General and Comparative Endocrinology 175 (2012) 27–33

Contents lists available at SciVerse ScienceDirect

General and Comparative Endocrinology

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

Review

Origin and convergent evolution of exendin genes

David M. Irwin *

Department of Laboratory Medicine and Pathobiology, 1 King's College Circle, University of Toronto, Toronto, Ont., Canada M5S 1A8
Banting and Best Diabetes Centre University of Toronto, Toronto, Ont., Canada M5S 1A8

ARTICLE INFO

Article history:
Received 5 July 2011
Revised 14 November 2011
Accepted 15 November 2011
Available online 25 November 2011

Keywords:
Exendin
Secretin hormones
Convergent evolution
Adaptive evolution
Genomic data
Vertebrate

ABSTRACT

Exendins are secretin hormone-like peptides that are components of the toxins from two venomous lizards, *Heloderma suspectum* (Gila monster) and *Heloderma horridium* (Mexican bearded lizard). Exendins-1 and -2 are vasoactive intestinal peptide (VIP)-like, both in sequence and function, while exendins-3 and -4 are glucagon-like peptide-1 (GLP-1)-like. The evolutionary origin of these peptides, and the genes that encode them, has been unclear. Recently, genes orthologous to exendin have been identified in reptiles, birds and amphibians. Analysis of the orthologous sequences demonstrates that the *Heloderma* exendins diversified by gene duplication from a common exendin ancestor on the *Heloderma* lineage after divergence from other reptiles, including the anole lizard and Burmese python. In addition, the exendin toxin peptide sequences, but not their pro or signal peptides, have evolved very rapidly on the *Heloderma* lineage, likely as they adapted to their new function as toxins. Exendins-1 and -2 not only evolved rapidly but their sequences have evolved convergently upon that of VIP, resulting in a doubling of its identity with VIP, while exendins-3 and -4 have retained an ancestral property of being more GLP-1-like sequences. These results suggest that the ancestral role of exendin, which is potentially still retained in some species, had greater similarity with proglucagon-derived peptides or GIP.

1. Introduction

The venomous lizards *Heloderma suspectum* (Gila monster) and *Heloderma horridium* (Mexican bearded lizard), which prey on small mammals, produce toxins that are composed of complex mixtures of proteins that are secreted from glands on the sides of the lower jaw [11,13,28,30]. Among the proteins found in the *Heloderma* toxins are four secretin hormone-like peptides, which are now collectively called exendins [12,17,31]. The pathological effects of these peptides are mediated by the relaxation of cardiac smooth muscle leading to hypotension [13,35]. The first characterized exendins, exendin-1 and -2, were originally named helospectin and helodermin, respectively, and were found to not only have immunological and biological properties similar to mammalian vasoactive intestinal peptide (VIP) and peptide histidine isoleucine (PHI) but also have high sequence similarity (52–56%) with these peptides in their primary amino acid sequences [26,36]. Both exendin-1 and exendin-2 were found to bind and activate mammalian VIP receptors [15,36]. Exendins-3 and -4, in contrast, activate mammalian GLP-1 receptors and have biological functions most similar to, and have highest sequence similarity (56%) with, mammalian glucagon-like peptide-1 (GLP-1) [9,10,16,29]. Indeed, the similarity in action of exendin-4 with that of GLP-1 has led to this peptide, and derivatives of it, being used in the treatment of diabetes [14,24]. An advantageous feature of exendin-4 is that it has a longer biological half-life than GLP-1 due to its resistance to degradation by the enzyme dipeptidyl peptidase 4 (DPP-4) [34]. Intriguingly, despite levels increasing after feeding, exendin-4 does not appear to effect glucose metabolism in the Gila monster [5].

