General and Comparative Endocrinology 195 (2014) 164–173

Contents lists available at ScienceDirect

General and Comparative Endocrinology

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

Evolution of the gastrin–cholecystokinin gene family revealed by synteny analysis

Délia Dupré, Hervé Tostivint *

UMR 7221 CNRS/MNHN Evolution des Régulations Endocriniennes, Muséum National d'Histoire Naturelle, 75231 Paris, France

ARTICLE INFO

Article history:
Received 14 May 2013
Revised 29 September 2013
Accepted 29 October 2013
Available online 11 November 2013

Keywords:
Gastrin
CCK
Evolution
Multigenic family
Gene duplication
Tetraploidization

ABSTRACT

Gastrin (GAST) and cholecystokinin (CCK) are two structurally and functionally related peptide hormones that exert many functions, including regulation of gastric and pancreatic secretion, feeding behaviour and energy homeostasis. *GAST* and *CCK* genes are assumed to have diverged from a common ancestral gene, over 500 million years ago in the vertebrate lineage. However, although a large number of GAST and CCK-related sequences have been identified both in vertebrate and non-vertebrate species, the evolutionary history of the GAST/CCK family remains little understood. To address this issue, we used extensive genome synteny comparisons of vertebrate chromosomes, in particular to evaluate the impact of whole-genome duplications. In the present study, we confirm that the GAST/CCK family in vertebrates is composed of two paralogous genes, namely *GAST* and *CCK*, and even three in teleosts, namely *GAST*, *CCK1* and *CCK2*. We also show that the *GAST* and *CCK* genes arose by duplications of a single ancestral gene through the 2R and that the two copies of the *CCK* gene found in teleosts have probably been generated through the 3R. Finally, our results suggest that the vertebrate ancestor possessed four members of the GAST/CCK family, of which two have likely been lost during evolution.

1. Introduction

Gastrin (GAST) and cholecystokinin (CCK) are two hormones that were initially identified by their ability to stimulate gastric acid and pancreatic juice secretions, respectively (Edkins, 1905; Ivy and Goldberg, 1928). Several decades later, their purification and sequencing revealed that they are two structurally related peptides characterized by their common C-terminus: Trp-Met-Asp-Phe-NH_2 (Gregory et al., 1964; Mutt and Jorpes, 1968). Subsequent studies have shown that the GAST and CCK molecules each include a number of N-terminal extensions of this invariant C-terminal tetrapeptide (Fig. 1). Up to now, it has been assumed that GAST and CCK evolved from a common ancestral gene (Baldwin et al., 2010; Larsson and Rehfeld, 1977).

GAST is the main regulator of gastric acid secretion and gastric mucosal growth, whereas CCK regulates gall bladder emptying, pancreatic enzyme secretion, and furthermore acts as a major neurotransmitter in the central and peripheral nervous systems (Dockray et al., 2005; Rehfeld, 2004). GAST and CCK exert their actions through two receptors that belong to the G protein-coupled receptor superfamily, called $CCKR_1$ (or $CCKR_A$) and $CCKR_2$ (or $CCKR_B$) Dufresne et al., 2006.

About one hundred GAST and CCK sequences have been reported so far from various vertebrate species including mammals, where they were initially identified, birds (Dimaline et al., 1986), reptiles (Jonson et al., 1990), amphibians (Rourke et al., 1997), teleosts (Peyon et al., 1998) and cartilaginous fish (Johnsen, 1998, for review). In teleost fish, many species have been shown to contain two CCK sequences (Kurokawa et al., 2003). More or less distant relatives of the GAST/CCK family have also been identified in tunicates (Johnsen and Rehfeld, 1990; Monstein et al., 1993), in insects (Nachmann et al., 1986) and in nematodes (Janssen et al., 2008), suggesting that this neuropeptide family emerged very early during the evolution of bilaterians (Baldwin et al., 2010; Mirabeau and Joly, 2013). Phylogenetic analysis showed that in vertebrates the GAST and CCK sequences cluster into two main groups, namely GAST and CCK and that, in the CCK cluster, teleost CCKs are themselves divided into two clusters, CCK1 and CCK2 (Baldwin et al., 2010; Kurokawa et al., 2003). All these data support the notion that the *GAST* and *CCK* genes diverged, at the latest, upon the appearance of gnathostomes (Johnsen and Rehfeld, 1990; Johnsen, 1998) and that a more recent duplication of the *CCK* gene occurred in the teleost lineage (Baldwin et al., 2010). However, such a scenario does not give any information regarding the mechanisms by which *GAST* and *CCK* genes were generated.

