that are strongly conserved throughout evolution. Recently, the genome of the holocephalan, *Callorhinchus milii* (elephant shark) has been sequenced. The phylogenetic position of *C. milii*, along with the relatively slow evolution of the cartilaginous fish suggests that neuropeptides in this species may resemble the earliest gnathostome forms. The genome of the elephant shark was screened, *in silico*, using the various conserved motifs of both the vertebrate CRF paralogs and the insect diuretic hormone sequences to identify the structure of the *C. milii* CRF/DH-like peptides. A similar approach was taken to identify the GnRH peptides using conserved motifs in both vertebrate and invertebrate forms. Two CRF peptides, a urotensin-1 peptide and a urocortin 3 peptide were found in the genome. There was only about 50% sequence identity between the two CRF peptides suggesting an early divergence. In addition, the urocortin 2 peptide seems to have been lost and was identified as a pseudogene in *C. milii*. In contrast to the number of CRF family peptides, only a GnRH-II preprohormone with the conserved mature decapeptide was found. This confirms early studies about the identity of GnRH in the Holocephali, and suggests that the Holocephali and Elasmobranchii differ with respect to GnRH structure and function.

## 1. Introduction

The recent sequencing of the holocephalan, *Callorhynchus milii* (elephant shark) represents the first of the chondrichthyan genomes to be characterized [35]. Holocephalans represent an early-derived lineage distinct from the much larger elasmobranch lineage within the cartilaginous fish [27,28–30]. Moreover, they possess one of the smallest genomes among species of the Chondrichtyes making this taxon an attractive model for genome studies [7,35]. Chondrichthyan fishes, in general, have a median genome size of over 7 Gb or about twice the size of the human genome making genomic studies for most cartilaginous fish particularly difficult. Cartilaginous fish are the most basal of the gnathostomes and are, therefore, a sister taxon to all other jawed vertebrates. Moreover, their genomes are more highly conserved than those of tetrapods or teleosts [25], and thus these genomes are likely to provide evidence on the structure of ancestral genes.

Because only a paucity of information was known about the genomes of cartilaginous fish, and indeed, the genomes of ancestral chordates, much of the understanding of the neuropeptide evolution is based on what was known in sarcopterygian (lobe-finned) and actinopterygian (ray-finned) lineages. However, the use of these lineages to understand basal vertebrate genome evolution is limited, because numerous genes have been lost or silenced, have undergone local genome duplications, or are highly derived. Thus, as more genomes are characterized, then gaps in knowledge from other genomes can be filled in by the development of a vertebrate-wide model of particular gene systems.

Holocephalans, or chimaerans (Order Chimaeriformes), are an enigmatic lineage of fishes that include ratfishes, rabbitfishes and elephant fishes. There are only about 35–40 holocephalan species worldwide. Since their initial description, their phylogeny has been debated [10]. Today, they are generally recognized as a separate subclass in the Class Chondrichthyes with the caveat that they evolved as a separate lineage early in the formation of the protochondrichthyan class of fishes. However, features such as single gill slits and a fin structure resembling an actinopterygian design is suggestive of a lineage that evolved later than the Chondrichthyes. Regardless of their exact placement in vertebrate evolution, they have had over 400 million years of independent evolution. Because of this, they may represent a particularly important transition stage during vertebrate evolution.

Gene expansion events leading to the paralogous development of neuropeptide genes leads to the formation of new neuroendocrine pathways that ultimately serve to favor adaptation of species

---

* Corresponding author. Fax: +1 416 978 8532.
*E-mail address:* david.lovejoy@utoronto.ca (D.A. Lovejoy).

0016-6480/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygcen.2011.02.001

to particular niches and habitats. Central to these adaptations is the interaction of the physiological stress response with reproduction. Pivotal to these systems are the functions associated with the corticotropin-releasing factor (CRF) and gonadotropin-releasing hormone (GnRH) families, respectively.

