General and Comparative Endocrinology 209 (2014) 61–73

Contents lists available at ScienceDirect

General and Comparative Endocrinology

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

Calcitonin receptor family evolution and fishing for function using *in silico* promoter analysis

Rute Martins, Flobela A. Vieira, Deborah M. Power *

Centro de Ciências do Mar, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

ARTICLE INFO

Article history:
Available online 8 May 2014

Keywords:
Calcitonin
Evolution
G protein coupled receptors
Promoter analysis
Synteny

ABSTRACT

In the present study the calcitonin receptor (CTR) sub-family of family B G-protein coupled receptors (GPCRs) in teleosts is evaluated and put in the context of the families overall evolution from echinodermates to vertebrates. Echinodermates, hemichordates, cephalochordates and tunicates have a single gene that encodes a receptor that bears similarity to the vertebrate calcitonin receptor (CTR) and calcitonin-like receptor (CTR/CLR). In tetrapods one gene encodes the calcitonin receptor (CALCR) and another gene the calcitonin receptor-like receptor (CALCRL). The evolution of CALCR has been under strong conservative pressure and a single copy is also found in fishes and high conservation of gene organisation and synteny exits from teleosts to human. A teleost specific CTR innovation that occurred after their divergence from holostei is the presence of several HBDs in the N-terminus. CALCRL had a different evolutionary trajectory from CALCR and although a single gene copy is present in tetrapods the sarcopterygii fish, the coelacanth, has 1 copy of CALCRL but also a fish specific form CALCRL3. The ray-finned fish, the spotted gar, has 1 copy of CALCRL and 1 of CALCRL3 but the teleost specific whole genome duplication has resulted in a CALCRL1 and CALCRL2 in addition to the fish specific CALCRL3. Strong conservation of CALCRL gene structure exists from human to fish. Promoter analysis *in silico* reveals that the duplicated CALCRL genes in the teleosts, zebrafish, takifugu, tetraodon and medaka, have divergent promoters and different putative co-regulated gene partners suggesting their function is different.

1. Introduction

The ultimobranchial gland a diffuse endocrine tissue located in the *septum transversum* is the site of calcitonin (CT) production in teleosts (Copp et al., 1962). In contrast, all other vertebrates produce the 32 amino acid peptide in the C cells of the thyroid gland (Rosenfeld et al., 1983). CT belongs to the calcitonin gene related peptide (CGRP) superfamily that in vertebrates includes amylin (AMY), adrenomedullin (AM), intermedin (IMD) and calcitonin receptor stimulating peptide (CRSP) (Katafuchi et al., 2009; Katafuchi and Minamino, 2004; Naot and Cornish, 2008; Ogoshi et al., 2006). The CGRP superfamily members are encoded by different genes, with the exception of CT and CGRP that are generated by alternative splicing of the calcitonin gene (*CALCI*). The first CGRP family member for which a physiological function was described is CT and from amphibians to mammals it acts as a hypocalcemic and hypophosphatemic hormone by inhibiting bone resorption (Azria, 1989). The other members of the CGRP superfamily have weak or no calcitropic actions and instead are neuroregulatory, thermoregulatory, vasodilatory and also regulate food intake (Katafuchi et al., 2009). The calcitropic action of CT in teleost fish has still not been convincingly demonstrated and it has been reported to be hypercalcaemic, hypocalcaemic or to have no effect (Bjornsson et al., 1989; Evans et al., 2005; Lafont et al., 2007; Mukherjee et al., 2004; Suzuki et al., 2000a,b). Furthermore although most of the members of the CGRP superfamily have been identified in fish (Ogoshi et al., 2006; Wong and Takei, 2009) and their tissue distribution has been described their physiological functions are still far from resolved (reviewed in Katafuchi et al., 2009; Lafont et al., 2007; Nobata et al., 2008; Ogoshi et al., 2008).