Two main types of gene duplications are classically distinguished: local duplications that generally implicate individual genes, and whole-genome duplications that involve all the chro-

* Corresponding author. Fax: +33 (0)1 40 79 36 18.
E-mail address: htostivi@mnhn.fr (H. Tostivint).

0016-6480/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygcen.2013.10.019

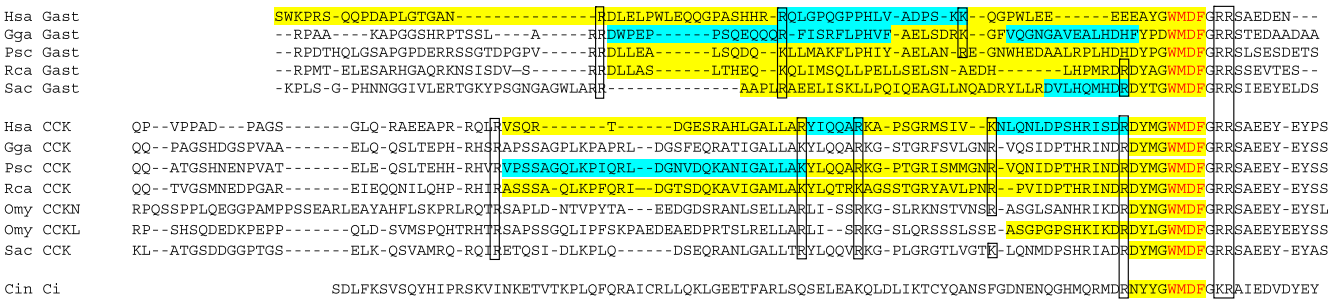

Fig. 1. Alignment of the amino acid sequences of several representative vertebrate GAST and CCK precursors. The sequence of the pro-cionin has been added for comparison. The sequences of the known peptides processed from these precursors are colored: human GASTs 71, 34 and 17; chicken GASTs 53, 30 21 and 7; turtle GAST 52; frog GAST 47; dogfish GASTs 49, 17 and 8; human CCKs 58, 39, 33, 22 and 8; chicken CCK 8; turtle CCKs 70 and 40; frog CCK 69; trout CCKL 21, trout CCKN 8; dogfish CCK 8; Cionin 8. The tetrapeptide amide common to all members of the GAST/CCK family is in red. Conserved putative cleavage sites are boxed. Hsa, *Homo sapiens*; Gga, *Gallus gallus*; Psc, *Pseudemys scripta*; Rca, *Rana catesbiana*; Omi, *Oncorhynchus mykiss*; Sac, *Squalus acanthias*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mosomes and hence all the genes they bear. It is now well established that vertebrates underwent two major rounds of whole-genome duplications during the course of their evolution (Dehal and Boore, 2005; Nakatani et al., 2007; Putnam et al., 2008). These tetraploidization events, referred to as 1R and 2R, occurred very early during the vertebrate evolution, likely before the emergence of cyclostomes (Smith et al., 2013). An additional round of tetraploidization, referred to as 3R, occurred in the stem lineage of teleost fish (Jaillon et al., 2004; Kasahara et al., 2007).

The vestiges of all these duplication events can be revealed in the genome of the extant vertebrate species by synteny analysis, the analysis of relative gene-order conservation between species (Larhammar et al., 2009). Thus, during the last years, synteny analysis has helped clarify the evolutionary history of numerous neuropeptide families in vertebrates, including opioid peptide (Sundström et al., 2010), GnRH (Kim et al., 2011; Smith et al., 2013; Tostivint, 2011), somatostatin (Liu et al., 2010; Tostivint et al., 2008), urotensin II (Parmentier et al., 2011; Quan et al., 2012), kisspeptin (Pasquier et al., 2012) and secretin (Hwang et al., 2013) families. The aim of the present study was to re-examine the evolution of the GAST/CCK family by using extensive genome synteny comparisons of vertebrate chromosomes and, in particular, to evaluate the impact of both local and whole-genome duplications on this evolution.