The origin of the genes for the exendin peptides from Gila monster and Mexican bearded lizard has been unclear [13,17,31]. It had been suggested that they were just highly divergent orthologs of the mammalian VIP and GLP-1 sequences, however this possibility was excluded by the characterization of cDNA sequences for exendins-2 and -4 and proglucagon, and the partial genomic sequences for the genes for PACAP, PHI, and VIP from *H. horridium* [4,27] (see Table S1 for the accession numbers for peptides and cDNAs). cDNA and genomic sequences demonstrated that exendins-4 was not encoded by the proglucagon cDNA, nor was exendin-2 encoded by the VIP, PHI, or PACAP genomic sequences [4,27]. Surprisingly, despite the greater similarity of exendins-1 and -2 with VIP [15,36] and exendins-3 and -4 with GLP-1 [9,10,29], the characterization of the exendin-2 and -4 cDNA sequences demonstrated that the genes for these two toxin genes are closely related [27]. While there are great differences in the exendin-2 and -4 toxin peptide sequences, the signal and pro-peptide regions of these genes are more similar to each other than to the corresponding region of

* Address: Department of Laboratory Medicine and Pathobiology, University of Toronto, 6207 Medical Sciences Building, 1 King's College Circle, Toronto, Ont., Canada M5S 1A8.
E-mail address: david.irwin@utoronto.ca

0016-6480/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygcen.2011.11.025

any other secretin-like hormone precursor [27]. These observation have led to several models for the evolution of the exendin peptides with the most likely models involving duplication of either (1) a glucagon gene followed by a additional duplications and convergent evolution of some exendin sequences to the VIP peptide sequence or (2) a VIP gene followed by additional duplications and convergent evolution of some sequences upon the GLP-1 peptide sequence [13]. The lack of sequences orthologous to exendins from other vertebrate species had prevented distinguishing between these two models. Recent advances with data obtained from genome projects allow us to now resolve this question.

2. Structure of vertebrate exendin genes

Recently, gene sequences orthologous to exendin were identified from the genome of another reptile, the anole lizard, and from the genomes of several birds (chicken, turkey, duck, and zebra finch) and the frog *Xenopus tropicalis* (also known as *Silurana tropicalis*) [20]. While the coding sequence and genomic locations of these genes were reported, the genomic structure of these exendin-like sequences and the evolution of the genes, and the peptides encoded by them, were not [20]. Searches of the assembled anole lizard genome in the Ensembl database (www.ensembl.org) for sequences that could predict protein sequences similar to glucagon or glucose-dependent insulinotropic polypeptide (GIP) resulted in the identification of an exendin-like sequence [20]. Blast searches had resulted in the identification of three glucagon-like genomic sequences, two of which were previously characterized as the proglucagon and GIP genes [18]. The third sequence (see Table S2) identified in these searches was roughly equally similar to both proglucagon and GIP.

The third lizard glucagon-like gene was characterized after downloading a 10 kb genomic sequence that was centered on the glucagon-like sequence. When this sequence was used in a Blastx search [1] of the NCBI protein database the most similar sequences identified were proglucagon-derived sequences from diverse vertebrates and exendin sequences from *Heloderma*. Further characterization of this 10 kb anole lizard genomic sequence identified a region 5′ to the glucagon-like sequence that had greatest similarity to the N-terminal signal and pro-peptide regions of the exendin sequences [20]. A gene could be predicted from the genomic sequence (Fig. S1), which has a signal and pro-peptide encoded by one exon, and an exendin/glucagon-like sequence is encoded by a second exon. Each of these exons are flanked by appropriate splice donor and acceptor sequences [23] with the two coding exons separated by a phase II intron, i.e., the coding sequence is interrupted by the intron after the second base of a codon, an intron phase that is identical to those of the homologous introns in the proglucagon and GIP genes [18,19]. A potential spice acceptor sequence is found immediately upstream of the start codon of the sequence (Fig. S1), suggesting that the exendin gene, like the proglucagon and GIP genes [18,19] has a 5′ non-coding exon with an initiator codon located in exon 2. Similarly, a splice donor site was found at the 3′ end of the exendin-encoding exons, with a predicted intron in phase II homologous to the introns found in the proglucagon and GIP genes. Searches of the sequence 3′ to the exendin exon suggested a possible exon encoding 5 amino acids, a stop codon, and a 3′ untranslated sequence (Fig. S1).