The CGRP superfamily members bring about their function when they bind to seven-transmembrane domain G-protein coupled receptors (GPCRs). The GPCRs to which CGRP superfamily members bind belong to the secretin family (a.k.a family B or 2), one of the largest receptor families for hormones and neuropeptides. GPCRs of the calcitonin subfamily include the calcitonin receptor (CTR) and calcitonin receptor-like receptor (CRL) and receptor activity modifying proteins (RAMPs) interact with receptors and modify their specificity (Girgis et al., 2013). The association between receptors and RAMPs give rise to seven distinct receptors and glycosylation of the extended N-terminal region is

* Corresponding author..
E-mail address: dpower@ualg.pt (D.M. Power).

http://dx.doi.org/10.1016/j.ygcen.2014.04.025
0016-6480/

important for receptor recognition and high affinity binding by CT (Lerner, 2006; McLatchie et al., 1998; Naot and Cornish, 2008). In mammals several isoforms of CTR have been identified and CTR1a, which is highly expressed in multinucleated osteoclasts, is the receptor that mediates the action of CT on bone (Lerner, 2006; Pondel, 2000). CLR shares greater than 50% amino acid identity with CTR and is recognized by CGRP and adrenomedullin. In addition to being involved in biological functions, like cardiovascular development, CLR also appears to play a role in the skeleton, since it is detected in both osteoblasts and osteoclasts (Dackor et al., 2006; Lerner, 2006; Naot and Cornish, 2008; Schinke et al., 2004; Uzan et al., 2004).

Mammalian genomes contain one gene for CALCR, one for CALCRL and three genes for RAMPs (1–3). In contrast, teleost genomes contain one gene for CALCR, three for CALCRL (1–3) and 5 genes encode the RAMPs (1–5) (Nag et al., 2006, 2007a; Parameswaran and Spielman, 2006; Poyner et al., 2002). The expansion of the CALCR/CALCRL gene family in teleosts is intriguing particularly since the function of the receptors is poorly established and the involvement of CT in bone and calcium homeostasis remains ambiguous. In the present study a comparative genomics approach is taken to provide insight into the evolution of the CALCR/CALCRL gene family in fish and to also procure novel receptor members in the vertebrates. Hypothesis about putative biological function of CALCR/CALCRL gene family members are generated using comparative promoter analysis. Computer modeling of the proximal promoter region of teleost CALRL1 and CALRL2 is used to identify potential regulatory regions and infer the functional context of gene expression of these duplicates. The generally highly conserved order and spacing of regulatory elements, such as transcription factor (TF) binding sites (TFBS) in promoters during evolution is exploited to gain insight into function (Cartharius et al., 2005; Cohen et al., 2006; Qiu, 2003; Werner, 2001; Werner et al., 2003). Promoter modules or frameworks, the ordered complex of TFBSs that are conserved in distance and orientation (Dohr et al., 2005; Fessele et al., 2002), are used to identify putative co-regulated genes that provide insight into putative functions of the CLR1 and CLR2.

2. Methods

2.1. CALCR and CALCRL sequence retrieval

CALCR and CALCRL gene sequences from vertebrates were procured in NCBI, SWISSPROT (www.ncbi.nlm.nih.gov) and EBI (www.ebi.ac.uk/) public databases and in the sea bass genome draft (Kuhl et al., 2010) using Human CTR and CLR (NP_001158209, NP_005786) and *Takifugu rubripes* CALCR (ENSTRUP00000039142) and CALCRL 1–3 sequences (ENSTRUP00000016740, ENSTRUP00000029916 and ENSTRUP00000029511, respectively) as bait. Database searches were directed at obtaining cDNA sequences from representatives of the phylogenetic groups, agnatha (lampreys and hagfish), gnathostomata (sharks), osteichthyes (teleosts, gars, bichirs and sturgeon) sarcopterygii (coelacanths), tetrapoda (frogs, lizards, birds, monotremes, marsupials and placentals). A similar search strategy to that outlined above was also used to identify putative CALCR and CALCRL gene homologues in echinodermata (sea urchin, *Strongylocentrotus purpuratus*), hemichordata (acorn worm, *Saccoglossus kowalevskii*), cephalochordata (amphioxus, *Branchiostoma floridae*) and tunicata (*Ciona intestinalis*). The sequences collected were used for subsequent sequence comparisons and phylogenetic analysis. For information on the accession numbers of CTR and CLRs proteins used in this study see Supplementary Table 1.