GnRH is a critical neuropeptide involved in reproductive development and function in vertebrates. It is released via the hypothalamic–hypophysial portal system to regulate the synthesis and release of pituitary gonadotropins which ultimately trigger steriodogenesis and stimulate gonadal maturation. Two forms, GnRH-I and GnRH-II, have been characterized in vertebrates and a third form, GnRH-III, appears to occur exclusively in teleosts [3,23,24]. However, in cartilaginous fish, depending on the species as many as five or seven different forms of GnRH have been postulated to exist [32]. However, in contrast, only a single form of GnRH (GnRH-II) has been confirmed in a holocephalan [18], although other immunoreactive forms have been found [22], suggesting that the control of reproduction in this lineage may differ fundamentally from other vertebrates. Similarly, the CRF family is well conserved among the jawed vertebrates but, only a single sequence of urotensin-1 (urocortin) is known from the dogfish *Scyliorhinus canicula* [36]. However, our recent finding of only a single form of a CRF-like peptide from the tunicate, *Ciona intestinalis* [16], in comparison to the four paralogous forms of CRF found in gnathostomes [15,17], suggests that four forms should be found in cartilaginous fish, consistent with the postulated two rounds of genome duplications that have occurred in the vertebrates (2R hypothesis). Moreover, given the conservation of the genome sequence within the cartilaginous fish, we might expect these forms to resemble the ancestral forms of the peptide found in vertebrates. More importantly, however, identification of all CRF and GnRH paralogues in the holocephalan and elasmobranch lineages helps to formulate a model of how paralogous neuropeptide formation changes the interaction between different key physiological systems, in this case, stress and reproduction. For example, GnRH-I and CRF are neuroanatomically juxtaposed in the forebrain and evidence suggests a direct interaction, whereas GnRH-II and urotensin-I/urocortin are similarly situated nearby in the midbrain [12,34,33], suggesting that, as part of the expansion of these gene systems, a similar functional interaction was retained. However, before such anatomical and functional studies are performed, it is first necessary to identify the expressed paralogs. With the recent sequencing of the *C. milii* genome, a genomic window has opened on an important vertebrate lineage. Thus, utilizing this new model, we sought to identify which paralogs of the GnRH and CRF family were present in the genome in order to develop a model of HPA and HPG interaction in phylogenetically older vertebrates.

## 2. Materials and methods

The *C. milii* genome, available at the elephant shark genome project website (http://esharkgenome.imcb.a-star.edu.sg/), was screened, *in silico*, using the tBLASTn function in the BLAST algorithm bundled with this website. The query sequences used for the genome screening were based on the various conserved motifs for the GnRH [11,19], and CRF-family [16,17] sequences described previously. Once a putative peptide sequence was found that matched the characteristics of the peptide family, the base sequence was retrieved from the Elephant Shark Genome Project database with a unique accession number and translated using the Expasy Proteomics Server (http://expasy.org/). The sequences can be retrieved from the elephant shark website using their corresponding Accession Nos: GnRH, AAVX01008228.1; U1frag, AAVX01644813.1; U3, AAVX0129862.1; CRF, AAVX01042359.1; CRF2, AAVX01072774.1. The translated sequence was examined for the presence of a start codon upstream of the peptide sequence as well as the presence of a signal peptide, appropriate cleavage and amidation sites, and other translational products in their proper location such as the associated cryptic peptides of both CRF and GnRH. After analysis, the sequence was classified as a real or pseudogene.

Orthologous and paralogous GnRH and CRF sequences from a diverse range of species were obtained from the Pubmed website (http://www.ncbi.nlm.nih.gov). The accession numbers of each species are provided in Section 3. These sequences, including those found in the *C. milii* genome, were aligned by Clustal W analysis [8] (http://myhits.isb-sib.ch/cgi-bin/index), using the Blosum matrix with an opening gap penalty of 10, extending gap penalty of 0.05, end gap penalty of 10 and a separation gap penalty of 0.05. Regions of consensus and conservation were determined.

## 3. Results

Only a single form of GnRH was clearly found in the genome of *C. milii* (Fig. 1). It bears all of the canonical structures of a GnRH-II peptide including a methionine start codon, a 24-residue hydrophobic, leucine-rich signal peptide followed by the 10-residue mature peptide, amidation motif (GKR) and a variable, less conserved cryptic peptide of 39 residues. A comparison across diverse species of vertebrates shows that orthologs of the GnRH-II mature peptide and amidation motif are 100% conserved (Fig. 2). The presence of the amidation motif is important for full biological activity among many bioactive peptides. The remainder of the prepropeptide is poorly conserved except for the leucine-rich region of the signal peptides. There is an insert region occurring at residue 50, which occurs in tetrapods but not teleosts, also occurs in the *C. milii* sequence although it is only three residues in length. However, like teleosts, the *C. milii* sequence is truncated relative to the tetrapod sequence. We could not find clear evidence for any other GnRH genes in the genome of *C. milii*, despite using multiple search parameters incorporating sequence elements of other GnRH forms. A number of GnRH-like sequences were found, but none possessed the expected conserved structure of the preprohormone. One sequence with the compelling structure of QNWSHGLRPG (AAVX01043360.1), which at face value possesses all of the features of a mature GnRH, was present in a sequence of 21 amino acids bracketed by stop codons. Moreover, none of the theoretical translated protein sequences with frame-shifted alignments showed a protein similar to that expected of the GnRH family.

In contrast, four forms of the CRF/urocortin family were identified in the *C. milii* genome (Fig. 3). Two unique forms of CRF were

```
                 10        20        30        40        50
          ---------|---------|---------|---------|---------|

GnRH-II   MALQRNLLLLLLVLLAINTQVSRAQHWSHGWYPGGKRELGQAQTPEVSPV
          AFFFKFVGLTEHLIIIILAFDIELYLVGETSQSASQAIL
```

**Fig. 1.** Primary structure of the *C. milii* GnRH-II preprohormone. The signal peptide is indicated by boxed region, the mature peptide is shown by the region in gray; the amidation motif is shown in bold. The cryptic peptide is indicated by regular text.