To better understand the origin and evolution of the exendin gene the predicted anole lizard exendin protein sequence was used to search for orthologous genomic sequences in the assembled genomes in the Ensembl database (www.ensembl.org) as well as the whole-genome shotgun reads (WGS), representing unassembled genome data, at the NCBI database (www.ncbi.nlm.nih.gov). From the Ensembl database searches, gene sequences that predict a protein similar to exendin were found in the chicken, turkey, duck, zebra finch and *X. tropicalis* genomes (Table S2 and Figs. S2 and S3), and their predicted protein sequences have been characterized [20]. Searches of the NCBI WGS database resulted in the identification of two non-overlapping genomic sequences from the unassembled Burmese python genome that aligned with two exons of the exendin gene from the anole lizard (Table S2 and Fig. S1). Each of these genomic sequences allowed the prediction of most, if not all, of the coding regions of the exendin genes (Fig. 1). The anole lizard, zebra finch, turkey and *Xenopus* exendin genes are located on large genomic scaffolds that allowed identification of neighboring genes while the exendin genes from the other species are located on shorter genomic sequences that did not predict any additional genes [20]. Comparison of the gene content of the different contigs showed that the neighboring genes were generally shared between species, suggesting that the exendin genes exist in a conserved genomic neighborhood, strongly indicating that these sequences are orthologous [20].

Genomic sequence can suggests the presence of a gene, however, demonstrating that the sequence is expressed, by identifying mRNAs, strengthens the likelihood that the gene is functional. To find evidence for expression, the NCBI non-redundant and EST databases were searched with each of the identified exendin genes resulting in the identification of mRNAs or ESTs (Table S2) from chicken, zebra finch, and *X. tropicalis* [20]. mRNAs and ESTs were also an aid in confirming the annotation of the structure of these genes by providing evidence for intron splicing and all of the identified mRNA and EST sequences were consistent with the predicted gene structures, and in the case of the *Xenopus* allowed the prediction of one additional intron in the C-terminal region (Figs. S2 and S3) [20]. Expression of the chicken and *Xenopus* exendin genes was also assessed using RT-PCR. Expression was found in several tissues in *X. tropicalis*, with strong expression in the brain, intestine and pancreas, while in the chicken, the highest level of expression was observed in the brain [20].

3. Evolution of exendin sequences

The protein coding sequences predicted from the exendin genomic sequences could be aligned with the introduction of only a few gaps (Fig. 1) [20]. All of the predicted protein sequences posses a potential proteolytic processing site [2], paired basic residues, at the N-terminus of the exendin hormone/toxin sequences, with the exendin sequences from species other than *Heloderma* having a second potential proteolytic processing site near the C-terminus that could generate a 29, or 28 in *Xenopus*, amino acid-long mature hormone. The exendin hormones have the potential to be secreted as the program SignalP [8] indicated that all of the precursor protein sequences predict signal peptides of 23 amino acids, except for zebra finch where it is 20 amino acids long. The zebra finch predicted a slightly different signal peptide, due to: (1) the peptidase cleavage site, being one amino acid residue N-terminal to those of the other sequence, likely due to the replacement of the tryptophan residue 22 with a cysteine; and (2) the deletion of two amino acids residues from the N-terminal region of the predicted signal peptide sequence. The portions of the precursor sequences that show the greatest variability in lengths are the pro-peptide and C-terminal regions (Fig. 1), regions that would be removed during the release of the potentially biologically active product. The C-terminal region, though, is poorly characterized, and in python and turkey sequences this region could not be identified (Figs. 1, S1, and S2). If the *Heloderma* exendin sequences were excluded, then the hormone portion of the coding sequence was best conserved, consistent with this region having a biological function. The signal peptides were less well conserved, however its hydrophobic nature was conserved.