2.2. Protein characterization and multiple sequence alignments

Characterization of conserved protein domains in echinodermata (sea urchin: XP_003725171), cephalochordata (amphioxus, a.k.a lancelet: XP_002586885) and vertebrata (Human, chicken CTR: XP_425985 and CLR: NP_001157122 and takifugu CTR and CLR1–3 sequences) was done using the simple modular architecture research tool (Smart, version 7) (Letunic et al., 2012). Multiple sequence alignments of deduced amino acid sequences for CTR and CLRs of fish and tetrapods were carried out using ClustalW v.2.0 (Larkin et al., 2007) and edited using GeneDoc version 2.7.0 (Nicholas et al., 2007). The transmembrane regions (TM) of deduced CTR and CLR protein sequences were predicted using TMHMM v. 2.0 (Sonnhammer et al., 1998). Protein kinase A (PKA) and C (PKC) phosphorylation and N-glycosylation sites were predicted using NetPhosK 1.0 and NetNGlyc 1.0 (Blom et al., 2004). The predicted N-glycosylation and protein kinase sites in CTR and CLR that had a statistical score ($p \geqslant 0.75$) were selected and only those conserved in the majority of the species analyzed were highlighted in the alignments. The percentage of identity of the hormone binding domain (HBD) identified in echinodermata, hemichordata, and chordata (cephalochordata and vertebrata) CTR and CLRs was estimated by generating an identity matrix with GeneDoc version 2.7.0 (Nicholas et al., 2007).

2.3. Genomic organization and synteny maps of vertebrate CALCR and CALCRL genes

The genomic organization of CALCR genes in human (*Homo sapiens*, CTR ENSG00000004948, cDNA NM_001164737), chicken (*Gallus gallus*, ENSGALG00000009509, cDNA XM_425985), takifugu (*T. rubripes*, ENSTRUG00000015319, cDNA ENSTRUT00000039283) and spotted gar (*Lepisosteus oculatus*, ENSLOCG00000010787, cDNA ENSLOCT00000013246) was characterized by aligning each genomic sequence with the respective cDNA using tblastn in the Ensembl database (www.ensembl.org). The genomic organization of CALCRL was established using a similar strategy in human (ENSG00000064989, cDNA NM_005795), chicken (ENSGALG00000002632, cDNA NM_001163650), takifugu (CALCRL1 ENSTRUG00000006816, CALCRL2 ENSTRUG00000011835 and CALCRL3 ENSTRUG00000011677; cDNA CALCRL1 ENSTRUT00000016813, CALCRL2 ENSTRUT00000030033 and CALCRL3 ENSTRUT00000029628), African coelacanth (*Latimeria chalumnae*, CLR ENSLACG00000003833 and CALCRL3 ENSLACG00000019098; cDNA CALCRL ENSLACT00000004346 and CALCRL3 ENSLACT00000025994) and the spotted gar (CALCRL ENSLOCG00000006075 and CALCRL3 ENSLOCG00000005174; CALCRL ENSLOCT00000007345 and CALCRL3 ENSLOCT00000006243). When full-length sequences of CALCR or CALCRL were not fully characterized in the Ensembl database, pairwise alignment of cDNA sequences using Spidey mRNA-to-genome software (www.ncbi.nlm.nih.gov) was used to characterize the complete gene organization. Furthermore, putative intron/exon boundary splice site consensus sequences (AG/GT) were identified and the amino acid and nucleotide sequences extracted from the Ensembl database. The gene organisation was graphically plotted using the software FancyGENE (Rambaldi and Ciccarelli, 2009).

2.4. Phylogenetic analysis

Phylogenetic analysis was performed using either the conserved HBD domain or the full sequence of CTR of CLR proteins including the regions spanning the HBD (HBDD in fish CTRs) up until the end of the 7 TM domains. Multiple sequence alignment of the predicted CTR and CLR protein sequences from echinodermata, hemichordata, cephalochordata, and vertebrates was generated using