2. Materials and methods

2.1. Phylogenetic analysis of the GAST/CCK family

The amino acid sequences of 68 selected prepro-peptides of the GAST/CCK family were aligned using Muscle (Edgar, 2004) then manually optimized (see Supplementary Fig. S1 for the complete list of these sequences and their accession numbers, and Supplementary Fig. S2 for the resulting alignment). The phylogenetic tree was built using PhyML with default parameters (Guindon and Gascuel, 2003) via the Seaview version 4 software (Gouy et al., 2010) then rooted using the sequence of the cionin precursor. The reliability of the tree was tested by the bootstrap procedure with five hundred replications.

2.2. Genomic synteny analysis

Genomic synteny analysis was first performed using Genomicus (version 72) (http://www.dyogen.ens.fr/genomicus-71.01/cgi-bin/search.pl) a genomic browser dedicated to the study of synteny among multiple genomes (Muffato et al., 2010). Synteny maps for the genomic neighbourhoods surrounding the *GAST* genes and the *CCK* genes in human (*Homo sapiens*), chicken (*Gallus gallus*), anole lizard (*Anolis carolinensis*), Western clawed frog (*Xenopus tropicalis*), medaka (*Oryzia latipes*), stickleback (*Gasterosteus aculeatus*), zebrafish (*Danio rerio*) and tetraodon (*Tetraodon nigroviridis*) were constructed with AlignView using human *HAP1* and *TRAK1* as reference genes, respectively. These two genes were chosen because they are present in most of the selected species. The *GAST* genes in chicken, lizard and teleosts were manually added because they did not appear in the initial views. They were retrieved by BLAST via the Ensembl server (http://www.ensembl.org/Multi/blastview). Note that we detected two tandemly arranged sequences encoding for GAST in the lizard genome. However, because their nucleotide sequences are completely identical, it is likely that they are due to an assembly problem.

Search for paralogous pair genes among the genes surrounding the *GAST* and *CCK* genes was then performed using Synteny Database (http://teleost.cs.uoregon.edu/synteny_db/), an automated system to identify conserved syntenic regions in a primary genome using as outgroup a genome that diverged from the investigated lineage before a whole genome duplication event (Catchen et al., 2009). The longest paralogon was obtained with the medaka *HAP1* gene as a query and the chicken genome as outgroup.

For the reconstruction of the tetraparalogon, i.e. the fourfold repeated regions carrying paralogous genes, which contains the *GAST* and *CCK* genes, we used the method recently proposed by Yegorov and Good, (2012). This method uses ancestral genome reconstructions to study orthologous and paralogous relationships among genes. It is specifically based on two different models, namely the Putnam model (or P-model) which reconstructed the karyotype of the chordate ancestor (Putnam et al., 2008) and the Nakatani-model (or N-model) which reconstructed that of the vertebrate ancestor (Nakatani et al., 2007). In the P-model, the number of protochromosomes equals 17 (numbered 1–17) while in the N-model, which illustrates a later stage in the evolution of the chordate genome, it is 10 (A–J). Thus, the genome of current vertebrates can be basically viewed as a mosaic of these 17 or 10 ancestral pieces twice duplicated through the 2R. Following the method of Yegorov and Good (2012), the two chromosomal segments bearing the *GAST* and *CCK* genes were matched with the reconstructed protochromosomes of the P- and N-models making it possible to identify the other chromosomal segments belonging to the same tetraparalogon. The Browser of Synteny Database was used to search for genes belonging to multigene families whose several members were specifically located on all these segments, namely for the P-model (Putnam et al., 2008) those referred into as 2.6 (segment 6 of human chromosome 2), 3.2 (segment 2 of human chromosome 3, containing the *CCK* gene), 7.2, 7.4, and 7.6 (segments 2, 4 and 6 of human chromosome 7), 10.2 (segment 2